KOREAN DIGIT RECOGNITION IN NOISE ENVIRONMENT USING SPECTRAL MAPPING TRAINING

Ki Young Lee , Chul Soo Bai , Hyun Seop Cho*

Dept.Electronic Comm.Eng.,Kwan Dong Univ.
Kang Neung Si, 210-701, S.KOREA
* Dept.Electric Eng.,Won Kwang Univ.
Iree Si, 570-749, S.KOREA

ABSTRACT This paper presents the Korean digit recognition method under noise environment using the spectral mapping training based on static supervised adaptation algorithm[1,2]. In the presented recognition method, as a result of spectral mapping from one space of noisy speech spectrum to another space of speech spectrum without noise, spectral distortion of noisy speech is improved, and the recognition rate is higher than that of the conventional method using VQ and DTW without noise processing, and even when SNR level is 0 dB, the recognition rate is 10 times of that using the conventional method. It has been confirmed that the spectral mapping training has an ability to improve the recognition performance for speech in noise environment.

1. INTRODUCTION

Recently, the wider the application range of the speech recognizer, the more the necessity of the one robust to noise, because of the speech overlapped with noise from the external machine or the outside environment, and distorted through microphone or other channel with noise.

According to the results of Acero's study[3], it is found that if speaker-independent speech recognizer was under noise environment, the recognition performance was very lowered. The conventional noisy speech recognizer has mainly used the noise reduction method using the statistical characteristics of speech and noise.

Boll[4] attempted to cancel noise from speech using the spectrum reduction method with DFT coefficients. Stockham[5] used the spectrum equivalent circuit as the noise compensation method. Van Compemolle[6] utilized both the spectrum reduction method and the spectrum equivalent circuit and won the good results. However, this method had a defeat that the independence for noise spectrum estimates must be assumed. Besides, the adaptive noise cancelling method has been studying, but this method requests many computations or multi-sensors.

We noticed the point of elevating the recognition performance for unknown speaker's speech through spectral mapping based on the static supervised adaptation algorithm[1,2].

Therefore, this paper presents the speech recognition method in noise environment using spectral mapping from one space of noisy speech spectrum to another space of speech spectrum without noise, so that spectral distortion of noisy speech is lessened and the reognition rate is higher than that of the conventional mehtod using VQ and DTW without noise processing. In the spectral mapping procedure, we use DTW[9] which is adaptive to several word lengths so that the correspondent relationship is more correctly. This recognition method has two merits of applying the recognizer without changing system under any noise environment through the training procedure according to noisy speech, and having a less influence on the noise characteristics in the recognition procedure because of the mapped codebook linearly combinated with speech spectra without noise.

2. SPCTRAL MAPPING TRAINING

In this paper, spectral mapping is a conversion work that obtains the mapping relationship from one space of noisy speech spectrum to another space of speech spectrum without noise. Because speech have an infinite spectrum space, VQ is used to make a finite space. Basic concept described above is shown is Fig.1.

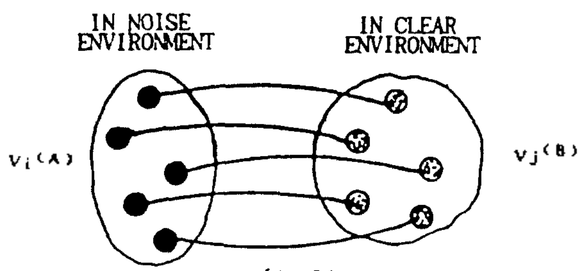


Fig. 1. Basic concept of spectral mapping

This mapping is to find correspondence between two codebooks from different environments. The correspondence is represented in the form of a mapped codebook which is obtained from a training procedure based on a static supervised adaptation algorithm. Fig.2 shows a block diagram of the training procedure for making a making a mapped codebook in noise environment

In the recognition procedure, DTW techniques is used with a distortion matrix between two codebooks, and the decision rule is kNN. Fig.3 shows a block diagram of word recognition in noise environment using a mapped codebook.

3. EXPERIMENTS AND RESULTS.

Speech data is korean digits and spoken by 2 rates 10 times in computer lab. room . Speech under noise environment is made by adding white noise whose SNR levels are 20, 10, 0 db. Training words was 10 digits first-time spoken. Fig.4. shows a adaptive adjustment window to lessen errors on the path searched by DTW. Analysis conditions are described in Table 1.

We use LBG[11] algorithm to make codebooks. To examine the effect according to phoneme spectrum of each digit, we need to classify Korean digits with vowels and consonants. The number of vowels and consonants in Korean digits is 13 { vowel : 이, 아, 오, 우, 유 // initial- consonants : ㅅ, ㅊ, ㄱ, ㅍ // final-consonant : ㄹ, ㅇ, ㅁ, ㄱ }. Therefore, the number of spectra characterised by vowels and consonants of Korean digits is better to be more than 13.

In Fig.5, the recognition results are shown according to each SNR level and commpared with each method, where the recognition rate using spectral mapping method is higher than that using WM method. When codebook size is 16, the rate is more improved than the one of any other case. This is because more correctly correspondent relationship is possible as considered in Fig.5. When codebook size is 32, the rate is not so much improved as when codebook 16, because the codebook have too many spectra damaged by noise, and the mapping relationship between two spectrum spaces is not correctly matched.

The reason of higher performance using NM method than using 1M method is because the more the iterative number, the less the noise effect because of the renewal mapped codebook consisted with speech spectrum without noise. When SNR is 0 dB, the recognition rate using NM method is improved 10 times of the one using WM method regardless of many codebook size.

4. CONCLUSION

Spectral mapping training for speech recognition under noise environment is presented and examined its validify through experiments using Korean digits adding several white noise levels. The results show that the lower SNR level, the higher improved recognition rate than the one using the conventional method. Therefore, it is confirmed that spectral mapping training has an ability to improve the recognition performance for speech in noise environment.

Table 1. Analysis conditions

sampling frequency	:	10 kHz
window length	:	20.0 ms
overlapped window length	:	10.0 ms
window function	:	Hamming window
preemphasis	:	$1 - 0.95z^{-1}$
LPD order	:	14

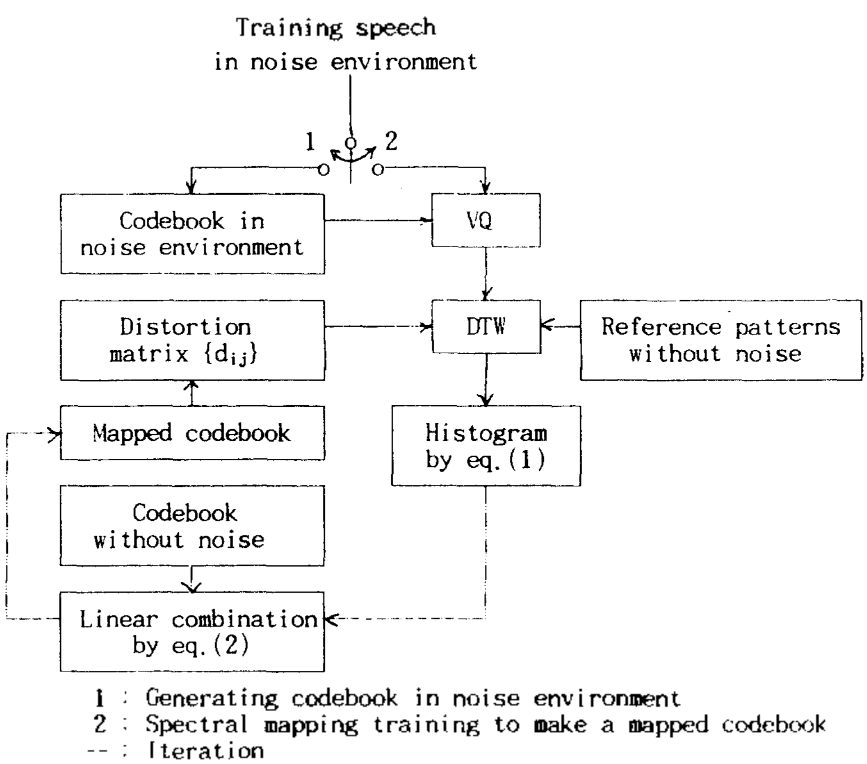


Fig. 2. Block diagram of training for making a mapped codebook in noise environment

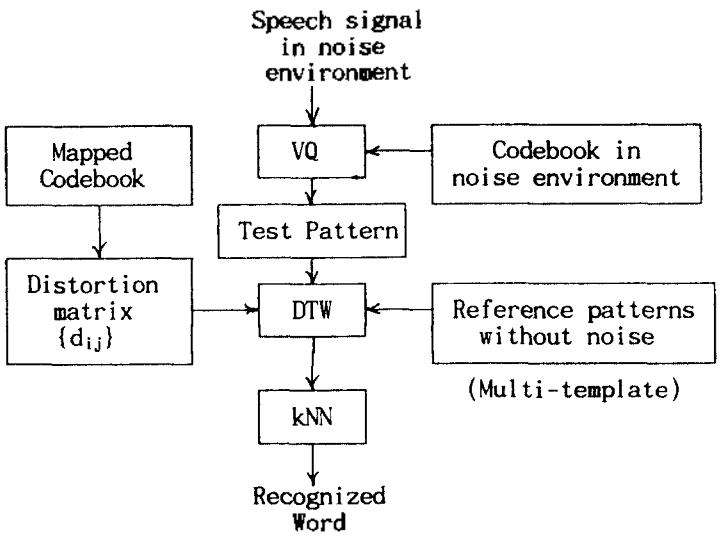


Fig. 3. Block diagram of word recognition in noise environment using a mapped codebook

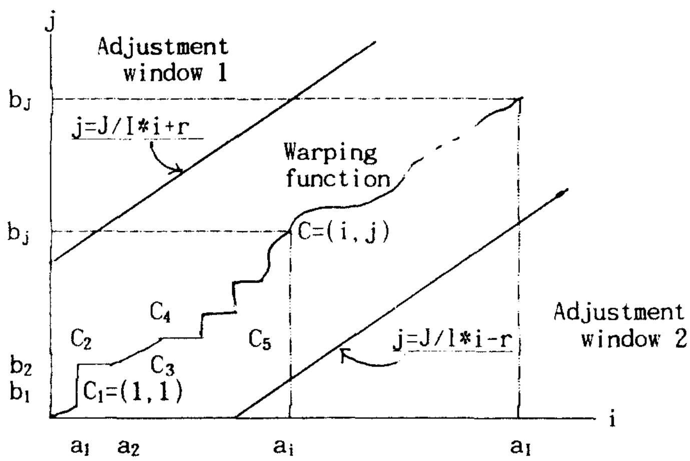
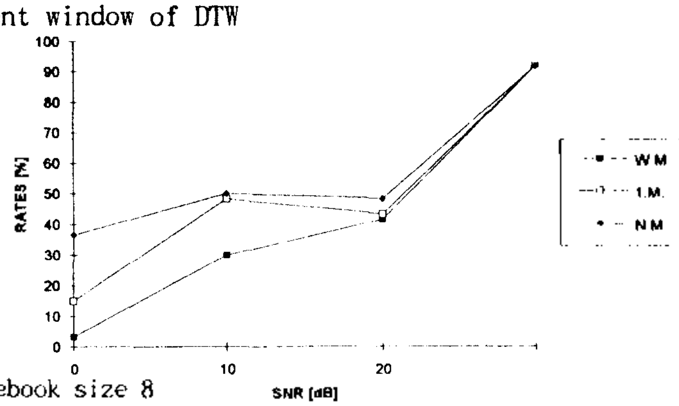


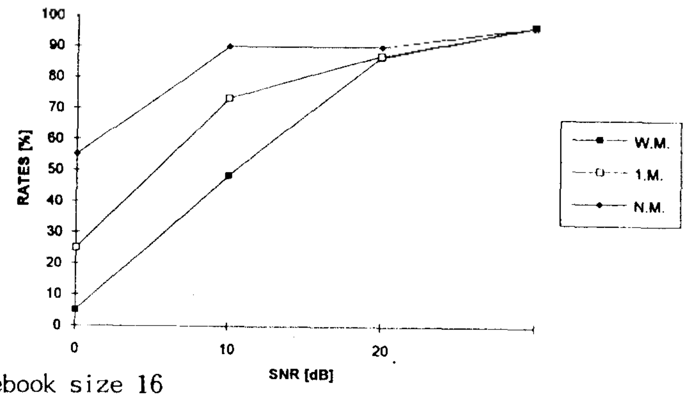
Fig. 4. Adaptive adjustment window of DTW

SNR / METHODS	0	10	20	∞
WM	3.3	30.0	41.6	91.6
IM	15.0	48.3	43.3	91.6
NM	36.6	50.0	48.3	91.6



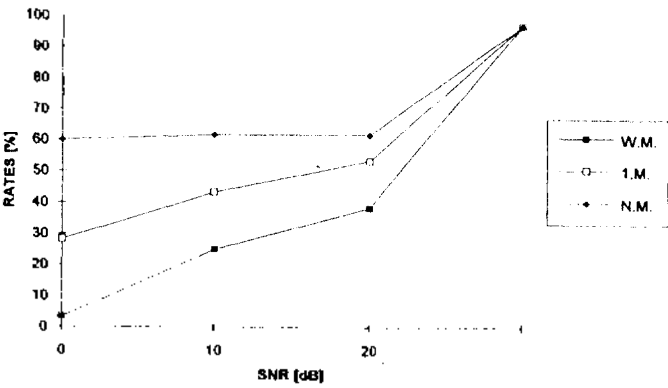
(a) Codebook size 8

SNR / METHODS	0	10	20	∞
WM	5.0	48.3	86.6	96.6
1M	25.0	73.3	87.0	96.6
NM	55.0	90.0	90.0	96.6



(b) Codebook size 16

SNR / METHODS	0	10	20	∞
WM	3.3	25.0	38.3	96.6
1M	28.3	43.3	53.3	96.6
NM	60.0	61.6	71.6	96.6



(c) Codeb

Fig. 5. Recognition rates for SNR levels

(where, WM : without spectral mapping,
1M : using 1 time spectral mapping,
NM : using n times spectral mapping.)

REFERENCES.

1. K.Shikano,K.F.Lee,R.Reddy,"Speaker Adaptation Through Vector Quantization," Proc. ICASSP 86,pp.49.5,1986
2. R.M.Stern,"Dynamic Speaker Adaptation for Feature-Based Isolated Word Recognition," IEEE Trans.on ASSP, pp.751-763, 1987
3. A.Acero,R.M.Stern,"Environmental Robustness in Automatic Speech Recognition," Proc.ICASSP 90,pp.849-852,1990
4. J.Porter,S.F.Boll,"Optimal Estimators for Spectra Restoration of Noisy Speech," Proc.ICASSP84, pp.18A.2.1-4., 1984
5. T.G.Stockham,et al.,"Blind Deconvolution Through Digital Signal Processing," Proc. IEEE,Vol.63, pp.678-692, 1975
6. D.Van Compemolle,"Noise Adaptation in a HMM Speech Recognition System," Computer, Speech and Language, Vol.3,pp.151-167, 1989
7. B.Widrow,et al.,"Adaptive Noise Cancelling Principles and Applicationa."Proc. IEEE. Vol.63, pp.1692-1716, 1975
8. V.R.Viswanathan,C.M.Henry,"Noise-immune Multisensor Speech Input:Formal Subjective Testing in Operational Conditions,"ICASSP89, pp.373-376, 1989
9. M.K.Brown,L.R.Rabiner,"An Adaptive, Ordered, Graph Search Technique for Dynamic Time Warping for Isolated Word Recognition," IEEE Trans. on ASSP,pp.535 -544, 1982