ESTIMATION OF LPC CEPSTRUM VECTOR OF SPEECH CONTAMINATED BY ADDITIVE NOISE AND ITS APPLICATION TO SPEECH ENHANCEMENT

Hidefumi KOBATAKE and Hideta SUZUKI

Graduate School of Bio-Applications and Systems Engineering
Tokyo University of Agriculture and Technology
Koganei, Tokyo 184, JAPAN
Tel. +81 423 88 7147, FAX: +81 423 85 5395, E-mail: kobatake@cc.tuat.ac.jp

ABSTRACT

This paper presents a new method for speech enhancement. It is well known that Wiener filtering is effective in reducing additive noises and the proposed method is based on it. This paper focuses on the design of Wiener filter, where we place emphasis on the recovery of original formant characteristics and the smooth transition of speech spectrum. Transformation method of LPC cepstrum vector extracted from noisy speech to reduce noise effects is given, which gives an estimated LPC cepstrum vector of original speech. Sharpening of formant peaks and eliminating false spectral peaks are necessary for high quality speech restoration and they are realized by the proposed method. Experiments of noise reduction have been performed, whose results show the effectiveness of the proposed method.

1. INTRODUCTION

With the advent of wide practical application age of digital speech processing technology, the role of noise suppression in speech processing problems has taken on an increased importance. Spectral subtraction [1] and Wiener filtering [2]-[6] are representative speech enhancement algorithms. An artificial experiment showed that Wiener filtering restores high quality speech if the Wiener filter is designed using true speech and noise spectra [4]. One of the authors has shown that the combination of comb filter and Wiener filter is an optimal noise suppression scheme in the sense of maximum likelihood estimation of speech waveform [5]. However, the pitch estimation of noisy speech is difficult and the application of comb filtering is not practical. We can say, after all, that the primary problem in the area of speech enhancement is how to design Wiener filter.

Formants represent acoustic characteristics of phonemes. Additive noise suppresses formant peaks in general. They are sometimes lost and moreover false spectral peaks appear. Sharpening formant peaks and eliminating false spectral ones are necessary for high quality speech restoration. This paper presents a new method to restore high quality speech waveform from noisy speech. It is based on Wiener filtering. This paper focuses on the design of Wiener filter, where we place emphasis on the recovery of original formant characteristics and the smooth transition of speech spectrum.

2. SYSTEM FORMULATION

2.1 Optimal System and Practical One

One of the authors showed that the optimal noise suppression system is the combination of comb filtering and Wiener filtering as shown in Figure 1 [5]. For voiced speech, comb filter to extract its harmonic components is necessary. For the purpose, precise pitch extraction is required. Figure 2 shows the relationship between the pitch estimation error and the degradation in SN ratio caused by the error. Precise pitch estimation is a difficult task for noisy speech and we can say that it does more harm than good to apply unreliable comb filter. In the following, we adopt only Wiener filter for speech enhancement.

If speech spectrum can be estimated precisely, Wiener filter can be designed as follows.

$$H(\omega) = \left(\frac{\hat{P}_s(\omega)}{\hat{P}_s(\omega) + \alpha \hat{P}_n(\omega)} \right)^{\beta}$$

where, P means the power spectrum and its suffixes s and n mean speech and noise, respectively. The problem is how to estimate speech spectrum precisely.

2.2 Estimation of Cepstrum Coefficients

Figure 3 shows the procedure to estimate speech spectrum. It is basically LPC cepstrum vector quantization system. In the following, an LPC cepstrum coefficient vector is called as an LPC vector in short.

Part (a) of Figure 3 shows the basic system configuration, in which two codebooks are used. Codebooks I and II are formed using clean speech database and noisy one, respectively, and every entry of the Codebook I has a corresponding entry in the Codebook II. The first processing step of this system is an LPC cepstrum vector quantization of input noisy

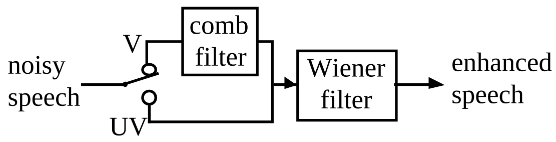

Figure 1 Optimal system for speech enhancement.

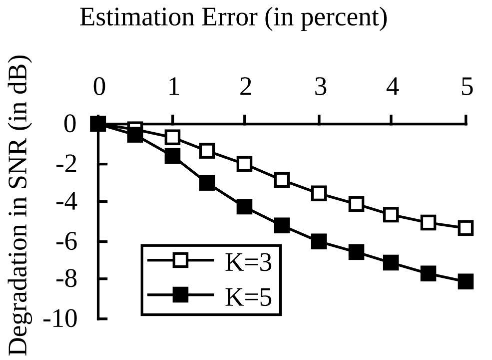

Figure 2 Degradation of SNR caused by pitch estimation error.

speech using the Codebook II which is formed from noisy speech database. The estimated code is replaced with the corresponding entry of the Codebook I. This replacement makes it possible to recover impaired spectral peaks and to eliminate false spectral peaks. The third processing step is the smoothing of the estimated code sequence. The smoothing algorithm is a simple averaging of seven LPC cepstrum coefficients obtained from successive seven analysis frames.

Part (b) of Figure 3 shows another system to restore speech waveform. In this system, only Codebook I is employed. In place of the vector quantization using the Codebook II, LPC vector is transformed to compensate noise effects. This compensated LPC vector is quantized using the Codebook I. As can be seen easily, the Codebook II depends on not only the signal-to-noise ratio (SNR) but also noise characteristics. It means that in the case of the system A, the Codebook II must be adapted to every noise environment. If the transformation of the system B works successfully, it may be advantageous over making the Codebook II adapted to various noise environments.

2.3 Compensation of Noise Effects

The dynamic range of noisy speech spectrum decreases as its SNR decreases. The squared norm of an LPC vector, which is called as the squared norm in short in the following, decreases monotonously as SNR decreases. It means, in average, that LPC vectors move to the origin of 20-dimensional space as SNR decreases. We can show the followings:

(a) The decrease in the squared norm depends on the original squared norm and SNR, but the larger the original squared norm is, the bigger its decrease is.

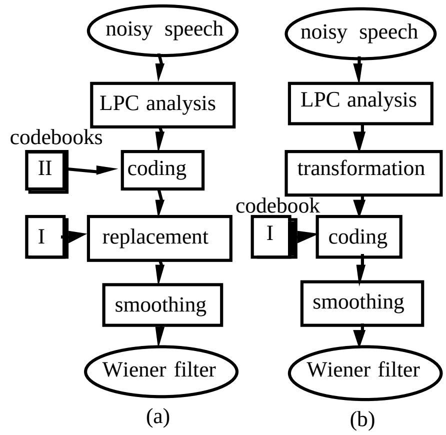

Figure 3 The basic system and modified one

(b) Approximate SNR of the noisy speech can be estimated using the squared norm. Experimental results show that the averaged decrease normalized by the standard deviation averaged squared norm and its standard deviation without information on SNR.

The following formula has been derived from the analysis described above.

$$c'_{mn} = \left[\frac{a}{\sigma}\left\{\left(bX + c\right) + d\left(X_m - X\right)\right\} + 1\right] c_{mn}$$

where, c_{mn} and c_{mn}' are the n-th cepstrum coefficient and its transformed one of the m-th analysis frame, respectively and S and Xm are the squared norm averaged over a short period and the squared norm of the m-th analysis frame, respectively. The parameter σ is the standard deviation of the squared norm. The other parameters are constant determined experimentally.

The above formula can be applied to white noise. If the additive noise not white but colored, additional processing is required as follows:

step 1: Whitening filter is applied to the noisy speech.
step 2: Transformation is applied to the LPC vector obtained from the filtered signal.
step 3: Deconvolution to obtain the compensated LPC vector of speech component.
step 4: Coding using the Codebook I.

3. EXPERIMENTS

3.1 Experimental Materials

ATR database consisting of 216 phoneme-balanced words was used to form the Codebook I. It includes utterances of 6 males and 6 females. Each utterance was sampled at 16 kHz with 16 bits accuracy. White noise is used as additive noise. Preliminary experiments showed that formant-weighted distance measure is desirable and then Root-Power Sum distance measure is adopted.

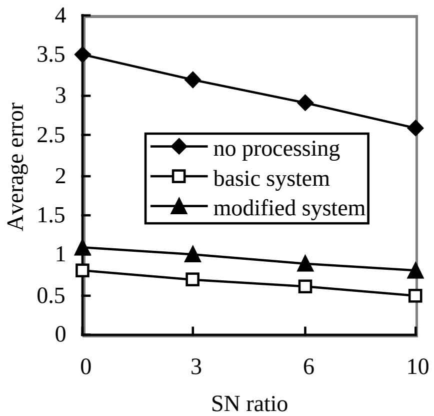

Figure 4 Estimation errors of LPC cepstrum vector.

3.2 Formation of Codebooks

Liner prediction analysis was applied to each analysis frame with 256 samples. The order of the linear prediction model is twelve and 12 LPC coefficients are transformed into 20 LPC cepstrum coefficients. The corpus of 20-dimensional vectors are clustered into 128 representative vectors which are the entries of the Codebook I.

The formation of the Codebook II is as follows. First, an analysis frame is searched which gives the minimum LPC cepstrum distance to one of entries of the Codebook I. Second, noise is added to the analysis frame to generate noisy speech and then its LPC vector is obtained. By changing additive noise waveforms, noisy LPC vectors are obtained. Their center of gravity in the 20-dimensional space gives the entry of the Codebook II. Such procedure was applied to all the entries of the Codebook I. Therefore, every entry of the Codebook II has its corresponding entry of the Codebook I.

3.3 Results

3.3.1 Errors of Estimated LPC Vector

Figure 4 shows estimated errors of LPC vectors. The error for the case where no noise suppression method is applied shows the average deviation of LPC vectors from those of the original clean speech by the additive noise. We can see that the error of the basic system is very small and LPC cepstrum vector can be recovered by the proposed system. The performance of the modified system is not so good as the basic system. In the case of the basic system, however, the Codebook II must be prepared adapted to each SNR. On the other hand, the modified system needs no *a priori* information on SNR. Therefore, we can say that the transformation of LPC cepstrum coefficients works satisfactorily.

3.3.2 Improvement in SNR

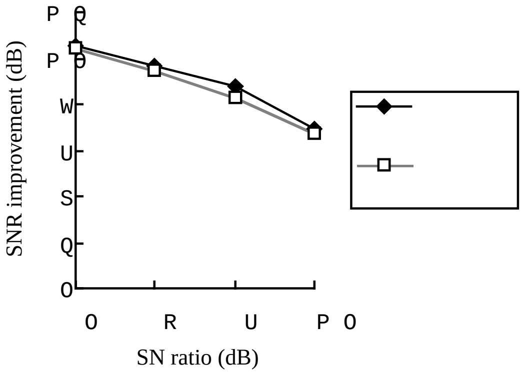

Figure 6 Improvement in SNR

Parts (a), (b), (c) and (d) of Figure 5 show power spectra of the original speech, its noisy speech whose SNR is 0 dB, the restored speech by the basic system, and that by the modified system, respectively. We can see that formant characteristics are well restored by the proposed method. Figure 6 shows the improvement in SNR by the proposed methods. About 10 dB improvement in SNR can be expected by the proposed method if the SNR of the observed signal is as low as 0 dB.

Sound files are stored in the CD-ROM. You can check the quality of the restored speech. The file names of the original noisy speech(SNR=6dB), those enhanced by the basic system and the modified one are A0809S01.WAV, A0809S02.WAV and A0809S03.WAV, respectively. You can also access my Home Page (http://www.tuat.ac.jp/~khbase) and evaluate the proposed method.

4. CONCLUSION

This paper proposed new speech enhancement methods based on LPC cepstrum vector quantization. One employs two LPC cepstrum coefficient codebooks whose entries make a pair to each other. The other one employs only one codebook and LPC cepstrum coefficients are transformed by an experimentally determined formula. Experimental results show that the proposed systems work well and formant characteristics can be well restored.

This work was supported in part by the Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture, Japan.

References

[1] S.L. Boll, IEEE Trans. on ASSP, Vol.ASSP-27, No.2, pp.113-120, 1979.
[2] J.S. Lim and A.V. Oppenheim, IEEE Trans. on ASSP,

Vol.ASSP-26, No.3, pp.197-210, 1978.
[3] B.L. Pellom and J.H.L. Hansen, Proc. ICASSP, 1996.
[4] Y. Ephraim, D. Malah and B.H. Juang, IEEE Trans. Acoust., Speech, and Signal Process., Vol.ASSP-37, pp.1846-1856, 1989.
[5] H. Kobatake and K. Gyoutoku, Proc. ICASSP, pp.973-976, May 1991.
[6] L. Arslan, A. McCree, and V. Viswanathan, Proc. ICASSP, 1995.

.

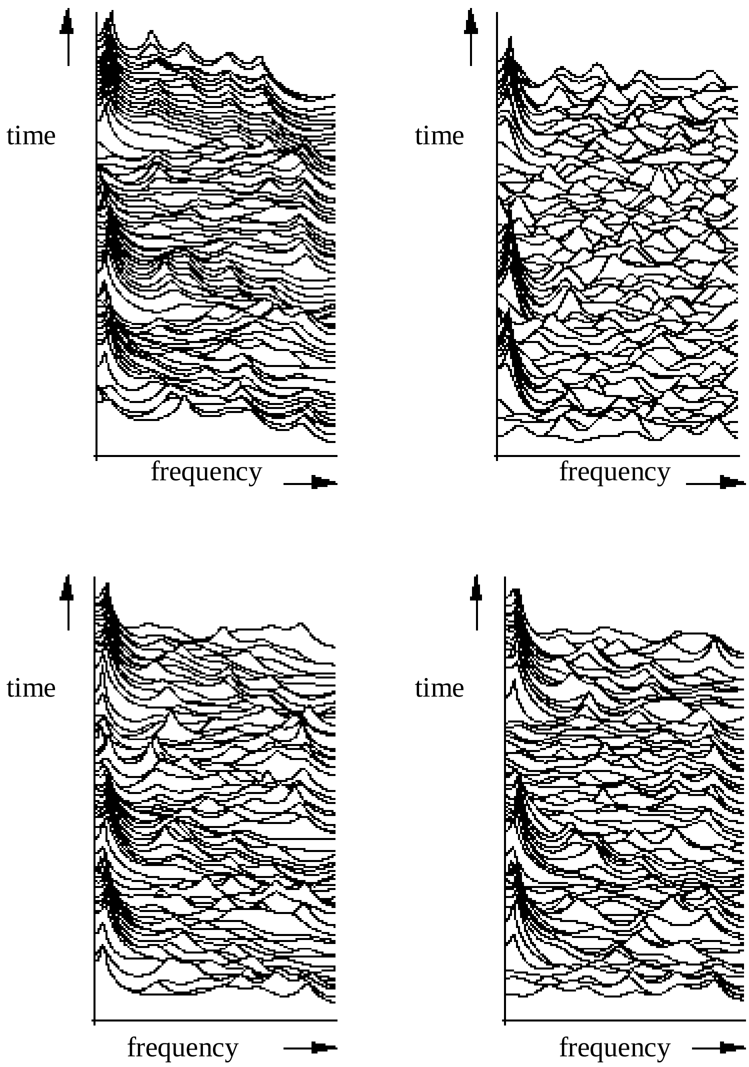

Figure 5 Spectral envelopes