Voice Activity Detection Using Laplacian Model and UMP Test

KEUN WON JANG, DONG KOOK KIM
Chonnam National University
Department of Electronics and Computer Engineering
300 Yongbong-dong, Buk-gu, Gwangju, 500-757
KOREA

JOON-HYUK CHANG
Inha University
Department of Electronic Engineering
253, Yonghyun-dong, Nam-gu, Incheon, 402-751
KOREA

Abstract: This paper presents a new voice activity detection (VAD) method using the Laplacian distribution and a uniformly most powerful (UMP) test. The UMP test is employed to derive the new decision rule based on likelihood ratio test (LRT). The proposed method provide the decision rule by comparing the sum of magnitude of real and imaginary parts of the noisy spectral component to the adaptive threshold estimated from the noise statistics. Experimental results show that the proposed VAD algorithm based on the Laplacian and the UMP test outperforms the conventional scheme at a low SNR.

Key–Words: Voice activity detection, Laplacian distribution, Likelihood ratio test, Uniformly most powerful test

1 Introduction

Voice activity detection (VAD) refers to the classical problem of distinguishing active speech from nonspeech and has various applications such as speech coding, speech enhancement, and echo cancelation. Recently, many statistical model-based VAD algorithms in which the likelihood ratio test (LRT) is applied to a set of hypotheses have been developed [1]-[4]. These methods adopt the statistical models which operate in the discrete Fourier transform (DFT) domain. Usually the distributions of noisy speech and noise spectra are assumed to be complex Gaussians [1]. Chang et. al., utilized the complex Laplacian and Gamma probability density functions (pdfs) to model the distributions of noisy speech and noise spectra, and showed that VAD based on these distributions was better than that based on the complex Gaussian distribution [2], [3]. Recently we proposed a new VAD technique based on the Gaussian and a uniformly most powerful (UMP) test which compares the magnitude of the noisy speech spectral to the threshold which depends only on the noise statistics and the false-alarm probability to detect the speech or nonspeech [4].

In this paper, we further extend the ideas of the previous UMP technique [4] by incorporating the Laplacian pdf instead of the Gaussian. We apply the UMP test to derive the new decision rule which requires the distribution for the sum of magnitude of real and imaginary parts of the noisy speech spectral. The proposed method depends only on the magnitude of real and imaginary parts of the noisy spectral component and the threshold based on the estimated noise statistics, which is computationally efficient in its implementation. Experimental results indicate that the proposed VAD algorithm based on the Laplacian and the UMP test shows better results compared with the conventional algorithms at a low SNR.

2 Likelihood Ratio Test for VAD

We assume that a noise n is added to a speech signal s, with their sum being denoted by x. Given two hypotheses H_0 and H_1, which, respectively, indicate speech absence and presence, it is assumed that

$$\begin{aligned} H_0\text{: speech absent } &: \mathbf{X}(t) = \mathbf{N}(t) \\ H_1\text{: speech present } &: \mathbf{X}(t) = \mathbf{N}(t) + \mathbf{S}(t) \end{aligned} \quad (1)$$

in which, $X(t) = [X_0(t), X_1(t), \ldots, X_{M-1}(t)]^T$, $N(t) = [N_0(t),\ N_1(t),\ \ldots,\ N_{M-1}(t)]^T$ and $S(t) = [S_0(t), S_1(t), \ldots, S_{M-1}(t)]^T$ are the DFT coefficients at frame t of the noisy speech, noise, and clean speech, respectively. Also, M is total number of frequency bins, and T is transposed matrix.

We consider the statistical model-based VAD using the Laplacian pdf for speech signal [2], [3]. 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 coefficient X_k. If both the real and imaginary parts have the same variance, their distributions are given by

$$p(X_{k(R)}) = \frac{1}{\sigma_x} \exp\left\{-\frac{2|X_{k(R)}|}{\sigma_x}\right\} \quad (2)$$

$$p(X_{k(I)}) = \frac{1}{\sigma_x} \exp\left\{-\frac{2|X_{k(I)}|}{\sigma_x}\right\} \quad (3)$$

where σ_x is the variances of X_k. If the real and imaginary parts of X_k are further assumed to be independent, the distribution $p(X_k)$ of X_k turns out to be

$$\begin{aligned} p(X_k) &= p(X_{k(R)}) \cdot p(X_{k(I)}) \quad (4) \\ &= \frac{1}{\sigma_x} \exp\left\{-\frac{2(|X_{k(R)}| + |X_{k(I)}|)}{\sigma_x}\right\} \quad (5) \end{aligned}$$

From (5), the distribution of the DFT coefficients under the respective hypotheses are given by

$$p(X_k|H_0)=\frac{1}{\lambda_{n,k}}\exp\left\{-\frac{2(|X_{k(R)}|+|X_{k(I)}|)}{\sqrt{\lambda_{n,k}}}\right\} \quad (6)$$

$$p(X_k|H_1)=\frac{1}{\lambda_{s,k}+\lambda_{n,k}}\exp\left\{-\frac{2(|X_{k(R)}|+|X_{k(I)}|)}{\sqrt{\lambda_{s,k}+\lambda_{n,k}}}\right\} \quad (7)$$

where $\lambda_{s,k}$ and $\lambda_{n,k}$denote the variances of S_k and N_k, respectively.

For VAD based on the assumed statistical models, the likelihood ratio (LR) for the kth frequency is given by

$$\Lambda_k=\frac{p\left(X_k|H_1\right)}{p\left(X_k|H_0\right)} \quad (8)$$

$$=\frac{1}{1+\xi_k}\exp\left\{2\bar{X}_k\left(\frac{\sqrt{\lambda_{s,k}+\lambda_{n,k}}-\sqrt{\lambda_{n,k}}}{\sqrt{\lambda_{s,k}+\lambda_{n,k}}\sqrt{\lambda_{n,k}}}\right)\right\} \quad (9)$$

where $\bar{X}_k = |X_{k(R)}| + |X_{k(I)}|$ and $\xi_k = \lambda_{s,k}/\lambda_{n,k}$, and ξ_k is called the *a priori* SNR [5]. The decision rule is constructed as the geometric mean of the LRs computed for the individual frequency bins such that

$$\log\Lambda = \frac{1}{M}\sum_{k=0}^{M-1}\log\Lambda_k \underset{H_0}{\overset{H_1}{\gtrless}} \eta \quad (10)$$

with η denoting the threshold of detection.

3 Likelihood Ratio Test Based on UMP Test

3.1 VAD Based on UMP Test

In this section, we present a new statistical model-based VAD method for the Laplacian pdf using the UMP test. To find the new decision rule, the decision statistic in (9) can be rewritten as

$$\Lambda_k = \frac{\lambda_{n,k}}{\lambda_{s,k}+\lambda_{n,k}}\cdot \exp\left\{2\bar{X}_k\left(\frac{\sqrt{\lambda_{s,k}+\lambda_{n,k}}-\sqrt{\lambda_{n,k}}}{\sqrt{\lambda_{s,k}+\lambda_{n,k}}\sqrt{\lambda_{n,k}}}\right)\right\} \quad (11)$$

where the value of $\lambda_{s,k}$ is unknown parameter, although a priori we know that $\lambda_{s,k} > 0$. We assume that the noise variance, $\lambda_{n,k}$ is known, which can be estimated during periods of nonspeech activity [5]. Then, the LRT is to decide H_1 if

$$\frac{\lambda_{n,k}}{\lambda_{s,k}+\lambda_{n,k}}\exp\left\{2\bar{X}_k\left(\frac{\sqrt{\lambda_{s,k}+\lambda_{n,k}}-\sqrt{\lambda_{n,k}}}{\sqrt{\lambda_{s,k}+\lambda_{n,k}}\sqrt{\lambda_{n,k}}}\right)\right\} > \eta\ . \quad (12)$$

Since it is known that $\lambda_{s,k} > 0$ and $\lambda_{n,k} > 0$, after taking the logarithm and some manipulation, we have finally the new decision rule as follows:

$$\bar{X}_k > \eta_k \quad (13)$$

where

$$\eta_k=\frac{\sqrt{\lambda_{s,k}+\lambda_{n,k}}\sqrt{\lambda_{n,k}}}{\sqrt{\lambda_{s,k}+\lambda_{n,k}}-\sqrt{\lambda_{n,k}}}\left(\ln\eta-\ln\left[\frac{\lambda_{s,k}+\lambda_{n,k}}{\lambda_{n,k}}\right]\right)\ . \quad (14)$$

Clearly, rather than comparing a test statistic like the LR to a threshold, the presented VAD method compares the sum of magnitude of real and imaginary parts of the noisy spectral component to a threshold η_k. But the key question is whether we can implement this detector without the estimate of the value of $\lambda_{s,k}$. The $\bar{X}_k$ does not depend on $\lambda_{s,k}$ but it appears that the threshold η_k does. If $\bar{X}_k$ is large, then the speech signal is probably present. But to prevent the noise from causing the large $\bar{X}_k$, we have to adjust the threshold η_k to control the probability of false alarm (FA), P_{FA}, with larger threshold values reducing P_{FA}. Over all possible detectors that have a given P_{FA}, this detector is optimal in that it yields the highest detection probability, P_D, for any value of $\lambda_{s,k}$, as long as $\lambda_{s,k} > 0$, This type of test, when it exists, is called a UMP test [6]. But the UMP test is not adopted in the case of $-\infty < \lambda_{s,k} < \infty$, which means that two-sided problems never produce the UMP tests. For the UMP test to exist, the parameter test must be one-sided test [6]. After finding the threshold based on the P_{FA}, the decision rule is established from the arithmetic mean of the magnitude and threshold of kth frequency bins, respectively, which is given by

$$\frac{1}{M}\sum_{k=1}^{M}\bar{X}_k \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{M}\sum_{k=1}^{M}\eta_k\ . \quad (15)$$

In order to determine the detection performance, the distribution for the $\bar{X}_k$ of the noisy spectral component should be found. The distribution of $\bar{X}_k$ for kth frequency bin can be derived as follows :

$$P(\bar{X}_k) = \frac{4\bar{X}_k}{\lambda_{x,k}}\exp\left(-\frac{2\bar{X}_k}{\sqrt{\lambda_{s,k}}}\right)U(\bar{X}_k) \quad (16)$$

where $U(\bar{X}_k)$ denotes the unit-step function. With this pdf, the distributions of the noisy spectral components conditioned on H_1 and H_0 are given by

$$p(\bar{X}_k|H_0) = \frac{4\bar{X}_k}{\lambda_{n,k}}\exp\left(-\frac{2\bar{X}_k}{\sqrt{\lambda_{n,k}}}\right)U(\bar{X}_k) \quad (17)$$

$$p(\bar{X}_k|H_1) = \frac{4\bar{X}_k}{\lambda_{s,k}+\lambda_{n,k}}\cdot \exp\left(-\frac{2\bar{X}_k}{\sqrt{\lambda_{s,k}+\lambda_{n,k}}}\right)U(\bar{X}_k)\ . \quad (18)$$

The key problem now becomes one of determining the optimal threshold in some best manner. Specifically, we describe how to find the threshold below.

3.2 Decision of Threshold

Since the threshold in (14) depends on the variances $\lambda_{s,k}$ and $\lambda_{n,k}$, it is very important to control the threshold according to the noise statistics to improve the detection performance. To find the threshold, we can use the false-alarm probability, P_{FA}, [6]. A false alarm is realized when the sum of magnitude of real and imaginary parts of the spectral component is larger than the threshold, given the null hypothesis is present. By considering the assumed pdf (17) under the null hypothesis, the false-alarm probability can be derived as

$$P_{FA} = Pr\left\{\bar{X}_k > \eta_k : H_0\right\} \quad (19)$$

$$= (\frac{2\eta_k}{\sqrt{\lambda_{n,k}}} + 1)\exp\left(-\frac{2\eta_k}{\sqrt{\lambda_{n,k}}}\right) . \quad (20)$$

By letting $\eta_k^{'} = 2\eta_k/\sqrt{\lambda_{n,k}}$, and rearranging terms, we have

$$\eta_k^{'} = -\ln(P_{FA}) + \ln(\eta_k^{'} + 1) . \quad (21)$$

To find the $\eta_k^{'}$, the fixed point iteration method [6] is used as follows :

$$\eta_{k,n+1}^{'} = -\ln(P_{FA}) + \ln(\eta_{k,n}^{'} + 1) . \quad (22)$$

For the given P_{FA}, we can find the final threshold $\eta_k^* = \eta_{k,N}^{'}\sqrt{\lambda_{n,k}}/2.0$ by iterating with $\eta_{k,0}^{'} = 1$ and the number of iteration, N. According to this result, we can discover that the threshold depends only on the noise statistics and false-alarm probability.

3.3 Discussion

The proposed UMP-based VAD approach is similar to that in [4]. [4] uses the Gaussian pdf for the speech signal while this scheme uses the Laplacian and the UMP test. Both methods requires the adaptive threshold which depends only on the noise statistics and the false-alarm probability. The UMP test-based decision rule derived from the Gaussian pdf is given by

$$|X_k| \;>\; \eta_k \quad (23)$$

where the threshold, $\eta_k = \sqrt{-2\lambda_{n,k}\ln(P_{FA})}$ [4]. This VAD method compares the magnitude of the noisy spectral component to a threshold η_k

In order to implement the proposed VAD algorithm for various noise environments, several factors such as the sum of magnitude, threshold, and the noise power need to be considered . It is well known that the smoothing parameter using a forgetting factor and a hang-over scheme improves the detection performance [1], [3]. Therefore, the $\bar{X}_k$ and threshold in (13) are modified to incorporate the forgetting factor schemes such that

$$\bar{X}_k(t) = (1 - \lambda_{\bar{X}_k}) \cdot \bar{X}_k(t) + \lambda_{\bar{X}_k} \cdot \bar{X}_k(t-1) \quad (24)$$

$$\hat{\eta}_k(t) = (1 - \lambda_{\eta_k}) \cdot \eta_k(t) + \lambda_{\eta_k} \cdot \hat{\eta}_k(t-1) \quad (25)$$

where $\lambda_{\bar{X}_k}$ and λ_{η_k} are the forgetting factors for the sum of magnitude of real and imaginary parts and threshold, respectively.

4 Experimental results

To verify the improved performance of proposed algorithms, we compared the speech detection and false-alarm probabilities (P_D and P_{FA}) for each VAD algorithm. We define P_D as the ratio of correct speech decisions to the hand-marked speech frames, while P_{FA} as that of false speech decisions to the hand-marked noise frames. To obtain P_D and P_{FA}, we made reference decisions for a clean speech material of 456 s long by labeling manually at every 10 ms frame. The percentage of the speech sounds is 58.2%, which consists of 44.8% voiced sounds and 13.4% unvoiced sounds. To make a noisy signal, we added the babble, factory and white noises from NOISEX-92 database [7] to the clean speech waveform with varying SNR. The VAD test was performed for each 10 ms frame.

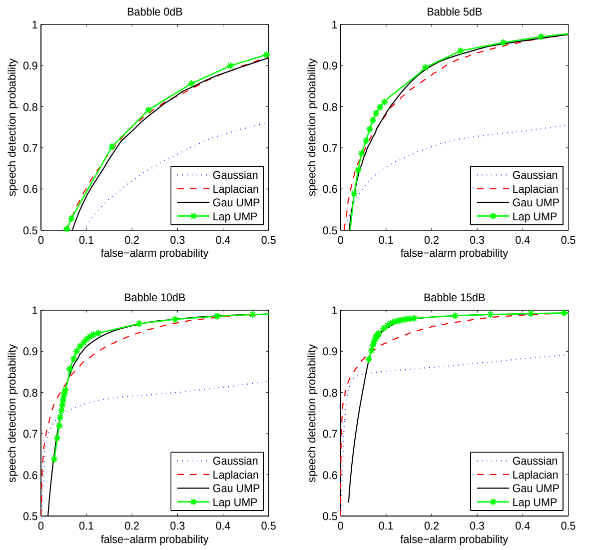

Figure 1: ROC curves for the Gaussian, Laplacian, Gaussian UMP and Laplacian UMP test approaches with the babble noise at 0 dB, 5 dB, 10 dB, 15 dB SNR, respectively.

The performance of each approach was measured in terms of the receiver operating characteristic (ROC) curve which shows the tradeoff between P_D and P_{FA}. The ROC curves are presented by figures comparing the proposed algorithm with the conventional algorithms such as Gaussian [1], Laplacian [2] and Gaussian UMP [4]. Figs. 1, 2 and 3 are ROC curves in babble, factory and white noises where the SNRs are 0 dB, 5 dB, 10 dB and 15 dB, respectively. The proposed VAD algorithm was implemented with the forgetting factor scheme while keeping $\lambda_{\bar{X}_k} = 0.9$ and $\lambda_{\eta_k} = 0.2$, through the various experiments. In the case of the babble and factory noises, which are shown in Figs.

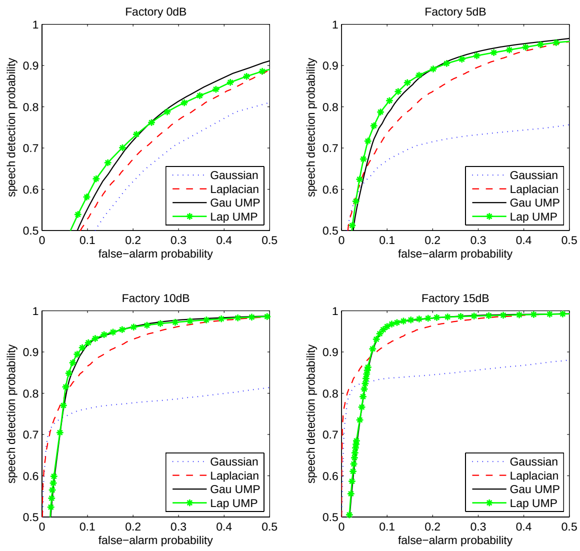

Figure 2: ROC curves for the Gaussian, Laplacian, Gaussian UMP and Laplacian UMP test approaches with the factory noise at 0 dB, 5 dB, 10 dB, 15 dB SNR, respectively.

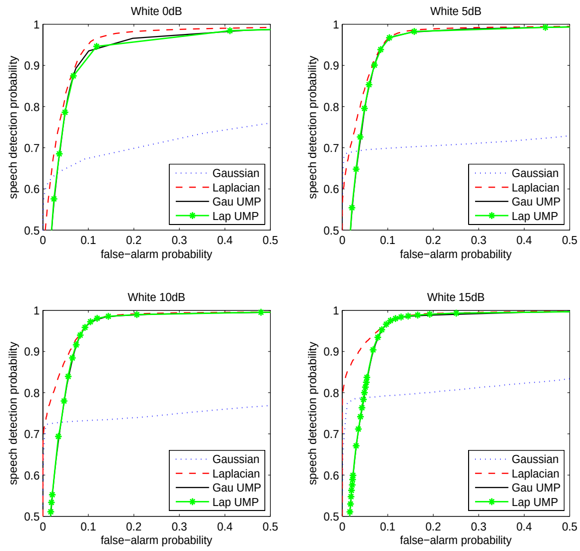

Figure 3: ROC curves for the Gaussian, Laplacian, Gaussian UMP and Laplacian UMP test approaches with the white noise at 0 dB, 5 dB, 10 dB, 15 dB SNR, respectively.

1 and 2, the UMP-based VAD algorithms outperformed the Gaussian and Laplacian approaches. In a high SNR, The ROC curves of UMP-based VAD tend to have lower performance than Gaussian-based and Laplacian-based methods in very small P_{FA} (below about 0.06) but we could confirm that UMP-based methods have much more excellent performance when $P_{FA} > 0.06$. From Fig. 3, which showed the results for the white noise, the UMP-based algorithms yielded a performance superior to the Gaussian method when $P_{FA} > 0.06$ at all SNRs while a similar performance to the Laplacian method. Also from the figures, we can find that the Laplacian-based UMP has a better performance improvement than the Gaussian-based UMP in a low SNR while it has a similar performance in a high SNR. From the experimental results, it is evident that the proposed method based on the UMP outperforms the conventional algorithm in the babble and factory noise environments when SNR is low.

5 Conclusion

We have presented a new VAD algorithm based on the Laplacian distribution and UMP test to detect the speech or nonspeech from the input noisy signal. We have applied the UMP test to find the new decision rule based on LRT as well as the distribution of the magnitude of the noisy spectral components. The efficient decision rule is derived to compare the arithmetic mean of the sum of magnitude of real and imaginary parts with the threshold which depends only on the noise statistics and the false-alarm probability. It has been found that the UMP test-based VAD algorithms using Laplacian outperformed the conventional algorithm in the environments such as babble and factory noise environments at a low SNR

References:

[1] J. S. Sohn, N. S. Kim and W. Y. Sung, A Statistical Model-Based Voice Activity Detection, *IEEE Signal Processing Lett.*, Vol. 6. No. 1, 1999, pp. 1-3.

[2] J. -H. Chang and N. S. Kim, Voice Activity Detection Based on Complex Laplacian Model, *IEE Electronics Lett.*, Vol. 39, No. 7, April 2003, pp. 632 - 634.

[3] J. -H. Chang, N. S. Kim and S. K. Mitra, Voice Activity Detection Based on Multiple Statistical Models, *IEEE Trans. Signal Processing*, Vol. 54, No. 6, June 2006, pp. 1965 - 1976.

[4] D. K. Kim, K. W. Jang and J. -H. Chang, A New Statistical Voice Activity Detection Based on UMP Test, *IEEE Signal Processing Lett.* (accepted), 2007.

[5] Y. Ephraim and D. Malah, Speech Enhancement Using A Minimum Mean-square Error Short-time Spectral Amplitude Estimator, *IEEE Trans. Acoust., Speech, Signal Proc.*, Vol. ASSP-32, Dec. 1984, pp. 1109-1121.

[6] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice Hall. 1998.

[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 Communication*, Vol. 12, No. 3, Jul. 1993, pp.247-251.