음성 활동 구간 검출을 위한 스펙트랄 엔트로피의 재구성 효과

권호민*, 한학용*, 이광석**, 고시영***, 허강인*
동아대학교*, 경일대학교**, 진주산업대학교***

Reconstruction Effect of the Spectral Entropy for the Voice Activity Detection

Ho-Min Kwon*, Hag-Yong Han*, Kwang-Seok Lee**, Si-Young Koh***, Kang-In Hur*

Dept. of Electronics, Dong-A University*
Dept. of Electronic & Information Engineering, Kyung-il University***
Dept. of Electronics, JinJu National University**
E-mail : kihur@mail.donga.ac.kr

Abstract – Voice activity detection is important problem in the speech recognition and communication. This paper introduces feature parameter which is reconstructed by the spectral entropy of information theory for the robust voice activity detection in the noise environment, analyzes and compares it with the energy method of voice activity detection and performance. In experiment, we confirmed that the spectral entropy is more feature parameter than the energy method for the robust voice activity detection in the various noise environment.

I. INTRODUCTION

VAD(Voice Activity Detection) which can exert the decisive effect on the speech recognition rate is the important pre-processing of speech recognition and communication. VAD detects the real speech section among the various noise and sounds. However, the implementation of VAD which is independent and stable in every sounds is difficult work.
Common VAD is based on the energy method[1][3]. In this method, it is only suitable in the good environment without noise. Specially, at the start point of the speech section, to detect consonants and vowels which have low energy is difficult. Furthermore, a cough sound of the outside and additional noise of breathing at the start and end point of the speech section should be removed. For those problems, commonly, when the sound section is longer than threshold holding time of speech, exceeding threshold value of short time average energy, speech activity, is detected and the start point of speech is located ahead of the particular time from the detected point of energy threshold value. Furthermore, this method is used with zero crossing for the more trusty VAD[4]. Another VAD method which use the spectral analysis detect the voice activity by using the difference between the input signal and reference noise spectrum.

In this paper, it is used the spectral entropy which is based on the entropy, the principle concept of information theory. This method is firstly used by J.L. Shen in speech processing for the first time. Through his experiment Shen showed a lot of difference between speech and non-speech of the spectral entropy[5].
In order to complement the spectral entropy, this

paper suggests new feature parameter which is reconstructed by the spectral entropy, analyzes and compares the suggested new feature with energy method, and confirm the application possibility of VAD parameter in various noise environment.

Ⅱ. ENTROPY

2.1. Entropy

Entropy which is based on the Shannon's information theory is the scale measuring the amount of information. According to information theory, the information derivable from outcome x_i depends on its probability. If the probability $P(x_i)$ is small, we can derive a large degree of information, because the outcome that is has occurred is very rare. On the other hand, if the probability is large, the information derived will be small, because the outcome is well expected. Thus, the amount of information is defined as follows:

$$I(x_i) = \log \frac{1}{P(x_i)} \qquad (1)$$

Suppose X is a discrete random variable taking value x_i(referred to as a symbol)from a finite or countable infinite sample space $S=\{x_1,x_2,...,x_i,...\}$(referred to as a symbol). The symbol x_i is produced from an information source with alphabet S, according to the probability distribution of the random variable X. One of the most important properties of an information source is the entropy H(S) of the random variable X, defined as the average amount of information (expected information):

$$\begin{aligned} H(X) &= E[I(X)] \\ &= \sum_S P(x_i)I(x_i) \\ &= \sum_S P(x_i)\log\frac{1}{P(x_i)} \qquad (2) \\ &= E[-\log P(X)] \end{aligned}$$

2.1 Spectral entropy

Spectral entropy process is consistec of calculating FFT of input signal, probability density of the power spectrum in the band limited speech signal, and entropy. The probability density of the spectrum is estimated in a method that has the normalization effect of kinds of frequency components.

$$p_i = \frac{s(f_i)}{\sum_{k=1}^{M} s(f_k)}, \qquad i=1...M$$

where, s(fi) is power spectrum of the frequency component fi, pi is the corresponding probability density, and N is the total number of frequency components in FFT.

Next step is calculating the entropy. However, the above process emphasize the entropy of the noise and non-speech section, therefore, we can emphasize the speech section through the entropy conversion, and then, last step, estimated entropy is reconstructed.

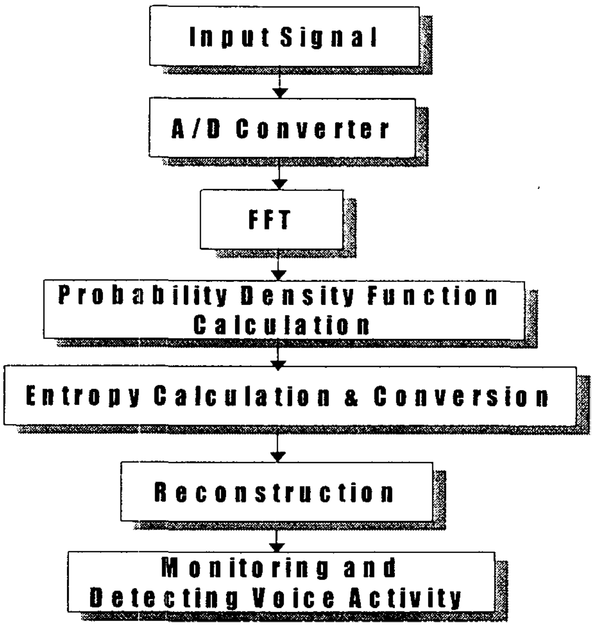

Input Signal
A/D Converter
FFT
Probability Density Function Calculation
Entropy Calculation & Conversion
Reconstruction
Monitoring and Detecting Voice Activity

Figure 1. The process of Spectral Entropy

2.2 Reconstruction of Spectral Entropy for VAD

Reconstruction of spectral entropy gives margin setting threshold value and stresses speech section only. In this paper, there are various feature parameters for VAD as follows.

Feature 1 : Entropy
Feature 2 : Entropy×Log Energy
Feature 3 : Entropy×(ZCR MAX - ZCR)
Feature 4 : Entropy×Log Energy×(ZCR MAX - ZCR)
Feature 5 : Entropy×gaussian distribution function of Speech Entropy
Feature 6 : Entropy×Gaussian distribution function of Speech Entropy × Log Energy

Gaussian distribution function of Feature 5, 6 which is applied by gaussian distribution function of spectral entropy and made by speech sample previously is to emphasize speech section. The applied gaussian distribution function is followed.

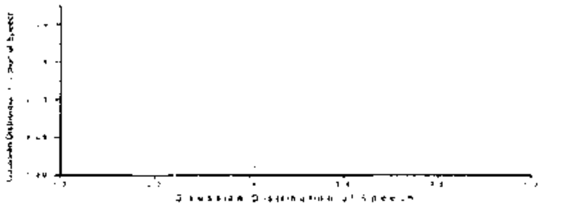


Figure 2 Gaussian Distribution of Speech Entropy

Ⅲ. EXPERIMENTAL RESULTS

3.1 Data Base

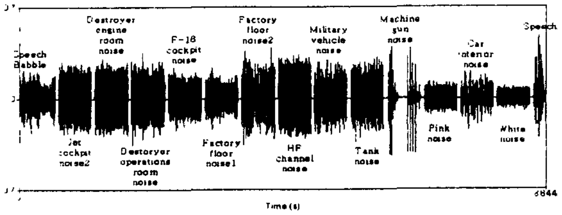
In the experiment this paper uses the NOISE-92[8] database which has the various noise, 19.98 KHz - 16 bit and filtering anti-aliasing, and that is changed to 16 KHz - 16 bit.

	Noisex-92
1	Speech babble
2	Jet cockpit noise2
3	Destroyer engine room noise
4	Destroyer operations room noise
5	F-16 cockpit noise
6	Factory floor noisel
7	Factory floor noise2
8	HF channel noise
9	Military vehicle noise
10	Tank noise
11	Machine gun noise
12	Pink noise
13	Car interior noise
14	White noise

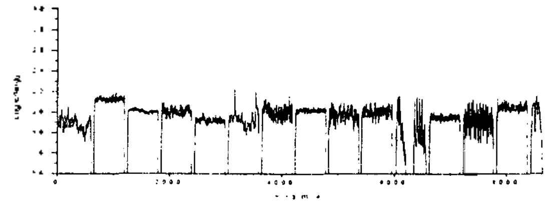
Table 1 NOISEX-92

3.2. Experiment result and inquiry

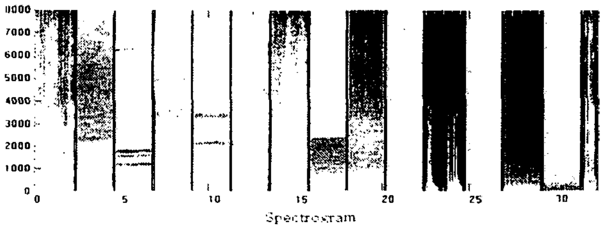
Figure 3 is sample data for estimating, which compares above-mentioned features, section 2.2, with energy.



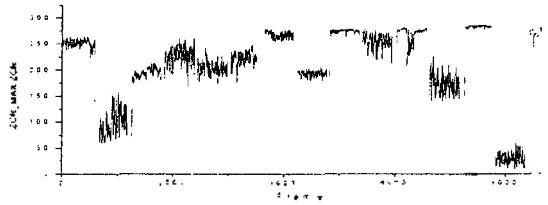
(a) Source Signal



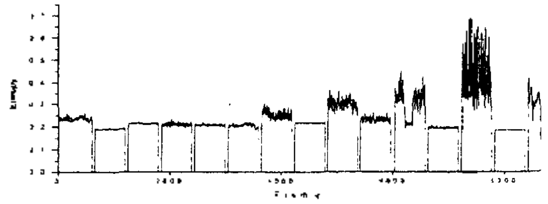
(b) Log Energy



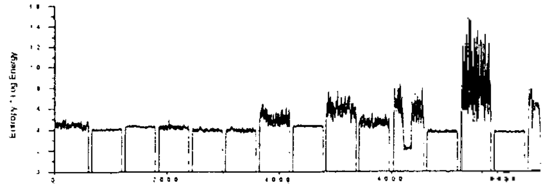
(c) Spectrogram



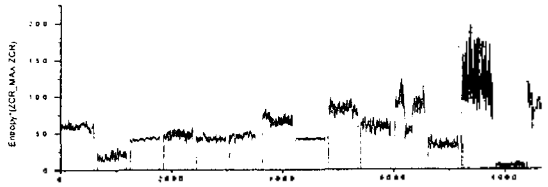
(d) ZCR_{MAX} - ZCR



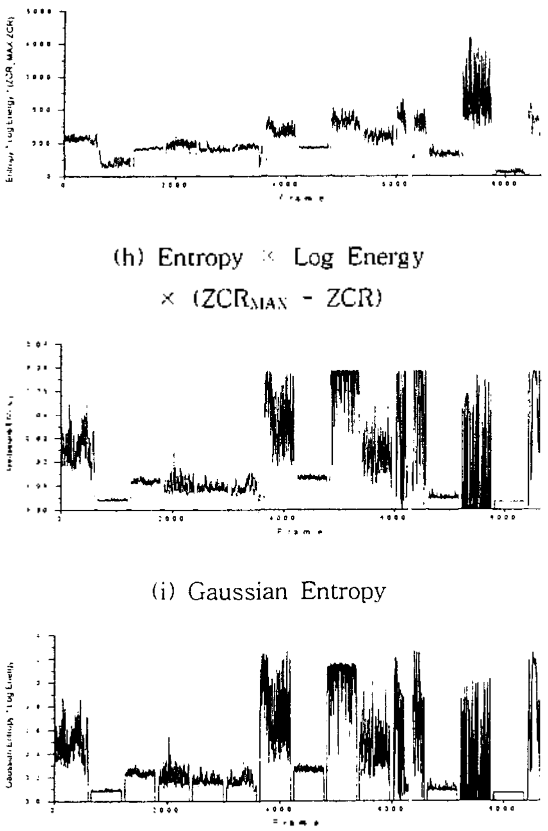
(e) Entropy



(f) Entropy * Log Energy



(g) Entropy * (ZCR_{MAX} - ZCR)



(h) Entropy × Log Energy × (ZCR$_{MAX}$ - ZCR)

(i) Gaussian Entropy

(j) Gaussian Entropy × Log Energy

Figure 3. Comparison of Log Enrgy and Reconstruction Feature I

Figure 4 shows averages of reconstructed features in each sample data during five second. Features and index of horizontal axis are section 2.2 and Table 1, respectively. All feature are normalizec as speech for comparing energy. we can verify that energy is difficult to set threshold value but reconstructed features are easily to set threshold value. Index 15 is speech and feature 5 is excellent.

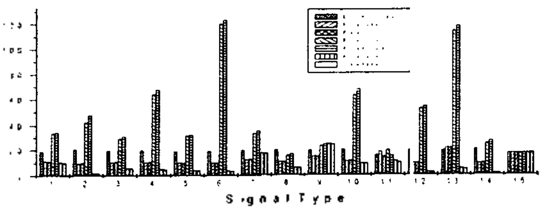


Figure 4. Averages of Log Energy and Reconstructed Features

Ⅳ . CONCLUSION

Recently, because of advancement of hardware process speed, diverse research is studied for more trusty and robust VAD based on not only the energy but also various noise and audio sound classification. In this paper spectral entropy which is new feature parameter and several reconstructed feature are used, compared and analyzed with the case by energy for robust VAD in the noise environment. Experimental results showed that the reconstructed spectral entropy features were more effectual than energy based on algorithms, specially, it is confirmed that appling gaussian distribution function of spectral entropy in speech has a good performance

Ⅴ. REFERENCES

[1] Sadaoki Furui : "Digital Speech Processing, Synthesis, and Recognition", MAECEL DEKKER, INC. 2001 pp. 248-249

[2] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon : "SPOKEN LANGUAGE PROCESSING", Prentice Hall 2001, pp120-130

[3] Nikos Doukas, Patrick Naylor and Tania Stathaki : "Voice Activity Detection Using Source Separation Techniques", Signal Processing Section, Proc. Eurospeech '97

[4] L.R.Rabiner, R.W.Schafer : "Digital Processing of Speech Signals", PRENTICE HALL

[5] J.L. Shen, J.Hung, L.S.Lee : "Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments", Preceeding of ICLP-98, 1998

[6] S. McClellan and J.D. Gibson : "Variable-rate celp based on subband flatness", in IEEE Transactions on Speech and Audio-Processing, vol. 5, pp. 120-130, 1997

[7] J.Sohn and W.Sung : "A voice activity detector employing soft decision based noise spectrum adaptation", in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 356-368, 1998

[8] J.D. Hoyt and H. Wechsler : "Detection of human speech in structured noise" in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 237-240, 1994

[9] http://spib.rice.edu/spib/select_noise.html