EURASIP Journal on Applied Signal Processing 2003:9, 922–929

Channel Effect Compensation in LSF Domain

An-Tze Yu
Department of Computer Science, National Chubei Senior High School, Chubei, Hsinchu, Taiwan 302, Taiwan
Email: yuat@cpshs.hcc.edu.tw

Hsiao-Chuan Wang
Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan 300, Taiwan
Email: hcwang@ee.nthu.edu.tw

Received 15 April 2003 and in revised form 9 May 2003

This study addresses the problem of channel effect in the line spectrum frequency (LSF) domain. LSF parameters are the popular speech features encoded in the bit stream for low bit-rate speech transmission. A method of channel effect compensation in LSF domain is of interest for robust speech recognition on mobile communication and Internet systems. If the bit error rate in the transmission of digital encoded speech is negligibly low, the channel distortion comes mainly from the microphone or the handset. When the speech signal is represented in terms of the phase of inverse filter derived from LP analysis, this channel distortion can be expressed in terms of the channel phase. Further derivation shows that the mean subtraction performed on the phase of inverse filter can minimize the channel effect. Based on this finding, an iterative algorithm is proposed to remove the bias on LSFs due to channel effect. The experiments on the simulated channel distorted speech and the real telephone speech are conducted to show the effectiveness of our proposed method. The performance of the proposed method is comparable to that of cepstral mean normalization (CMN) in using cepstral coefficients.

Keywords and phrases: line spectrum frequency, channel distortion, channel effect compensation, robust speech recognition.

1. INTRODUCTION

Channel distortion is always a serious problem in speech recognition systems. Channel distortion may drastically degrade the performance of speech recognition [1, 2, 3]. The channel effect in the cepstral domain has been extensively studied. Many approaches have been proposed for eliminating the influence of channel distortion to speech recognition performance [4, 5, 6, 7, 8, 9]. However, few studies aim at the channel effect in the line spectrum frequency (LSF) domain. LSFs are usually the parameters used for low bit-rate speech transmission (e.g., ITU-T G.723.1, G.728, G.729, TIA IS-96, IS-127,...). A speech or speaker recognition algorithm based on LSFs is of interest in mobile communication and Internet systems [10, 11, 12, 13, 14]. Although the LSF parameters show the poor performance in a large vocabulary continuous speech recognition (LVCSR) system, they can obtain comparable performance as cepstral coefficients do in connected digits recognition or small vocabulary speech recognition systems [12, 13]. Since the LSF parameters can be extracted directly from the bit stream of encoded speech, they are the very promising features for speech recognition in some simple applications.

The effect of codec process is another factor to influence the speech quality [15]. Since the encoded speech parameters are the only available information we can use, it is hard to compensate this nonlinear channel effect. If the bit error rate in the transmission of encoded speech is negligibly low, the channel distortion comes mainly from the microphone or the handset. In this study, we deal with only the linear channel distortion due to transducers. However, the effect of codec process on recognition performance will be evaluated for comparison.

LSFs are alternative representations of linear prediction coefficients (LPCs) and have been extensively used in speech coding and synthesis [16, 17, 18, 19]. The use of LSFs directly extracted from the encoded bit stream for speech recognition is preferred since it will become unnecessary to decode the encoded speech into a waveform [10, 13, 14]. Some researches have reported that features obtained in this way are more robust in adverse environments than those from decoded speech waveform [10, 20].

In this study, we formulate the speech signal in terms of inverse filter derived from linear prediction (LP) analysis. When the speech signal is represented by the phase of inverse filter, the channel distortion can be expressed in terms of the channel phase [21]. Further derivation shows that the

mean subtraction performed on the phase of inverse filter can minimize the channel effect. Based on this finding, an iterative algorithm is proposed to remove the bias on LSFs due to channel effect.

Two series of experiments are conducted herein. The first series of experiments use simulated channel distorted speech to examine the channel effect on a digital communication system due to the handset distortion and the effect of codec process. The second series of experiments are performed on a real telephone speech to demonstrate the effectiveness of the proposed method. The experimental results show that the performance degradation caused by the codec process is worse than that by the handset distortion. The combination of the codec process and handset distortion yields the worst performance. Nevertheless, the proposed method yields significant improvements in the performance of speech recognition.

This paper is organized as follows. Section 2 briefly reviews the fundamentals of LSFs. Section 3 describes the channel effect on the phase of inverse filter and in the LSF domain. Section 4 introduces the mean normalization on the phases of inverse filters to minimize the channel effect. An iterative algorithm is then derived for removing the bias on LSFs due to channel effect. Section 5 illustrates some experimental results to show the effectiveness of our proposed methods. Section 6 draws the conclusion.

2. A BRIEF REVIEW OF LSFs

2.1. Linear prediction

In LP analysis, the speech production is modeled as a discrete-time equation,

$$x(n) = \sum_{i=1}^{M} a(i)x(n-i) + Ge(n), \tag{1}$$

where $a(1), a(2), \ldots, a(M)$ are the LPCs, M is the system order, $e(n)$ is the excitation source, and G is the gain of the excitation. Equation (1) in the z-domain is

$$X(z) = \frac{GE(z)}{A(z)}, \tag{2}$$

where

$$A(z) = 1 - \sum_{i=1}^{M} a(i)z^{-i} \tag{3}$$

is the inverse filter, and $X(z)$ and $E(z)$ are the signal and the excitation, respectively. The $G/A(z)$ is called the LP model, and is often used to characterize the spectral envelope of a speech signal.

2.2. Line spectrum frequencies

LSFs can be obtained from the LP model by defining a symmetrical polynomial $P(z)$ and an antisymmetrical polynomial $Q(z)$ in terms of the inverse filter $A(z)$:

$$\begin{aligned} P(z) &= A(z) + z^{-(M+1)}A(z^{-1}), \\ Q(z) &= A(z) - z^{-(M+1)}A(z^{-1}). \end{aligned} \tag{4}$$

The zeros of $P(z)$ and $Q(z)$ are on the unit circle and are interlaced. These zeros are complex conjugates and their angles are the LSFs.

LSF can also be computed by formulating a ratio filter as

$$R(z) = z^{(M+1)}\frac{A(z)}{A(z^{-1})}. \tag{5}$$

In radian frequency, the phase of the ratio filter is given by

$$\phi(\omega) = (M+1)\omega + 2\theta(\omega), \tag{6}$$

where $\phi(\omega)$ and $\theta(\omega)$ represent the phase of ratio filter $R(e^{j\omega})$ and the phase of inverse filter $A(e^{j\omega})$, respectively. The LSFs are frequencies at which the phase of ratio filter is equal to a multiple of π-radians; that is,

$$\phi(\omega_k) = k\pi, \quad k = 1, 2, \ldots, M. \tag{7}$$

Therefore, (6) provides another approach for calculating LSFs. In this study, (6) and (7) serve as the basis to investigate the channel effect.

3. CHANNEL EFFECT ON LSFs

3.1. Channel effect on the phase of ratio filter

For a speech signal $x(n)$, the channel distorted signal is expressed as $y(n) = x(n)^{*}h(n)$ in time domain, where $h(n)$ is the impulse response of the channel $H(z)$. By expressing the speech signal and the distorted signal in terms of inverse filters, we obtain the following relation:

$$\frac{G_y}{A_y(z)} = \frac{G_x H(z)}{A_x(z)}, \tag{8}$$

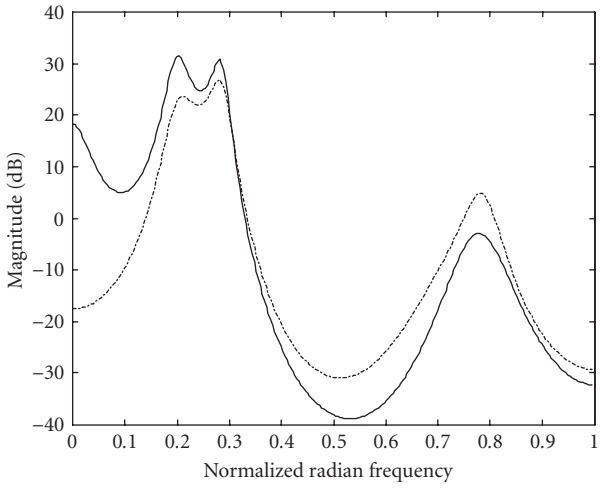
where $A_y(z)$ and $A_x(z)$ are the inverse filters of the channel distorted speech $y(n)$ and the original speech $x(n)$, respectively; G_y and G_x are the gains in the LP analysis of $y(n)$ and $x(n)$, respectively. In radian frequency, the phase of inverse filter $A_y(e^{j\omega})$ is expressed by

$$\theta_y(\omega) = \theta_x(\omega) - \theta_h(\omega), \tag{9}$$

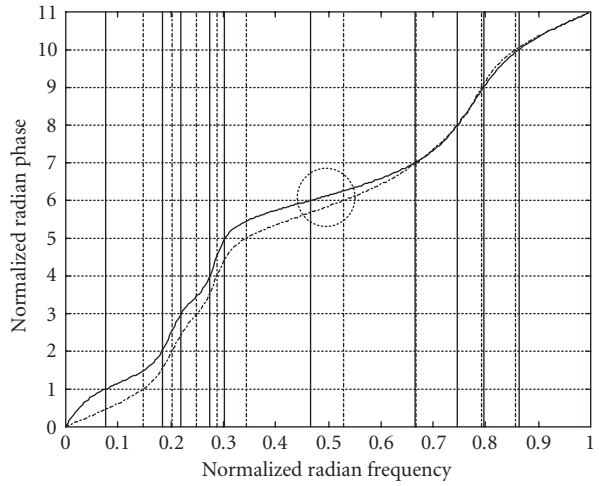
where $\theta_x(\omega)$ and $\theta_h(\omega)$ are the phases of $A_x(e^{j\omega})$ and $H(e^{j\omega})$, respectively. By the definition of (6), the phase of ratio filter for $y(n)$ is expressed as

$$\begin{aligned} \phi_y(\omega) &= (M+1)\omega + 2\theta_y(\omega) \\ &= (M+1)\omega + 2\theta_x(\omega) - 2\theta_h(\omega) \\ &= \phi_x(\omega) - 2\theta_h(\omega), \end{aligned} \tag{10}$$

where $\phi_x(\omega)$ is the phase of ratio filter for $x(n)$. This equation indicates that the channel effect causes a bias to the phase of ratio filter. Figure 1 shows an example of the channel effect on the power spectrum and the phase of ratio filter.



(a)



(b)

Figure 1: Channel effect on spectrum and phase of ratio filter for the vowel /a/. The solid curve represents the clean speech and the dotted curve represents the distorted speech. (The radian frequency is normalized by π, i.e., k means $k\pi$.) (a) Channel effect on the spectrum. (b) Channel effect on the phase of ratio filter.

3.2. Channel effect on LSFs

Starting from the channel effect on the phase of ratio filter, we want to derive the channel effect on LSFs. At first, we look at the curve of phase of ratio filter for $y(n)$ and $\phi_y(\omega)$. The mean slope of the curve between ω_k^x and ω_k^y is defined by

$$s_y(\omega_k^x, \omega_k^y) = \frac{\phi_y(\omega_k^y) - \phi_y(\omega_k^x)}{\omega_k^y - \omega_k^x}, \tag{11}$$

where ω_k^x and ω_k^y are the kth LSFs for $x(n)$ and $y(n)$, respectively (see Figure 2). According to (7), we find that

$$\phi_y(\omega_k^y) = \phi_x(\omega_k^x). \tag{12}$$

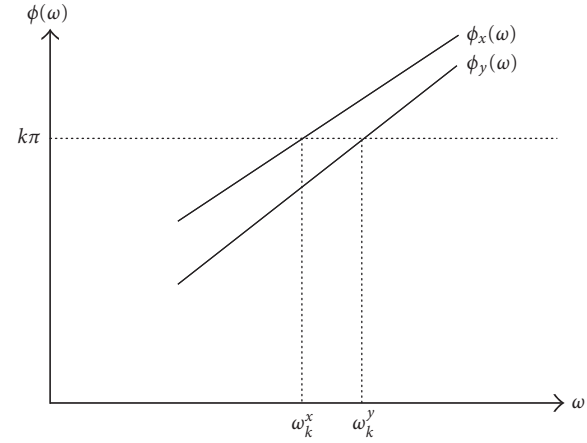


Figure 2: The shift of the phase of ratio filter due to the channel effect (see the circled region of Figure 1b).

Substituting (12) into (11) and applying the relationship of (10), we rewrite (11) as

$$\begin{aligned} s_y(\omega_k^x, \omega_k^y) &= \frac{\phi_x(\omega_k^x) - \phi_y(\omega_k^x)}{\omega_k^y - \omega_k^x} \\ &= \frac{2\theta_h(\omega_k^x)}{\omega_k^y - \omega_k^x}. \end{aligned} \tag{13}$$

Rearranging (13), we get

$$\omega_k^y = \omega_k^x + \frac{2}{s_y(\omega_k^x, \omega_k^y)}\theta_h(\omega_k^x). \tag{14}$$

The above equation states that the channel effect on LSFs is a bias which is in terms of the slope and the channel phase.

4. COMPENSATION OF CHANNEL EFFECT

Equation (14) indicates that the bias of LSFs resulted from channel effect can be compensated if the slope $s_y(\omega_k^x, \omega_k^y)$ and the channel phase $\theta_h(\omega_k^x)$ are available. However, the channel phase is hard to be estimated.

We assume that the channel effect is stationary in an utterance. By taking the average over the whole utterance on (9), we obtain

$$\begin{aligned} \overline{\theta}_y(\omega) &= \frac{1}{L}\sum_{m=1}^{L}\theta_{y,m}(\omega) \\ &= \frac{1}{L}\sum_{m=1}^{L}\theta_{x,m}(\omega) - \theta_h(\omega) \\ &= \overline{\theta}_x(\omega) - \theta_h(\omega), \end{aligned} \tag{15}$$

where m is the frame index and L is the number of frames in an utterance. If we subtract the mean from each phase of inverse filter for $y(n)$, it comes out that

$$\begin{aligned}\hat{\theta}_{y,m}(\omega) &= \theta_{y,m}(\omega) - \overline{\theta}_y(\omega)\\ &= \theta_{y,m}(\omega) - \overline{\theta}_x(\omega) + \theta_h(\omega)\\ &= \theta_{x,m}(\omega) - \overline{\theta}_x(\omega)\\ &= \hat{\theta}_{x,m}(\omega).\end{aligned} \tag{16}$$

The result is exactly the mean subtracted phase of inverse filter for $x(n)$. It implies that the mean subtraction on the phase of inverse filter will eliminate the channel phase. By using the mean subtracted phase of inverse filter to find LSFs, the channel effect on LSFs will be minimized. Hence we formulate the equation as follows to solve LSFs:

$$\begin{aligned}\hat{\phi}_{y,m}(\omega) &= (M+1)\omega + 2\hat{\theta}_{y,m}(\omega)\\ &= (M+1)\omega + 2\theta_{y,m}(\omega) - 2\overline{\theta}_y(\omega)\\ &= \phi_{y,m}(\omega) - 2\overline{\theta}_y(\omega).\end{aligned} \tag{17}$$

The resulted LSFs are the frequencies that satisfy the following equation:

$$\hat{\phi}_{y,m}(\hat{\omega}^y_{k,m}) = k\pi. \tag{18}$$

The following description is to show how to achieve $\{\hat{\omega}^y_{k,m}\}$ starting from $\{\omega^y_{k,m}\}$. It results in an iterative algorithm to remove the bias on LSFs due to channel effect.

Similar to the derivation of (13), we consider the curve of $\phi_y(\omega)$. The mean slope of the curve between $\omega^y_{k,m}$ and $\hat{\omega}^y_{k,m}$ is defined by

$$s_y(\omega^y_{k,m}, \hat{\omega}^y_{k,m}) = \frac{\phi_y(\hat{\omega}^y_{k,m}) - \phi_y(\omega^y_{k,m})}{\hat{\omega}^y_{k,m} - \omega^y_{k,m}}. \tag{19}$$

Applying the equality $\phi_y(\omega^y_{k,m}) = \hat{\phi}_y(\hat{\omega}^y_{k,m})$ and (17), we obtain that

$$\begin{aligned}s_y(\omega^y_{k,m}, \hat{\omega}^y_{k,m}) &= \frac{\phi_y(\hat{\omega}^y_{k,m}) - \hat{\phi}_y(\hat{\omega}^y_{k,m})}{\hat{\omega}^y_{k,m} - \omega^y_{k,m}}\\ &= \frac{2\overline{\theta}_y(\hat{\omega}^y_{k,m})}{\hat{\omega}^y_{k,m} - \omega^y_{k,m}}.\end{aligned} \tag{20}$$

Rearranging (20), we get

$$\hat{\omega}^y_{k,m} = \omega^y_{k,m} + \frac{2}{s_y(\omega^y_{k,m}, \hat{\omega}^y_{k,m})}\overline{\theta}_y(\hat{\omega}^y_{k,m}). \tag{21}$$

In order to solve (21) for $\hat{\omega}^y_{k,m}$, an iterative scheme based on Newton-Raphson method [22] is applied. At first we define the following quantity:

$$g(\hat{\omega}^y_{k,m}) = \omega^y_{k,m} - \hat{\omega}^y_{k,m} + \frac{2\overline{\theta}_y(\hat{\omega}^y_{k,m})}{s_y(\omega^y_{k,m}, \hat{\omega}^y_{k,m})}, \tag{22}$$

where $\omega^y_{k,m}$ and $\hat{\omega}^y_{k,m}$ are the kth LSF at frame m for without and with phase mean subtraction, respectively. The LSF $\omega^y_{k,m}$ can be extracted from the bit stream of encoded speech or calculated from performing the LP analysis on the channel distorted speech in frame m. Let $\hat{\omega}^y_{k,m}[n]$ denote the value of $\hat{\omega}^y_{k,m}$ at nth iteration. The recursion formula is given as follows:

$$\hat{\omega}^y_{k,m}[n+1] = \hat{\omega}^y_{k,m}[n] - \eta\frac{g(\hat{\omega}^y_{k,m}[n])}{g'(\hat{\omega}^y_{k,m}[n])}, \tag{23}$$

where η is a scalar factor for adjusting the step size and $g'(\hat{\omega}^y_{k,m}[n])$ is the derivative of $g(\hat{\omega}^y_{k,m})$ with respect to $\hat{\omega}^y_{k,m}$ evaluated at $\hat{\omega}^y_{k,m} = \hat{\omega}^y_{k,m}[n]$. The calculation for $g'(\hat{\omega}^y_{k,m}[n])$ is formulated as

$$\begin{aligned}g'(\hat{\omega}^y_{k,m}[n]) = -1 + 2\Bigg(&\overline{\theta}'_y(\hat{\omega}^y_{k,m}[n]) - \overline{\theta}_y(\hat{\omega}^y_{k,m}[n])\\ &\times \frac{\phi'_y(\hat{\omega}^y_{k,m}[n]) - s_y(\omega^y_{k,m}, \hat{\omega}^y_{k,m}[n])}{\phi_y(\hat{\omega}^y_{k,m}[n]) - \phi_y(\omega^y_{k,m})}\Bigg)\\ &\times \frac{1}{s_y(\omega^y_{k,m}, \hat{\omega}^y_{k,m}[n])},\end{aligned} \tag{24}$$

where $\phi'_y(\hat{\omega}^y_{k,m}[n])$ and $\overline{\theta}'_y(\hat{\omega}^y_{k,m}[n])$ are the derivatives of $\phi_y(\hat{\omega}^y_{k,m}[n])$ and $\overline{\theta}_y(\hat{\omega}^y_{k,m}[n])$, respectively. They can be approximated on the functions $\phi_y(\omega)$ and $\theta_y(\omega)$ nearby $\omega = \hat{\omega}^y_{k,m}[n]$. The initial guess is given as

$$\hat{\omega}^y_{k,m}[0] = \omega^y_{k,m} - \delta\,\mathrm{sgn}\,(\overline{\theta}_y(\omega^y_{k,m})), \tag{25}$$

where δ is a small value and $\mathrm{sgn}(\cdot)$ is the sign function.

5. EXPERIMENTS

Two series of experiments are conducted herein. The first series of experiments use simulated channel distorted speech to examine the channel effect due to handset distortion and also the effect of codec process. The second series of experiments are performed on the real telephone speech.

5.1. Experiment 1

The TI digits database is used in this series of experiments. The "train" part of TI digits (112 speakers, each uttering 77 digit strings) is used to train the word models. The "test" part of TI digits (113 speakers, each uttering 77 digit strings) is to evaluate the speech recognition performance. The original sampling rate of speech signal in TI digits is 16 kHz. The sampling rate is lowered to 8 kHz in the following experiments. The frame size is 240 samples with an overlap of 120 samples. The Hamming window is applied in each frame. The features consist of 10 LSFs and one log energy, and their first- and second-order time derivatives. Hence, a feature vector of 33 dimensions is computed. Twelve word models (zero/oh, one, two, ..., nine, and silence) are used in the experiment. Each word model is represented by a 7-state HMM with six Gaussian mixtures in each state.

This experiment examines the channel effect due to

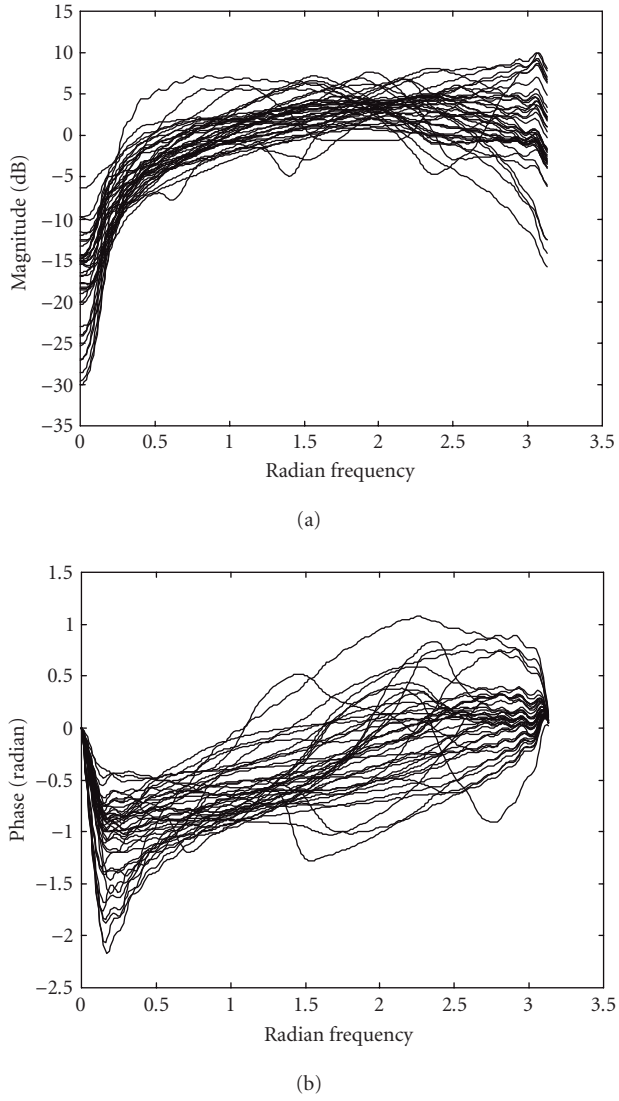


Figure 3: Characteristics of 41 handsets used in experiments. (a) Magnitude. (b) Phase.

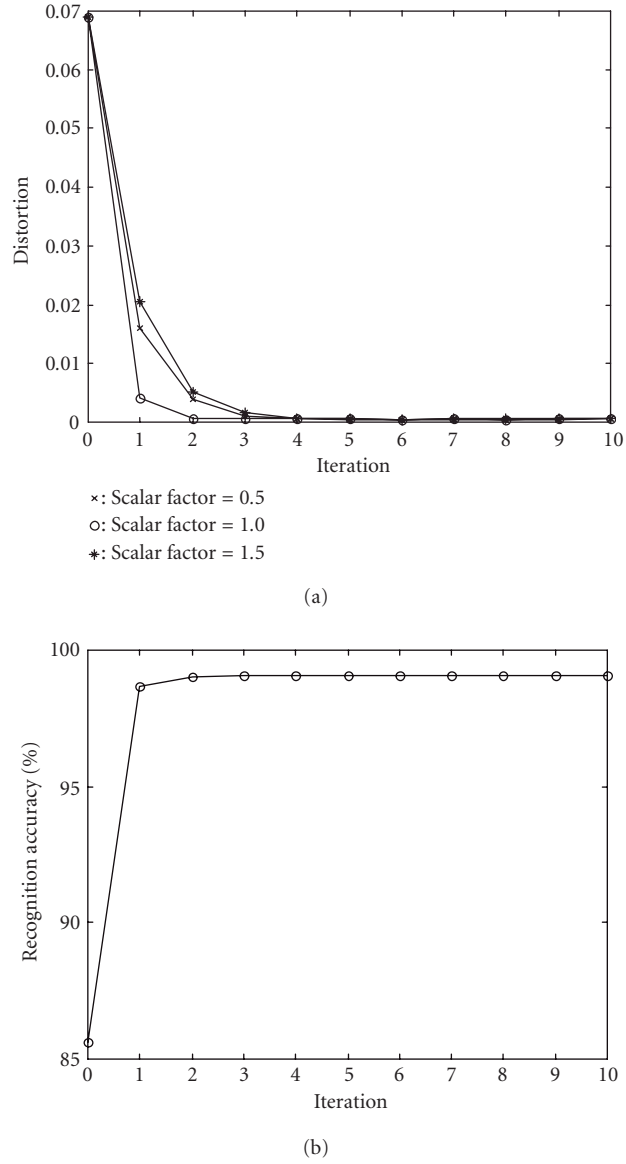


Figure 4: (a) The learning behavior of the proposed algorithm. (b) The effect of the number of iterations on the recognition performance.

handset distortion and also the effect of the codec process. Figure 3 shows the characteristics of the 41 handsets used in the experiments. The codec process is the algorithm of ITU G.723.1. The channel distorted speech is simulated as follows to evaluate the channel effect.

(1) In the case of handset distortion, the speech signal is convoluted with a randomly selected handset before feature extraction is performed. LSFs are calculated for the 50% overlapping frames, based on LP analysis.

(2) In the case of the codec process, the speech signal is fed into a G.723.1 CELP (code excited linear prediction) encoder to produce an encoded bit stream. The LSFs are extracted directly from the bit stream without decoding the speech into a waveform. Linear predictive derived cepstral coefficients (LPCCs) parameters are obtained through a conversion from LSFs. Since the encoder performs without frame overlapping, the number of extracted frames is inconsistent with that of overlapped frames in comparison. An interpolated frame is inserted into each pair of consecutive frames to overcome this inconsistency. The linear interpolation is applied to determine the average feature vector from each pair of features.

(3) In the case of the combination of handset distortion and codec process, the speech signal is first convoluted with a randomly chosen handset and then fed into the CELP encoder to generate an encoded bit stream.

At first, the learning behavior of the proposed iterative algorithm is investigated. Figure 4a displays the learning

Table 1: Recognition rates obtained using LSFs (with speech models trained on clean speech).

	Distortion			
	Clean	Handset	Codec	Handset + Codec
Baseline	99.42%	85.37%	69.58%	54.65%
The proposed method	99.30%	99.13%	85.34%	71.65%

Table 2: Recognition rates obtained using LPCCs (with speech models trained on clean speech).

	Distortion			
	Clean	Handset	Codec	Handset + Codec
Baseline	99.64%	85.78%	70.94%	55.98%
Cepstral mean subtraction	99.31%	99.10%	84.57%	72.86%

Table 3: Recognition rates obtained using LSFs (with models trained on encoded speech).

	Distortion	
	Clean	Handset
Baseline	99.32%	85.62%
The proposed method	99.17%	99.08%

Table 4: Recognition rates obtained using LPCCs (with models trained on encoded speech).

	Distortion	
	Clean	Handset
Baseline	99.30%	85.42%
Cepstral mean subtraction	99.14%	99.02%

behavior with various scalar factors. The distortion is measured by the average distance between LSFs of before and after channel compensation. The resulted curves show that the iterative scheme converges quickly within first two iterations. For the case of $\eta = 1$, the relationship between the iteration number and the recognition performance of simulated channel distorted speech is illustrated in Figure 4b. It shows that the satisfactory performance can be achieved within two iterations. This is very promising for real-time applications.

Table 1 displays the performance of using LSFs in speech recognition with speech models trained on clean speech. It shows that the three kinds of distortions substantially degrade the performances. The performance degradations are about 14%, 29%, and 45% for cases of being affected by handset distortion, codec process, and the combination of handset and codec process, respectively. It is obvious that the performance degradation caused by codec process is much worse than that caused by handset distortion. The combination of handset distortion and the codec process results in the worst performance. Significant improvement can be obtained when the proposed channel effect compensation method is applied to the case of handset distortion. However, the performance is less improved for speech distorted by the codec process or the combination of handset distortion and codec process.

For comparison, the performance of using LPCCs derived from LSFs is evaluated and listed in Table 2. Comparing Table 1 with Table 2, we find that the proposed channel effect compensation method gives comparable performance as the CMN method in using LPCCs. Inconsistency in feature extraction substantially degrades the performances for both LSFs and LPCCs.

Tables 1 and 2 also show that the codec process causes the unacceptable performance. The bad performance is due to the mismatches generated by the nonlinear operation of the codec process and the inconsistent feature extraction. The proposed channel effect compensation method cannot effectively compensate for these mismatches. Hence, the speech models are retrained using speech features directly extracted from encoded bit stream. Since both training and testing data are processed by the same codec algorithm, these retrained models give much better performance. Similarly, we also retrain the models in using LPCCs for comparison. Tables 3 and 4, respectively, show the performance of using LSFs and LPCCs with speech models trained on encoded speech. The result indicates that the performance obtained using encoded speech models is substantially enhanced. Although handset distortion significantly degrades the performance in this case, the proposed channel compensation method can effectively recover the performance. The performance is close to that of using LPCCs with speech models trained on encoded speech.

5.2. Experiment 2

The subdatabase MATDB-2 of database Mandarin Across Taiwan-2000 (MAT-2000) is used in this series of experiments. MAT-2000 comprises telephone Mandarin speech of 1005 male and 1227 female speakers recorded in Taiwan telephone network. The MATDB-2 contains numbers pronounced in five different ways (including telephone number, date, time, money, and car plate number). The 4400 utterances from 500 male and 600 female speakers are used to train the word models. The 4528 utterances from 505 male and 627 female speakers are used to evaluate the performance. The speech data is coded in 16-bits PCM and the sampling rate is 8 kHz. The frame size is 256 samples with an overlap of 128 samples. The Hamming window is applied in each frame. The features consist of 12 LSFs and one log energy, and their first- and second-order time derivatives. Hence, a feature vector of 39 dimensions is computed. On the other hand, to compare the effectiveness of the proposed method, recognition on LPCC features is also performed. The experiment uses 26 word models. Each word model is represented by a 7-state HMM with eight Gaussian mixtures in each state.

Table 5: Recognition rates using telephone speech.

	Without compensation	With compensation
LPCC	91.01%	92.51%
LSFs	91.02%	92.54%

Table 5 displays the recognition results of using LSFs and LPCCs. It shows that when the cepstral mean subtraction and the proposed channel effect compensation method are not performed, the recognition performance is about 91% for LPCCs and LSFs. When they are performed, the performances are enhanced to about 92.5%. The results suggest that the proposed channel effect compensation method is effective and its performance is comparable to that of using CMN method in LPCCs.

6. CONCLUSIONS

This work focuses on the compensation of channel effect in LSF domain. When a speech signal is represented in terms of the phase of inverse filter derived from LP analysis, the channel distortion can be expressed in terms of the channel phase. Further derivation shows that the mean subtraction performed on the phase of inverse filter can minimize the channel effect. Based on this finding, an iterative algorithm is proposed to compensate the channel effect. To demonstrate the effectiveness of the proposed methods, two series of experiments on the simulated channel distorted speech and the real telephone speech are conducted. The experimental results show that the proposed methods yield significant improvements for both situations. The performance of the proposed method is comparable to that of CMN in using cepstral coefficients.

ACKNOWLEDGMENT

This research was partially sponsored by the National Science Council, Taiwan, under Contract NSC-90-2213-E-007-028.

REFERENCES

[1] R. A. Bates, "Reducing the effects of linear channel distortion on continuous speech recognition," M.S. thesis, Boston University, Boston, Mass, USA, 1996.

[2] S. Lerner and B. Mazor, "Telephone channel normalization for automatic speech recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 261–264, San Francisco, Calif, USA, March 1992.

[3] H. A. Murthy, F. Beaufays, L. P. Heck, and M. Weintraub, "Robust text-independent speaker identification over telephone channels," *IEEE Trans. Speech, and Audio Processing*, vol. 7, no. 5, pp. 554–568, 1999.

[4] S. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 29, no. 2, pp. 254–272, 1981.

[5] F. H. Liu, R. M. Stern, X. Huang, and A. Acero, "Efficient cepstral normalization for robust speech recognition," in *Proc. ARPA Speech and Nat. Language Workshop*, pp. 69–74, Princeton, NJ, USA, March 1993.

[6] J. D. Veth and L. Boves, "Comparison of channel normalisation techniques for automatic speech recognition over the phone," in *Proc. Fourth International Conference on Spoken Language Processing*, pp. 2332–2335, Philadelphia, Pa, USA, October 1996.

[7] A. Sankar and C. H. Lee, "A maximum-likelihood approach to stochastic matching for robust speech recognition," *IEEE Trans. Speech, and Audio Processing*, vol. 2, no. 3, pp. 190–202, 1996.

[8] H. Hermansky and N. Morgan, "RASTA processing of speech," *IEEE Trans. Speech, and Audio Processing*, vol. 2, no. 4, pp. 578–589, 1994.

[9] J. T. Chien and H. C. Wang, "Telephone speech recognition based on Bayesian adaptation of hidden Markov models," *Speech Communication*, vol. 22, no. 4, pp. 369–384, 1997.

[10] H. K. Kim and R. V. Cox, "A bitstream-based feature extraction for wireless speech recognition on IS-136 communications system," *IEEE Trans. Speech, and Audio Processing*, vol. 9, no. 5, pp. 558–568, 2001.

[11] K. K. Paliwal, "A study of LSF representation for speaker-dependent and speaker-independent HMM based speech recognition systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 804–807, Albuquerque, NM, USA, April 1990.

[12] K. K. Paliwal, "A study of line spectrum pair frequencies for speech recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 485–488, Seattle, Wash, USA, May 1998.

[13] A. T. Yu and H. C. Wang, "A study on the recognition of low bit-rate encoded speech," in *Proc. International Conf. on Spoken Language Processing*, pp. 1523–1526, Sydney, Australia, November–December 1998.

[14] A. T. Yu and H. C. Wang, "Effect of noise on line spectrum frequency and a robust speech recognition method for the low bit-rate encoded speech," in *Proc. Int. Conf. on Phonetic Science*, San Francisco, Calif, USA, August 1999.

[15] Intel Corporation and France Telecom, "ITU-T G.723.1 floating point speech coder ANSI C source code. Version 5.1F," 1995.

[16] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *Journal of the Acoustical Society of America*, vol. 57, no. Suppl.1, pp. S35, 1975.

[17] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech, and Audio Processing*, vol. 1, no. 1, pp. 3–14, 1993.

[18] L. M. Arslan and D. Talkin, "Voice conversion by codebook mapping of line spectral frequencies and excitation spectrum," in *Proc. Eurospeech '97*, Rhodes, Greece, September 1997.

[19] R. Laroia, N. Phamdo, and N. Farvardin, "Robust and efficient quantization of speech LSP parameters using structured vector quantisers," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 641–644, Toronto, Ontario, Canada, May 1991.

[20] B. Raj, J. Migdal, and R. Singh, "Distributed speech recognition with codec parameters," in *IEEE Automatic Speech Recognition and Understanding Workshop*, Trento, Italy, December 2001.

[21] A. T. Yu and H. C. Wang, "Compensation of channel effect on line spectrum frequencies," in *Proc. International Conf. on Spoken Language Processing*, Denver, Colo, USA, September 2002.

[22] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, USA, 1965.

An-Tze Yu received the B.S. degree in industrial education from National Changhua University of Education, Changhua, Taiwan, in 1988 and the M.S. degree in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1993. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering at National Tsing Hua University, Hsinchu, Taiwan. He joined the Department of Computer Science, National Chupei Senior High School, Hsinchu, Taiwan, in 1993. Yu has been the Chair of the Department of Computer Science (August 1993–July 1995). His current research interests include speech recognition and speech coding.

Hsiao-Chuan Wang received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1969, and the M.S. and Ph.D. degrees in electrical engineering from the University of Kansas, Lawrence, Kansas, in 1973 and 1977, respectively. He joined the Department of Electrical Engineering at National Tsing Hua University, Hsinchu, Taiwan, in 1977. He has been Chair of the Department of Electrical Engineering (August 1986–July 1992), Director of the University Library (August 1998–July 2000), and Director of the Computer & Communication Center (August 1998–July 2000). He is a life member of the Chinese Institute of Electrical Engineering (CIEE). He has served on the editorial board of the Journal of CIEE (a technical journal in English) since 1993 and was the Editor-in-Chief (1993–1995) and then the Chair of the editorial board (1996–1999). He is a Senior Member of IEEE and has served as the Chair of IEEE Taipei Section (1997–1999) and Associate Editor of IEEE Transactions on Speech and Audio Processing (March 1999–February 2002). He has been President of Association of Computational Linguistics and Chinese Language Processing (ACLCLP) (December 1999–December 2001), and is currently a member thereof. His current research interests include speech recognition, speech coding, audio processing, and digital signal processing.