2025 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)

GCI detection and glottal wave estimation based on TV-CAR speech analysis

Keiichi FUNAKI*
* University of the Ryukyus
E-mail: funaki@cc.u-ryukyu.ac.jp

***Abstract*—Accurate detection of Glottal Closure Instances (GCIs) is crucial for speech processing applications. The SEDREAMS method detects GCIs by identifying peaks in LPC residual within candidate intervals determined from mean-based signals. However, LPC analysis may not adequately separate glottal excitation from vocal tract characteristics, especially in low frequencies. This paper proposes replacing LPC residual signals with complex ones from Time-Varying Complex AR (TV-CAR) analysis of analytic signals. TV-CAR provides improved low-frequency estimation by eliminating negative frequency components. Experiments on the CMU Arctic database demonstrate that TV-CAR-based SEDREAMS achieves better identification detection rates than conventional LPC-based methods, particularly for male speakers. This paper proposes a method to estimate the GCI using the TV-CAR method.**

I. INTRODUCTION

The speech signal is generated when the periodic waveform produced by the vocal cords passes through the vocal tract filter, whose frequency response is determined by the shape of the vocal tract. The model that describes the voice production process as a direct connection between a glottal source and a vocal tract filter is called a source-filter model. To avoid estimating the characteristics of the glottal source as the characteristics of the vocal tract filter, in the past, methods such as covariance LPC analysis of short-term intervals in the closed interval of the glottal excitation, which was called the short-term interval analysis method[1], and sample selective LPC analysis[2], which selects the interval of the zero source for LPC analysis, were proposed. Still, in practice, the automatic obtaining of closed intervals was challenging. Various attempts have been made to estimate the glottal excitation from the speech signal. A method for estimating the vocal tract filter, which is the spectral component of the speech signal, and estimating the glottal source by inverse filtering has been studied from [3]. It is called Glottal Inverse Filtering (GIF). P.Alku has developed a more accurate GIF by repeating the inverse filtering process four times, named IAIF(Iterative Adaptive Inverse Filtering)[4]. Furthermore, a method has been proposed to estimate the model parameters of the glottal excitation model, such as the LF-model[5], from the residual signal obtained by the GIF with the vocal tract characteristics adequately removed. Repetition is necessary because if the spectral analysis extracts even the glottal source spectrum other than the vocal tract filter, the glottal source information is removed from the source obtained by the inverse filter, so a method of gradual subtraction is used. However, it is difficult to remove the glottal source from speech. Kane et al. [6][7] have proposed a more effective glottal inverse filtering method with DP matching that is matched so that parameters do not vary significantly between adjacent pitches. Since the 1980s, simultaneous estimation of both glottal excitation and vocal tract filter has been investigated to solve this problem. In this method, glottal excitation is modeled by the glottal excitation model, including the LF-model[5]. Starting with Per Hedelin's Glottal LPC analysis[8], Fujisaki et al. proposed GARMA analysis[9], LPC analysis by inverse filtering of the glottal excitation[10] and ARX (auto-regressive exogenous) analysis[11], and Glottal-ARMAX analysis, solving the simultaneous optimization problem with Genetic Algorithm (GA) and Simulated Annealing (SA), and improving the accuracy by introducing phase correction[12] and multirate processing[13]. Subsequently, Kaburaki et al. proposed a method using a constrained polynomial model[14]. Fu et al.[15] presented a two-step optimization in which a simpler glottal source model, RK model, was used to estimate the preliminary values for a more complex glottal source model, LF model, and then the ARX model, as the vocal tract model, was combined for joint optimization. Recently, Li et al.[16] have proposed a more effective ARX analysis named ARX-LF analysis that introduces a preliminary selection of the LF-model parameters using [6][7] and final selection using ARX model identification. In the ARX-LF analysis, GCI (Glottal Closure Instance) estimation is essential to determining the location of the glottal excitation, and the SEDREAMS method cite sedreams proposed by Drugman et al. is chosen as a GCI estimation method. The SEDREAMS decides the GCI and GOI (Glottal Open Instance) using LPC residual. The LPC residual removes formant components that only interfere with the estimation.

On the other hand, we have proposed a Time-Varying Complex AR (TV-CAR) speech analysis for an analytic signal and applied it to speech processing. The analytical signal is a complex signal with the observed speech signal in the real part and its Hilbert transform signal in the imaginary part. Since the negative frequency region spectrum is zero, it can be thinned by a ratio of 2. As a result, the accuracy of spectrum estimation in the low-frequency domain is improved. A TV-CAR model is introduced, in which an arbitrary complex basis function expansion describes the complex AR coefficients. The TV-CAR analysis is proposed to estimate complex parameters from the analytic signal, including Minimizing Mean Squared

979-8-3315-7206-8/25/$31.00

Error (MMSE) analysis[18], robust analysis[19], ℓ_1-norm regularization method[20], ℓ_2 norm regularization method[21][22], and Total Least Squares (TLS) method. It is applied for the estimation of speech F_0[23][24][25] and the robust ASR front-end [26]. In particular, the estimation F_0 using TV-CAR residual improves the estimation accuracy because formant components are removed more appropriately. Thus, TV-CAR residual provides fewer formant components than the LPC residual does. We have already shown the effect on the estimation F_0. As a result, it is expected that the use of the TV-CAR residual in the SEDREAMS method rather than the LPC residual can improve the accuracy of the estimation of GCI and GOI. In this paper, we construct the SEDREAMS method using the TV-CAR residual and compare it using the LPC residual using objective evaluation criteria. Moreover, the TV-CAR analysis makes it more accurate to realize the IAIF to estimate the differential glottal excitation. The IAIF, based on the TV-CAR analysis, is also proposed.

II. GCI ESTIMATION USING LPC RESIDUAL

The SEDREAMS method [17] consists of the steps described in Subsections A and B.

A. Interval Determination from Mean-Based Signal

In the first step, a mean-based signal is calculated; from this, the intervals where speech events, GCI, and GOI are expected to occur are extracted.

The mean-based signal $p(n)$ is calculated using the following equation:

$$p(n) = \frac{1}{2N+1} \sum_{m=-N}^{N} w(m) y(n+m) \tag{1}$$

where $y(n)$ is the speech signal, and $w(m)$ is a window function of length $2N+1$. A Blackman window is used as the window function, and its length is set to 1.75 times the average pitch period. The mean-based signal has a characteristic that changes with the local pitch cycle. The GCI is expected to exist between the minimum value of the mean-based signal and the subsequent positive zero-crossing. The GOI is expected to exist between the maximum value of the mean-based signal and the subsequent negative zero-crossing. A margin of 0.25 ms is added on both sides of the interval for the GOI.

B. GCI and GOI Position Estimation from LPC Residual

In the second step, the exact position of the speech event (GCI or GOI) within the interval obtained in the previous step is determined.

The exact position of the event within each interval is determined using the LPC residual. For the GCI, the maximum peak of the LPC residual within the interval is assumed to be the GCI position. A similar method is applied for the GOI. However, the excitation at the GOI is less intense and more diffuse, making it less distinct compared to the GCI. This method allows for the accurate and unique detection of the GCI and GOI.

These two consecutive steps effectively detect GCI and GOI from speech signals. However, the detection accuracy for GOI tends to be lower than that of GCI[17].

III. IAIF METHOD

The IAIF method[4], proposed by P. Alku in the 1990s, is a widely used technique worldwide that estimates glottal excitation by iterating LPC analysis four times. It is represented by a block diagram consisting of processes 1 through 10 shown in Figure 1. 1. LPC analysis of order 1: Performs first-order LPC analysis. 2. Inverse filtering: Performs inverse filtering using the first-order AR filter obtained in step 1. This serves as adaptive pre-emphasis. 3. LPC analysis of order t_1 that performs LPC analysis of order t_1. 4. Inverse filtering that performs inverse filtering using the order t_1 AR filter obtained in step 3. 5. Integration: Performs integration to eliminate lip radiation characteristics. 6. LPC analysis of order g_2 that performs LPC analysis of order g_2. 7. Inverse filtering that performs inverse filtering using the order t_1 AR filter obtained in step 6 to estimate the glottal excitation $g_1(n)$. 8. LPC analysis of order t_2 that performs LPC analysis of order t_2. 9. Inverse filtering that performs inverse filtering using the order t_1 AR filter obtained in step 8. 10. Integration that performs integration to eliminate lip radiation characteristics and estimates the glottal excitation $g_a(n)$.

IAIF performs adaptive pre-emphasis and repeats the process of LPC analysis, inverse filtering, and integration twice to estimate the glottal excitation. The first adaptive pre-emphasis uses a first-order PARCOR coefficient, while the second adaptive pre-emphasis uses an AR filter obtained through LPC analysis of a slightly higher order. This is followed by LPC analysis, inverse filtering, and integration to estimate the glottal excitation.

IV. TV-CAR MODEL

Time-Varyng Complex AR(TV-CAR) model is defined by Eq.(2).

$$\begin{aligned} y^c(t) &= -\sum_{i=1}^{I} a_i^c(t) y^c(t-i) + u^c(t) \\ &= -\sum_{i=1}^{I} \sum_{l=0}^{L-1} g_{i,l}^c f_l^c(t) y^c(t-i) + u^c(t) \end{aligned} \tag{2}$$

where $a_i^c(t)$ is the ith-order AR coefficient at time t and $g_{i,l}^c$ denotes the complex parameter, and $f_l^c(t)$ denotes an arbitrary complex basis and L denotes the complex basis order and $y^c(t)$ is the analytic signal at time t and $u^c(t)$ is the complex input at time t. Eq.(3) calculates the complex residual signals.

$$r^c(t) = y^c(t) + \sum_{i=1}^{I} \sum_{l=0}^{L-1} \hat{g}_{i,l}^c f_l^c(t) y^c(t-i) \tag{3}$$

where $\hat{g}_{i,l}^c$ is an estimate of $g_{i,l}^c$. Eq.(2) and Eq.(3) are denoted by the vector-matrix notation as in Eq.(5) and Eq.(4), respectively.

$$\begin{aligned}\mathbf{r}_c &= \mathbf{y}_c + \mathbf{\Phi}_c\hat{\theta} &(4)\\ &= [r^c(I), r^c(I+1), r^c(I+2), \cdots, r^c(N-1)]^T\end{aligned}$$

where $\hat{\theta}$ is the estimated vector of θ and T means transpose.

$$\begin{aligned}\mathbf{y}_c &= -\mathbf{\Phi}_c\theta + \mathbf{u}_c &(5)\\ \theta &= [\mathbf{g}_0^T, \mathbf{g}_1^T, \cdots, \mathbf{g}_l^T, \cdots, \mathbf{g}_{L-1}^T]^T &(6)\\ \mathbf{g}_l &= [g_{1,l}^c, g_{2,l}^c, \cdots, g_{i,l}^c, \cdots, g_{I,l}^c]^T\\ \mathbf{y}_c &= [y^c(I), y^c(I+1), y^c(I+2), \cdots, y^c(N-1)]^T\\ \mathbf{u}_c &= [u^c(I), u^c(I+1), u^c(I+2), \cdots, u^c(N-1)]^T\\ \mathbf{\Phi}_c &= [\mathbf{S}_0^c, \mathbf{S}_1^c, \cdots, \mathbf{S}_l^c, \cdots, \mathbf{S}_{L-1}^c]\\ \mathbf{S}_l^c &= [\mathbf{s}_{1,l}^c, \mathbf{s}_{2,l}^c, \cdots, \mathbf{s}_{i,l}^c, \cdots, \mathbf{s}_{I,l}^c]\\ \mathbf{s}_{i,l}^c &= [y^c(I-i)f_l^c(I), y^c(I+1-i)f_l^c(I+1),\\ &\qquad \cdots, y^c(N-1-i)f_l^c(N-1)]^T\end{aligned}$$

V. TV-CAR ANALYSIS

A. MMSE algorithm

Since the ℓ_2 norm minimization method, MMSE analysis minimizes MSE, it is defined by (7).

$$\hat{\theta}_{MMSE} = \arg\min_{\theta} \mathbf{r}_c^H \mathbf{r}_c \tag{7}$$

The MMSE analysis is realized by Eq.(8).

$$\left(\mathbf{\Phi}_c^H \mathbf{\Phi}_c\right) \hat{\theta}_{MMSE} = -\mathbf{\Phi}_c^H \mathbf{y}_c \tag{8}$$

where H is the Hermite operator. The reason why H is applied is that the matrix and vector contain the complex-valued signals.

VI. PROPOSED METHOD

This paper proposes a modified SEDREAMS, and IAIF method introducing the TV-CAR analysis for the analytic signal instead of LPC analysis for speech signal. The TV-CAR analysis analyzes the analytic signal, and the complex AR residual is calculated by applying the inverse filter of the estimated TV-CAR filter to the analytic signal using Eq.(3). We have proposed several TV-CAR algorithms. In this case, non-noise corrupted speech is applied. For this reason, this paper estimates $g_{i,l}^c$ using MMSE-based TV-CAR analysis denoted as Eq.(8).

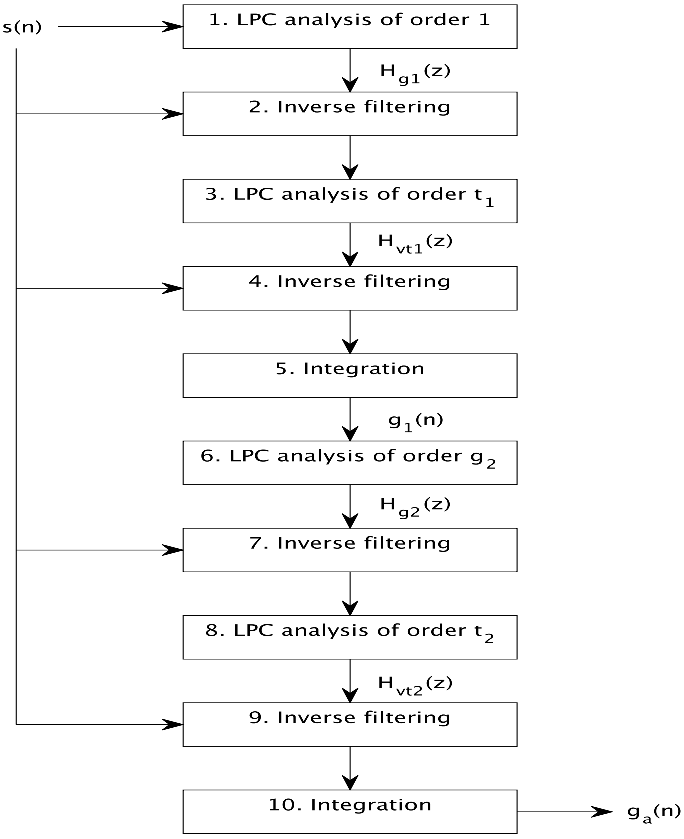

Fig.1: IAIF algorithm

VII. EXPERIMENTS

By using the CMU arctic database[27] with EGG signals, the TV-CAR residual signals are compared with LPC residual on the SEDREAMS method.

· RAR: Time-Invariant Real AR Analysis,
· TV-RAR: Time-Varying Real AR Analysis,
· CAR: Time-Invariant Complex AR Analysis,
· TV-CAR: Time-Varying Complex AR Analysis,

The analysis conditions are shown in Table 1.

Table 1: Experimental Conditions

Speech data	CMU Arctic EGG Database[27] Male sentences bdl 2 and jmk Female sentences slt
Sampling	16kHz/16bit
Analysis window	Window Length: 25[ms] Shift Length: 5[ms]
LPC	$I = 24$(Time-Invariant)
RAR	$I = 24, L = 1$(Time-Invariant)
TV-TAR	$I = 24, L = 2$(Time-Varying)
CAR	$I = 12, L = 1$(Time-Invariant)
TV-CAR	$I = 12, L = 2$(Time-Varying)
Basis	$f_l^c(t) = t^l/l!$
Pre-emphasis	None

In Figure 2:

(a) shows the waveform,

(b) shows the mean-based signal calculated from the speech signal using Eq.(1),
(c) shows the GCI interval calculated from (b),
(d) shows the GOI interval calculated from (b),
(e) shows the corresponding EGG signal,
(f) shows the LPC residual,
(g) shows the RAR residual,
(h) shows the TV-RAR residual,
(i) shows the CAR residual,
(j) shows the TV-CAR residual.

In (e) through (j), circles (○) represent estimated GCIs, and crosses (×) represent estimated GOIs. Figure 2 suggests that TV-CAR analysis provides the most accurate GCI estimation.

The evaluation uses the objective evaluation values of IDR, MR, FAR, and IDA proposed by Naylor et al.[28]. These indexes are as follows. The pitch cycle is determined from the mean of the reference GCI/GOI. Only one estimated GCI/GOI on the pitch cycle exists. It is regarded as a successful estimation.

· IDR (Identification Detection Rate): succeed pitch cycle rate
· MR (Miss rate): The rate when no estimated GCI/GOI on the pitch cycle exists.
· FAR (False alarm rate): The rate when more than two estimated GCI/GOI on the pitch cycle exists.
· IDA (Identification Accuracy): On the successful pitch cycle, the standard deviation of the absolute error between the estimated and reference GCI/GOI.

Reference GCIs and GOIs were determined from the differential EGG signals in the database. The results are shown in Tables 2 and 3. Tables 2 and 3 show the GCI and GOI estimations results, respectively. Table 2 shows that the TV-CAR analysis performs better than LPC in terms of GCI, while Table 3 shows that LPC performs better than TV-CAR in terms of GOI. GCI is more important than GOI; thus, TV-CAR performs better than LPC overall.

Figure 3 shows the experimental results of the IAIF. The experimental conditions are the same as in Table 1, with $t_1 = I$, $g_2 = 8(real), 4(complex)$, and $g_2 = I$. In the figure:

(a) Shows the estimated glottal excitation $g_1(n)$ using adaptive pre-emphasis with LPC analysis
(b) Shows the estimated glottal excitation $g_a(n)$ using IAIF with LPC analysis
(c) Shows the estimated glottal excitation $g_1(n)$ using adaptive pre-emphasis with RAR analysis
(d) Shows the estimated glottal excitation $g_a(n)$ using IAIF with RAR analysis
(e) Shows the estimated glottal excitation $g_1(n)$ using adaptive pre-emphasis with TV-RAR analysis
(f) Shows the estimated glottal excitation $g_a(n)$ using IAIF with TV-RAR analysis
(g) Shows the estimated glottal excitation $g_1(n)$ using adaptive pre-emphasis with CAR analysis
(h) Shows the estimated glottal excitation $g_a(n)$ using IAIF with CAR analysis
(i) Shows the estimated glottal excitation $g_1(n)$ using adaptive pre-emphasis with TV-CAR analysis
(j) Shows the estimated glottal excitation $g_a(n)$ using IAIF with TV-CAR analysis

Figure 3 suggests that TV-CAR analysis can estimate glottal estimation with more accuracy.

Table 2 Comparison of GCI estimation methods

(1)Male speakers

	IDR	MR	FAR	IDA
LPC	97.57%	1.53%	0.90%	0.54
RAR	98.04%	1.30%	0.66%	0.46
TV-RAR	97.75%	1.44%	0.81%	0.46
CAR	97.75%	1.44%	0.81%	0.47
TV-CAR	97.75%	1.44%	0.81%	0.47

(b)Female speaker

	IDR	MR	FAR	IDA
LPC	95.49%	4.35%	0.16%	0.24
RAR	95.49%	4.35%	0.16%	0.26
TV-RAR	95.49%	4.35%	0.16%	0.26
CAR	95.49%	4.35%	0.16%	0.24
TV-CAR	95.49%	4.35%	0.16%	0.24

(c)Male and Female speakers

	IDR	MR	FAR	IDA
LPC	96.53%	2.94%	0.53%	0.39
RAR	96.76%	2.82%	0.41%	0.36
TV-RAR	96.62%	2.90%	0.48%	0.36
CAR	96.62%	2.90%	0.48%	0.35
TV-CAR	96.62%	2.90%	0.48%	0.36

Table 3 Comparison of GOI estimation methods

(1)Male speakers

	IDR	MR	FAR	IDA
LPC	97.35%	2.04%	0.61%	0.50
RAR	97.47%	1.92%	0.61%	0.75
TV-RAR	97.47%	1.92%	0.61%	0.76
CAR	97.47%	1.92%	0.61%	0.81
TV-CAR	97.47%	1.92%	0.61%	0.81

(b)Female speaker

	IDR	MR	FAR	IDA
LPC	95.28%	4.26%	0.47%	0.42
RAR	95.11%	4.26%	0.63%	0.54
TV-RAR	95.11%	4.26%	0.63%	0.53
CAR	95.11%	4.26%	0.63%	0.54
TV-CAR	95.11%	4.26%	0.63%	0.54

(c)Male and Female speakers

	IDR	MR	FAR	IDA
LPC	96.31%	3.15%	0.54%	0.46
RAR	96.29%	3.09%	0.62%	0.64
TV-RAR	96.29%	3.09%	0.62%	0.65
CAR	96.29%	3.09%	0.62%	0.68
TV-CAR	96.29%	3.09%	0.62%	0.68

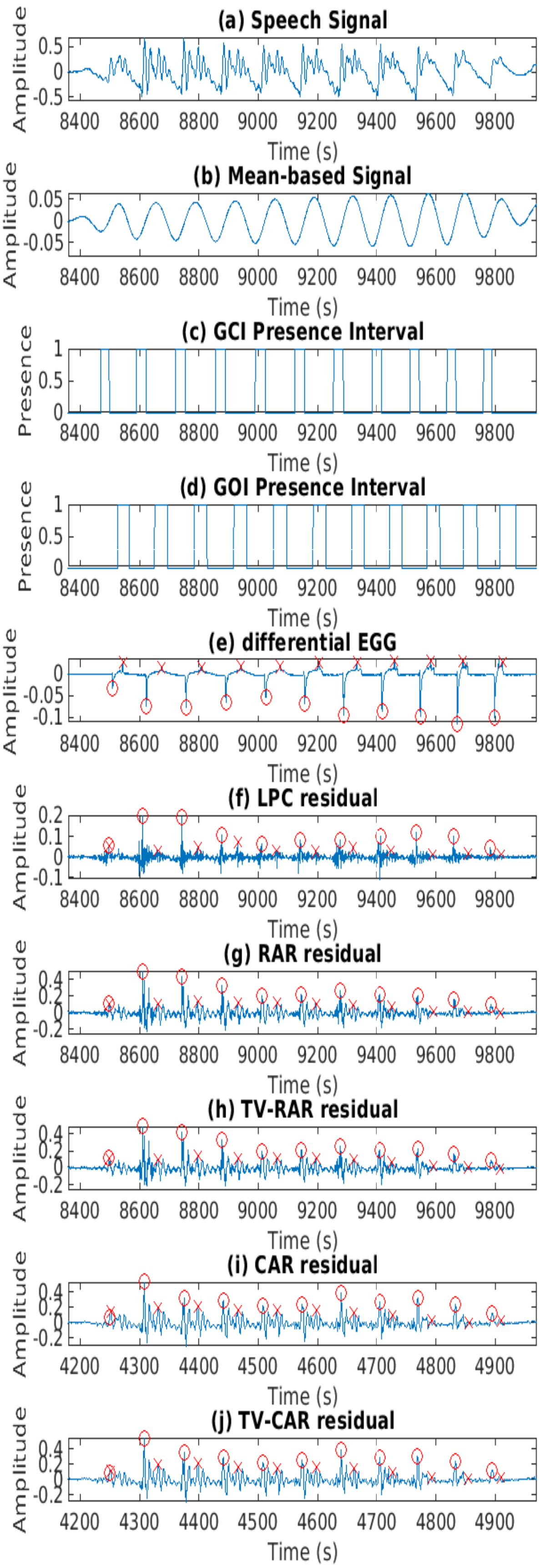

Fig.2: GCI/GOI estimation

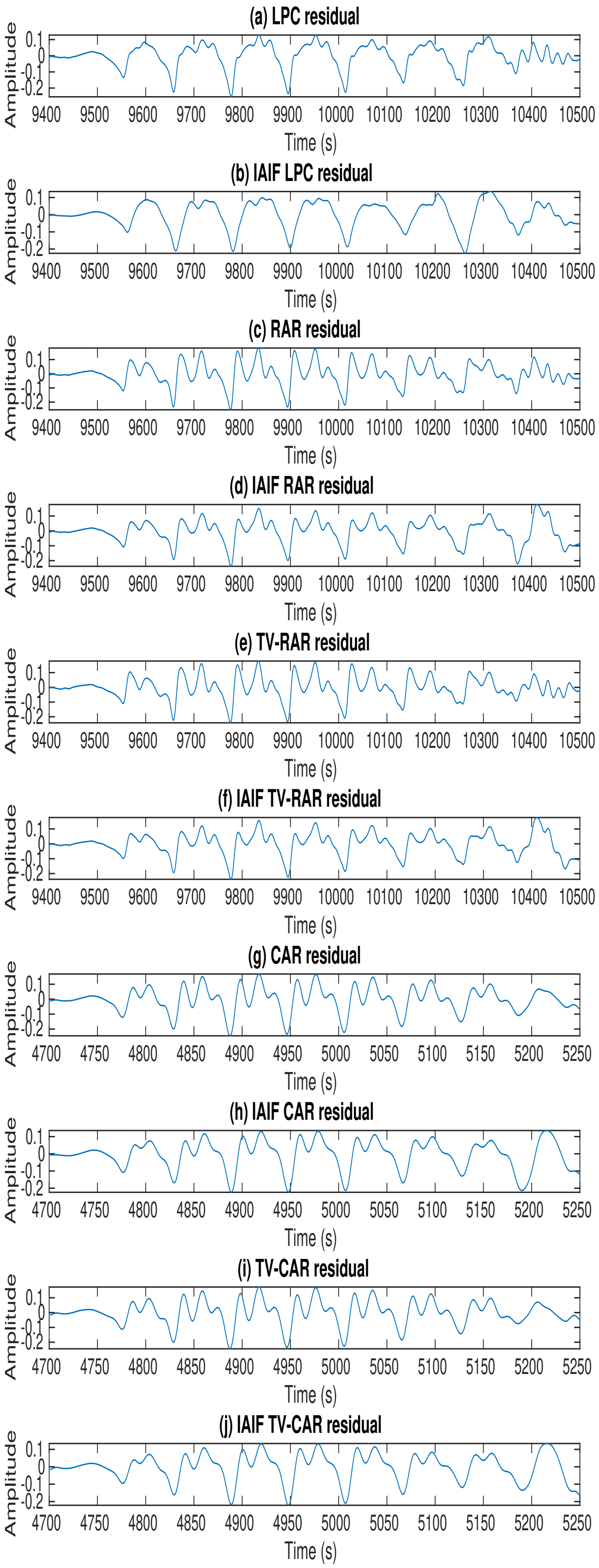

Fig.3: Glottal estimation

VIII. CONCLUSION

We want to develop the novel ARX-LF method that simultaneously estimates both glottal excitation and vocal tract model parameters from speech signals. We have shown the history, and the ARX-LF requires accurate estimating GCI and residual as differential glottal excitation using GCI. The best method to estimate GCI is the SEDREAMS method based on LPC residual peak picking. This paper proposed the methods using TV-CAR analysis instead of LPC analysis in the SEDREAMS method for estimating GCI and GOI and IAIF for estimation of glottal excitation. A comparative study of LPC residual and four types of AR residual was conducted using the CMU Arctic database with EGG signals. The objective evaluation using IDR, MR, FAR, and IDA showed that TV-CAR was better at estimating GCIs than LPC. Moreover, the TV-CAR-based IAIF performs better than LPC. It will be proven by using LF model estimation of [6].

REFERENCES

[1] H. Kawahara, K. Tochinai and K. Nagata, "On the Linear Predictive Analysis using a Small Analysis Segment and its Error Evaluation," The Journal of the Acoustical Society of Japan Vel.33 No. 9, 1977.(in Japanese)

[2] Y. Miyoshi, K. Yamato, R. Mizoguchi, M. Yanagida and O. Kakusho, "Analysis of speech signals of short pitch period by a sample-selective linear prediction," IEEE Trans. on ASSP., Vol.35, No.9, 1987.

[3] D. Wong, J. Markel and A. Gray, "Least squares glottal inverse filtering from the acoustic speech waveform," IEEE Trans. Acoust. Speech Signal Process., vol. 27, no. 4, pp. 350-355, Aug. 1979.

[4] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," Speech Commun., vol. 11, no. 2/3, pp. 109-118, 1992.

[5] G. Fant, J. Liljencrants and Q. G. Lin, "A four-parameter model of glottal flow," STL-QPSR, vol. 4, no. 1985, pp. 1-13, Apr. 1985.

[6] J. Kane and C. Gobl, "Automating manual user strategies for precise voice source analysis," Speech Commun., vol. 55, no. 3, pp. 397-414, 2013.

[7] J. Kane and C. Gobl, "Evaluation of automatic glottal source analysis," in LNAI, vol. 7911, pp. 1-8, 2013.

[8] P. Hedelin, "A glottal LPC-vocoder," IEEE proc.ICASSP84, Mar.1984.

[9] H. Fujisaki and M. Ljungqvist, "Proposal and evaluation of models for glottal source waveform," Proc. IEEE Int. Conf. Acoust. Speech Signal Process., pp. 1605-1608, 1986.

[10] K. Funaki and Y. Mitome, "A speech analysis method based on a glottal source model.," Proc.ICSLP90, Kobe, 1990.

[11] W. Ding, H. Kasuya and S. Adachi, "Simultaneous estimation of vocal tract and voice source parameters based on an ARX model," IEICE Trans. Inf. Syst., vol. 78, no. 6, pp. 738-743, Jun. 1995.

[12] K. Funaki, Y. Miyanaga and K. Tochinai, "Recursive armax speech analysis based on a glottal source model with phase compensation," Signal Process., vol. 74, no. 3, pp. 279-295, 1999.

[13] K. Funaki, Y. Miyanaga and K. Tochinai, "On subband analysis based on glottal-ARMAX speech model," Speech synthesis workshop, 1998.

[14] T. Kaburagi and K. Kawai, "Analysis of voice source characteristics using a constrained polynomial model," Proc.EUROSPEECH-2003, Genove, Swissland, Sep. 2003.

[15] Fu Qiang and P. Murphy, "Robust glottal source estimation based on joint source-filter model optimization," IEEE Trans. Audio Speech Lang. Process., vol. 14, no. 2, pp. 492-501, Mar. 2006.

[16] Y. Li , J. Tao, D. Erickson, B. Liu, and M. Akagi, "F0-noise-robust glottal source and vocal tract analysis based on ARX-LF Model," IEEE/ACM Trans. on ASLP, Vol.29, 2021.

[17] T. Drugman and T. Dutoit, "Glottal closure and opening instant detection from speech signals," in Proc. Interspeech-2009, 2009.

[18] K. Funaki, Y. Miyanaga and K. Tochinai, "On a Time-varying Complex Speech Analysis," Proc. EUSIPCO-98, Rhodes, Greece, Sep.,1998.

[19] K. Funaki, "A time-varying complex AR speech analysis based on GLS and ELS method," Proc. Eurospeech2001, Aalborg, Denmark, Sep. 2001.

[20] K. Funaki, "Sparse Time-Varying Complex AR(TV-CAR) speech analysis based on Adaptive LASSO," IEICE, Trans. on Fundamentals, Vol.E102-A, No.12, Dec.2019.

[21] K. Funaki, "TV-CAR speech analysis based on Regularized LP," Proc. EUSIPCO-2019, Spain, Sep. 2019.

[22] K. Funaki, "TV-CAR speech analysis based on the l2-norm regularization in the time-domain and frequency domain," Proc. APSIPA-2020, Dec. 2020(remote)

[23] T. Kinjo and K. Funaki, "Robust F_0 Estimation Based on Complex LPC Analysis for IRS Filtered Noisy Speech," IEICE Trans., E90-A, No.8, 2007.

[24] K. Funaki and T. Kinjo, "Robust F_0 Estimation Using ELS-Based Robust Complex Speech Analysis," IEICE Trans. Vol. E91-A, 2008.

[25] K. Funaki and T. Higa, "F_0 Estimation using SRH based on TV-CAR Speech Analysis," Proc.EUSIPCO-2012. Bucharest, Romania, Aug., 2012.

[26] K. Higa and K. Funaki, "Robust ASR Based on ETSI Advanced Front-End Using Complex Speech Analysis," IEICE Trans. Vol.E98-A, No.11, 2015.

[27] [Online], CMU ARCTIC speech synthesis databases, http://festvox.org/cmu arctic/

[28] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm," IEEE Trans. Audio Speech Lang. Processing, vol.15, pp.34-43, 2007.