

# COMPENSATION OF CHANNEL DISTORTION IN LINE SPECTRUM FREQUENCY DOMAIN

*An-Tze Yu and Hsiao-Chuan Wang*

Department of Electrical Engineering  
National Tsing Hua University, Hsinchu, Taiwan, ROC  
[yuat@ms25.hinet.net](mailto:yuat@ms25.hinet.net), [hcwang@ee.nthu.edu.tw](mailto:hcwang@ee.nthu.edu.tw)

## ABSTRACT

This paper addresses the problem of channel effect in the line spectrum frequency (LSF) domain. The channel effect can be expressed in terms of the channel phase. The speech signal is represented by its inverse filter derived from LP analysis. Then the mean normalization on the inverse filters is introduced for removing the channel distortion. Further study indicates that the mean normalization on the inverse filters becomes the mean subtraction in phase domain. Based on this finding, two methods are proposed to compensate the channel effect. Experiments on simulated channel distorted speech are conducted to evaluate the effectiveness of the proposed methods. The experimental results show that the proposed methods can give significant improvements in speech recognition performance. The performance of the proposed methods is comparable to that of CMN in using cepstral coefficients.

## 1. INTRODUCTION

Channel distortion is always a serious problem in speech recognition systems [1]. The channel effect on the cepstral domain has been extensively studied. Many approaches have been proposed for eliminating the influence of channel distortion to speech recognition performance [2]. However, few studies aim at the channel effect on line spectrum frequency (LSF) domain. LSFs are the parameters used in low bit-rate coding for digital speech transmission. A speech or speaker recognition algorithm based on the LSFs is of interest in mobile communication and Internet systems [3]. Several studies have shown that LSF parameters may not be good enough for a large vocabulary continuous speech recognition (LVCSR) system. However, LSFs are still good as comparing with MFCCs in connected digits recognition or small vocabulary systems [6-7].

The channel effect is an additive term in the cepstrum domain. Cepstral mean subtraction (CMS), also called cepstral mean normalization (CMN), is a simple but effective method to remove the channel effect in speech recognition. The line spectrum (LS) is an alternative representation of linear prediction (LP) analysis. LSFs have been extensively used in speech

coding and synthesis [3] because of their robustness in parameter quantization. The use of LSFs directly extracted from the encoded bit stream is preferred since it becomes unnecessary to decode the encoded speech into a waveform [3][6]. Some researches also reported that features obtained in this way are more robust in adverse environments than those from decoded speech waveform [6].

This paper focuses on compensating the channel effect for LSFs. The channel effects on phase domain and LSF domain are investigated and formulated [7]. Let the speech signal be represented by its inverse filter derived from LP analysis. The mean normalization on the inverse filters is a way for removing the channel distortion. Further study indicates that the mean normalization on inverse filters becomes phase mean subtraction in phase domain. Based on this finding, two methods are proposed to compensate the channel effect. The first one calculates the phase mean of whole utterance. Then the phase mean subtraction is applied to each frame before LSFs are computed. The second method is to find the LSF difference between the cases of with and without phase mean subtraction. An iterative algorithm based on this LSF difference is used for removing the channel effect.

Experiments on simulated channel distorted speech are conducted to evaluate the effectiveness of the proposed methods. The experimental results show that the proposed methods give significant improvements in speech recognition performance. The performance of the proposed methods is comparable to that of CMN in using cepstral coefficients.

## 2. CHANNEL EFFECT ON LSFs

### 2.1. Channel effect in phase domain

For a speech signal  $x(n)$  and a channel filter  $h(n)$ , the distorted speech is expressed as  $y(n)=x(n)*h(n)$  in time domain. The spectral envelope of  $y(n)$  is modeled [7] as

$$\frac{G_y}{A_y(z)} = \frac{G_x H(z)}{A_x(z)}, \quad (1)$$

where  $A_y(z)$  and  $A_x(z)$  are the inverse filters of the distorted speech  $y(n)$  and the clean speech  $x(n)$ , respectively.  $G_y$  and  $G_x$  are the gains in the LP analysis for  $y(n)$  and  $x(n)$ , respectively.  $H(z)$  is the z-transform of channel filter  $h(n)$ .

Based on Eq. (1), the phase of  $A_y(z)$  equals,

$$\theta_y(\omega) = \theta_x(\omega) - \theta_h(\omega), \quad (2)$$

where  $\theta_x(\omega)$  and  $\theta_h(\omega)$  are the phases of  $A_x(\omega)$  and  $H(\omega)$ , respectively.

In [7], the ratio phase of  $y(n)$  is expressed as

$$\phi_y(\omega) = (M+1)\omega + 2\theta_y(\omega). \quad (3)$$

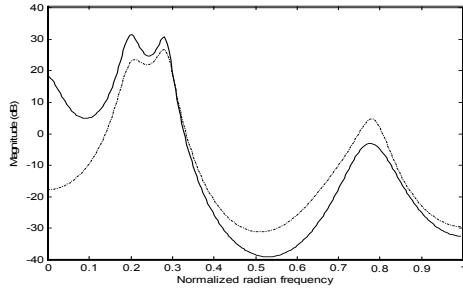
Applying Eq.(2) yields

$$\phi_y(\omega) = (M+1)\omega + 2\theta_x(\omega) - 2\theta_h(\omega). \quad (4)$$

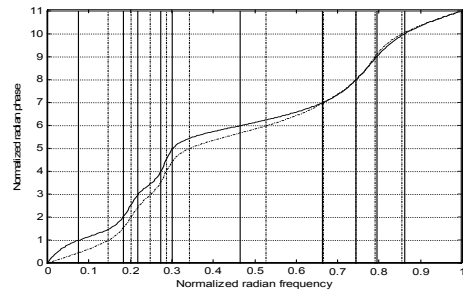
Equation (4) can be further expressed as

$$\phi_y(\omega) = \phi_x(\omega) - 2\theta_h(\omega), \quad (5)$$

where  $\phi_x(\omega)$  is the ratio phase of  $x(n)$ . This equation indicates that the ratio phase  $\phi_y(\omega)$  deviates from the ratio phase  $\phi_x(\omega)$  by  $-2\theta_h(\omega)$ . Figure 1 shows an example of the channel effect on the power spectrum and the ratio phase.



(a)



(b)

Figure 1: Channel effect on spectral domain and ratio phase function for the vowel /a/. The solid line represents clean speech and the dotted line represents distorted speech.

## 2.2. Channel effect on LSFs

The mean slope of the ratio phase  $\phi_y(\omega)$  between  $\omega_k^x$  and  $\omega_k^y$  is defined as follows.

$$m_y(\omega_k^x, \omega_k^y) = \frac{\phi_y(\omega_k^y) - \phi_y(\omega_k^x)}{\omega_k^y - \omega_k^x}, \quad (6)$$

where  $\omega_k^x$  and  $\omega_k^y$  are the  $k$ -th LSFs respectively for  $x(n)$  and  $y(n)$ . Substituting the following equality [7]

$$\phi_y(\omega_k^y) = \phi_x(\omega_k^x), \quad (7)$$

into Eq. (6) yields,

$$m_y(\omega_k^x, \omega_k^y) = \frac{\phi_x(\omega_k^x) - \phi_y(\omega_k^x)}{\omega_k^y - \omega_k^x}. \quad (8)$$

Applying Eq. (5), Eq. (8) becomes,

$$m_y(\omega_k^x, \omega_k^y) = \frac{2\theta_h(\omega_k^x)}{\omega_k^y - \omega_k^x}. \quad (9)$$

Rearranging Eq. (9) yields

$$\omega_k^y = \omega_k^x + \frac{2}{m_y(\omega_k^x, \omega_k^y)} \theta_h(\omega_k^x). \quad (10)$$

## 3. CHANNEL COMPENSATION

Equation (10) indicates that the deviation of LSFs resulted from the channel effect can be compensated if the mean slope,  $m_y(\omega_k^x, \omega_k^y)$ , and the channel phase,  $\theta_h(\omega_k^x)$ , are known. However, an effective channel phase estimation is hard to be implemented. The alternative approach must be developed.

For an utterance, Eq. (1) in frequency domain for a specific frame is expressed as

$$A_{y,m}(\omega) = \frac{A_{x,m}(\omega)}{H(\omega)}, \quad m=1..L, \quad (11)$$

where  $A_{x,m}(\omega)$  and  $A_{y,m}(\omega)$  are the inverse filters of the clean speech and the distorted speech at frame  $m$ .  $L$  is the number of frames of the utterance. Prediction gains in Eq. (1) are dropped for its no relation to the following derivation.

The geometrical mean of the inverse filters of the distorted speech is computed as follows,

$$\bar{A}_y(\omega) = \left( \prod_{m=1}^L A_{y,m}(\omega) \right)^{\frac{1}{L}}. \quad (12)$$

Applying Eq. (11) yields,

$$\bar{A}_y(\omega) = \frac{\left(\prod_{m=1}^L A_{x,m}(\omega)\right)^{\frac{1}{L}}}{H(\omega)} = \frac{\bar{A}_x(\omega)}{H(\omega)}. \quad (13)$$

$\left(\prod_{m=1}^L A_{x,m}(\omega)\right)^{\frac{1}{L}}$  is the geometrical mean of the inverse filters of the clean speech.

In order to eliminate the channel effect, let's define the mean normalized inverse filters of the distorted speech as

$$\hat{A}_{y,m}(\omega) \equiv \frac{A_{y,m}(\omega)}{\bar{A}_y(\omega)}. \quad (14)$$

Substituting Eq. (11) and (13) into Eq. (14) yields

$$\hat{A}_{y,m}(\omega) = \frac{A_{x,m}(\omega)}{\bar{A}_x(\omega)} = \hat{A}_{x,m}(\omega). \quad (15)$$

The equivalence of the mean normalized inverse filters of the clean speech and the distorted speech implies that the channel effect is removed. Referring to Eq. (14), the phase function of the mean normalized inverse filter of the distorted speech is calculated as

$$\hat{\theta}_{y,m}(\omega) = \theta_{y,m}(\omega) - \bar{\theta}_y(\omega), \quad (16)$$

where  $\theta_{y,m}(\omega)$  and  $\bar{\theta}_y(\omega)$  are the phase functions of  $A_{y,m}(\omega)$  and  $\bar{A}_y(\omega)$  respectively.  $\bar{\theta}_y(\omega)$  is computed as

$$\bar{\theta}_y(\omega) = \arg(\bar{A}_y(\omega)), \quad (17)$$

where  $\arg(\cdot)$  is the argument function.

Substituting Eq. (12) into Eq. (17) yields

$$\bar{\theta}_y(\omega) = \frac{1}{L} \sum_{m=1}^L \theta_{y,m}(\omega). \quad (18)$$

Equation (18) suggests that the phase function of the geometrical mean of the inverse filters equals the mean phase functions of the inverse filters. Re-examining the derivation of Eq. (16) shows that the mean normalization on the inverse filters becomes phase mean subtraction in phase domain. Based on this finding, two methods are proposed to compensate the channel effect for LSFs.

### 3.1. Method 1: Phase Mean Subtraction

Based on the previous discussion, we can compute LSFs using mean subtracted phases. The LSFs will

be immune from the channel effect. Hence we define the ratio function in terms of mean subtracted phases as

$$\hat{\phi}_{y,m}(\omega) = (M+1)\omega + 2\theta_{y,m}(\omega) - 2\bar{\theta}_y(\omega). \quad (19)$$

Solving  $\hat{\phi}_{y,m}(\hat{\omega}_{k,m}^y) = k\pi$ , we obtain the LSFs,  $\hat{\omega}_{k,m}^y$ , which are free of channel effect.

### 3.2. Method 2: Iterative Channel Deconvolution

If LSFs of distorted speech are available, it would be attractive to remove channel effect in LSFs domain directly. The relationship of ratio phases among ratio functions derived with and without phase mean subtraction is formulated as

$$\hat{\phi}_{y,m}(\omega) = \phi_{y,m}(\omega) - 2\bar{\theta}_y(\omega). \quad (20)$$

Similar to the derivation of Eq. (10), the following relationship exists.

$$\hat{\omega}_{k,m}^y = \omega_{k,m}^y + \frac{2\bar{\theta}_y(\hat{\omega}_{k,m}^y)}{m_y(\omega_{k,m}^y, \hat{\omega}_{k,m}^y)}. \quad (21)$$

The LSF based on performing phase mean subtraction,  $\hat{\omega}_{k,m}^y$ , is obtained by solving the equation,

$$g(\hat{\omega}_{k,m}^y) = \omega_{k,m}^y - \hat{\omega}_{k,m}^y + \frac{2\bar{\theta}_y(\hat{\omega}_{k,m}^y)}{m_y(\omega_{k,m}^y, \hat{\omega}_{k,m}^y)} = 0, \quad (22)$$

where  $\omega_{k,m}^y$  and  $\hat{\omega}_{k,m}^y$  are the  $k$ -th LSF at frame  $m$  respectively for without and with phase mean subtraction.  $\omega_{k,m}^y$  can be calculated from performing the LP analysis on the distorted speech.  $\hat{\omega}_{k,m}^y$  should be obtained by solving the above equation,  $g(\hat{\omega}_{k,m}^y) = 0$ . Applying Newton's method,  $\hat{\omega}_{k,m}^y$  is iteratively computed using the following formula,

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

where  $\hat{\omega}_{k,m}^y[n]$  represents the value of  $\hat{\omega}_{k,m}^y$  at iteration  $n$ , and  $\eta$  is the step-size factor.  $g'(\hat{\omega}_{k,m}^y[n])$  is the derivative of  $g(\hat{\omega}_{k,m}^y)$  with respect to  $\hat{\omega}_{k,m}^y$  evaluated at  $\hat{\omega}_{k,m}^y = \hat{\omega}_{k,m}^y[n]$ . The initial guess is given as,

$$\hat{\omega}_{k,m}^y[0] = \omega_{k,m}^y - \delta \operatorname{sgn}(\bar{\theta}_y(\omega_{k,m}^y)), \quad (24)$$

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

#### 4. EXPERIMENTS

Experiments on TI-Digits database are conducted to evaluate the proposed methods. 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 performance. The original sampling rate of speech in TI-Digits is 16 kHz. This sampling rate is lowered to 8 kHz in the following experiments.

To simulate the transmission of speech in a digital communication system, the speech signal is first convoluted with a channel filter (randomly chosen from 40 channels [7]) and then fed into the ITU G.723.1 CELP encoder to generate an encoded bit stream. The LSFs used as recognition features are extracted directly from the bit stream without decoding the speech into waveform. LPCC parameters are obtained through a conversion from extracted LSFs. An interpolated frame is inserted into each pair of consecutive frames. The feature vectors consist of ten LSFs and one log energy, and their first and second order time derivatives. 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.

Table 1. Recognition rates obtained using LSFs

	Accuracy
Baseline	84.70%
Phase mean subtraction	99.07%
Iterative deconvolution	98.54%

Table 2. Recognition rates obtained using LPCCs

	Accuracy
Baseline	85.42%
Cepstral mean subtraction	99.09%

Table 1 shows the performance utilizing two proposed methods. It is clear that the channel distortion substantially degrades the performances. Significant improvement can be obtained when the proposed compensation methods are applied. Table 2 shows the performances of using derived LPCCs. Comparing the performance shown in table 1 with that in table 2, we can see that the performance of using LSFs directly is very close to that of using LPCCs with cepstral mean subtraction.

#### 5. CONCLUSIONS

This work aims at the compensation of channel effect in LSF domain. The channel effect on phase domain and LSF domain are investigated and formulated. They show that the channel effect can be expressed in terms of the channel phase. Then the mean normalization on inverse filters is introduced. Further study indicates that the mean normalization on inverse filters becomes phase mean subtraction in phase domain. Based on this finding, two methods are proposed to compensate the channel effect. Experiments on simulated channel distorted speech are conducted to evaluate the performance of proposed methods in encoded speech data. The experimental results show that the performance of the proposed methods is comparable to that of CMN in using cepstral coefficients.

#### ACKNOWLEDGEMENT

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

#### REFERENCES

- [1]. Bates R. A., "Reducing The Effects Of Linear Channel Distortion On Continuous Speech Recognition", Thesis, Boston University, 1996
- [2]. Liu F. H., Stern R. M., Huang X., Acero A., "Efficient Cepstral Normalization for Robust Speech Recognition", Proceedings ARPA Speech and Nat. Language Workshop, Princeton, NJ, pp. 69-74, 1993
- [3]. Kim, H. K. and Richard V. Cox, "Bit stream-Based Feature Extraction For Wireless Speech Recognition," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing'00*, pp. 1207-1210, 2000
- [4]. Furui S., "Cepstral Analysis Technique for Automatic Speaker Verification", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 29, no. 2, pp. 254-272, 1981
- [5]. Itakura, F., "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol. 57, Suppl., no. 1, S35, 1975
- [6]. Yu, A. T. and Wang, H. C., "A study on the recognition of low bit-rate encoded speech," in *Proc. International Conference on Spoken Language Processing*, Sydney, Australia, vol. 4, pp. 1523-1526, 1998
- [7]. Yu, A.T. and Wang, H.C., "Compensation of channel effect on line spectrum frequencies," *Proc. International Conference on Spoken Language Processing, ICSLP2002*, Denver, Colorado, 2002