Low Power and Real Time LSI Design of Noise Robust Parallel Speech Recognition System

Yoshikazu Miyanaga, Wataru Takahashi and Shingo Yoshizawa
Hokkaido University, Sapporo, 060-0814, Japan

Abstract **— This report introduces recent noise robust speech recognition systems and proposes new advanced speech analysis techniques named MSC (Modulation Spectrum Control)/DRA (Dynamic Range Adjustment).**

I. Introduction

Some robust speech recognition methods have been already developed under additive noises [1]-[7]. Among them, a spectral subtraction has been widely applied in some circumstances. In addition, a robust method for convolute noises, a Cepstrum Mean Subtraction (CMS) has been developed. In CMS, an estimated noise spectrum on the log spectrum domain is eliminated from an observed log spectra where a convolute noise can be represented as an additive component in the log spectrum.

In addition to the above conventional methods, we have already developed other noise robust technique [8]-[10]. This method is based on FIR filtering which is used in running spectrum domain. It was called as a RSF method. Using this method, we can realize high noise robust speech recognition compared with any other conventional methods. In particular, under low SNR circumstances, its performance is recognized as higher than others.

In this report, our developed noise robust speech recognition methods have been explained first. In addition, some current and new results are presented in this report. For example, it is shown that the result of recognition rate is around 90% in case of high speed running car noise with 0dB SNR.

In addition, we have developed a new LSI suitable for the propose system. This designed LSI can realize low power consumption and the real time processing for speech recognition simultaneously.

II. Noise Robust Phase-Based Speech Recognition

In this report, the automatic speech recognition (ASR) system suitable for the isolated words and short phrases is introduced. In particular, Fig.1 shows its important specifications in our designed ASR. The system can accept any distance (a few cm – 5 m) between a microphone and a sound source. In addition, under many and various noise circumstances, the ASR system can recognize its target words correctly.

The ASR system consists of 3 main blocks. These blocks are described in Fig.2. The block of "Automatic Speech Detection" can detect a part of speech sound and eliminates the other parts, i.e., non-speech and only noise signals.

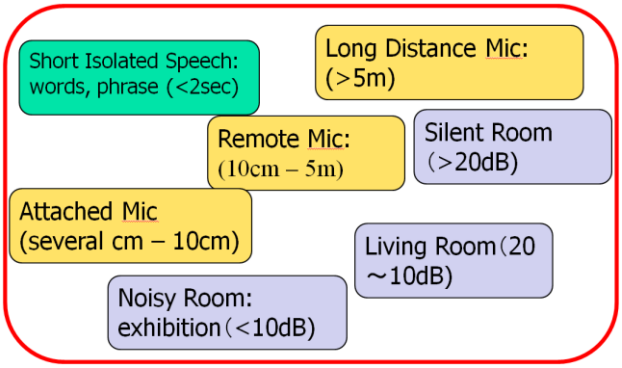

Fig.1 Specification of a Designed ASR System

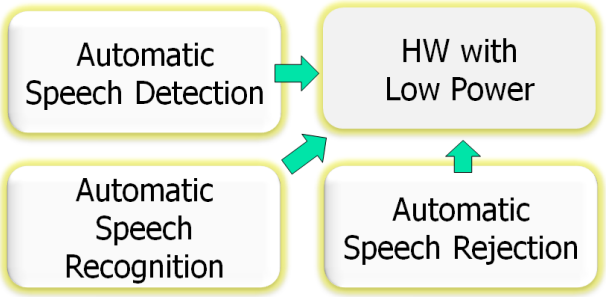

Fig.2 The Functions Used in the Designed ASR System

Using this block, a user does not need to mention its start and end point of its speech sound. The block of "ASR" recognizes a given input speech by using noise robust feature estimation methods and HMM. In noise robust feature estimation methods, RSF which is an algorithm as MSC and DRA are implemented [8],[9].

The block of "Automatic Speech Rejection" can select target keywords. In this system, some target keywords up to 1,000 phrases are registered in ASR blocks. However, we assume any inputted speech sound can be given to this system. In other words, when an given speech is out of target phrases, it should be rejected from the final recognition result. Since the block of automatic speech recognition cannot realize such rejection mechanism, the new block of automatic speech rejection is designed and added to this total system. This block consists of the evaluation of likelihoods in all HMMs and the decision algorithm based on GMM criterions.

The total system has been implemented by 0.25micron CMOS LSI technology. Fig.3 shows its total gates. In this hardware design, the HMM calculation is implemented in parallel/pipelined form. If the number of parallel processing increases, its calculation time becomes short. However, we need many gates for its design.

No. Parallel Processing	32	8
HMM	491,600	116,980
RSF/DRA	11,910	
MFCC	39,670	
System Control	18,310	
Bus Control	1,310	
SRAM	63,400	
Total	626,200	251,580

Fig.3 Total Gate Number of Standard Cell LSI Design

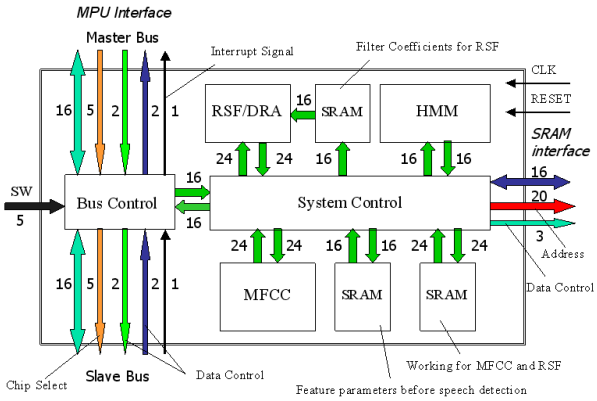

Fig.4 Block Diagram of ASR System LSI Hardware

Fig.4 shows the block diagram of the designed ASR LSI. The LSI can recognize 1,000 word/phrase in real time. In order to execute its processing in real time, the system in which 8 parallel processing in HMM with 10 MHz clock rate is used. At that condition, its power consumption is estimated lower than 100mW. The HMM training data are memorized into the SRAM which is attached from the outside. This system can be controlled by a master-slave mechanism. In other words, if we have several ASR systems, one of them can be master mode and others should be slave mode. Using these modes, all processing can be executed in parallel. If we have 5 chips, the total system can recognize 5,000 word/phrase simultaneously.

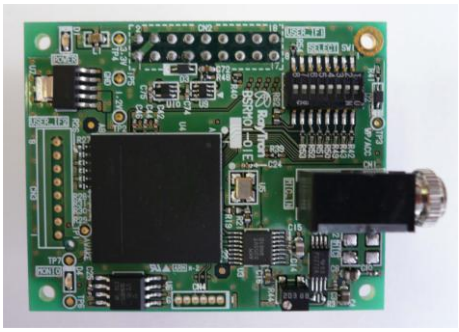

Fig.5 ASR System Board with 55mm×44mm size

The system has been also implemented into a small board shown in Fig.4. The board implements 16 bits AD/DA interface and a serial port connecting to any other peripheral.

Some experiments in real fields can show the high performance of its ASR system shown in Fig.6.

Environment	Noise Level	Correctness
Meeting Room	50dB	96.4%
Elevator	50dB	95.0%
Stairs	45dB	85.1%
Car A (Idling, No-Moving)	50dB	99.4%
Car B (High Speed, Open Window)	75dB	93.3%
Car C (High Speed, Audio ON(FM))	75dB	88.9%
Total		93.0%
※Cruiser Board (Outside, high speed)	80dB	82.7%

Fig.6 Speech Recognition Accuracy in Real Fields

III. Conclusions

In this report, the ASR LSI system has been introduced. In our designed system, noise robust speech recognition mechanism is implemented. In addition, its LSI can realize real time processing and low power consumption.

References

[1] M. Omologo, P. Svaizer and M. Matassoni "Environmental Conditions and Acoustic Transduction in Hands-Free Speech Recognition," Speech Communication, vol. 25, pp. 75-95, 1998.

[2] M. J. Gales, "Cepstral parameter compensation for HMM recognition in noise," Speech Communication, vol.12, pp. 231-240, 1993.

[3] S. Furui, "Speaker-Independent isolated word recognition using dynamic features of speech spectrum", IEEE Trans. on Acoust., Speech, and Signal Process., vol. 34, pp. 52-59, 1986.

[4] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. on ASSP, vol. 27, pp. 113-120, 1979.

[5] N. Kanedera, T. Arai and T. Funada, "Robust automatic speech recognition emphasizing important modulation spectrum," IEICE Trans. on Information and Communication Engineers, vol. J84-D-2, pp. 1261-1269, 2001.

[6] H. Hermansky, "RASTA processing of speech," IEEE Trans. on Speech Audio Processing, vol. 2, pp. 578-589, 1994.

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

[8] Yoshizawa S, Hayasaka N, Wada N and Miyanaga Y, "Cepstral Amplitude
Range Normalization for Noise Robust Speech Recognition," IEICE Trans. on Information and Systems, vol. E87-D-8, pp. 2130-2137, 2004.

[9] Noboru Hayasaka, Kham Khankhavivone, Yoshikazu Miyanaga, Kraisin Songwatana, "New Robust Speech Recognition By Using Nonlinear Running Spectrum Filter", Proceedings of IEEE International Symposium on Communications and Information Technologies 2006, Vol.1, W3A-6, pp.133-136, October 2006.

[10] Noriyuki Ohtsuki, Shingo Yoshizawa, Yoshikazu Miyanaga, "Improvement of Speech Quality using RSF Noise Reduction and its HW Design", Proceedings of The 2008 International Conference on Embedded Systems and Intelligent Technology (ICESIT2008), pp.193-197, February 2008.