Noise-Reduced Complex LPC Analysis for Formant Estimation of Noisy Speech

Takuma Kaneko and Tetsuya Shimamura
Graduate School of Science and Engineering Saitama University, Saitama, Japan
Email: {kaneko, shima} @sie.ics.saitama-u.ac.jp

Abstract—**In this paper we present a complex linear prediction analysis method for estimating the formant frequencies of noisy speech. The proposed method effectively utilizes the signal (being the analytic signal) which is ignored in the conventional complex linear prediction analysis to achieve noise reduction. Also, the covariance and forward-backward linear prediction (FBLP) methods are compared, and the FBLP method is deployed for predictive coefficients estimation in the proposed method. Experimental results show that the proposed method yields better performance of formant estimation when compared to some conventional methods in white noise**.

***Index Terms*—LPC, formant, noise reduction, analytic signal**

I. Introduction

Free resonances of the vocal-tract system are called formants. Formants are associated with peaks in the smoothed power spectrum of speech [1]. The peak locations, that is, formant frequencies play a fundamental role in speech synthesis, recognition and compression. For example, formant frequencies serve as an important acoustic feature and offer a phonetic reduction in speech recognition [2] . They are also crucial in the design of some hearing aids [3]. From these reasons, we need accurate formant frequency estimation from the speech signal. Among different formant estimation techniques, linear prediction coding (LPC) based methods have received considerable attention [4] [5] [6].

LPC is the most commonly used technique for speech analysis, which can estimate the spectral envelope of the voiced speech signal by modeling it by a set of parameters closely related to the speech production transfer function. The transfer function of the LPC modeling is expressed by.

$$H(z) = \frac{G}{\sum_{i=1}^{M} a_i z^{-i}} \tag{1}$$

Being an all-pole filter where G is the gain and a_i are the predictive coefficients.

However, LPC analysis suffers from some drawbacks. One of the most famous ones is that the predictive coefficients can be accurately estimated and the voiced speech signal can also be represented accurately in noise-free environment. LPC method, however, becomes very difficult to estimate the predictive coefficients in noisy environment. The accuracy of the method will significantly degrade in the presence of additive noise. On the other hand, complex linear prediction coding (CLPC) has been proposed [7]. CLPC is the method for linear prediction using the complex signal called analytic signal. When performing formant estimation, CLPC shows results better than LPC. Especially, the difference between CLPC and LPC is large in the low frequency region. When analyzing the speech signal corrupted by additive noise, however, the estimation accuracy of CLPC becomes lower than that of LPC. Therefore, noise reduction is required to maintain the excellent performance of CLPC in noisy environment.

When implementing CLPC, decimation by a factor of 2 is needed for the analytic signal. In that process, half of the analytic signal is discarded. That is, without using the data being half of the analytic signal, linear prediction is implemented. In this paper, we present a new method of noise reduction by using the lost data when performing the decimation in CLPC. The adverse noise which corrupts the speech signal is assumed to be white noise. We perform an average operation between the two signals generated by the decimation so as to obtain a new enhanced signal. The new enhanced signal could not only emphasis the speech signal but also suppress the effect of the corrupted noise.

The remainder of this paper is organized as follows. Section II explains the proposed method. In Section III, through experiments, we verify the effectiveness of the proposed method by comparing with some other methods. Finally in Section IV, we conclude the paper.

II. Proposed Method

In this section, we explain the CLPC method briefly and then derive the proposed method. The CLPC method performs linear prediction using the analytic signal obtained by converting the original signal.

The analytic signal is one of complex signals in which the imaginary part signal is the Hilbert transform of the real part signal. It is well known that the real part signal and the imaginary part signal have orthogonality. Let the observed real signal be $x(m)$ and let its Hilbert transform be $x_h(m)$. Then the analytic signal, $z(m)$, is given by

Manuscript received November 6, 2013; revised February 8, 2014.

doi: 10.12720/ijeee.2.2.90-94

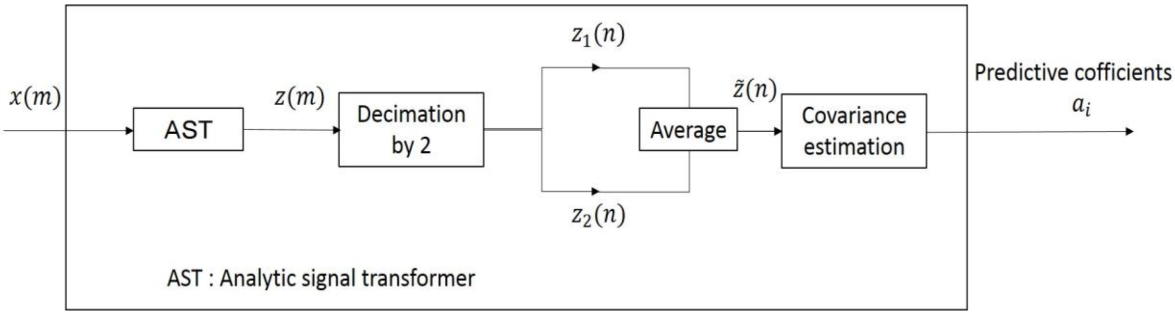

Figure 1. Block diagram of proposed method

$$z(m) = x(m) + jx_h(m) \tag{2}$$

In (2), x(m) and $x_h(m)$ are related by

$$x_h(m) = \sum_{k=-\infty}^{\infty} h(k)\, x(m-k) \tag{3}$$

where $h(k)$ is the impulse response of the Hilbert transform, which is given by

$$h(k) = \begin{cases} \frac{2}{\pi k} sin^2\left(\frac{\pi k}{2}\right) & (k \neq 0) \\ 0 & (k = 0) \end{cases} \tag{4}$$

The analytic signal has an interesting property in the frequency domain. Let the Fourier transform of $z(n)$, $x(n)$, and $x_h(n)$ be $Z(e^{j\omega})$, $X(e^{j\omega})$, and $X_h(e^{j\omega})$, respectively. Then,

$$Z(e^{jw}) = X(e^{jw}) + jX_h(e^{jw}) \begin{cases} 2X(e^{jw}) & (\omega > 0) \\ 0 & (\omega \leq 0) \end{cases} \tag{5}$$

Consequently, the analytic signal does not have the negative frequency component, but twice the positive frequency component of the real signal [7].

Let the speech signal be $x(m)$. When the speech signal $x(m)$ is passed through an analytic signal transformer (AST), the corresponding analytic signal $z(m)$ is generated. Then, the analytic signal is decimated by a factor of 2 to expand the spectrum, which is required for the CLPC to obtain an accurate estimation result [7] [8]. Here two signals, $z_1(n)$ and $z_2(n)$, are generated. z(m) , $z_1(n)$and $z_2(n)$ are related by

$$z_1(n) = z(2m) \tag{6}$$

$$z_2(n) = z(2m+1) \tag{7}$$

where it is noticed that *n* and *m* are positive integers such as *n* = 0,1,2⋯ and *m* = 0,1,2⋯. The conventional CLPC method [9] uses only the signal $z_2(n)$ to estimate the predictive coefficients a_i However, the proposed method uses a new analytic signal $\hat{z}(n)$ generated by averaging $z_1(n)$ and $z_2(n)$. The new analytic signal can be expressed by

$$\hat{z}(n) = \frac{z_1(n)+z_2(n)}{2} \tag{8}$$

$$\hat{z}(n) = \frac{z_1(2m)+z_2(2m+1)}{2} \tag{9}$$

The analytic signal $z_1(n)$ is almost similar to $z_2(n)$ because the delay between the signals $z_1(n)$ and $z_2(n)$ is only $\frac{1}{fs}$, where *fs* is the sampling frequency. This suggests that $\hat{z}(n)$ results in

$$\hat{z}(n) = \frac{z_1(n)+z_2(n)}{2} \tag{10}$$

$$\hat{z}(n) \approx z_2(n) \tag{11}$$

In the proposed method we utilize $\hat{z}(n)$ instead of $z_2(n)$.

In the presence of noise $v(m)$, the observed speech signal is given by

$$y(m) = x(m) + v(m) \tag{12}$$

where $v(m)$ is assumed to white noise with zero mean and variance σ^2. The analytic signal of the noisy speech $y(m)$ can be expressed by

$$z_y(m) = z(m) + z_v(m) \tag{13}$$

where *z*(*m*) and $z_v(m)$ are the analytic signals of clean speech and white noise, respectively. Based on the decimation by a factor of 2, the noisy analytic signal $z_y(m)$ is divided into

$$z_{y1}(n) = z_1(n) + z_{v1}(n) \tag{14}$$

$$z_{y2}(n) = z_2(n) + z_{v2}(n) \tag{15}$$

where $z_{v1}(n)$ and $z_{v2}(n)$ are white noise with zero mean and variance σ^2, respectively, which are uncorrelated each other. Even in noisy environment, the conventional CLPC method utilizes the $z_{v2}(n)$ to estimate the prediction coefficients a_i . However, in this case, the conventional CLPC method can not reduce the effect of the adverse noise. In this paper, we produce the new analytic signal $\hat{z}_y(n)$ to estimate the predictive coefficients, and set out to reduce the noise effect.

The new analytic signal $\hat{z}_y(n)$ is expressed by

$$\hat{z}_y(n) = \frac{z_1(n)+z_2(n)+z_{v1}(n)+z_{v2}(n)}{2} \tag{16}$$

$$\hat{z}_y(n) \approx z_2(n) + \frac{z_{v1}(n)+z_{v2}(n)}{2} \tag{17}$$

As mentioned above, $z_{v1}(n)$ and $z_{v2}(n)$ are uncorrelated white noise with zero mean and variance σ^2. The variance of the noise component $\frac{z_{v1}(n)+z_{v2}(n)}{2}$ in (17) results in

$$\Phi\left(\frac{z_{v1}(n)+z_{v2}(n)}{2}\right)=\frac{\sigma^2}{2} \quad (18)$$

where Φ denotes variance. Comparing $\hat{z}_y(n)$ with $z_{y2}(n)$, the noise power in the new analytic signal $\hat{z}_y(n)$ becomes half of the noise power in $z_{y2}(n)$. Hence, utilizing the new analytic signal to estimate the predictive coefficients we could reduce the effect of the adverse noise. Fig. 1 represents a block diagram of propose method.

In this paper, based on the new analytic signal $\hat{z}_y(n)$, we utilize the Forward-Backward Linear Prediction (FBLP) method [9] [10], being a developed version of the covariance method, to estimate the predictive coefficients. As we know, compared with the autocorrelation method, the covariance method can estimate more accurate prediction coefficients, while it cannot ensure the stability of the resulting all-pole filter. However, it is not necessary to ensure the stability of the all-pole filter in estimating the formant frequencies. In addition, for the complex linear prediction analysis, the target signal is decimated by a factor of 2 from the original signal. This means that the number of the target signal samples becomes half. In this case, the FBLP approach is more effective than the conventional autocorrelation method based approach in [7] to estimate the predictive coefficients from the signal with less length.

In the next session, we verity the effectiveness of the FBLP in a short length signal and the superior performance of the proposed method.

TABLE I. Experimental Parameter Specification

Sampling frequency	12kHz
Analysis window	Hamming
LPC order	12
Additive noise	White

TABLE II. AE Values for Synthetic Vowel /O/

synthetic vowel		FBLP(256)	FBLP(128)	Covariance(256)	Covariance(128)
20dB	F1	0.0232	0.0232	0.0224	0.0235
	F2	0.0172	0.0192	0.0177	0.0262
	F3	0.0053	0.0053	0.0026	0.0031
15dB	F1	0.0511	0.0511	0.0511	0.0520
	F2	0.0378	0.0378	0.0378	0.0396
	F3	0.0065	0.0065	0.0065	0.0079
10dB	F1	0.1085	0.1085	0.1023	0.1433
	F2	0.0606	0.0606	0.0684	0.0696
	F3	0.0128	0.0142	0.0121	0.0175

III. Experiments

A. Covariance and FBLP Methods for LPC

At first, we investigated the use of the FBLP and covariance methods instead of the autocorrelation method. Only real valued linear prediction, which is not complex-valued linear prediction, was conducted. A synthetic vowel /o/[1] was generated and used for the experiment. For the performance assessment, we computed the Absolute Error (AE) at different noise levels. The AE is defined by

$$\delta(i)=\frac{|f_e(i)-f_t(i)|}{f_t(i)} \quad (19)$$

where $f_t(i)$ and $f_e(i)$ are the true formant frequency and its estimate of the i-th formant frequency, respectively.

Table I shows the experimental parameter specification. The analysis frame length was set to 256 and 128 data samples with 50% overlap. For each noise level, 100 individual trials to generate white noise were conducted and the calculation of (19) was averaged. In Table II, the evaluated AE is shown for the performance comparison between the FBLP and covariance methods at SNR=20dB, SNR=15dB and SNR=10dB. It is observed that the FBLP method can keep the estimation accuracy when the speech signal is short. Table II suggests that the FBLP method can be extended to a complex-valued version and the complex FBLP method is suitable for complex linear prediction analysis in the proposed method.

B. Results Using Synthetic and Real Vowels

We investigated the performance of the proposed method with the complex FBLP using the synthetic vowel /o/ and five real natural vowels. We compared experimentally the performance of the proposed method with that of the CLPC [7] and INCM [11] in white noise environment. INCM is an iterative noise compensated method, in which the estimation of noise power is computed by using a simplified noise power spectrum estimator proposed by Martin [12].

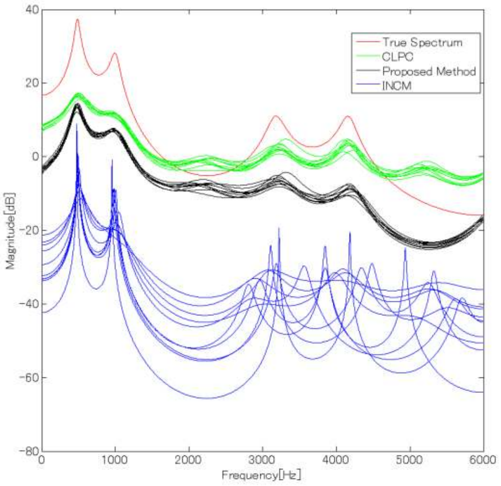

Figure 2. LPC spectra for synthetic vowel /o/ corrupted by white noise at SNR=10dB

For each noise level, 100 individual trials to generate white noise were conducted again, and the calculation of (19) was averaged. In Table III, for the synthesized vowel /o/ the evaluated AE is shown at SNR=20dB, SNR=15dB and SNR=10dB. It is observed that the proposed method is able to provide reduced AE at any SNR. It should be noticed here that comparing the column of the proposed method in Table III with that of the FBLP method with

[1] The synthetic vowel /o/ was generated from an impulsive sequence excitation through the transfer function in (1) with the following parameters: G = 0.134, a_1 =0.134, a_2 =0.97789, a_3 =-1.48396, a_4 =1.78023, a_5 =-0.71707, a_6 =-0.73514, a_7 =0.76348, a_8 =-0.12135, a_9=-0.15552, a_{10}=1.78143

256 samples in Table II, the proposed method with the complex FBLP provides more accurate formant frequency estimation with reduced AE values. Fig. 2 shows LPC spectra for the synthetic vowel /o/ corrupted by white noise at SNR=10dB. LPC spectra of the proposed method are most similar to the true spectrum. INCM can not estimate the F3 well.

Next, we investigated in natural vowels /a/, /i/, /u/ ,/e/ and /o/. In the experiment of natural vowels, the true frequency $f_t(i)$ was defined as the estimate of formant frequencies obtained by each method in noise free environment. The evaluated AE values of formant frequencies are shown in Table IV. Table IV shows that the formant estimation accuracy of the proposed method is significant and holds the best performance regardless to SNR. Especially, the estimates of F1 by the proposed method provide high accuracy.

From Tables III and IV, it is commonly observed that the proposed method behaves more robustly against noise in lower SNR conditions.

TABLE III. AE VALUES FOR SYNTHETIC VOWEL /O/

synthetic vowel		Prop	CLPC	INCM
20dB	F1	0.0090	0.0089	0.0193
	F2	0.0093	0.0095	0.0163
	F3	0.0032	0.0103	0.0137
15dB	F1	0.0162	0.0265	0.0213
	F2	0.0104	0.0123	0.0131
	F3	0.0086	0.0143	0.0239
10dB	F1	0.0290	0.0300	0.0502
	F2	0.0411	0.0647	0.0684
	F3	0.0425	0.0924	0.1144

TABLE IV. AE VALUES FOR NATURAL VOWELS

synthetic vowel		Prop	CLPC	INCM
20dB	F1	0.0446	0.0577	0.0687
	F2	0.0347	0.0525	0.0495
	F3	0.0032	0.0104	0.0138
15dB	F1	0.0588	0.1354	0.1314
	F2	0.0421	0.0638	0.0595
	F3	0.0671	0.0632	0.2032
10dB	F1	0.0997	0.1532	0.1456
	F2	0.0472	0.1478	0.0870
	F3	0.0685	0.0655	0.2835
5dB	F1	0.1351	0.1739	0.1509
	F2	0.1293	0.2434	0.1771
	F3	0.0754	0.0923	0.3688

IV. CONCLUSION

In this paper, we have presented a new CLPC analysis method, in which the signal which is ignored in the conventional CLPC analysis is effectively utilized to achieve noise reduction. We investigated the performance of the proposed method using synthetic and natural vowels. As a result, the formant estimation accuracy of the proposed method is significant and holds the best performance in comparing with that of the conventional methods regardless to SNR.

APPENDIX A ANALYSIS OF THE ANALYTIC SIGNAL

In this Appendix, the new analytic signal (18) is discussed more.

The speech signal $x(m)$ is considered as a combination of multiple sine waves. In the same way, the analytic signals $z_1(n)$ and $z_2(n)$ also have the same characteristics. The analytic signals $z_1(n)$ and $z_2(n)$ can be expressed by the equations belows respectively.

$$z_1(n) = \sum_{i=0}^{\infty} a_i cos\,(2\pi i f_0(2m) + \theta_i) \tag{20}$$

$$z_2(n) = \sum_{i=0}^{\infty} a_i cos\,(2\pi i f_0(2m+1) + \theta_i) \tag{21}$$

where $f_0 = \frac{1}{T_0}$ is the fundamental frequency, T_0 is the pitch period. The analytic signal $\hat{z}(n)$ used by the proposed method is generated by averaging $z_1(n)$and $z_2(n)$. The analytic signal $\hat{z}(n)$ can be expressed by

$$\hat{z}(n) = \frac{z_1(n)+z_2(n)}{2} \tag{22}$$

$$= \sum_{i=0}^{\infty} a_i cos \frac{(2\pi i f_0(2m)+\theta_i)+(2\pi i f_0(2m+1)+\theta_i)}{2}$$

$$\frac{(2\pi i f_0(2m+1)+\theta_i)-(2\pi i f_0(2m)+\theta_i)}{2} \tag{23}$$

$$= \sum_{i=0}^{\infty} a_i cos \frac{(4\pi i f_0(2m)+2\pi i f_0+2\theta_i)}{2} cos \frac{2\pi i f_0}{2} \tag{24}$$

$$= \sum_{i=0}^{\infty} a_i cos 2\pi i f_0 \left(2m + \frac{1}{2}\right) cos \pi i f_0 \tag{25}$$

The larger the i is, the closer to 0 the value of the $cos\pi i f_0$ is. It means that the components of the $cos\pi i f_0$can suppress the spectrum in the high frequency region. The effect of the $cos\pi i f_0$ is equivalent to play a role of low-pass filter in the frequency domain. In this case, the locations of formants included in the analytic signal $\hat{z}(n)$ do not change. Therefore, (25) suggests that in a noiseless case, the proposed method does not provide a performance degradation.

REFERENCES

[1] L. Deng, A. Acero, and I. Bazzi,"Tracking vocal tract resonances using a quantized nonlinear function embedded in a temporal constraint," *IEEE Trans. Audio Speech Lang. Processing*, vol. 14, no. 2, pp. 425-434, 2006.

[2] L. Welling and H. Ney, "Formant estimation for speech recognition," *IEEE Trans. Speech Audio Process*, vol. 6, no. 1, pp. 36-48, 1998.

[3] K. Mustafa and I. C. Bruce, "Robust formant tracking for continuous speech with speaker variability," *IEEE Trans. Audio Speech Lang. Processing*, vol. 14, no. 2, pp. 435-444, 2006.

[4] R. C. Snell and F. Milinazzo, "Formant location from LPC analysis data," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 2, pp. 129-134, 1993.

[5] G. Duncan and M. A. Jack, "Pole focusing: A new approach to LPC based speech analysis offering superior formant resolution," in *Proc. IEEE ICASSP*, 1988, pp. 2256-2259.

[6] S. Mc Candless, "An algorithm for automatic formant extraction using linear prediction spectra," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-22, pp.135-141, 1974.

[7] T. Shimamura and S. Takahashi, "Complex linear prediction method based on positive frequency domain," *Electronics and Communication in Japan Part 3*, vol. 73, no. 9, pp. 68-77, 1990.

[8] S. M. Kay, "Maximum entropy spectral estimation using the analytic signal," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-26, no. 10, pp. 467-469, 1978.

[9] S. L. Marple, *Digital Spectral Analysis with Applications*, Prentice Hall, 1987.

[10] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 2001.

[11] A. Trabelsi and M. Boukadoum, "Improving LPC analysis of speech in additive noise," in *Proc. IEEE NEWCAS*, 2007, pp. 93-96.

[12] R. Martin, "Noise power spectral estimation based on optimal smoothing and minimum statistics," *IEEE Trans. on Speech Audio Process*, vol. 9, no. 2, pp. 504-512, 2001.

Takuma Kaneko was born in Hokkaido, Japan, on July 16, 1990. He received the B.E. degree information engineering from Saitama University, Saitama, Japan, in 2013. He is currently pursuing the M.E. degree at Saitama University. His research interests include noise reduction and speech signal processing.

Tetsuya Shimamura received the B.E, M.E., and Ph.D degrees in electrical engineering from Keio University, Yokohama, Japan, in 1986, 1988, and 1991, respectively. In 1991, he joined Saitama University, Saitama, Japan, where he is currently a Professor. He was a visiting researcher at Longhborough University, U.K. in 1995 and at Queen's University of Belfast, U.K. in 1996, respectively. Prof. Shimamura is an author and coauthor of 6 books. He serves as an editorial member of several international journals and is a member of the organizing and program committees of various international conferences. His research interests are in digital signal processing and its application to speech, image, and communication systems.