Pseudo-Cepstral Representation of Speech Signal and Its Application to Speech Recognition

음성 신호의 의사 켑스트럼 표현 및 음성 인식에의 응용

Hong-Kook Kim*, Hwang-Soo Lee*
김 홍 국*, 이 황 수*

Abstract

In this paper, we propose a pseudo-cepstral representation of line spectrum pair(LSP) frequencies and evaluate speech recognition performance with cepstral lift using the pseudo-cepstrum. The pseudo-cepstrum corresponding to LSP frequencies is derived by approxmating the relationship between LPC-cepstrum and LSP frequencies. Three cepstral liftering procedures are applied to the pseudo-cepstrum to improve the performance of speech recognition. They are the root-power-sums lifter, the general exponential lifter, and the bandpass lifter. Then, the liftered psedudo-cepstra are warped into a mel-frequency scale to obtain feature vectors for speech recognition. Among the three lifters, the general exponential lifter results in the best performance on speech recognition. When we use the proposed pseudo-cepstra feature vectors for recognizing noisy speech, the signal-to-noise ratio (SNR) improvement of about 5~10dB over LSP is obtained.

요 약

본 논문에서는 line spectrum pair (LSP)의 의사 켑스트럼 표현을 제안하고 이 의사 켑스트럼에 켑스트럼 lifter를 적용하여 얻은 특징 벡타를 이용하는 음성 인식 시스템의 성능을 평가한다. 의사 켑스트럼 표현은 LSP와 LPC 켑스트럼 사이의 관계로부터 근사적으로 유도된다. 이때 음성 인식 시스템의 성능을 더욱 향상 시키기 위하여 root-power-sums lifter, general exponential lifter (GEL), 그리고 bandpass lifter 등과 같은 켑스트럼 lifter가 의사 켑스트럼에 적용된다. 또한 mel 주파수로의 변환도 행해진다. 인식 실험 결과, GEL로 liftering된 mel 주파수 의사 켑스트럼이 가장 좋은 성능을 나타내며, LSP에 비해 5~10dB정도의 신호대잡음비의 개선을 얻을 수 있다.

I. Introduction

A line spectrum pair (LSP) representation of speech signal has been introduced by Itakura [1] as an alternation of linear predictive coding (LPC) and widely used in the speech processing areas including speech coding, synthesis, and recognition [2]. The LSP representation has better quantization property than other LPC representations, such as the log area ratio (LAR) and the partial correlation (PARCOR) coefficients. Since LSP frequencies are frequency domain

*한국과학기술원 정보및 통신공학과
접수일자 : 1994년 1월 24일

parameters, the error on an LSP parameter gives rise to the spectral distortion in the neighborhood of a specific frequency corresponding to the LSP parameter [3]. It has been reported experimentally [4] that the number of bits for LSP quantization can be reduced to 70 ~ 80% of those for PARCOR to achieve the same spectral distortion. Moreover, the LSP paramenters have well behaved dynamic range and good interpolation characteristics. The stability of an LSP synthesis filter is also preserved after quantizing LSP parameters and can be checked simply.

In speech recognition based on LSP frequencies, Paliwal [5] studied several LSP distance measures for speaker-dependent steady-state vowel recognition and found that the weighted LSP distance measure shows the best performance among various measures based on the LSP frequencies and slightly better than other linear prediction distance measures. In the HMM-based isolated word recognition, the performance of the LSP representation is comparable to that of the cepstral representation [6]. Gurgen *et al.* [7] reported similar results that the LSP representation is superior to the cepstral representation for speaker-independent DTW-based isolated word recognition. From these results and the fact that a cepstral liftering procedure results in more improved recognition accuracy [8]-[9], we predict that the performance of an LSP-based recognition system can be more improved if a procedure like liftering on the cepstral representation is applied to LSP frequencies. In order to apply the weighting procedure to LSP frequencies, we first derive a relationship between cepstral coefficients and LSP frequencies. We define a pseudo-cepstral representation by approximating this relationship, and then a liftering procedure is applied to the pseudo-cepstral coefficients. In this paper, we consider three kinds of lifters : the root-power-sums lifter (RPS), the general exponential lifter (GEL), and the bandpass lifter (BPL).

Following this introduction, we review the LSP analysis method in Section 2. In Section 3, we derive the pseudo-cepstral representation and explain the cepstral liftering procedures applied in this work. Also, mel-scaled warping of pseudo cepstrum is described. The performance results on recognition experiments for the task of recognizing 10 confusable syllables are shown in Section 4. The recognition results obtained by using the pseudo-cepstrum with/without liftering procedures are compared with those by using LSP frequencies. The effects of mel-scaling of the pseudo-cepstrum and LSP on recognition performance are also obtained.

Ⅱ. LSP Representation

A short-time frame of speech signal is modeled by a pth order all-pole filter $H_p(z)=\frac{\alpha_p}{A_p(z)}$ in LPC analysis, where α_p is a filter gain corresponding to the root mean value of error residuals. The pth order linear prediction filter (or inverse filter) $A_p(z)$ is described as

$$A_p(z)=1+a_1z^{-1}+\cdots+a_pz^{-p} \tag{1}$$

where $\{a_1,\cdots,a_p\}$ are the linear predictive coefficients of order p. Instead of a direct form implementation of the inverse filter, it is possible to implement the filter in lattice form using the pth reflection coefficient, k_p, and the $(p-1)$th analyis filter $A_{p-1}(z)$ as follows.

$$A_p(z)=A_{p-1}(z)+k_pB_{p-1}(z), \tag{2}$$

$$zB_p(z)=k_pA_{p-1}(z)+B_{p-1}(z), \tag{3}$$

with

$$A_0(z)=zB_0(z)=1, \tag{4}$$

where $B_p(z)$ is a pth order backward linear prediction filter which estimates the current sample

based on future samples, and has the relationship, $B_p(z) = z^{-(p+1)} A_p(z^{-1})$.

To obtain an LSP representation of order p, the lattic filter of order $(p+1)$ can be exended from the pth order filter by setting the $(p+1)$th reflection coefficient k_{p+1} to the two extreme values of ± 1. This means that the vocal tract at the sound source of the lossless tube model is completely closed or completely open and the power in the vocal tract is zero. Thus a symmetric polynomial $P_{p+1}(z)$ and an anti-symmetric polynomial $Q_{p+1}(z)$ are obtatined as follows.

$$P_{p+1}(z) = A_p(z) + B_p(z), \tag{5}$$

$$Q_{p+1}(z) = A_p(z) - B_p(z). \tag{6}$$

Real zeros of the polynomials $P_{p+1}(z)$ and $Q_{p+1}(z)$ are $z=-1$ and $z=+1$, respectively, and all the other zeros are complex. These complex zeros detemine the LSP frequencies of order p. $P_{p+1}(z)$ and $Q_{p+1}(z)$ have the following important properties :

(1) All zeros of $P_{p+1}(z)$ and $Q_{p+1}(z)$ lie on the unit circle.

(2) Zeros of $P_{p+1}(z)$ and $Q_{p+1}(z)$ are interlaced with each other.

(3) Minimum phase property of $A_p(z)$ is easily preserved after quantization of the zeros of $P_{p+1}(z)$ and $Q_{p+1}(z)$.

Since the zeros of $P_{p+1}(z)$ and $Q_{p+1}(z)$ are on the unit circle, they can be expressed as $\{e^{j\theta_i}\}$ $(i=1, 3, \cdots, \frac{p}{2}-1)$ or $\{e^{j\theta_i}\}$ $(i=2, 4, \cdots, \frac{p}{2})$ for both $P_{p+1}(z)$ and $Q_{p+1}(z)$ when p is odd and even, respectively, and these $\{\theta_i\}$ are called the LSP frequencies of order p.

To calculate $\{\theta_i\}$ from $\{a_i\}$, $A_p(z)$ is first converted to $P_{p+1}(z)$ and $Q_{p+1}(z)$, and then a search procedure is required in the frequency axis after applying the fast Fourier transform (FFT) or discrete cosine transform (DCT) to $P_{p+1}(z)$ and $Q_{p+1}(z)$. Another method is to project the polynomials of (5) and (6) on the real axis with $x = \cos\theta = \frac{z+z^{-1}}{2}$, and then the roots of the projected polynomials are approximately found and converted into the LSP frequencies with $\theta = \cos^{-1} x$.

The synthesis filter $H_p(z)$ can also be obtained from the polynomials of $P_{p+1}(z)$ and $Q_{p+1}(z)$ as follows :

$$H_{p+1}(z) = \frac{\alpha_p}{A_p(z)} = \frac{\alpha_p}{1+\{(P_{p+1}(z)-1)+(Q_{p+1}(z)-1)\}/2} \tag{7}$$

By using the LSP frequencies, we can construct a synthesis filter without LPC coefficients.

Ⅲ. Pseudo-cepstral Representation

3.1 Pseudo-cepstral Conversion from LSP Frequencies

The polynomials $P_{p+1}(z)$ and $Q_{p+1}(z)$ of (5) and (6) can be rewritten in terms of LSP frequencies $\{\theta_i\}$ as follows [11]. When p is even,

$$P_{p+1}(z) = (1-z^{-1}) \prod_{i=2, even}^{p} (1-2\cos\theta_i z^{-1}+z^{-2}), \tag{8}$$

$$Q_{p+1}(z) = (1+z^{-1}) \prod_{i=1, odd}^{p} (1-2\cos\theta_i z^{-1}+z^{-2}). \tag{9}$$

When p is odd,

$$P_{p+1}(z) = (1-z^{-2}) \prod_{i=2, even}^{p} (1-2\cos\theta_i z^{-1}+z^{-2}), \tag{10}$$

$$Q_{p+1}(z) = \prod_{i=1, odd}^{p} (1-2\cos\theta_i z^{-1}+z^{-2}), \tag{11}$$

By multiplying $P_{p+1}(z)$ and $Q_{p+1}(z)$, we obtain the following equation for any $p \geq 1$.

$$P_{p+1}(z)\, Q_{p+1}(z) = A_p^2(z)[1-R_{p+1}^2(z)]$$

$$= (1-z^{-2}) \prod_{i}^{p} (1-e^{j\theta_i} z^{-1})(1-e^{j\theta_i} z^{-1}), \tag{12}$$

where $R_{p+1}(z) = z^{-(p+1)} A_p(z^{-1})/A_p(z)$. Taking the logarithm on both sides of (12) gives

$$2\ln A_p(z) + \ln(1 - R^2_{p+1}(z)) = \ln(1 - z^{-1}) + \sum_{i=1}^{p} [\ln(1 - e^{j\theta_i} z^{-1}) + \ln(1 - e^{-j\theta_i} z^{-1})]. \quad (13)$$

Since the right hand side of (13) has zeros on the unit circle, the zeros should be shifted radially by a factor of $\alpha (0 < \alpha < 1)$ [12]. In other words, a different contour for the computation of inverse z-transform must be used. Therefore both sides of (13) becomes

$$2\ln A_p(\alpha^{-1} z) + \ln(1 - R^2_{p+1}(\alpha^{-1} z)) = \ln(1 - \alpha^{-1} z^{-1}) + \sum_{i=1}^{p} [\ln(1 - \alpha e^{j\theta_i} z^{-1}) + \ln(1 - \alpha e^{-j\theta_i} z^{-1})]. \quad (14)$$

And by using the power series expansion

$$\ln(1 - \alpha z^{-1}) = -\sum_{n=1}^{\infty} \frac{\alpha^n}{n} z^{-n}, \quad |z| > |\alpha|, \quad (15)$$

the inverse z-transform of right side of (15) has the form

$$-\sum_{n=1}^{\infty} \frac{\alpha^n}{n} z^{-n} - \sum_{n=1}^{\infty} \frac{(-\alpha)^n}{n} z^{-n} - \sum_{i=1}^{p} \sum_{n=1}^{\infty} \frac{\alpha^n}{n} [e^{j\theta_i n} + e^{-j\theta_i n}] z^{-n}, \quad (16)$$

and that of left side of (15) by using $\ln A_p(z) = -\sum_{n=1}^{\infty} c_n z^{-n}$ becomes

$$-2\sum_{n=1}^{\infty} \alpha^n c_n z^{-n} + 2\sum_{n=1}^{\infty} \alpha^n R_n z^{-n}, \quad (17)$$

where $\alpha^n R_n$ is defined as the inverse z-transform of

$$\frac{1}{2} \ln(1 - R^2_{p+1}(\alpha^{-1} z)). \quad (18)$$

From (16) and (17), the inverse z-transform of (14) is given as

$$\alpha^n c_n - \alpha^n R_n = \frac{\alpha^n}{2n} + \frac{(-\alpha)^n}{2n} + \sum_{i=1}^{p} \frac{\alpha^n}{2n} [e^{j\theta_i n} + e^{-j\theta_i n}]$$

$$= \frac{\alpha^n}{2n} (1 + (-1)^n) + \frac{\alpha^n}{n} \sum_{i=1}^{p} \cos \theta_i n, \; n \geq 1. \quad (19)$$

Therefore the cepstral coefficients $\{c_n\}$ can be rewritten as a function of $\{\theta_i\}$ and R_n as follows.

$$c_n = \frac{1}{2n} (1 + (-1)^n) + \frac{1}{n} \sum_{i=1}^{p} \cos \theta_i n + R_n, \; n \geq 1. \quad (20)$$

Now we will express R_n completely in terms of α and $\{c_n\}$. The phase of $R_{p+1}(e^{j\theta})$, $\Psi_{p+1}(\theta)$ is a function of $A_p(e^{j\theta})$ and becomes

$$\Psi_{p+1}(\theta) = (p+1)\theta + 2 \arg[A_p(e^{j\theta})] \quad (21)$$

$$= (p+1)\theta + 2 \sum_{n=1}^{\infty} c_n \sin \theta_n. \quad (22)$$

The second term of (22) results from

$$\ln A_p(e^{j\theta}) = \ln |A_p(e^{j\theta})| + j \arg A_p(e^{j\theta}), \quad (23)$$

and

$$\sum_{n=1}^{\infty} c_n e^{jn} = \sum_{n=1}^{\infty} c_n \cos \theta n + j \sum_{n=1}^{\infty} c_n \sin \theta n. \quad (24)$$

Since (23) and (24) are equal cepstral coefficients $\{c_n\}$ are real, $\arg[A_p(e^{j\theta})] = \sum_{n=1}^{\infty} c_n \sin \theta n$.

If we consider the shifting factor α in $R_{p+1}(e^{j\theta})$ can be given

$$R_{p+1}(\alpha^{-1} e^{j\theta}) = \alpha^{p+1} e^{-j\dot{\Psi}_{p+1}(\theta)}, \quad (25)$$

where

$$\dot{\Psi}_{p+1}(\theta) = (p+1)\theta + 2 \sum_{n=1}^{\infty} \alpha^n c_n \sin \theta n. \quad (26)$$

Next we expand $\ln(1 - R^2_{p+1}(\alpha^{-1} e^{j\theta}))$ by power seres expansion. That is,

$$\ln(1 - R^2_{p+1}(\alpha^{-1} e^{j\theta})) = -\sum_{k=1}^{\infty} \frac{1}{k} R^{2k}_{p+1}(\alpha^{-1} e^{j\theta}). \quad (27)$$

By substituting (25) into (27), we obtain

$$\frac{\cdots}{k=1}\frac{k}{k}\quad \cdots \quad \cdots \quad \cdots$$

$$-\sum_{k=1}^{p}\frac{1}{k}\alpha^{2k(p+1)+n}e^{-j2k\sum_{i=1}^{p}\theta_i}e^{-jk\sum_{i=1}^{p}\theta_i}\sum_{l=1}^{p}\alpha^{l}c_l\sin\theta l. \quad (28)$$

The inverse Fourier transform is applied to (28) and we obtain R_n as

$$R_n=\sum_{k=1}^{p}\frac{\alpha^{2k(p+1)+n}}{4\pi k}\int_{-\pi}^{\pi}\cos[(2k(p+1)+n)\theta$$

$$-4k\sum_{l=1}^{p}c_l\alpha^{l}\sin\theta l]d\theta. \quad (29)$$

For the speech signal whose spectrum has sharp peaks as nasals and vowels, the frequencies of poles are more important than the radii of them in speech recognition. So we can assume that the effect of R_n is negligible in the vowel and nasal sounds since R_n provides the magnitude information about the inverse filter $A_p(z)$. The first term of (20) is a constant. If the distance measure for $\{c_n\}$ does not use the cross correlation of each coefficient, we can ignore the constant. As a result, we only consider the mid-term of (20) which is defined as the pseudo-cepstrum (PC) of LSP frequencies. That is, the pseudo-cepstral coefficients $\{\hat{c}_n\}$ become

$$\hat{c}_n=-\frac{1}{n}\sum_{i=1}^{p}\cos n\theta_i. \quad (30)$$

$\{\hat{c}_n\}$ give the root-power-sums [13] of the filter $\hat{A}_p(z)=1/\prod_{i=1}^{p}(1-e^{j\theta_i}z^{-1})(1-e^{j\theta_i}z^{-1})$ divided by the quefrency n, where $\{\theta_i\}$ are the LSP frequencies.

Fig. 1(b) and (c) show cepstrally smoothed logarithmic spectra obtained from the LPC-cepstral coefficients and the pseudo-cepstral coefficients, respectively for a vowel speech segment of Fig. 1(a). As shown in the figures, PC shows a similar spectral representation to the LPC-cepstrum with slightly enhanced spectral peaks.

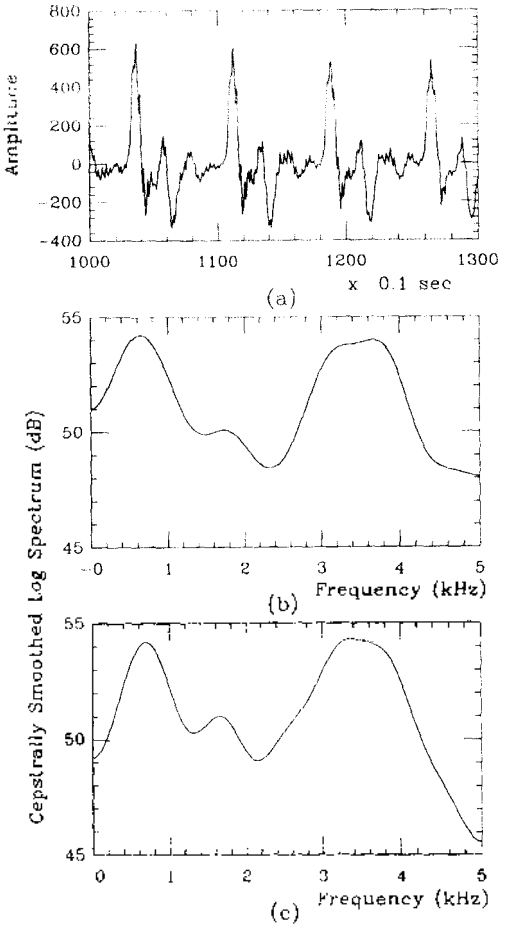


Fig. 1. Smoothed log spectra : (a) speech segement, (b) cepstrum computed from LPC cepstral coefficients, and (c) pseudo-cepstrum.

3.2 Liftered Pseudo-cepstral Representation

Cepstral lifters have been applied to cepstral coefficients in order to improve the performance of speech recognition systems based on cepstral representation [8] [9]. In general, the effect of liftering procedures is to normalize variances of cepstral coefficients which are inversely proportional to the square of quefrency, the index of cepstral coefficients [10]. Liftered cepstral coefficients can be written in a general form as follows.

$$w'_n = w_n \cdot l_n \tag{31}$$

where $\{c_n\}$ and $\{c_n'\}$ are the cepstral coefficients before and after applying a cepstral lifter, w_n respectively. The cepstral lifters are classified as a different name according to w_n. In summary,

$$w_n = \begin{cases} n & (RPS) \\ n^s & (GEL) \\ 1 + h \sin \dfrac{\pi n}{L} & (BPL), \end{cases} \tag{32}$$

Especially $s = 0.6$ is used in GEL ($G_{0.6}$) and $h = 6$ and $L = 12$ in BPL (B_{12}).

In this work we apply cepstral lifters of (32) to the pseudo cepstrum and denote each liftered pseudo cepstral coefficients as follows

$$c_{rps,n} = \sum_{i=1}^{p} \cos n\theta_i,$$

$$c_{gel,n} = n^{0.6} \sum_{i=1}^{p} \cos n\theta_i,$$

$$c_{bpl,n} = \left(1 + 6 \sin \frac{\pi n}{12}\right) \sum_{i=1}^{p} \cos n\theta_i, \; 1 \le n \le 12. \tag{33}$$

Fig. 2 shows the spectra after applying the lifters to the pseudo cepstrum. Comparison of the figure with that of Fig. 1(c) reveals that the lifters enhance spectral peaks in the cepstrally smoothed logarithmic spectra, and RPS makes a larger enhancement than other lifters.

In any cepstral representation, the distance between $\{c_i^k\}$ and $\{c_i^l\}$ is defined as

$$D(k, l) = \sum_{i=1}^{p} (c_i^k - c_i^l)^2, \tag{34}$$

where $\{c_i^k\}$ and $\{c_i^l\}$ are the cepstral representations for the kth and the lth analysis frame of speech, respectively. In the next section we examine the performance of cepstral representations including LSP frequencies, pseudo-cepstrum, and four types of liftered pseudo-cepstra for speech recognition.

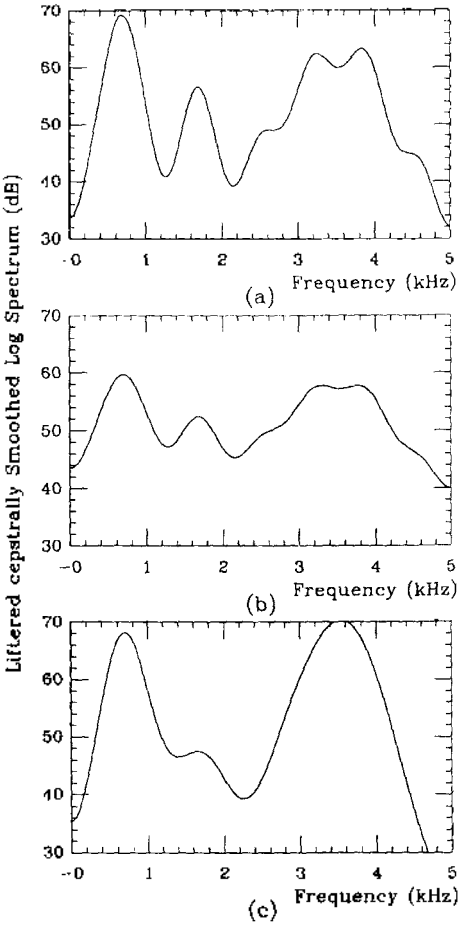


Fig 2. Pseudo-cepstrally smoothed log spectra obtained by applying (a) root-power-sums lifter, (b) geneal exponential lifter, and (c) bandpass lifter to the pseudo-cepstrum of Fig. 1.

3.3 Mel-frequency Pseudo-Cepstrum

By warping linear-scaled frequency into mel-scaled or Bark-acaled frequency, the first an and effective second formants of speech spectrum can be well represented by a rather order LPC analysis [14]. A bilinear transform introduced by Oppenheim and Johnson [15] is a method to implement the warping by expanding the low-frequency axis and compressing the high-frequency axis [16]. An all-pass filter, $H(z)$, for the frequency warping is given

$$H(z)=\frac{z^{-1}-a}{1-az^{-1}},\quad -1<a<1, \tag{35}$$

where a is a parameter of frequency warping. Therefore a warped frequency, $\hat{\theta}$ becomes

$$\hat{\theta}=\theta+2\tan^{-1}\frac{a\sin\theta}{1-a\cos\theta} \tag{36}$$

For a sampling frequency of 10kHz, $\hat{\theta}$ results in a very good approimation to the Bark-scale or mel-scale if $a=0.47$ [14]. Since the LSP frequencies are the values on the frequency axis, a bilinear transformed LSP frequency can be easily obtained by (36). Also, the bilinear transformed pseudo-cepstral coefficients can be given by

$$\hat{c}_{mel,n}=\frac{1}{n}\sum_{i=1}^{p}\cos n\hat{\theta}_i, \tag{37}$$

where $\hat{\theta}$ is a mel-scaled LSP frequency. We call $\{\hat{c}_{mel,n}\}$ the mel-scaled pseudo-cepstral coefficients (MPCC).

Ⅳ. Recognition Experiments and Discussions

4.1 Speech Data and Recognition System

The speech data for isolated speech recognition consists of 10 confusable words, spoken by 10 male and 10 female speakers. The words are the nasal consonants /m, n/ followed by one of the vowels /a, ʌ, o, u, i/. Each speaker uttered each word 10 times in a noise-free condition, which yields a total of 100 utterances for eack speaker. Each utterance was low-pass-filtered up to 4.7kHz, then sampled at 10kHz with 12 bit quantizer. End-point detection was done manually to include the whole nasal sounds.

Fig. 3 shows a block diagram of the feature extraction procedures used in this work. The speech signal is preemphasized with a factor of 0.98 and the Hamming-window with a length of 30ms at a frame rate of 100Hz is applied to the speech signal. And then the LPC analysis of order 8 to 14 with a step of 2 is applied to the speech signal. The sequences of LSP frequencies are obtained from LP coefficients by using method in [19]. The pseudo-cepstrum is extracted by (37). And then lifters are applied to the pseudo-cepstrum. The warping procedure is applied to the LSP before finding the pseudo-cepstrum. Also lifters are applied to the mel-frequency pseudo-cepstral coefficients (MPCC).

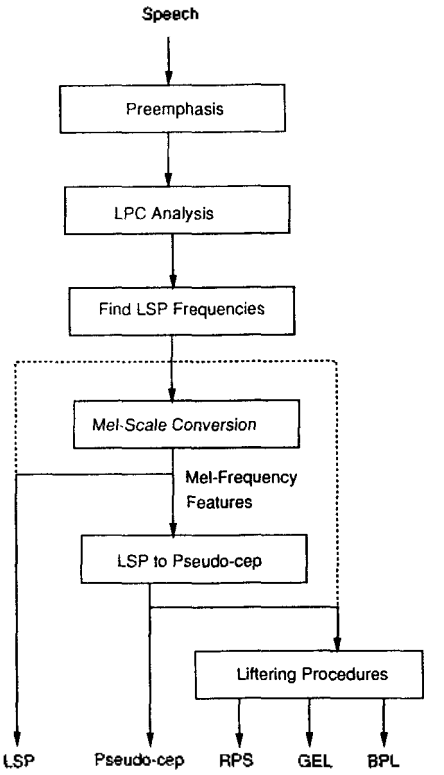


Fig 3. A block diagram of feature extraction procedures for speech recognition.

A DTW-based recognition system with the symmetric Sakoe-Chiba path constraints [17] is implemented in speaker-dependent mode. Two utterances of each word from every speaker are used as reference patterns and the other 80 utterances from each speaker as test patterns. The recognition accuracies obtained in this paper are averaged over 20 speakers. To simulate noisy

zero mean white Gaussian noise is ... the test utterances except reference ... The amount of additive noise is controlled ... segmental signal to noise ratio(SNR).

4.2 Recognition Results and Discussions

Recognition using the pseudo cepstrum and the LSP frequencies are compared in Fig. 4. The recognition results are obtained by varying the analysis order from 8 to as well as SNR's. The recognition system employing the pseudo cepstrum shows better recognition performance than that using the LSP frequencies when SNR is above 30dB. However, the results are reversed when SNR is below 20 dB. These results are common for all the analysis orders. This is because the spectrum does not show spectral peaks markedly when SNR is low. The effect of R_n in (20) increases and the pseudo-cepstrum fails to represent the spectrum effectively for speech recognition. This problem can be overcome by applying cepstral lifters to the pseudo-cepstrum since cepstral lifters enhance the spectral peaks and make the spectra be more robust in noise.

In Fig. 5, we show the average recognition accuracies for the LSP frequencies and liftered

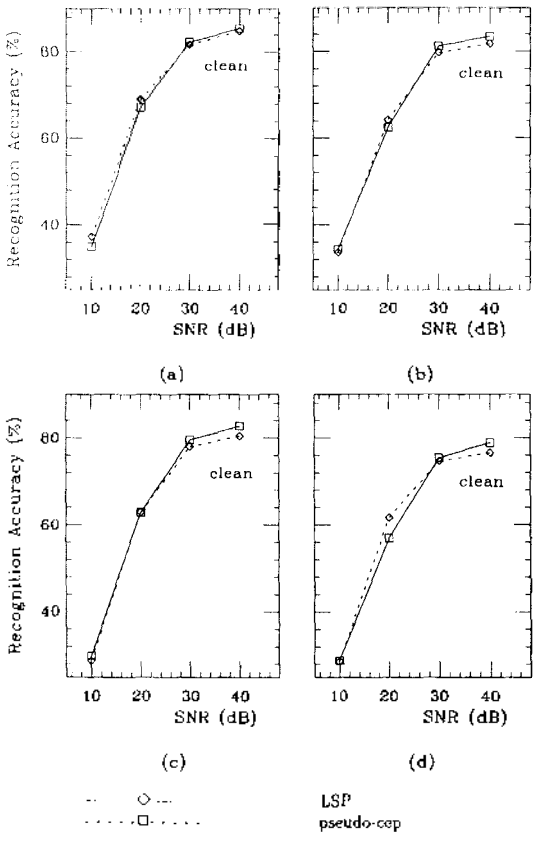


Fig 4. Average recognition of the recognition systems with pseudo-cepstrum and LSP versus SNR for 20 speakers. Orders of analysis are (a) 14, (b) 12, (c) 10, and (d) 8.

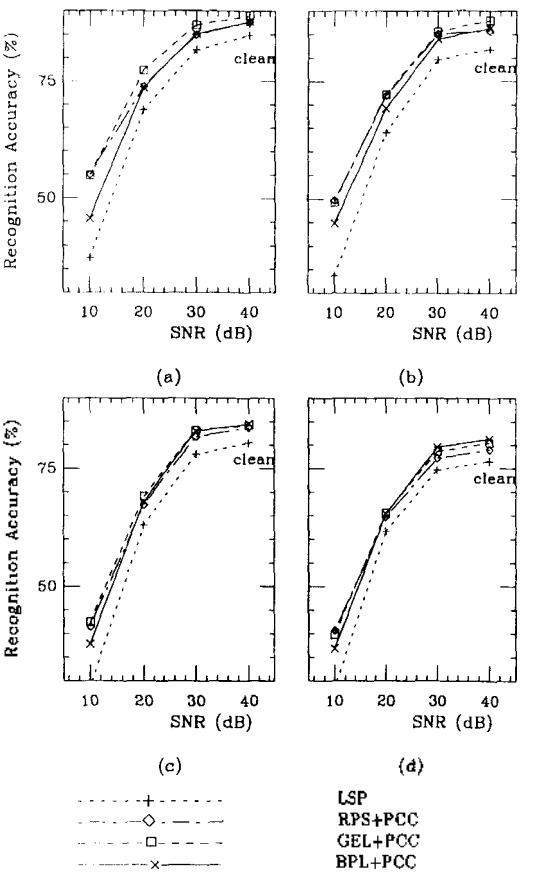


Fig 5. Average recognition results using LSP frequencies, the liftered pseudo-cepstra with analysis orders of (a) 14, (b) 12, (c) 10, and (d) 8 for 20 speakers.

pseudo cepstrum for different analysis orders from 14 to 8 with a step of 2 for 20 speakers. RPS, GEL and BPL are used to obtain the liftered pseudo-cepstra. The liftered pseudo cepstral representations provide higher recognition results than the LSP frequencies and the pseudo-cepstrum for all the analysis orders and SNR's. Especially, GEL shows the superior performance to other lifters when the analysis orders are 14, 12, and 10, and the similar performance at the analgsis order of 8. The SNR improvement of about 5 dB can be obtained by applying the liftering procedures to the pseudo-cepstrum. For comparison, we summarize the results in Table 1.

Table 1. Comparison of recognition results at the analysis order of 14

SNR	LSP	PCC	RPS+PCC	GEL+PCC	BPL+PCC
Clean	84.88	85.56	87.69	89.00	87.75
30dB	81.69	82.31	84.88	87.06	85.13
20dB	68.88	67.00	73.91	77.38	73.69
10dB	37.38	35.00	54.94	54.75	45.63

Next we extract the mel-frequency pseudo-cepstral coefficients (MPCC) by letting the warping paramenter a be 0.47. In Fig. 6, we compare the performances of PCC and MPCC. MPCC provides the higher recognition accuracy than PCC except SNR = 10dB since the noise signal enhances the high frequency of speech spectrum. Thus no warping or a slight warping is desirable in this case. Figs. 7 and 8 show performance results of Mel-LSP and liftered MPCC's at the warping parameter of 0.47 and 0.2, respectively. Liftered MPCC's always yield higher performance than Mel-LSP. The warping procedure at a = 0.2 is more effective than that at a = 0.47 except BPL since a strong warping procedure decreases the higher-order PCC in that BPL gives more weight on the lower-order PCC relatively. The results are summarized in Table 2. As a concluding remark, the MPCC liftered by GEL yields the best performance among other representations and a slight warping for PCC shows better performance.

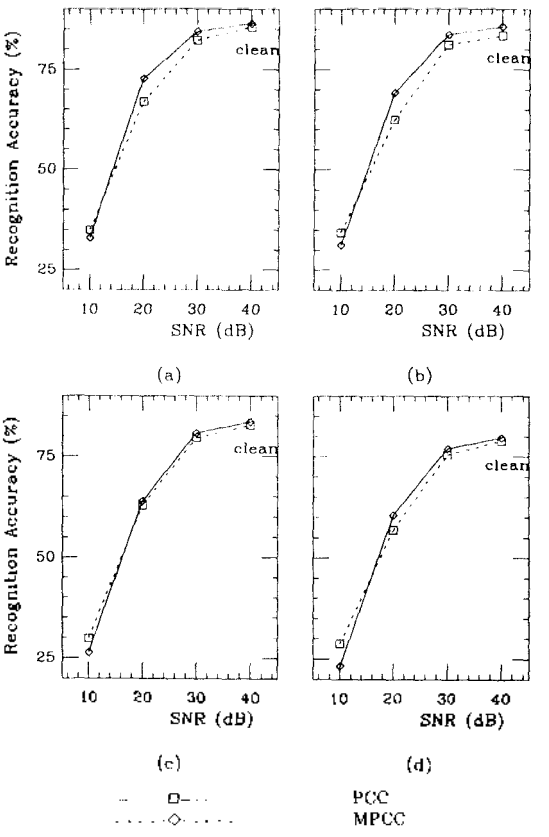
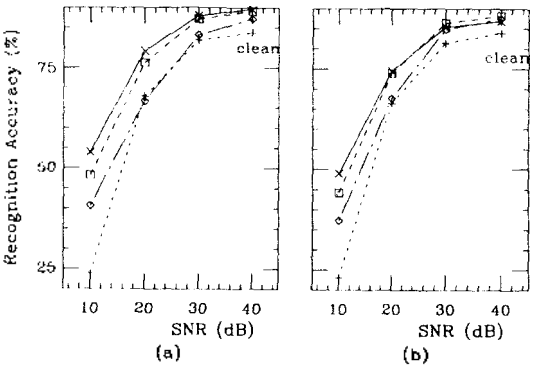


Fig 6. Average recognition results using PCC and MPCC warped at $a = 0.47$ with the analysis orders of (a) 14, (b) 12, (c) 10, and (d) 8 for 20 speakers.



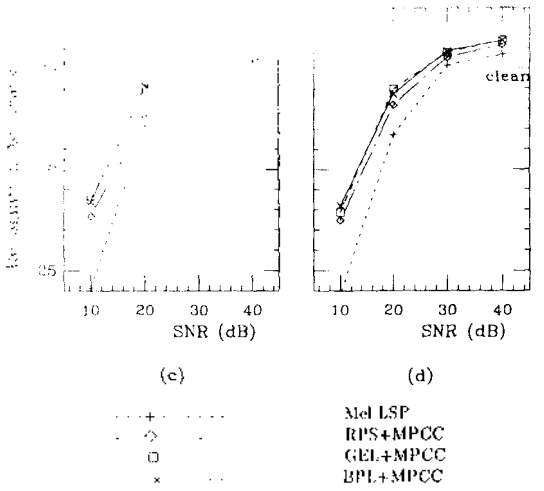


Fig 7. Average recognition results using mel-scaled LSP frequencies at $a=0.47$ and the liftered MPCC with the analysis order of (a) 14, (b) 12, (c) 10, and (d) 8 for 20 speakers.

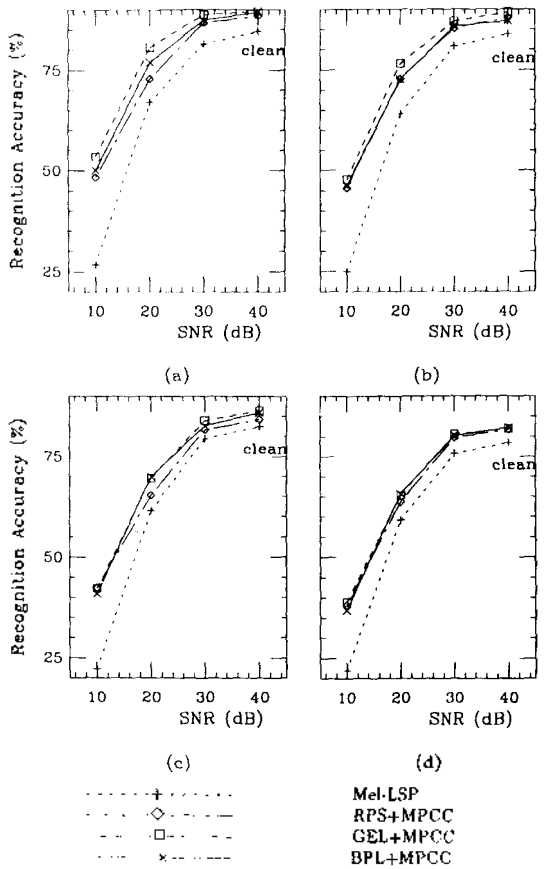


Fig 8. Average recognition results using mel-scaled LSP frequencies at $a=0.2$ and the liftered MPCC with the analysis order of (a) 14, (b) 12, (c) 10, and (d) 8 for 20 speakers.

Table 2. Comparison of recognition results at the analysis order of 14 with the warping parameter of 0.2

SNR	Mel-LSP	MPCC	RPS+MPCC	GEL+MPCC	BPL+MPCC
Clean	84.63	87.13	88.44	89.63	89.50
30dB	81.44	84.88	86.75	88.69	87.31
2dB	67.06	72.13	72.88	80.69	76.94
10dB	26.69	32.69	48.31	53.50	50.06

V. Conclusions

To improve the performance of speech recognition system based on LSP frequencies, a pseudo-cepstral representation is proposed. The pseudo-cepstrum is derived by approximating the relationship between the LPC-cepstrum and the LSP frequencies. A recognition experiment for 10 nasal-vowel sounds shows that the speaker-dependent recognition system using the pseudo-cepstrum gives higher recognition performance than that of the LSP frequencies when SNR is above 30dB. Cepstral lifters including RPS, GEL, and BPL are applied to the pseudo-cepstrum to improve the performance of the pseudo-cepstrum. All the lifters provide better recognition results for all analysis orders of 8 to 14 and SNR's, and also give SNR improvement of about 5dB over the LSP frequencies and the pseudo-cepstrum. Also mel-frequency warping is applied to the pseudo-cepstrum and the lifters are applied to the mel-frequency pseudo-cepstrum. The recognition system employing the mel-frequency pseudo-cepstrum liftered by GEL shows the best recognition accuracy.

Acknowlegement

Thanks to Dr. K. C. Kim, senior research scientist at KAIST, for his help and discussions.

References

1. F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Am.*, Vol. 57, S35(A), 1975.
2. S. Saito, "*Speech science and technology (ed.)*," Ohmsha, Tokyo, 1992.
3. G. S. Kang and L. J. Fransen, "Application of line-spectrum pairs to low bit rate speech coders," in *ICASSP-85 Proc.*, pp. 7.3.1-4.
4. N. Sugamura and F. Itakura, "Speech data compression by LSP speech analysis techniques," *Trans. on IEICE*, Vol. 64-A, No. 8, pp. 599-606, 1981.
5. K. K. Paliwal, "A study of line spectrum pair frequencies for vowel recognition," *Speech Commun.*, Vol. 8, pp. 27-33, 1989.
6. K. K. Paliwal, "A study of LSF representation for speaker-dependent and speaker-independent HMM-based speech recognition," in *ICASSP-90 Proc.*, pp. 801-804.
7. F. S. Gurgen, S. Sagayama, and S. Furui, "Line spectrum frequency-based distance measures for speech recognition," in *ICASSP-90 Proc.*, pp. 521-524.
8. K. K. Paliwal, "On the performance of the quefrency-weighted cepstral coefficients in vowel recognition," *Speech Commun.*, pp. 151-154, May 1982.
9. J. C. Junqua and H. Wakita, "A comparative study of cepstral lifters and distance measures for all pole models of speech in noise," in *ICASSP'89 Proc.*, Glasgow, Scottland, 25-28.
10. B. H. Juang, L. R. Rabiner, and J. G. Wilpon, "On the use of band-pass liftering in speech recognition," *IEEE Trans. Acoust., Speech, and Signal Process.*, Vol. 35, No. 7, pp. 947-954, July 1987.
11. S. Saito and K. Nakata, *Fundamentals of speech signal processing*, Academic Press, 1984.
12. A. V. Oppenheim and R. W. Schafer, *Digital signal processing*, Prentice-Hall, 1975.
13. M. R. Schroeder, "Direct (nonrecursive) relations between cepstrum and predictor coefficients," *IEEE Trans. Acoust. Speech, and Signal Process.*, Vol. 29, No. 2, pp. 287-301, Apr. 1981.
14. H. W. Strube, "Linear prediction on a warped frequency scale," *J. Acoust. Soc. Am.*, Vol. 68, No. 4, pp. 1071-1076, Oct. 1980.
15. A. V. Oppenheim and D. H. Johnson, "Discrete representation of signals," *Proc. of IEEE*, Vol. 60, No. 6, pp. 681-691, June 1972.
16. K. Shikano, *Evaluation of LPC spectral matching measures for phonetic unit recognition*, CMU-CS-86-108, Carnegie Mellon Univ., Pittsburgh, PA, 1986.
17. H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, and Signal Process.*, Vol. 26, No. 1, pp. 43-49, Feb. 1978.
18. H. K. Kim and H. S. Lee, "A new extraction method and ordering properties of LSP parameters," in *JTC-CSCC Proc.*, Kyungju, Korea, pp. 25-28, 1992.

▲Hong-Kook Kim
Ph. D. Candidate
Department of Information and Communication Engineering, KIAST See Vol.12 No. 1E

▲Hwang-Soo Lee
Professor
Department of Information and Communication Engineering, KAIST See Vol.12 No. 1E