A TIME-VARYING COMPLEX AR SPEECH ANALYSIS
BASED ON GLS AND ELS METHOD

Ketichi FUNAKI
Computing and Networking Center
University of The RYUKYUS
Senbaru 1, Nishihara, Nakagami, OKINAWA, 903-0213, JAPAN
funaki@cc.u-ryukyu.ac.jp

ABSTRACT

We have already developed three kinds of time-varying complex AR (TV-CAR) parameter estimation algorithms for analytic speech signal, which are based on minimizing mean square error (MMSE), Huber’s robust M-estimation and Instrumental Variable (IV) method. This paper presents novel robust TV-CAR model parameter estimation algorithms on the basis of a Generalized Least Square (GLS) and Extended Least Square (ELS) method, in which the equation error is modeled by complex AR model with white Gaussian input to whiten the equation error. The experiments with natural speech corrupted by white Gaussian demonstrate that the proposed methods achieve robust spectral estimation against additive white Gaussian.

1. INTRODUCTION

LP C methods[1][2] and their derivatives are the most widely adopted speech analysis methods in speech processing and these methods are essential for low bit rate speech coding. The LPC methods, however, can not extract time-varying features from speech signal owing to the assumption of local stationary of speech although speech signal provides time-varying feature. Several speech analysis methods which can estimate the time-varying model parameters have already been proposed[3][4]. These methods are non-recursive analysis methods in which the model parameters are modeled by first order or second order polynomial function.

On the other hand, complex LPC methods for an analytic signal have already been proposed[5][6]. Analytic signal is a complex-valued signal whose real part is an observed signal and whose imaginary one is a Hilbert transformation of the observed signal. Since the spectrum of analytic signals is zero in negative frequency domain (−, 0), analytic signals can be decimated by a factor two with no degradation. For this remarkable feature, these complex-valued methods for the analytic signal take some advantages over conventional real-valued LPC methods, i.e., more accurate spectral estimation, smaller errors in terms of computation with finite precision as well as quantization of the coefficients, and so on. We have already developed a non-recursive complex speech analysis based on minimizing mean square error (MMSE) for analytic signal by introducing a time-varying complex AR (TV-CAR) model, in which the parameters are modeled by complex basis expansion[7]. In this MMSE-based method the complex AR coefficients can be efficiently estimated by solving linear equation by means of a complex-valued LDU decomposition. The method can extract time-varying features from speech signal with non-recursive processing. Although an MMSE method can achieve optimal estimation when underlying observationss are Gaussian, it is well known that an MMSE-based method suffers from biased estimation in noisy environment, especially for non-Gaussian nature of signal with heavy-tailed outliers. To cope with this problem, we have already developed the TV-CAR speech analysis method based on Huber’s robust M-estimation[8]. Moreover, we have proposed the robust TV-CAR speech analysis based on an Instrumental Valuable (IV) approach, which may be robust against noisy environment[9]. The experiments with natural speech corrupted with white Gaussian indicated that the IV-based TV-CAR model can realize more robust estimation of parameters in noisy environment. However, the IV-based method requires input estimation. It is very difficult to estimate the accurate input, as a result, unstable spectral estimation sometimes occurs in the IV-based method.

In this paper, two kinds of new robust TV-CAR speech analysis methods are presented based on Generalized Least Square (GLS) and Extended Least Square (ELS). In these approaches, the equation error of MMSE estimation is whitened by an AR filter. Therefore, the simultaneous estimation of the model parameters and whitened AR filter coefficients realizes unbiased estimation owing to the whiteness of the equation errors. The iterative method is proposed to estimate the sub-optimal estimation of the TV-CAR model parameters as well as whitened AR coefficients.

This paper is organized as follows. In section 2, the time-varying complex AR (TV-CAR) model is explained briefly. In section 3, the MMSE-based TV-CAR parameter estimation algorithm is shown. In section 4, robust TV-CAR parameter estimation algorithms based on GLS and ELS method are then explained, respectively. In section 5, experiments with natural speech signals corrupted by white Gaussian are shown to prove the effectiveness of the proposed method.
2. SPEECH PRODUCTION MODEL

Target signal of the time-varying complex AR (TV-CAR) method is an analytic signal that is complex-valued signal defined by

\[ y^c(t) = \frac{y(2t) + jy_H(2t)}{\sqrt{2}} \]  

where \( y^c(t) \), \( y(t) \), and \( y_H(t) \) denote an analytic signal at time \( t \), an observed signal at time \( t \), and a Hilbert transformed signal for the observed signal \( y(t) \), respectively.

Since analytic signals do have the spectra only over the range \((0, \pi)\), analytic signals can be decimated by a factor two. In Eq.(1), the term of \( \sqrt{2} \) is divided in order to adjust the power of an analytic signal with that of the observed one.

The TV-CAR model is defined as follows.

\[
a_i^c(t) = \sum_{i=0}^{L-1} g_{i,d} f_i^c(t) \\
y^c(t) = - \sum_{i=0}^{L-1} a_i^c(t) y^c(t-i) + u^c(t) \\
H(z,t) = \frac{1}{1 + \sum_{i=0}^{L-1} g_{i,d} f_i^c(t) z^{-i}}
\]

where \( u^c(t) \), \( H(z,t) \), \( a_i^c(t) \), \( L \), and \( f_i^c(t) \) are taken to be a complex-valued input, a transfer function of the model, \( i \)-th complex AR coefficient at time \( t \), AR order, an order of complex basis expansion, and a complex-valued basis function, respectively. In the TV-CAR model, the complex AR coefficient is expressed by a finite number of any complex basis such as complex Fourier basis \( \{ f_i^c(t) = e^{-j2\pi it/N} \} \), first order polynomial \( \{ f_i^c(t) = 1 \} \), \( f_i^c(t) = t \), or wavelet function, or so on. Note that superscript \( ''c'' \) denotes complex value in this paper.

3. MMSE ALGORITHM

Eq.(3) can be expressed by the following matrix expression.

\[
y = \begin{bmatrix} g & a \end{bmatrix} \begin{bmatrix} v \end{bmatrix} + \begin{bmatrix} \bar{g}^c \end{bmatrix} \\
\bar{g}^c = \begin{bmatrix} \bar{g}_{1}^c, \bar{g}_{2}^c, \ldots, \bar{g}_{L}^c \end{bmatrix} \\
y^c = \begin{bmatrix} y^c(1), y^c(2), \ldots, y^c(N) \end{bmatrix} \\
a^c = \begin{bmatrix} a^c(1), a^c(2), \ldots, a^c(N) \end{bmatrix} \\
\bar{a} = \begin{bmatrix} \bar{a}_0, \bar{a}_1, \ldots, \bar{a}_L \end{bmatrix} \\
\bar{D}_0 = \begin{bmatrix} \bar{d}_0, \bar{d}_1, \ldots, \bar{d}_L \end{bmatrix} \\
\bar{a}_{i,d} = \begin{bmatrix} g^c(i), y^c(1), \ldots, y^c(N-i) \end{bmatrix}
\]

where \( g \) is \((N-I,1)\) column vector, \( \bar{g} \) is \((L \cdot I,1)\) column vector, \( \bar{a} \) is \((N-I, L \cdot I)\) matrix. Superscript \( ''c'' \) denotes transposition.

Minimizing the MMSE criterion for Eq.(5) with respect to \( \bar{a}_{i,j} \) leads to the following MMSE algorithm.

\[
(\bar{g}_{i,j}^c)^* \bar{a}_{i,j} = -\bar{g}_{i,j}^c g 
\]

where superscript \( '*' \) denotes Hermitian transposition.

By using the estimated parameter \( \bar{a} \), the equation error is calculated as follows.

\[
r^c(t) = g + \bar{a}^c \\
r^c = \begin{bmatrix} r^c(1), r^c(2), \ldots, r^c(N) \end{bmatrix}^T
\]

where \( r^c(t) \) is the equation error at time \( t \).

4. ROBUST ALGORITHM

If the equation error shown in Eq.(7) is white Gaussian, the MMSE estimation is optimal, however, it is rare case. Consequently, an MMSE estimation suffers from biased estimation. In GLS and ELS method, a complex AR filter is adopted to whiten the equation error as follows.

\[
r^w(t) = - \sum_{k=1}^{K} \tilde{b}_k r^c(t-k) + \varepsilon^c(t)
\]

where \( \tilde{b}_k \) is \( k \)-th coefficient of the white AR filter whose order is \( K \) and \( \varepsilon^c(t) \) is 0-mean white Gaussian.

Since the equation error is white Gaussian, simultaneous estimation of both parameters achieves unbiased estimation. Two kinds of algorithms are presented in this paper, i.e., GLS and ELS-based algorithm.

4.1. GLS Algorithm

The model parameters are determined so as to minimize the following criterion \( V \).

\[
\begin{align*}
V &= \frac{1}{2\pi} \int_{|z|=1} |W(z)(Y^c(z) - R(z))|^2 \frac{dz}{z} \\
&= \frac{1}{2\pi} \int_{|z|=1} |W(z)(A(z,t)Y^c(z) - E(z))|^2 \frac{dz}{z} \\
A(z,t) &= 1 + \sum_{i=1}^{L-1} a_i^c(t) z^{-i} \\
&= 1 + \sum_{i=0}^{L-1} \sum_{t=0}^{L-1} g_{i,d} f_i^c(t) z^{-i} \\
B(z) &= \sum_{k=1}^{K} \tilde{b}_k z^{-k} \\
R(z) &= \frac{E(z)}{B(z)}
\end{align*}
\]

where \( W(z) \), \( A(z,t) \), \( B(z) \), \( R(z) \) and \( E(z) \) are an weighted function, a transfer function of the inverse TV-CAR model, a transfer function of the inverse whiten filter, \( z \) transform of \( r^c(t) \), and \( z \) transform of \( \varepsilon^c(t) \), respectively.
In GLS, \( W(z) \) is selected as \( B(z) \), thus, Eq.(9) can be modified to as follows.

\[
V = \frac{1}{2\pi j} \int_{|z|=1} |A(z, t) Y(z) B(z)|^2 \frac{dz}{z}
\]  
(10)

\( A(z, t) \) can be estimated by MMSE estimation using the inverse filtered analytic speech, \( Y(z) B(z) \), after estimating the whitening filter \( 1/B(z) \).

\( B(z) \) can be estimated so as to minimize the following criteria by means of MMSE estimation.

\[
\frac{1}{2\pi j} \int_{|z|=1} |R(z) B(z)|^2 \frac{dz}{z}
\]  
(11)

The inverse filtered analytic speech \( y^*(t) \) is calculated by inverse filtering for analytic speech with \( B(z) \).

By using \( y^*(t) \), the TV-CAR model parameters are estimated by MMSE estimation shown in Eq.(12).

\[
(\tilde{\theta}^H \tilde{\theta}) \tilde{\theta} = -\tilde{\theta}^H \tilde{y}
\]
(12)

where,

\[
\tilde{\theta} = [\tilde{\theta}_0, \tilde{\theta}_1, ..., \tilde{\theta}_N] \]

\[
\tilde{\theta}^T = [y^*(I), y^*(I + 1), y^*(I + 2), ..., y^*(N - 1)]
\]

\[
\tilde{\theta}_k = [\tilde{\theta}_k, \tilde{\theta}_k, ..., \tilde{\theta}_k]
\]

\[
\tilde{\theta}^T_k = [y^*(I - k), y^*(I + 1 - k), ..., y^*(N - 1 - k)]
\]  
(13)

The simultaneous estimation of \( A(z, t) \) and \( B(z) \) is not possible since the equation error \( r^*(t) \) is calculated by \( \tilde{A}(z, t) \). For this reason an iterative procedure is required to estimate both parameters.

The iterative algorithm is explained as follows.

1. Initial \( A(z, t) \) is estimated by MMSE (Eq.(6)).
2. The equation error \( r^*(t) \) is calculated by Eq.(7).
3. \( B(z) \) is estimated so as to minimize Eq.(11) using \( r^*(t) \).
4. The inverse filtered analytic signal \( y^*(t) \) is calculated using the estimated \( B(z) \) and \( y^*(t) \).
5. The \( A(z, t) \) is estimated by Eq.(12).
6. Go to (2).

The iteration is repeated with the pre-determined number to get the convergence.

4.2. ELS Algorithm

ELS algorithm is more sophisticated than GLS. The estimated parameter by MMSE method, \( \tilde{\theta} \), contains the biased parameter, \( \tilde{\theta}_{bias} \). In ELS, the biased parameter is estimated and unbiased parameter is obtained by subtracting the biased parameter from the MMSE one. The biased parameter \( \tilde{\theta}_{bias} \) is calculated by the following equation.

\[
(\tilde{\theta}^H \tilde{\theta}) \tilde{\theta}_{bias} = -\tilde{\theta}^H \tilde{r}
\]  
(14)

where,

\[
r^T = [r^*(I), r^*(I + 1), r^*(I + 2), ..., r^*(N - 1)]
\]

\[
r^*(t) = -\sum_{k=1}^{K} \hat{\theta}_k r^*(t - k).
\]  
(15)

The unbiased parameter \( \tilde{\theta} \) is calculated by \( \tilde{\theta} = \tilde{\theta} - \tilde{\theta}_{bias} \). For the same reason as in GLS, the ELS requires an iterative method. The iterative algorithm is shown as follows.

1. Initial \( A(z, t) \) is estimated by MMSE (Eq.(6)).
2. The equation error \( r^*(t) \) is calculated by Eq.(7).
3. \( B(z) \) is estimated so as to minimize Eq.(11) using \( r^*(t) \).
4. The biased parameter \( \tilde{\theta}_{bias} \) is calculated by Eq.(14).
5. The unbiased parameter is calculated by \( \tilde{\theta} = \tilde{\theta} - \tilde{\theta}_{bias} \).
6. Go to (2).

(1),(2),(3) are the exactly same as in GLS. The iteration is repeated with the pre-determined number to get the convergence.

5. EXPERIMENTS

In order to evaluate the proposed method, running spectrum of natural speech has been estimated by proposed methods. The testing natural speech /jibui/ is 10KHz sampled speech that is converted from 20KHz sampled ATR database data and its speaker is adult female, FKN. Figure 1 shows the results for speech /jibui/ without any additive noise. (SNR=\infty[dB]) shown in Figure 1(a). Figure 2 shows the results for speech /jibui/ added with white Gaussian (SNR=0[dB]) shown in Figure 2(a). In order to compare the performance, two conventional methods were also tested, LPC method and the MMSE-based TV-CAR method.

Table 1 shows the analysis conditions of each method.

<table>
<thead>
<tr>
<th>Method</th>
<th>( I )</th>
<th>( L )</th>
<th>( N )</th>
<th>( S )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b) LPC</td>
<td>14</td>
<td>-</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>(c) MMSE TV-CAR [7]</td>
<td>7</td>
<td>2</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>(d) GLS-based TV-CAR</td>
<td>7</td>
<td>2</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>(e) ELS-based TV-CAR</td>
<td>7</td>
<td>2</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

In all methods, the analysis length \( N \) and analysis shift length \( S \) are 20[msec] and 10[msec], respectively. In (b), Hamming window is operated and in otherwise no window is operated. Analysis order is 14 in the real valued method (b) and 7 in the complex-valued methods (c)-(e). In the TV-CAR methods (c)-(e), \( L \) is set to be 2 and first order polynomial is adopted as basis function. In the proposed methods (d) and (e), the iteration number is 5 and the white AR filter order is 5. Moreover, (20,20) IR filter [10] is adopted to realize Hilbert transform.

In both figures, spectrum is drawn at every 2[msec], (b) can only estimate one spectrum for one analysis frame, therefore, the same spectrum is repeatedly drawn within the same analysis frame.

Figure 1 demonstrates that in clear environment the proposed robust methods can estimate accurate speech spectrum in high frequencies although the MMSE method cannot extract accurate speech spectrum in high frequencies.

Figure 2 demonstrates that even in low SNR environment the proposed robust methods can extract speech spectrum although the MMSE method and LPC method only can estimate distorted and flattened speech spectrum.
6. CONCLUSIONS

Novel robust TV-CAR speech analysis methods have been proposed based on Generalized Least Square (GLS) and Extended Least Square (ELS) method, in which the equation error is modeled by AR model in order to whiten the equation error and to realize unbiased estimation. Two kinds of robust algorithms with iterative parameter estimation, viz., GLS-based and ELS-based algorithm, are proposed. The experimental results with natural speech and natural speech corrupted by white Gaussian demonstrate that the proposed methods can estimate more accurate speech spectrum in high frequencies in clear environment than the MMSE-based method and can estimate less distorted speech spectrum in low SNR environment than the MMSE and LPC method. The difference between two proposed methods is under investigation and these methods will have to be evaluated in speech coding and speech recognition.

7. ACKNOWLEDGEMENT

This work is partially supported by Grant-in-Aid of Japan Society for the Promotion of Science.

8. REFERENCES


