International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.6, No.6 (2013), pp.27-40
http://dx.doi.org/10.14257/ijsip.2013.6.6.03

A Real-Time DSP-Based System for Voice Activity Detection: Design and Implement

Xincheng Gao[1,2], Houbin Cao[1], Jianfeng Zhang[1, 2], Jinping Bai[1, 2], Tianhang Zhang[1] and Lihong Jia[1]

[1]*School of Aeronautic & Astronautics,*
[2]*Laboratory of Aerial Robotics,*
University of Electronic Science and Technology of China,
Chengdu, P.R.China

lar_uestc@163.com

Abstract

Most of the noise in speech communication lines can be considered as Gaussian white noise. Voice activity detection (VAD) in noisy environment is an important process in many speech signal processing algorithms. Unlike the other VAD algorithms, this paper proposes a simple and novel VAD algorithm based on the probability distribution function (PDF) of FFT magnitudes of both clean speech and Gaussian white noise. When the signal-to-noise ratio (SNR) is high enough, the method using Gamma distribution to detect the speech performs well, while the method using Rayleigh distribution under lower SNR can be complementary. In addition, the threshold to determine which method to use is presented based on the tests under different SNR. Simulation results show that the proposed algorithm is efficient. Both the hardware and software of a low cost system for VAD are introduced, with the proposed algorithm achieved in a digital signal processor (DSP). Each detection takes on less than 100 ms, which can be used for real-time processing.

Keywords: *VAD, FFT magnitudes, Gamma distribution, Rayleigh distribution, DSP*

1. Introduction

VAD plays an important role in speech signal processing algorithms, such as improving the speech intelligibility [1] and speech recognition [2]. Besides, as a real time VAD can distinguish the speech and noise on-line, it has been widely used in speech compression [3] and noise estimation [4].

Traditional VAD algorithms are typically based on the threshold of instantaneous or short term power, zero-crossing rate or spectral differences between the speech and noise. A neat summary of these strategies are found in [5]. In the last 20 years, with the developing models of speech and noise signals, VAD algorithms based on statistical models [6-16] have been proposed to get a higher performance than the traditional methods, *i.e.*, Gaussian distribution [6-7, 9, 15-16], Laplacian distribution [8-9] and Gamma distribution [9-13]. To be approximations of the real signal, these assumed statistical models give a deep comprehension of the speech and noise. Currently, algorithms based on statistical models have been the main trends for speech signal processing.

Among them, Sohn *et al.*, [6] assumed the speech signal to be complex Gaussian and proposed a VAD algorithm based on likelihood ratio test (LRT). Following this, numerous VAD algorithms [7-10, 15-16] have been proposed to further improve the VAD performance.

ISSN: 2005-4254 IJSIP

All of these algorithms used the LRT but modified models, such as generalized Gamma distribution model for clean speech spectra [10], Gaussian distribution model for the complex coefficients derived from complex exponential atomic decomposition of a signal [15] and Gaussian mixture model (GMM) for speech with transient noise [16], leading to modified and improved VAD outputs. As shown and widely accepted, a super-Gaussian model fits the speech signals better than Gaussian model. For instance, based on a two-sided generalized Gamma distribution, Almpanidis *et al.*, [12] modeled the speech signals and used the Bayesian Information Criterion to determine the boundaries of the speech in noisy environment. Unlike the above algorithms that using various models to fit the speech, Li Yu *et al.*, [14] assumed the power spectral density of noise to be Rayleigh distribution and proposed an adaptive threshold to detect voice activity indirectly. However, the computation complexities of these algorithms were both too large for real-time implementation.

Differing with the above algorithms that use only one model but cannot perform well in a large range of the SNR, this paper proposes an algorithm choosing statistical model on the basis of the transient SNR to detect voice activity with noise present in speech communication lines. The magnitudes of fast Fourier transform (FFT) of speech signal are assumed to follow Gamma distribution. The noise is regarded as Gaussian white noise, with its FFT magnitudes assumed to be Rayleigh distribution. The proposed algorithm is based on three simple observers and the computational complexity is small enough for real time voice activity detecting.

This paper is organized as follows. Section 2 describes the principle of the proposed algorithm and gives simulation results and analysis using MATLAB. In Section 3, both hardware architecture and software implementation of a DSP-based system are introduced. Finally, remarks and conclusions are provided in Section 4.

2. Proposed Method for VAD

In this Section, VAD methods using Gamma and Rayleigh distribution are discussed respectively. As shown, the two methods are complementary with each other under different SNR conditions. Three observers are set in this Section. The first two observers are used for VAD and the 3rd one is used for choosing a suitable method. All of the simulations and tests are achieved in MATLAB 2011A and each frame is selected to have 256 points.

2.1. VAD Using Gamma Distribution

In this paper, the speech and noise are considered independent, and an additive-noise signal model is simplified as:

$$y(k,m) = s(k,m) + n(k,m) \tag{1}$$

Where $y(k,m)$, $s(k,m)$ and $n(k,m)$ denote the m-th sample in the k-th frame of the noisy speech, clean speech and noise respectively.

Implement FFT for $y(k,m)$, we have:

$$Y(k,m) = S(k,m) + N(k,m) \tag{2}$$

Where $Y(k,m)$, $S(k,m)$ and $N(k,m)$ are the FFT magnitudes of $y(k,m)$, $s(k,m)$ and $n(k,m)$.

Comparing the Gaussian with Gamma distribution assumptions, we use a section of clean speech from TIMIT database to test the histogram of FFT magnitudes. As illustrates in Figure

1, the estimated PDF of Gamma distribution fit the true signal better than that of Gaussian distribution.

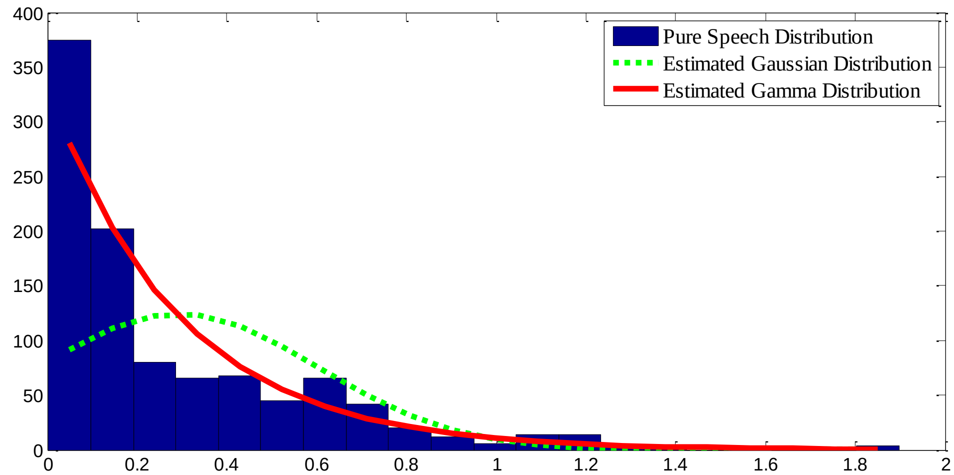


Figure 1. Estimated Distributions for Speech Signal

Assumed as Gamma distribution, the FFT magnitudes of clean speech have a PDF as:

$$f_{Gamma}(S(k,m);\alpha,\beta)=\frac{\beta^{\alpha}S(k,m)^{\alpha-1}}{\Gamma(\alpha)}e^{-\beta S(k,m)},S(k,m)>0 \quad (3)$$

Where $\Gamma(\cdot)$ denotes the Gamma function, α and β are shape parameter and scale parameter of the PDF.

Find the maximum likelihood estimations (MLE) of α and β. For n samples, the log-likelihood function is defined as:

$$l(\alpha,\beta)=\ln L(\alpha,\beta)=(\alpha-1)\sum_{m=1}^{n}\ln S(k,m)+n\alpha\ln\beta-n\ln\Gamma(\alpha)-\beta\sum_{m=1}^{n}S(k,m) \quad (4)$$

Thus, the maximum likelihood estimations [17] are:

$$\hat{\alpha}\approx\frac{\sqrt{(3-t)^2+24t}+3-t}{12t} \quad (5)$$

$$\hat{\beta}=\frac{n\hat{\alpha}}{\sum_{m=1}^{n}S(k,m)} \quad (6)$$

Where

$$t=\ln\sum_{m=1}^{n}\frac{S(k,m)}{n}-\frac{\sum_{m=1}^{n}\ln S(k,m)}{n}$$

Set $\hat{\alpha}$ as the first observer:

$$H_0:\hat{\alpha}>threshold \text{ , speech absent}$$

$$H_1 : \hat{\alpha} \leq threshold \text{ , speech present}$$

As shown in [13], the shape parameter of Gamma-distributed speech signal is less than one. Once the voice is non-active, the MLE of α increases to be more than 1, because the FFT magnitudes distribution of the Gaussian white noise is Rayleigh distribution, as shown in Figure 2.

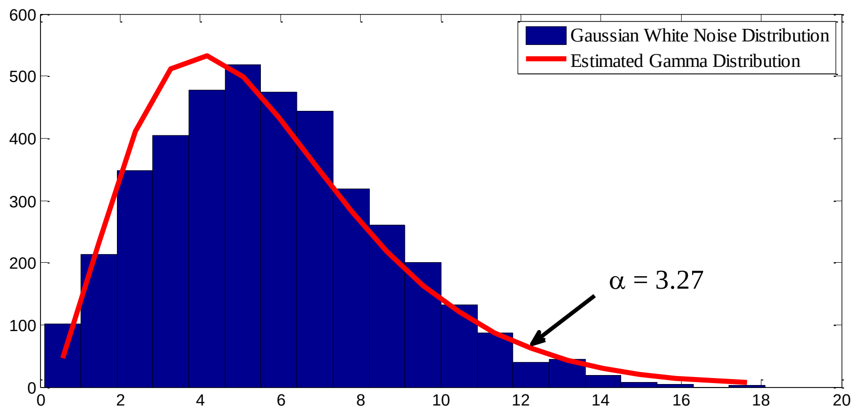


Figure 2. Estimated PDF of Gamma Distribution when Voice is Non-Active

To examine the observer, non-stationary noise is added to a speech signal. Simulation results are shown in Figure 3 and Figure 4.

Under high SNR, the proposed method based on the first observer performs well in non-stationary noise (as in shown in Figure 3), but misses some voice active sections when SNR decreases (as shown in Figure 4). As a result, this method is limited to be used under high SNR.

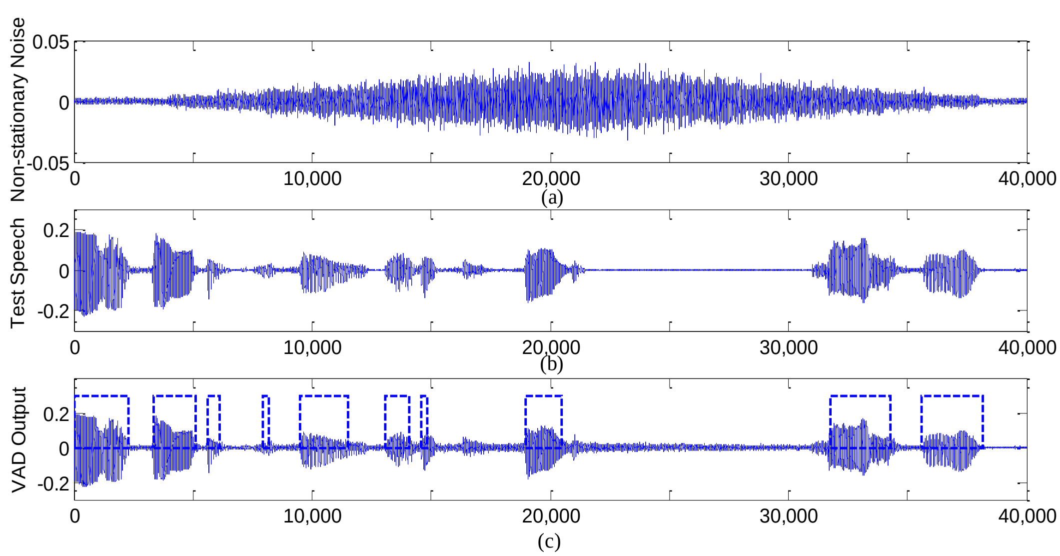


Figure 3. (a) The Additive Non-stationary, Low Level Noise. (b) The Test Speech. (c) The VAD Output on Noisy Speech

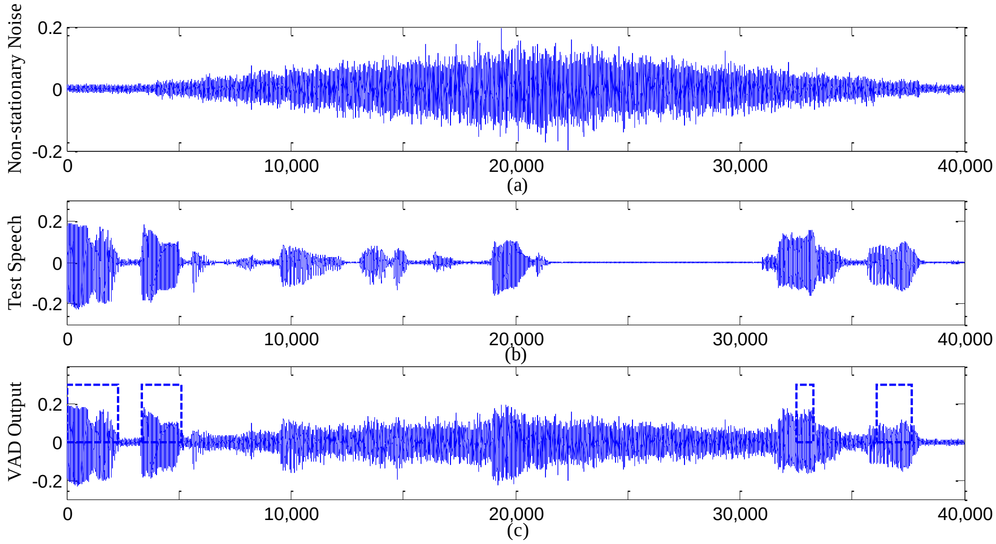


Figure 4. (a) The Additive Non-stationary, High Level Noise. (b) The Test Speech. (c) The VAD Output on Noisy Speech

2.2 VAD Using Rayleigh Distribution

The proposed method presented in Section 2.1 is unable to fight against with a low SNR. That is due to the misalignment of the Gamma distribution model and noisy speech. As the noise in speech communication lines is assumed to be Gaussian white, the real and image components of the FFT coefficients are assumed independent and identically distributed (i.i.d). Hence, $N(k,m)$ follows a Rayleigh distribution [18] (as shown in Figure 5), that is:

$$N(k,m) = \sqrt{N_R(k,m)^2 + N_I(k,m)^2} \tag{7}$$

Where $N_R \sim N(0,\sigma^2)$, $N_I \sim N(0,\sigma^2)$, $N(k,m) \sim Rayleigh(\sigma)$.

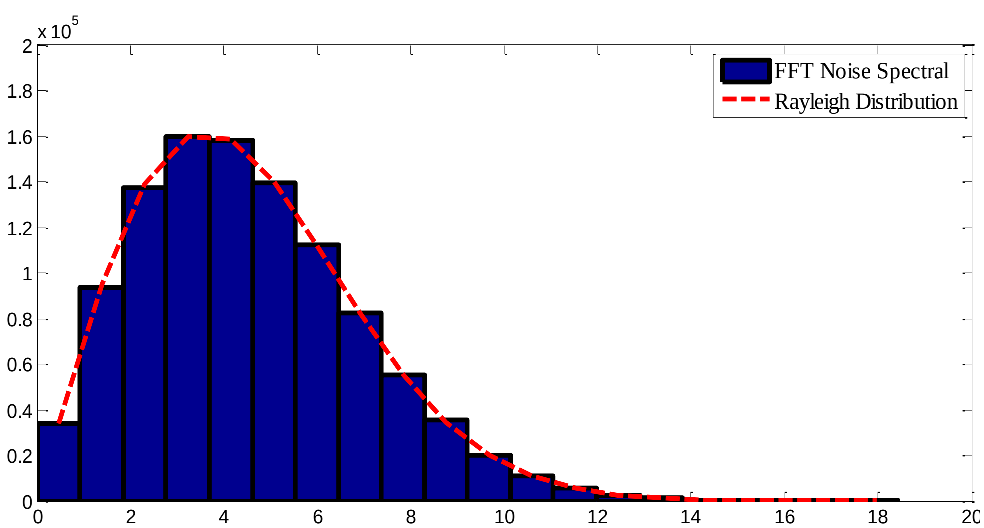


Figure 5. Estimated Rayleigh Distribution of Noise FFT Magnitudes

Therefore, the FFT magnitudes of noise have a PDF as:

$$f_{Rayleigh}(N(k,m);\sigma)=\frac{N(k,m)}{\sigma^2}e^{\frac{-N(k,m)^2}{2\sigma^2}} \tag{8}$$

For n samples, the MLE of σ is given [19]:

$$\hat{\sigma}=\sqrt{\frac{\sum_{m=1}^{n}N(k,m)^2}{2n}} \tag{9}$$

The mathematical expectation of the FFT magnitudes of n samples can be computed using:

$$\hat{E}[N]=\hat{\sigma}\sqrt{\frac{\pi}{2}} \tag{10}$$

Because the mean of the samples is the unbiased estimation of the mathematical expectation, when the voice is non-active, the mean of the FFT magnitudes must be approximately equal to the estimated expectation, that is:

$$r(N)=\frac{\hat{E}[N]}{\overline{N}}=\frac{\hat{\sigma}\sqrt{\frac{\pi}{2}}}{\frac{1}{n}\sum_{m=1}^{n}N(k,m)}=\frac{\sqrt{n\pi\sum_{m=1}^{n}N(k,m)^2}}{2\sum_{m=1}^{n}N(k,m)}\approx 1 \tag{11}$$

Hence, we set $r(N)$ as the second observer:

$$H_0 : r(N) < threshold \text{, speech absent}$$

$$H_1 : r(N) \geq threshold \text{, speech present}$$

An off-line test is done to find the appropriate value and the threshold is set to be 1.1. To test the proposed method, as doing in Section 2.1, both high level and low level non-stationary noise are added to the test speech signal. Simulation results are shown in Figure 6 and Figure 7.

As shown in Figure 6, this method performs well in a low level noise. But the results are not reliable, compared with the results shown in Figure 3. Most of voice active regions detected in Figure 6 have included a little voice non-active intervals, namely, this method can mistake the voice non-active region for active region. However, as shown in Figure 7, when the SNR is lower, the voice active region can be indirectly detected using Rayleigh distribution to detect which region is voice non-active. The method performs well unless the voice has submerged in the noise completely.

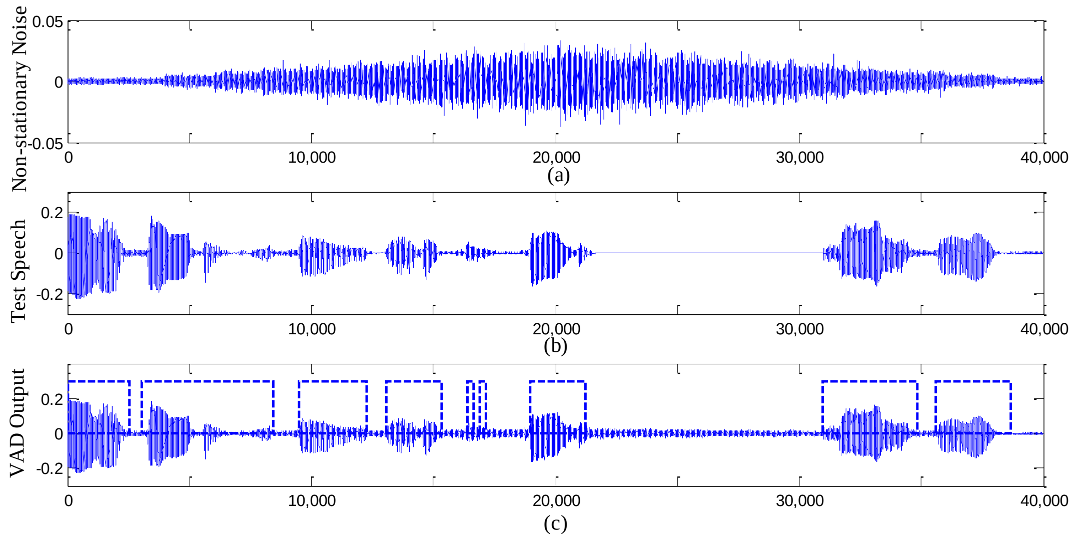


Figure 6. (a) The Additive Non-stationary, Low Level Noise. (b) The Test Speech. (c) The VAD Output on Noisy Speech

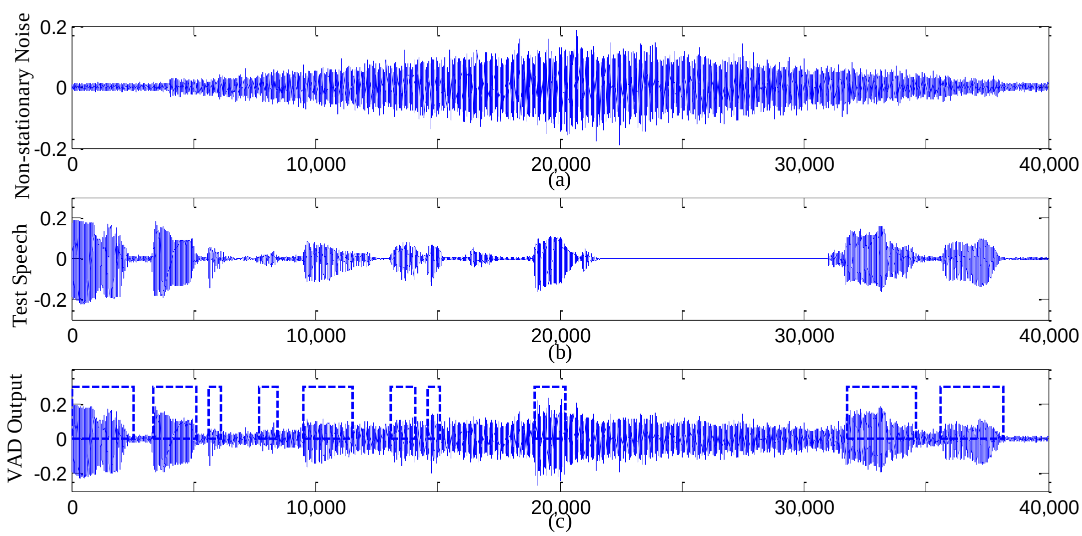


Figure 7. (a) The Additive Non-stationary, High Level Noise. (b) The Test Speech. (c) The VAD Output on Noisy Speech

2.3. VAD Using the Both Gamma and Rayleigh Distribution

Methods in Section 2.1 and Section 2.2 turn out to be complementary. The first one performs well under high SNR and the second one can be used under low SNR. To integrate the two methods can improve the proposed algorithm to be applied to a large range of SNR. Then, the problem comes to be the threshold to determine which method to use.

Another test is done to find the threshold. We select a section of the speech randomly, add noise in different levels and note the maximum size of the detected voice active region. The true value is about 2000, and the performances of the two methods are shown in Figure 8.

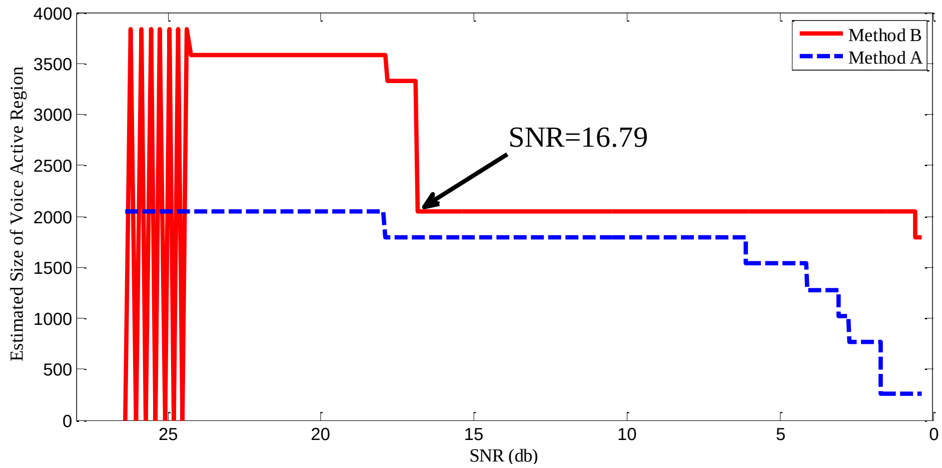


Figure 8. Performance Comparison of the Two Methods

In Figure 8, SNR with a range for 0 db to 30 db is shown. Method A denotes the method proposed in Section 2.1 and Method B is the method in Section 2.2. As is presented, Method A is stable and close to the true value when SNR is high (more than 18 db), but misses more and more voice active sections when the SNR decreases. On the contrary, Method B is not stable when SNR is more than 24 db. When SNR is between 16.79 db and 24 db, the results seem to be stable, but much bigger than the true value. This means that the VAD outputs contain voice non-active sections. However, when SNR is under 16.79 db, the VAD outputs become close to the true value and stable.

In this paper, we set the threshold to be 16.79 db. Note that 16.79 is not the unique solution, because the performances of the two methods under a SNR between 5 db to 16.69 db do not have big differences.

The simulation using both Gamma and Rayleigh distribution is shown in Figure 9. To implement the method, a priori SNR needs to be estimated. To reduce the computational complexity, in this paper, we only use the latest samples when voice is non-active to estimate the power of the noise. The transient SNR (TSNR) is computed using:

$$TSNR = 10\log \frac{\left| \sum_{m=1}^{n} (s(k,m) - \overline{s(k)})^2 - \hat{P}_n \right|}{\hat{P}_n} \tag{12}$$

Where $\overline{s(k)}$ is the mean of the current signal samples, $\hat{P}_n$ denotes the priori estimated power of noise.

Then, the third observer is:

$H_0 : TSNR < 16.79$, use Rayleigh distribution model of noise

$H_1 : TSNR \geq 16.79$, use Gamma distribution model of speech

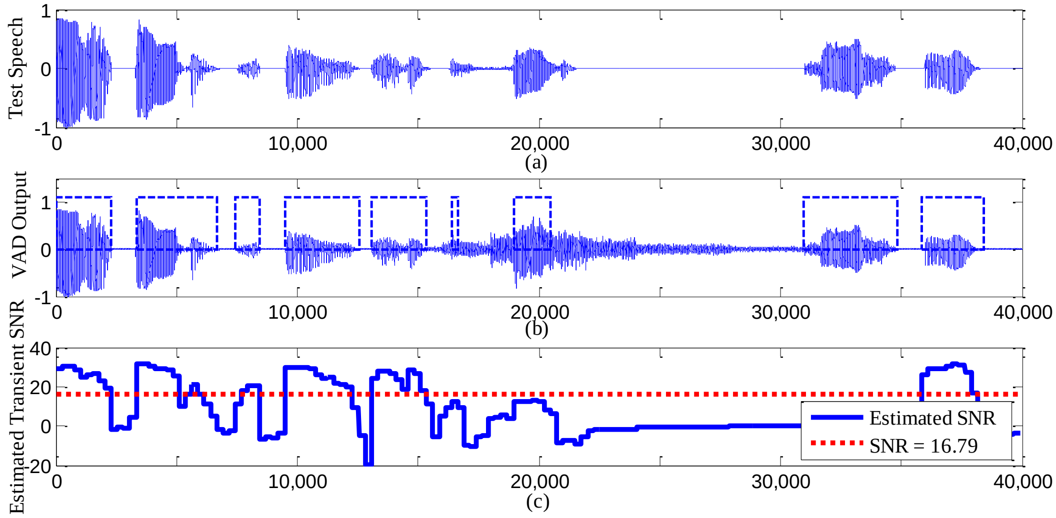


Figure 9. VAD outputs Using Both Gamma and Rayleigh Distribution

3. Implementation of the Proposed Method

In this section, a DSP-based system is introduced to implement the proposed algorithm. The algorithm is achieved in a DSP processor. Both hardware architecture and register set up are described in the section.

3.1. Design of Hardware Architecture

Taking the computing speed and cost into account, we use two low cost micro processors to control the system and achieve the proposed VAD algorithm. One of the two micro processors acts as the system control unit and another one is used as the kernel algorithm processing unit. The hardware architecture and the practicality diagram are shown in Figure 10 and Figure 11 respectively.

The system control unit (SCU) uses a 51-kernel micro control unit (MCU) STC89LE58RD+, which has up to 80M Hz system clock, 32K on-chip flash and a UART interface. The kernel algorithm processing unit uses a DSP TMS320VC5402, which has rich hardware resources, including a 16K×16-bit on-chip RAM, two multi-channel buffered serial ports (McBSPs), an enhanced 8-bit parallel host-port interface (HPI) and 10-ns single-cycle fixed-point instruction execution time (100 MIPS) but not an on-chip flash. To execute the program without a simulator, the 8-bit HPI provides the SCU an interface to load the code to the dual random memory (DRAM), which is called HPI boot method [20].

A level translator MAX3232 is used for downloading the code to SCU via DB9 from the software named STC-ISP. The two McBSPs of TMS320VC5402 are used for data collection (McBSP1) and VAD outputs (McBSP0). To record the important data, a 64×16-bit static random memory (SRAM) is used to exchange data with TMS320VC5402.

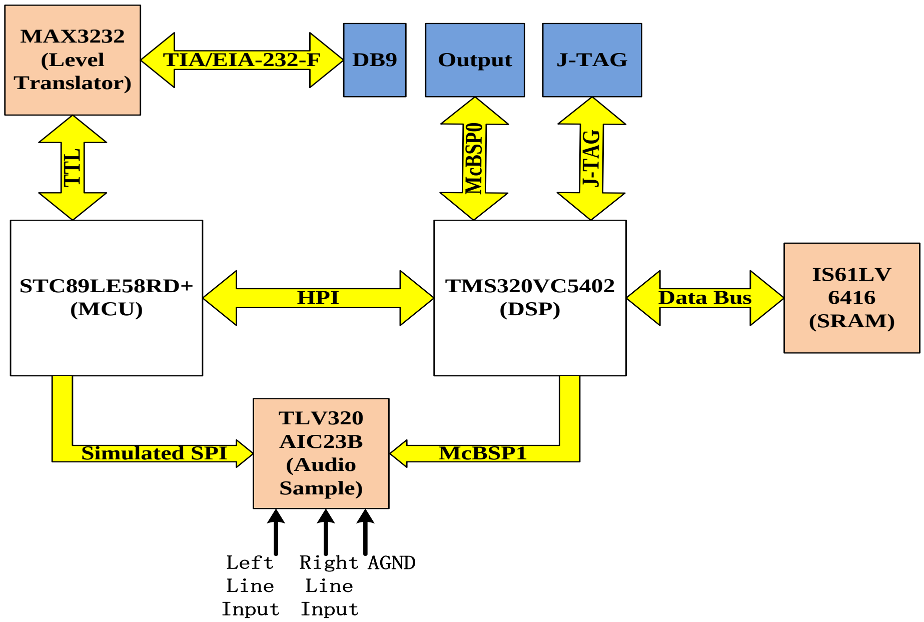
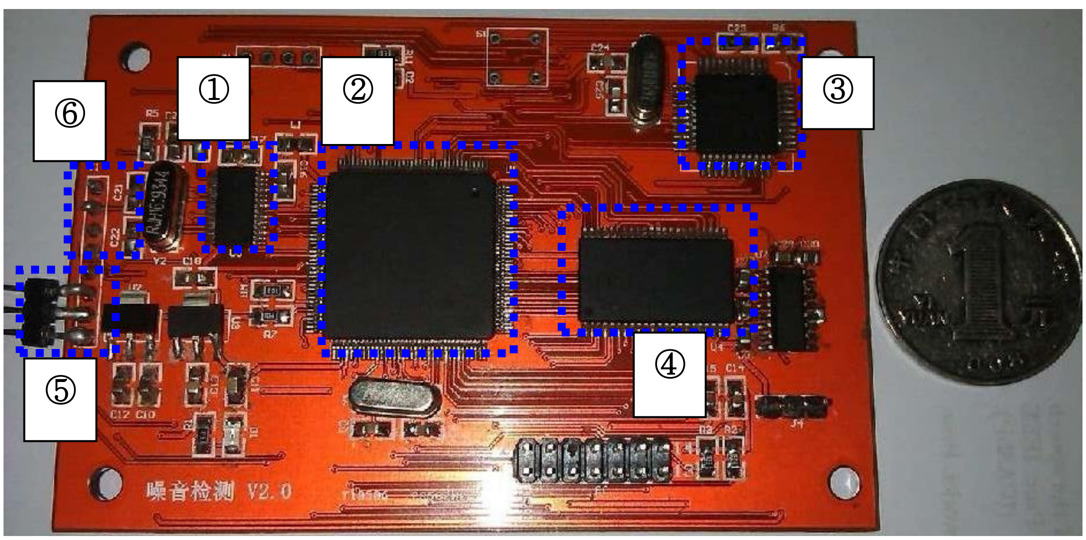


Figure 10. Hardware Architecture of the System for Real-Time VAD



①: TLV320AIC23B ②: TMS320VC5402 ③: STC89LE58RD+ ④: IS61LV6417
⑤: Speech Input Interface ⑥: VAD Output Interface

Figure 11. The Practicality Diagram of the System for Real-Time VAD

In order to sample the speech signal, a high-performance stereo audio codec TLV320AIC23B is used in the designed system. TLV320AIC23B provides two line inputs (left line and right line), but the right line input is reserved here. As the two McBSPs have been used, we use four general purpose input/output (GPIO) interfaces of STC89LE58RD+ as simulated serial peripheral interface (SPI) to configure TLV320AIC23B.

3.2. Software Design and Implementation

The proposed method in this paper contains a lot of float point operations, such as log functions and exacting roots. However, TMS320VC5402 is a fixed-point DSP. When computing, truncation errors even fault results may come out if not well treated. A method called Q.15 format [21] can do float point operations in fixed-point DSP, at the cost of abandoning the relatively smaller value. Similarly, in this paper, the float-point results are multiplied by 10,000 and stored in long int type, and divided by 10,000 when used. This is simple, efficient and easy to be realized, but less accurate than Q.15 format.

The code achieved in DSP is compiled in Code Composer Studio V3.3, and the code achieved in MCU is compiled in Keil uvision2. The flow chart of the algorithm in DSP is shown as Figure 12.

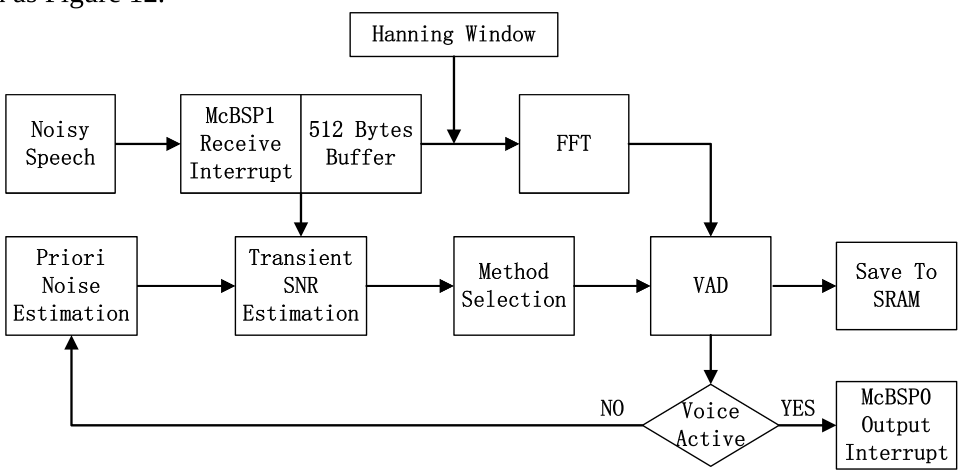


Figure 12. The Flow Chart of Software in DSP

Data collection, VAD algorithm operations and VAD outputs are implemented in the receive interrupt handler of McBSP1. TLV320AIC23B is configured to work in master mode and DSP data format so that it can provide the communication clock and connect easily with TMS320VC5402. The main configurations of TMS320VC5402 and TLV320AIC23B are listed in Table 1 and Table 2.

Table 1. The Main Configuration of TMS320VC5402

Register	Value	Remark
Clock mode register	0x40c7	System clock is 100M Hz
Software wait-state register	0x4240	2 wait except for on-chip program
Processor mode status register	0x00a0	MP/MC = 0, IPTR = 001,ovly=1
Bank-switching control register	0x0802	PS-DS = 1 and Bus holder enable
Serial port control register (SPCR11)	0x0001	The serial port McBSP1 receiver is enabled
Serial port control register (SPCR21)	0x0201	Free running mode and the serial port McBSP1 transmitter is enabled
Pin control register (PCR1)	0x0001	Receive data sampled on rising edge of CLKR
Receive control register (RCR11)	0x0140	16 bits per word and 2 words per frame
Interrupt Mask Register	0x0400	Enable McBSP1 Receive Interrupt

Table 2. The Main Configuration of TLV320AIC23B

Register	Value	Remark
Left Line Input Channel Volume Control	0x013	Default volume
Analog Audio Path Control	0x002	ByPass Disabled, Line input for ADC, microphone muted
Power Down Control	0x00A	All On but DAC and microphone input Off
Digital Audio Interface Format	0x043	Master Mode, DSP data format, the input bit length is set to be 16-bit
Sample Rate Control	0x02e	8.021K Hz Sample rate, Normal clock mode and 384 times oversampling
Digital Interface Activation	0x001	Active interface

In addition, to avoid trigonometry in FFT, we generate a sine table of 5/4 cycles (640 points) off-line. Based on our tests, each complete operation of the proposed algorithm can be done in less than 100 ms, which is suitable for real-time VAD.

4. Remarks and Conclusions

In this paper, Gamma distribution for speech signal and Rayleigh distribution for noise are both used for voice activity detection. Results show that the two methods are complementary with each other. Based on this principle with three observers, this paper proposes a simple but efficient VAD algorithm. What's more, to implement the proposed algorithm, a real-time DSP-based system is introduced with hardware architecture and software flow described in details. The proposed algorithm is simple, fast and has small computational complexity, performs well in a large range of SNR.

It should be noted that, the estimation of noise is executed during the interval of voice non-active. This is reasonable, because there must exist a short interval of voice non-active when a person gives a speech. During the short interval, noise will not change greatly in practical. Hence, the estimated TSNR in this paper is priori, but meaningful.

Acknowledgements

This work is partially supported by the National Training Programs of Innovation and Entrepreneurship for Undergraduates of UESTC. Engineers in Chengdu LinKon Communication Device Co., Ltd help us a lot in hardware design and functional test. In addition, the authors would like to thank Yong Zeng and Feng Li from the school of Aeronautic & Astronautics of UESTC for their constructive discussions.

References

[1] B. Cornelis, M. Moonen and J. Wouters, "Speech intelligibility improvements with hearing aids using bilateral and binaural adaptive multichannel Wiener filtering based noise reduction", J Acoust Soc. Am, vol. 131, no. 6, **(2012)**, pp. 4743-4755.

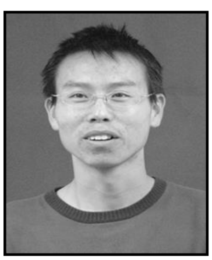
[2] A. Narayanan and D. Liang Wang, "The role of binary mask patterns in automatic speech recognition in background noise". J. Acoust. Soc. Am, 133, 5 **(2013)**, pp. 3083-3093.

[3] W. Q. Syed and Hsiao-Chun Wu, "Speech waveform compression using robust adaptive voice activity detection for nonstationary noise", EURASIP", Journal on Audio, Speech, and Music Processing, **(2008)**.

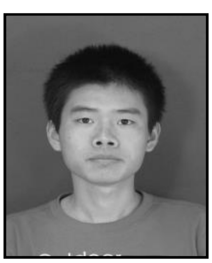
[4] A. R. Fukane and S. L. Sahare, "Noise estimation algorithms for speech enhancement in highly non-stationary environments", International Journal of computer science Issues, vol. 8, no. 2, **(2011)**.

[5] H. Zheng and D. Li, "Research of voice activity detection algorithm", Proceedings of Computational and Information Sciences (ICCIS), **(2011)** October 21-23; Chengdu, China.
[6] J. Sohn, N. S. Kim and W. Sung, "A statistical model-based voice activity detection", Signal Processing Letters, IEEE, vol. 6, no. 1 **(1999)**, pp. 1-3.
[7] D. K. Kim, K. W. Jang and J.-H. Chang, "A new statistical voice activity detection based on UMP test", Signal Processing Letters, IEEE, vol. 14, no. 11, **(2007)**, pp. 891-894.
[8] J.-H. Chang, J. W. Shin and N. S. Kim, "Likelihood ratio test with complex laplacian model for voice activity detection", Proceedings of 8th European Conference on Speech Communication and Technology, Eurospeech, Geneva, Switzerland, **(2003)** September 1-4.
[9] H. Farsi, M. A. Mozaffarian and H. Rahmani, "Modifying voice activity detection in low SNR by correction factors", Proceedings of the 3rd WSEAS International Conference on Circuit, System, Signal and Telecommunications (CISST), Athens, Greece, **(2009)** December, pp. 29-31.
[10] J. W. Shin, J.-H. Chang, H. S. Yun and N. S. Kim, "Voice activity detection based on generalized gamma distribution", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, PA, USA, **(2005)** March, pp. 18-23.
[11] T. H. Dat, K. Takeda and F. Itakura, "Generalized gamma modeling of speech and its online estimation for speech enhancement", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, PA, USA, **(2005)** March, pp. 18-23.
[12] G. Almpanidis and C. Kotropoulos, "Voice activity detection using generalized gamma distribution", Proceeding of IEEE International Conference on Multimedia and Expo, Toronto, Ont., **(2006)** July, pp. 9-12.
[13] N. Mohammadiha, R. Martin and A. Leijon, "Spectral domain speech enhancement using HMM-state dependent super-Gaussian priors", Signal Processing Letters, IEEE, vol. 20, no. 3, **(2013)**, pp. 253-256.
[14] Y. Li, J. Chen and H. Tan, "Voice activity detection under Rayleigh distribution", Journal of electronics (China), vol. 26, no. 4, **(2009)**, pp. 552-556.
[15] S. Deng and J. Han, "Voice activity detection based on conjugate subspace matching pursuit and likelihood ratio test", EURASIP Journal on Audio, Speech, and Music processing, **(2011)**.
[16] S. Mousazadeh and I. Cohen, "Voice activity detection in presence of transient noise using spectral clustering", IEEE Trans on Audio, Speech, and Language Processing, vol. 21, no. 6, **(2013)**, pp. 1261-1271.
[17] T. P. Minka, "Estimating a Gamma distribution", Microsoft Research, Cambridge, UK, Tech. Rep, **(2002)**.
[18] D. Lu, "Stochastic Process and Application", (Chinese), TsingHua University Press, Beijing, **(1986)**, pp. 480-481.
[19] J. Sijbers, A. J. den Dekker, E. Raman and D. Van Dyck, "Parameter estimation from magnitude MR images", International Journal of Imaging Systems and Technology, vol. 10, no. 2, **(1999)**, pp. 109–114.
[20] Texas Instruments Inc, "Bootloading the TMS320VC5402 in HPI mode", Dallas:Texas Instruments, **(2002)**.
[21] W.-S. Gan and S. M. Kuo, "Teaching DSP software development: from design to fixed-point implementations", IEEE Transactions on Education, vol. 49, no. 1, **(2006)**, pp. 122-131.

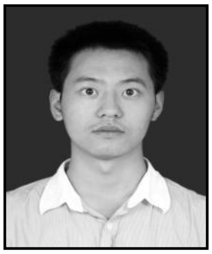
Authors



Xincheng Gao was born in Nantong, China in 1989 and received his bachelor's degree from UESTC, China in 2011. Recently, he is a postgraduate in the School of Aeronautic & Astronautics and Laboratory of Aerial Robotics of UESTC. His current research interests are digital signal processing, speech recognition and embedded systems.



Houbin Cao received his bachelor's degree in Mechatronics Engineering form Heze University ,China in 2010. He is currently studing for a master's degree in UESTC. His research interests are digital signal process, speech recognition and speech enhancement.



Jianfeng Zhang was born in Neijiang, China in 1990 and received his bachelor's degree from UESTC, China in 2011. Recently, he is a postgraduate in the School of Aeronautic & Astronautics and Laboratory of Aerial Robotics of UESTC. His research interests are digital signal processing, low cost GPS/INS navigation and embedded systems.

Jinping Bai is an associate professor in the School of Aeronautic & Astronautics of UESTC. Also, he is the leader of Laboratory of Aerial Robotics. In addition, he is a senior software engineer and an expert of Chinese Association for Quality (CAQ). His current research interests are including asseseement of systems dependability, integration of intelligent system, pattern recognition, digital image process and Internet of Things (IOT).

Tianhang Zhang was born in Chaohu, China in 1990 and received his bachelor's degree from AHUT, China in 2011. Recently, he is a postgraduate in the School of Aeronautic & Astronautics. His current research interests are system simulation and pattern recognition.

Lihong Jia was born in Shanxi, China in 1988 and received her bachelor's degree from Zhengzhou University in 2011. Recently, she is a postgraduate in the School of Aeronautic & Astronautics of UESTC. Her research interests are digital signal process, digital image processing and pattern recognition.

