A Real Time Noise-Robust Speech Recognition System

Naoya Wada, **Shingo Yoshizawa**, and **Yoshikazu Miyanaga**, Non-members

ABSTRACT

This paper introduces the extraction of speech features realizing noise robustness for speech recognition. It also explores advanced speech analysis techniques named RSF (Running Spectrum Filtering)/DRA (Dynamic Range Adjustment) in detail. The new experiments on phase recognition were carried out using 40 male and female speakers for training and 5 other male and female speakers for recognition. The result of recognition rate is improved from 17% to 63% under car noise at -10dB SNR for example. It shows the high noise robustness of the proposed system. In addition, the new parallel/pipelined LSI design of the system is proposed. It considerably reduces the calculation time. Using this architecture, the real time speech recognition can be developed. For this system, both of full-custom LSI design and FPGA design are introduced.

1. INTRODUCTION

A speech recognition system has been widely explored as one of human interfaces. There are two major approaches for speech recognition. One is continuous speech recognition and the other is word/phase speech recognition. While continuous speech recognition can recognize various long utterances, its accuracy is not as enough as word/phase speech recognition. Since word/phase recognition system can learn all speech articulations in speech models, this system provides higher recognition rate than the other generally speaking.

When we consider the interface of home electronics, mobile navigations and robots, the keyword command and key-phase command systems are valuable in real circumstances. In order to develop such system, a noise robust word/phase speech recognition system should be required. In addition to such noise robustness, real time response must be also demanded.

For noise robustness, we have developed sophisticated filtering on running speech spectrum, i.e., RSF (Running Spectrum Filtering)/DRA (Dynamic Range Adjustment) [1], [2]. Although these techniques require high calculation cost, highest noise robustness can be obtained under various noise circumstances. In order to also realize real time processing, we have developed a parallel architecture of the above system and developed its LSI system.

Our previous system which is designed with FPGA [2] has been implemented into a small board as shown in Fig. 1. This has been already used in some robots with speech recognition and answering mechanism.

In this paper, we introduce sophisticated noise robust speech recognition system, i.e., RFS/DRA speech recognition system, and explore it in detail. This paper also proposes new architecture of this system. The new architecture is designed with 0.18-μm CMOS standard cell and a 128-MHz clock frequency. It results drastically higher response. This paper also develops its FPGA based speech recognition system. This is also suitable for testing our system and implementing any mobile systems.

2. STATEMENT OF PROBLEM

As one of issues for the design of a robust speech recognition system, the extraction of robust speech features should be considered. It is known that cepstrum data are usually corrupted by noise. Various noise robust methods have been developed such as noise-robust LPC analysis [3], [4], Hidden Markov Model (HMM) decomposition and composition [5], [6], [7], and the extraction of dynamic cepstrum, [8], [9] etc. In spite of such research activities, the useful noise-robust technique is still limited as a spectral subtraction (SS) method [10].

The SS is useful for many noises. However it is only used for time invariant noises since the noise property should be estimated as a prior information. If noise circumstances change with time progress, the SS without adaptation to noise may cause deterioration of speech features such as musical noise. It means the estimation of an accurate noise status by SS becomes difficult in some circumstances. In this paper, we explore the robustness of speech features and propose new speech recognition techniques.

We have developed noise robust speech processing techniques, i.e. RSF and DRA. RSF employs FIR filtering and extracts speech components from noisy speech more effectively than RASTA [11]. DRA normalizes the maximum amplitudes of feature parameters and corrects the differences of dynamic ranges between that of trained data and observed speech data. Then RSF applies FIR filtering to noisy speech and

Manuscript received on January 17, 2006.

The authors are with Hokkaido University, Graduate School of Information Science and Technology, Sapporo 060-0814, Japan. E-mail: miya@ist.hokudai.ac.jp

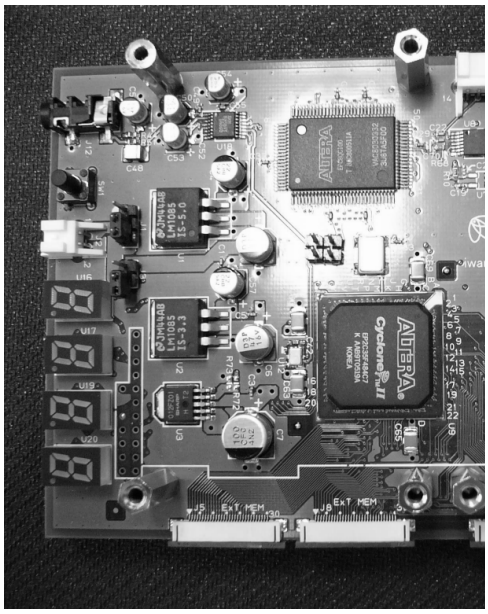


Fig.1: *Noise robust speech recognition system board. This has been developed by RayTron, inc. (http://www.raytron.co.jp).*

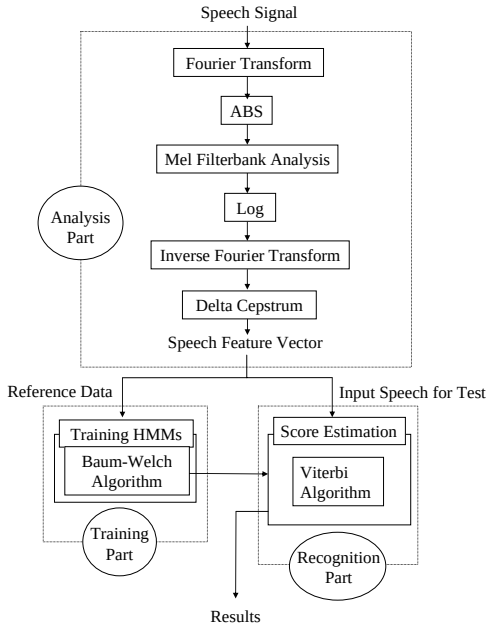


Fig.2: *Procedure of speech recognition system.*

emphasizes speech frequency bands in running spectrum domain. This paper compares RSF and RASTA and estimates their performances and DRA. Noise robustness of each method is evaluated by phase speech recognition experiments using HMM.

In addition, when the real application of phase speech recognition is considered, the real time response must be demeaned. In other words, if its calculation cost and its complexity are high, the special hardware should be developed.

In this paper, the hardware implementation of our system is proposed. There are two purposes in this implementation. The one is to realize real time speech recognition by reducing calculation time. The other is to realize the tiny hardware used in mobile electronics devices by reducing the size of devices and the power dissipation. In this paper, the 50msec response time of the recognition system by a 0.18-μm CMOS standard cell with a 128-MHz clock frequency is proposed. This system can recognize 800 phase speech at the same time. The implementation of this system is also tested on a FPGA device.

3. CONVENTIONAL SYSTEM AND CAUSES OF NOISE CORRUPTION

Figure 2 shows the process of the conventional recognition system. The former part of this figure shows the procedure of the speech feature extraction which consists of ordinary feature extraction based on Mel-Frequency Cepstral Coefficient (MFCC) [12], [13] and HMMs. MFCC is one of speech features and is based on the human's perception in frequency domain. The latter part shows a training part using Baum-Welch Algorithm and a recognition part using Viterbi Algorithm. In noisy speech, the stationary power of noise causes serious differences from noise-free speech.

Noisy speech signal $y(t)$ is converted to spectrum by DFT as

$$y(t) = h(t) \otimes (x(t) + a(t)) \quad (1)$$

$$Y(n,f) = H(n,f)X(n,f) + H(n,f)A(n,f) \quad (2)$$

where n denotes the frame number, $x(t)$ denotes the signal component, $h(t)$ denotes the system noise and $a(t)$ denotes the environmental noise.

Figure 3 compares time trajectories of power spectrum obtained from noise-free speech and noisy speech. This trajectory is given from short time windowing speech flames and its DFT spectra. Note that this trajectory is obtained from running speech spectrum at the fixed frequency [2]. There are differences on DC components of the trajectories. In order to obtain cepstrum, log power spectrum is required. Figure 3 (b) shows the difference of time trajectory between log power spectra of clean speech and that of noisy speech. In this conversion, power gains from the input utterance become different. It leads the reduction of dynamic ranges on cepstrum. Therefore, we have to consider two serious corruptions on DC components and gains from utterances.

4. RUNNING SPECTRUM FILTERING (RSF)

RSF utilizes the rhythm of syllables. There is a specific constant rhythm on the changes of sylla-

bles. On the other hand, noise components do not change so radically and there are generally no specific rhythms in noises. Therefore, if there is a parameter associated with the rhythm of changes in speech components, both components can be separated when we estimate their rhythm and filter them.

If we use modulation spectrum domain, the above rhythm can be represented separately. It is obtained as follows: At first, two dimensional data which contains time-versus-frequency information of spectrum is obtained by accumulating short-time spectrum [2]. In running spectrum domain, the time trajectory at a specific frequency is obtained by tracing its values in each flame. From the time trajectory, we can get a modulation spectrum by applying DFT to this trajectory.

It has been reported [11], [14] and [2] that speech components in modulation frequency domain are dominant around 4 Hz and the range from 0 to 1 Hz and from 7 Hz can be regarded as noise. Therefore, speech components can be extracted by applying band-pass filtering on running spectrum.

RASTA is the well known method focusing on modulation spectrum. RASTA employs IIR band-pass filters and removes noise components. However, IIR filters may be unstable and cause phase distortion. On the other hand, RSF employs FIR filters instead of IIR filters. It makes RSF stable and free from phase distortion. However, RSF requires high-order FIR filters to realize sharp modulation frequency cut-off and such high order of FIR filters causes many delay boxes. For example, in order to realize the modulation frequency properties of RSF shown in Fig. 4, 240 taps are required. Then, the required length of non-speech periods $l[sec]$ before and after the input speech is given by

$$l = \frac{\text{the number of taps} * \text{frame-shift}}{2 * \text{sampling rate}}. \qquad (3)$$

When the conditions of the speech analysis follow Table 1, about 1.4 second non-speech periods are required before and after the input speech. In our method, several non-speech frames are put into the front and the back of speech frames in a certain length so that enough filtering orders are obtained. Thus RSF realizes effective feature extraction and can be applied in practical speech recognition system.

The process of RSF is as follows: In (2), $H(n,f)A(n,f)$ is additive noise component. When the property of $H(n,f)X(n,f)$ is considered, it may be located from 1 Hz to 7 Hz in modulation spectrum domain. Accordingly, if you apply band-pass filter to the running speech spectrum, we can reduce noise components. In the first step, low-pass filtering is applied to reduce all higher frequency noise components.

The logarithmic power spectrum without the additive noise component is approximately written as

$$\begin{aligned} log|Y(n,f)| &= log|H(f)X(n,f)| \\ &= log|X(n,f)| + log|H(n,f)| \end{aligned} \qquad (4)$$

This system noise component $H(n,f)$ can be removed by applying band-pass filtering to the time trajectory of logarithmic power spectrum.

Using RSF influences of the differences in the spectral fine structure are eliminated as shown in Fig. 5 (e). This process removes unnecessary parts of speeches for speech recognition such as characteristics of speakers and noise influences and consequently eliminates the differences on DC components.

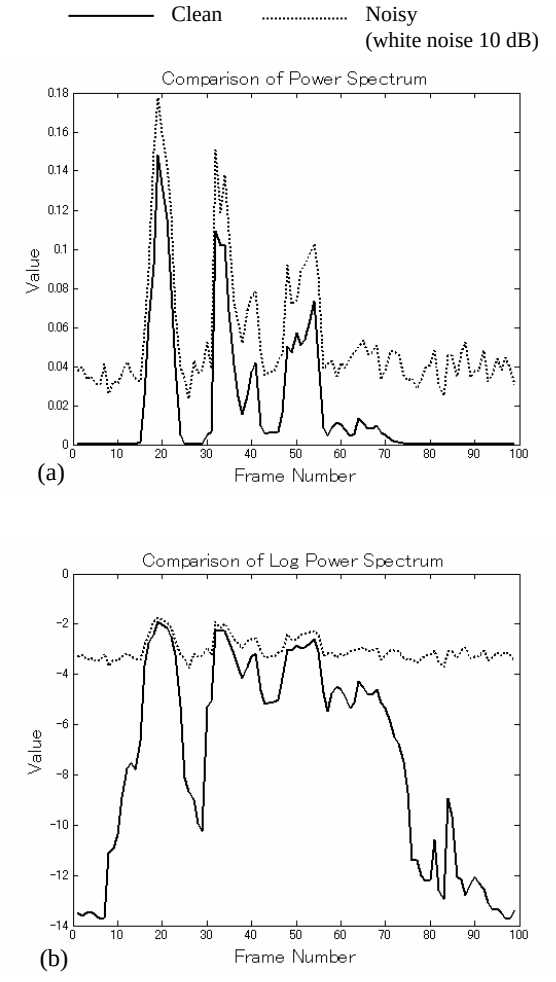


Fig.3: *Comparison of power spectra and log power spectra. (a): power spectra, (b): log power spectra.*

5. DYNAMIC RANGE ADJUSTMENT (DRA) ON CEPSTRUM

The dynamic range of cepstrum indicates the difference between maximum and minimum of cepstral values. As noted in Section 3, power gains from utterances decrease because of additive noise. It causes decrease of cepstral dynamic ranges. As a result, it seriously degrades the speech recognition performance

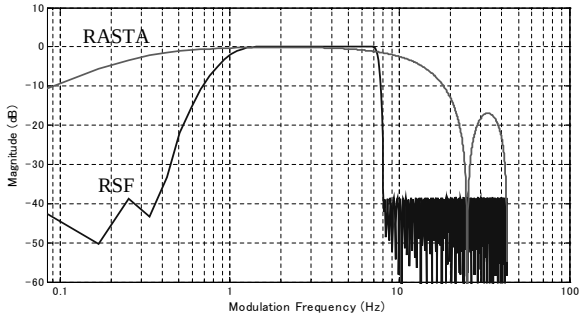


Fig.4: *Modulation frequency properties in RASTA and RSF.*

since both maximum and minimum values represent the characteristics of speech and they are corrupted by noise. Figure 6 shows distributions of the number of dynamic ranges of cepstra and proves that dynamic range is usually reduced by additive noise even if RASTA or RSF is applied.

DRA normalizes amplitudes of a speech feature vector with the maximum amplitude. In DRA, the amplitude of a cepstrum is adjusted in proportion to its maximum amplitude as

$$\tilde{f}_i(n) = f_i(n) / \max_{j=1,\cdots,m} |f_j(n)| \quad (i = 1, \cdots, m), \tag{5}$$

where $f_i(n)$ denotes an element of the cepstrum, m denotes the dimension and n denotes the frame number. Using (5), all amplitudes are adjusted into the range from -1 to 1.

With DRA/RSF, speech analysis is refined as shown in Fig. 7. Then using DRA the difference of cepstral dynamic range is adjusted as shown in Fig. 5 (f) and the cepstrum of noisy speech is adjusted to the one of clean speech.

6. EXPERIMENTS

6.1 Word Recognition Results

In order to evaluate the noise robustness of the proposed techniques, isolated word speech recognition using HMM [15] has been carried out. The recognition system is based on the conventional one shown in Fig. 2. The recognitions part is implemented using the MATLAB software. The acoustic models are thirty-two-state one-mixture-per-state HMMs. The whole database is Japanese common voice data 'Chimei' which means the names of places. They are presented by the Japan Electric Industry Development Association. The database consists of 100 Japanese isolated words spoken four times by 90 persons. The data are 11.025 kHz and 16 bit sampling speech. Other conditions are described in Table 1.

RASTA, RSF and DRA are applied. Several combinations of recognition are evaluated in these conditions. Speech feature vectors have 38-dimensional parameters which consist of 12 cepstral coefficients, 12 delta-cepstral coefficients, 12 delta-delta-cepstral coefficients, delta-logarithmic power and delta-delta-logarithmic power. Recognition results are shown in Table 2 and 3. At the first glance, same tendency of recognition rates are obtained in both white and running-car noise environments. Each noise robust speech feature extraction method, i.e., DRA, RASTA and RSF, improves recognition performance except DRA at higher SNR. Comparing recognition performances of RASTA and RSF, RSF is a little superior to RASTA in both noise environments. Then by combining DRA, both methods shows better performances. RSF with DRA shows the best performance among all methods. Especially in the running-car noise environment at -10dB SNR, DRA improves the recognition rate with RSF by 31.15% while DRA improves that with RASTA by 20.75% only.

6.2 Consideration

A reason why the combination of RSF and DRA shows the best performance is derived from the difference between the IIR filtering of RASTA and the FIR filtering of RSF. IIR filtering is not stable and causes phase distortion. The differences of recognition rates show the advantage of FIR filtering and RSF.

The other reason is derived from the DC offset. It makes serious influence on cepstrum by the increase or the decrease of whole values. Fig. 5. compares the original first order MFCC and the ones after RASTA and RSF. In the original cepstrum, DC offset occurs and the cepstral values of clean and noisy speeches in non-speech frames are much different. DRA can adjust the cepstral dynamic range but cannot correct the difference of position of whole waveform. Moreover, it can causes eccentric maximum amplitude.

Such higher maximum amplitude causes excessive adjustment in (5) because DRA uses maximum amplitude for normalizing. Excessive adjustments make speech characteristics flatter and degrade recognition performances. Therefore, DC offset should be removed for making use of DRA effectively. Comparing coefficients using RASTA and RSF, RSF can eliminate DC offset and both values of clean and noisy speeches in non-speech frames are almost same. On the other hand, DC offset remains in RASTA. Therefore, DRA may not work correctly with RASTA.

7. HARDWARE IMPLEMENTATION

7.1 VLSI Design of RSF/DRA

This section describes the hardware development of the RSF/DRA based robust speech recognition system. The goal of the hardware development is to provide real time processing and low power dissipation for a complete recognition processing. Mobile phones and PDAs require not only high recognition accuracy but also a long battery lifetime. Human-robot inter-

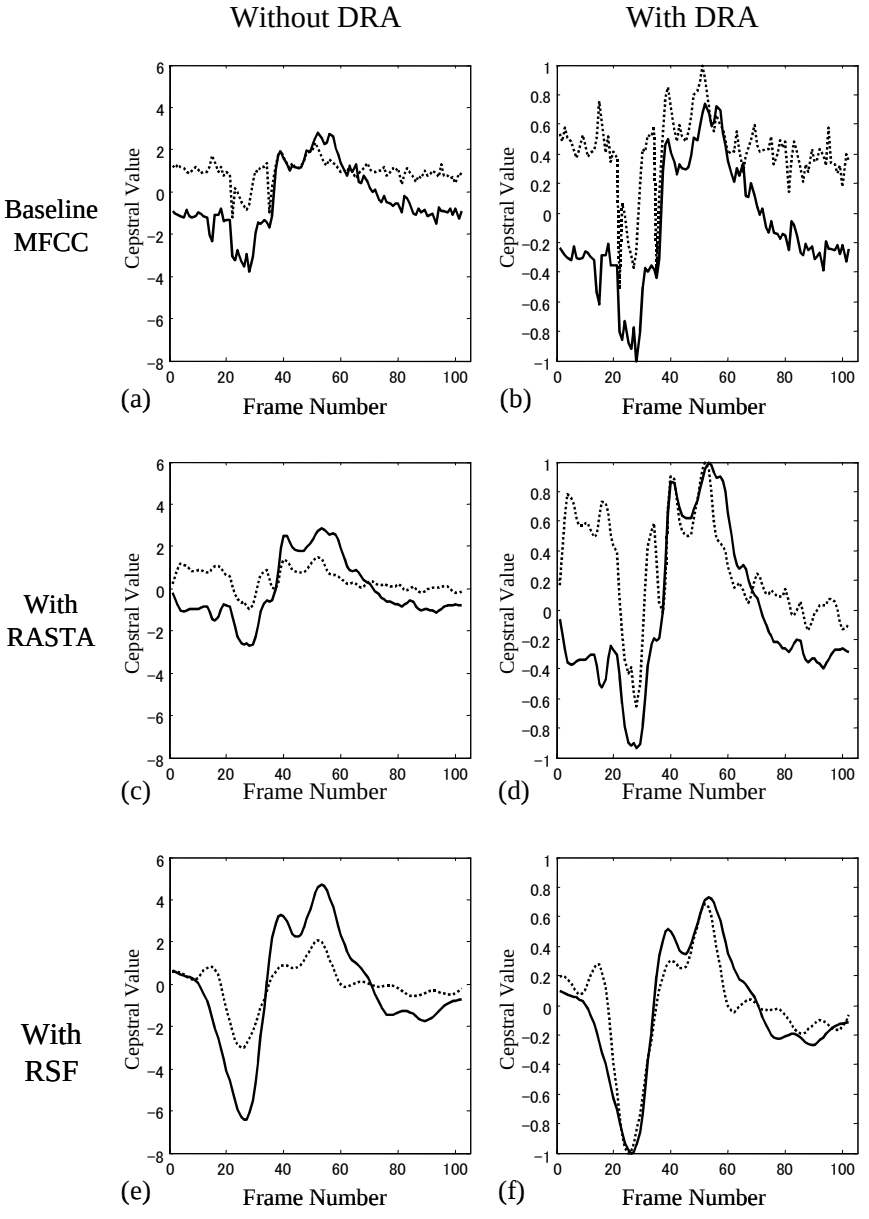


Fig.5: *A comparison of trajectories of the 1st order cepstra among baseline MFCC, MFCC after RASTA and MFCC after RSF. The solid lines show cepstrum of clean speech and the dash lines show one of noisy speech (running-car noise, 0dB SNR). The sample speech is /Kitami/ in Japanese. Used methods are as follows; (a): Baseline MFCC (b): MFCC after DRA (c): MFCC after RASTA (d): MFCC after RASTA and DRA (e): MFCC after RSF (f): MFCC after RSF and DRA.*

faces regard a short-time response as important to use recognition results for various actions.

A main stream of hardware design is classified into two categories, i.e., a processor and a custom hardware. A custom hardware has the advantages of circuit area and power consumption. In particular, if a complete recognition system (that includes speech analysis, robust processing, and recognition processing) is implemented into a single chip, a pure custom hardware can reduce redundant parts and achieves lower power than a hybrid of a DSP and a custom hardware. Hence, we have developed a complete recognition that executes all the processing of speech recognition. The designed system is embedded into a CMOS chip and a field programmable gate array (FPGA) board.

Figure 8 shows a whole structure of the recognition system. The system consists of speech recognition (illustrated as "HMM"), speech analysis ("MFCC"), robust processing ("RSF/DRA"), and system control. We described the VLSI implementation of a recognition part in [2]. The block diagram of speech analysis and robust processing parts is illustrated in Fig. 9. The MFCC circuit consists of a 512-point FFT, a logarithm arithmetic unit, and a 12-point IDCT (Inverse Discrete Cosine Transform). The RSF/DRA

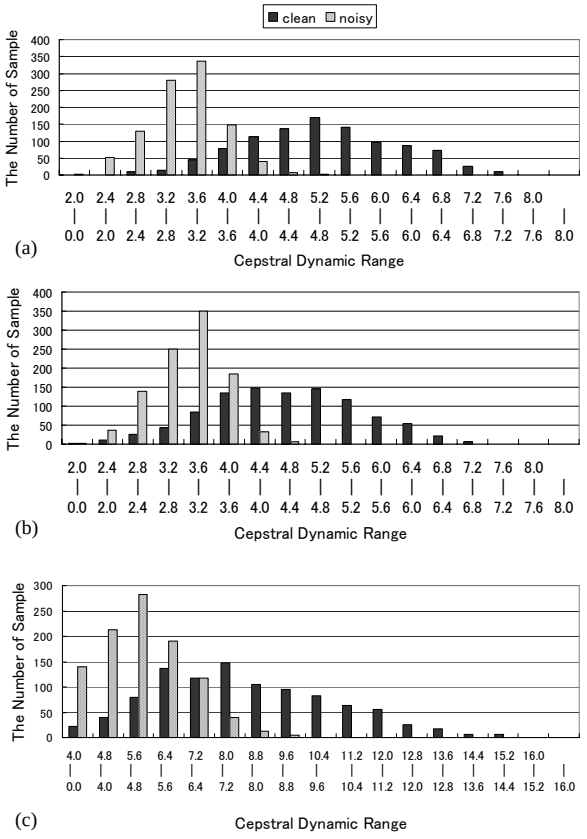


Fig.6: *Distributions of dynamic ranges of the 1st cepstra obtained from the analysis of 100 Japanese isolated words spoken two times by 5 male speakers. Additive noise is white noise at 10dB SNR. (a) is obtained from original cepstra, (b) is obtained from cepstra after RASTA and (c) is obtained from cepstra after RSF.*

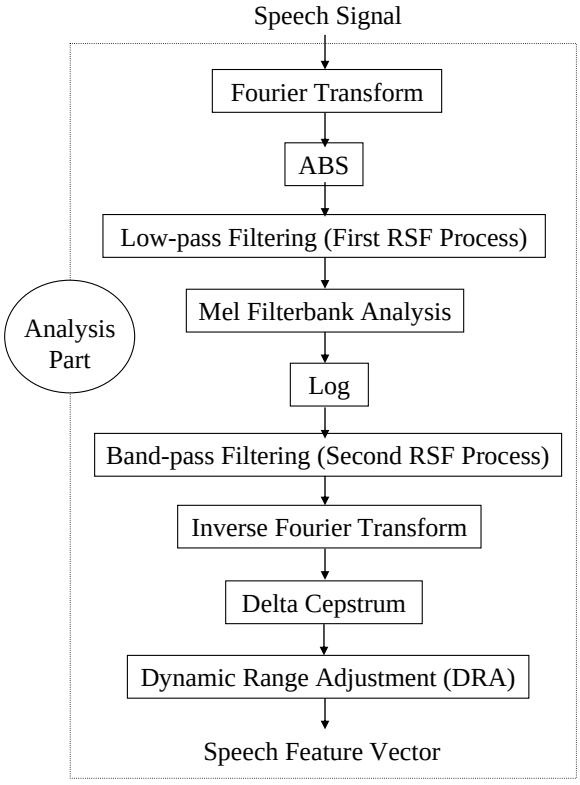


Fig.7: *Analysis method with DRA/RSF.*

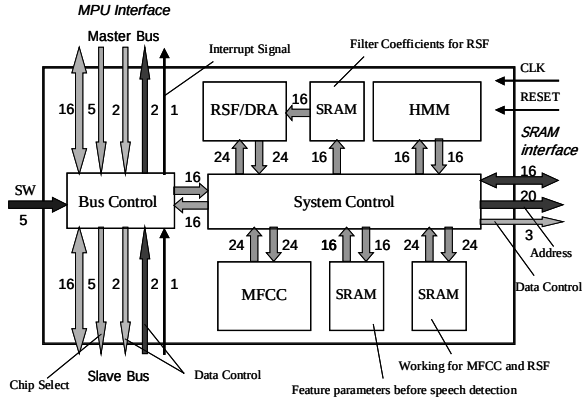


Fig.8: *Structure of the recognition system.*

circuit consists of a FIR filter, a divider and memory units. We have adopted a fixed point format for all the arithmetic operations. The word lengths of arithmetic units are minimized by iterative software simulations. The output data of the RSF/DRA circuit is given by an 8-bit word length with dynamic scaling. Figure 10 illustrates a detailed structure of the RSF/DRA circuit. The RSF/DRA circuit executes FIR filtering, calculating delta cepstral coefficients, and normalizing cepstral parameters in amplitudes. The divider is used for calculating reciprocal numbers in the DRA processing. The maximum amplitudes of cepstral parameters are calculated simultaneously with FIR filtering in the RSF processing. The RSF coefficients can be exchanged if those data are stored in an external memory. The dynamic scaling extracts 8 bits from 24 bits in cepstral data where those scaling factors are given by the HMM training data.

Table 4 shows processing time in the recognition system at a 0.18-μm CMOS standard cell and a 128-MHz clock frequency. The processing time of recognition is proportional to the number of word models. For example, an 800-word vocabulary task takes 28.6 ms in the total. Note that speech analysis processing is processed during an utterance. The response time after an utterance amounts to 34.9 ms in this task. Since this result is more than enough for achieving real time processing, it could minimize power dissipation by decreasing a clock speed. See [2] for the comparison of power dissipation between the custom hardware and a standard DSP.

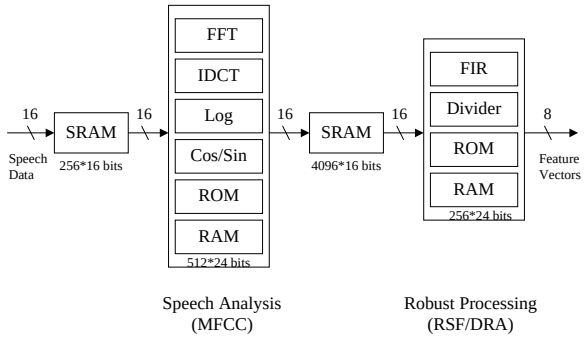


Fig.9: *Block diagram of speech analysis and robust processing.*

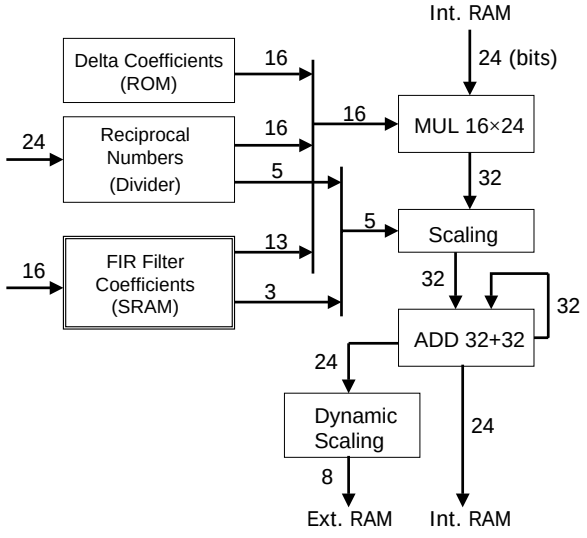


Fig.10: *RSF/DRA circuit.*

7.2 FPGA Board

The recognition system has been implemented to a FPGA board to verify circuit behavior and test actual recognition performance under real environments. Figure 11 shows the block diagram of FPGA board. The sampling clock generator, the A/D converter, the serial port interface, and the external SRAM are connected to the FPGA board. The sampling rate is 11.025 kHz with 12-bit quantization. The sequential control unit substitutes for a microprocessor. Speech detection starts when a switch on the board is pushed and ends automatically, after 1.5 seconds. Users should utter a word during this period. After the detection, recognition results are displayed as word numbers on an LED. Table 5 denotes the implementation results of the recognition system in a FPGA device of Altera APEX20KE. The clock speed can be changed to 5, 10, and 20 MHz. Due to the limitation of FPGA resources, we reduced the number of parallel arithmetic units to 1/4 in the recognition part. For a 40-word vocabulary with a speaker independent task, the FPGA board provided about 97% in recognition accuracy under real environments

Table 1: *The condition of speech recognition experiments.*

Recognition Task	Isolated 100 words vocabulary
Speech Data	100 Japanese place names from JEIDA
Sampling	11025 Hz, 16-bit
Window Length	23.2 ms (256 points)
Frame Period	11.6 ms (128 points)
Window Function	Hanning window
Pre-emphasis	1-0.97z^{-1}
Baseline Speech Feature Vector	38th order, based on MFCC (12-dimensional MFCC, 12-dimensional delta MFCC, 12-dimensional delta-delta MFCC, delta log-energy, delta-delta log-energy)
Acoustic Model	32-states continuous word HMMs
Training Set	40 female and 40 male speakers, 3 utterances each
Tested Set	Speaker-independent, 5 female and 5 male speakers, 2 utterances each
Noise Varieties	White noise at 10, 20, or 30 dB SNR, Running-car noise at -10, 0, or 10 dB SNR

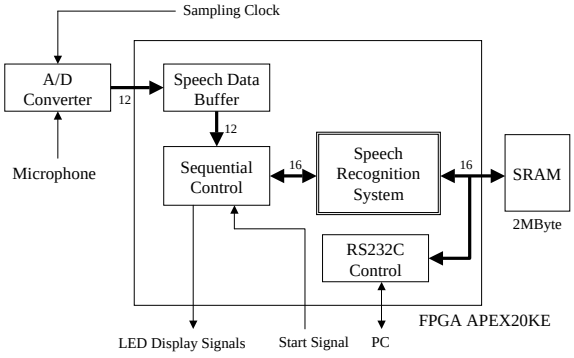


Fig.11: *Block diagram of the FPGA board.*

including distortions caused by a microphone and an A/D converter. Since the time length of speech detection is fixed in the current system, we consider that use of better detection algorithms would improve recognition performance.

8. CONCLUSIONS

In this paper, the techniques for noise suppression, DRA and RSF are explored in detail. RSF emphasizes speech frequency bands by applying the FIR filtering. DRA normalizes the maximum amplitudes of the cepstrum. The effectiveness is evaluated in new speech recognition experiments. Note that both data of males and females are used in training and recognition phases. It is noted that the combination of RSF and DRA shows the best performance. This result indicates that RSF extracts speech characteristics more effectively than RASTA and a synergistic

Table 2: *Recognition rates versus white noises for the estimation of feature extraction.*

Speech Feature / SNR	Rec. Rates [%] 10dB	20dB	30dB
Conventional	57.30	96.70	99.35
DRA	70.15	96.05	99.25
RASTA	70.45	96.95	99.20
RSF	74.55	97.05	99.25
RASTA+DRA	80.90	97.25	99.35
RSF+DRA	85.05	97.10	99.15

Table 3: *Recognition rates versus running car noises for the estimation of feature extraction.*

Speech Feature / SNR	Rec. Rates [%] -10dB	0dB	10dB
Conventional	17.10	77.80	95.55
DRA	25.40	76.90	95.25
RASTA	27.80	90.80	98.35
RSF	32.35	90.20	98.50
RASTA+DRA	48.55	89.90	97.80
RSF+DRA	63.50	93.75	98.35

Table 4: *Processing time in the recognition system at a 128-MHz clock frequency and an 800-word vocabulary task.*

Speech Analysis	12.5 ms ($145\mu s$ / frame)
Robust Processing	6.3 ms
Speech Recognition	28.6 ms at 800 words ($35.7\mu s$ / word)

effect should exist between DRA and RSF.

In addition, we have developed the total speech recognition system with 0.18μm CMOS LSI in order to realize real time processing. It can be designed by around 400k gates. The total power consumption is fairly less than that of DSP based embedded speech recognition systems. In this paper, the FPGA system design has been also introduced. It is quite suitable for testing the system.

Table 5: *Results of the FPGA implementation; the percentages denote a rate of utilization for a FPGA device.*

	Logic Elements	Memory Bits
HMM	10,427	36,032
RSF/DRA	3,235	0
MFCC	11,580	0
System Control	3,733	0
Bus Control	114	0
SRAM	0	36,864
Others	1,577	4,096
Total	30,666 (59%)	76,992 (17%)

Acknowledgment

The authors would like to thank Research and Development Headquarters, Yamatake Corporation for fruitful discussions. This study is supported in parts by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (B2) (15300010), 2003.

References

[1] N. Wada, Y. Miyanaga, N. Yoshida and S. Yoshizawa, "A consideration about an extraction of features for isolated word speech recognition in noisy environments," *ISPACS2002*, Vol. DSP2002-33, pp. 19-22, Nov. 2002.

[2] Shingo Yoshizawa and Yoshikazu Miyanaga, "Robust recognition of noisy speech and its hardware design for real time processing," *ECTI Transaction on Electrical Eng., Electronics, and Communications (EEC)*, Vol. 3, No. 1, pp. 36-43, Feb. 2005.

[3] Tierney J., "A study of LPC analysis of speech in additive noise," *IEEE Trans. on Acoust., Speech, and Signal Process.*, Vol. ASSP-28, No. 4 , pp. 389-397, Aug. 1980.

[4] Kay S.M., "Noise compensation for autoregressive spectral estimation," *IEEE Trans. on Acoust., Speech, and Signal Process.*, Vol. ASSP-28, No. 3 , pp. 292-303, Mar. 1980.

[5] Varga A. and Moore R., "Hidden Markov model decomposition of speech and noise," *Proc IEEE ICASSP*, pp. 845-848, 1990.

[6] Gales M.J.F. and Young S.J., "Cepstral parameter compensation for HMM recognition in noise," *Speech Communication*, Vol. 12, No. 3, pp. 231-239, 1993.

[7] Martin F., Shikano K., Minami Y. and Okabe Y., "Recognition of noisy speech by composition of hidden Markov models," *IEICE Technical Report*, Vol. SP92-96, pp. 9-16, Dec. 1992.

[8] Aikawa K. and Saito T., "Noise robustness evaluation on speech recognition using a dynamic cepstrum," *IEICE Technical Report*, Vol. SP94-14, pp. 1-8, June 1994.

[9] Aikawa K., Hattori H., Kawahara H. and Tohkura Y., "Cepstral representation of speech motivated by time-frequency masking: an application to speech recognition," *J. Acoust. Soc. Am.*, Vol. 100, No. 1, pp. 603-614, July 1996.

[10] Boll S., "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. ASSP*, Vol. ASSP-27, No. 2, pp. 113-120, 1979.

[11] Hermansky H. and Morgan N., "RASTA processing of speech," *IEEE Trans. Speech and Audio Process*, Vol. 2, pp. 578-579, Oct. 1994.

[12] Furui S., "Speaker-independent isolated word recognition using dynamic features of speech spec-

trum," *IEEE Trans. on Acoust., Speech, and Signal Process.*, Vol. ASSP-34, No. 1, pp. 52-59, Feb. 1986.

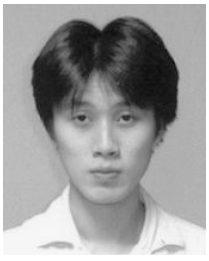
[13] Davis S.B. and Mermelstein P., "Comparison of parametric representations for mono-syllabic word recognition in continuously spoken sentences," *IEEE Trans. on Speechand Signal Processing*, pp. 357-366, 1980.

[14] Hayasaka N., Miyanaga Y. and Wada N., "Running spectrum filtering in speech recognition," *SCIS Signal Processing and Communications with Soft Computing*, Oct. 2002.

[15] Rabiner L.R., "A Tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, Vol. 77, No.2 , Feb. 1989.

[16] S. Yoshizawa, N. Wada, N. Hayasaka and Y. Miyanaga, "Noise robust speech recognition focusing on time variation and dynamic range of speech feature parameters," *ISPACS2003*, pp. 484-487, 2003.

[17] S. Yoshizawa, N. Wada, N. Hayasaka and Y. Miyanaga, "Scalable architecture for word HMM-based speech recognition," *Proc. IEEE ISCAS2004*, pp. 417-420, 2004.



Naoya Wada received the B.E. and M.E. degrees in Electrical Engineering from Hokkaido University, Japan in 2001 and 2003, respectively. He is currently studying at Graduate School of Information Science and Technology, Hokkaido University. His research interests are digital signal processing, speech analysis, and speech recognition.

Shingo Yoshizawa received the B.E., M.E. and Dr. Eng. degrees in Electrical Engineering from Hokkaido University, Japan in 2001 2003 and 2005, respectively. He is currently working at Graduate School of Information Science and Technology, Hokkaido University as a research fellow of the Japan Society for the Promotion Science. His research interests are speech processing, wireless communication systems, and VLSI architecture.

Yoshikazu Miyanaga was born in Sapporo, Japan, on December 20, 1956. He received the B.S., M.S., and Dr. Eng. degrees from Hokkaido University, Sapporo, Japan, in 1979, 1981, and 1986, respectively. He was a Research Associate at the Institute of Applied Electricity, Hokkaido University from 1983 to 1987, a lecturer of Electronic Engineering at Faculty of Engineering, Hokkaido University from 1987 to 1988 and an Associate Professor of Electronic Engineering at Faculty of Engineering, Hokkaido University from 1988 to 1997. He is currently a Professor of Laboratory for Information Communication Networks, Division of Media and Network Technologies at Graduate School of Information Science and Technology, Hokkaido University. His research interests are in the areas of adaptive signal processing, nonlinear signal processing and parallel/pipelined VLSI system. Dr. Miyanaga is a member of the Institute of Electrical and Electronics Engineers (U.S.A.), the Institute of Electronics, Information and Communication Engineers (Japan) and the Acoustical Society of Japan.