TV-CAR speech analysis based on Regularized LP

Keiichi Funaki
Computing & Networking Center, University of the Ryukyus
Nishihara. Okinawa, 903-0213, Japan
funaki@cc.u-ryukyu.ac.jp

***Abstract*—Linear Prediction (LP) analysis is speech analysis to estimate AR (Auto-Regressive) coefficients to represent the all-pole spectrum that is applied in speech synthesis recently besides speech coding. We have proposed l_2-norm optimization-based TV-CAR (Time-Varying Complex AR) speech analysis for an analytic signal, MMSE (Minimizing Mean Square Error) or ELS (Extended Least Square) method, and we have applied them into the speech processing such as robust ASR or F_0 estimation of speech. On the other hand, B.Kleijn et al. have proposed Regularized Linear Prediction (RLP) method to suppress pitch related bias that is an overestimation of the first formant. In the RLP, l_2-norm regularized term that is the norm of spectral changes in the frequencies is introduced to suppress the rapid spectral changes. The RLP estimates the parameter so as to minimize l_2-norm criterion added by the l_2-norm regularized penalty term. In this paper, the RLP-based TV-CAR speech analysis is proposed and evaluated with the F_0 estimation of speech using IRAPT (Instantaneous RAPT) with Keele Pitch Database under noisy conditions.**

***Index Terms*—Time-Varying Complex AR (TV-CAR) analysis, Analytic signal, l_2-norm regularization, F_0 estimation of speech**

I. INTRODUCTION

Linear Prediction (LP) analysis [1] proposed in the 1960s is commonly used in speech processing, especially in speech coding implemented on a smartphone, Skype, LINE, or so on. LP analysis assumes speech production model as an AR (Auto-Regressive) model and it can estimate the AR spectrum using 10 to 20 number of AR coefficients. LP analysis cannot estimate the AR spectrum efficiently by solving the 10 to 20 order linear equation using the auto-correlation or covariance function based on l_2-norm optimization, but also the coefficients can be effectively quantized with the LSP (Line Spectrum Pair) domain by using Vector Quantization (VQ). The LP is used in CELP (Code Excited Linear Prediction), ACELP [2] or RCELP [3] in which the long-term prediction error for the LP residual is quantized by using an innovation, multi-pulse, or VQ codebook. The LP is also used in MPEG-4 ALS [4] in which the LP residual is quantized by an entropy coding to implement the lossless coding. Moreover, the LP analysis is useful for Fundamental Frequency (F_0) estimation. The LP residual provides less formant structure than speech signal does since the formant structure is removed from the speech signal by an inverse LP filtering. For this reason, the performance of the F_0 estimation can be significantly improved by applying the LP residual instead of speech [5]. Moreover, the LP residual is useful also in speech enhancement. For example, the Iterative Wiener Filter (IWF) [6] introduces the LP spectrum to design the Wiener filter. In the ETSI (European Telecommunications Standards Institute) Advanced Front-End (AFE) [7] standardized as a front-end of ASR, FFT spectrum is applied to design the filter instead of the LP spectrum, however, it has already been reported that the LP spectrum can improve the performance [8]. In speech synthesis, although speech waveform concatenation method or HMM speech synthesis have been studied, recently revolution was occurred by appearing WaveNet [9] that can improve the speech quality drastically. The WaveNet is RNN based speech synthesis inspired from the concept of the LP analysis. The WaveNet provides a drawback that introduces white Gaussian noise. P.Alku has proposed GlotNet [10] in which glottal excitation is generated by the WaveNet with estimated glottal excitation and speech is synthesized by the LP synthetic filter. Since the LP synthetic filter offers a masking effect, the GlotNet can suppress the white Gauss noise, resulting in making the speech quality improved. As mentioned above, LP analysis is important methodology even in these days. For more than half a century, the extended versions of LP analysis has been proposed to cope with the drawbacks, for example, the ARMA method [11], the time-varying analysis [12] [13], the complex analysis [14] [15], short-term LP analysis for glottal closure interval [16] [17], and the simultaneous estimation of ARMA and glottal excitation model parameters [18] [19] [20] [21] [22].

We have been studying Time-Varying Complex AR (TV-CAR) speech analysis for an analytic signal, MMSE [23] and robust methods [24] [25] [26], and we have shown that the TV-CAR analysis can improve the performance on robust ASR and F_0 estimation of speech [5] [8] [27] [28] [29] [30] [31]. These TV-CAR analysis methods are based on l_2-norm optimization, however, sparse estimation is focused mainly in image processing [32] [33] in these days. The sparse estimation is realized by l_0-norm, l_1-norm optimization. l_1-norm based sparse LP analysis methods have already been proposed [34] [35] [36]. We have also proposed sparse TV-CAR analysis based on an adaptive LASSO (Least absolute shrinkage and selection operator) [37] [38] in which the LASSO is realized by IRLS (Iterative Reweighted Least Square) [32]. It can be considered that the sparse analysis cannot perform well in spite of its large amount of computation since an AR analysis can be regarded as a sparse estimation.

On the other hand, B.Kleijn et al. have proposed l_2-norm regularized LP (RLP) analysis to avoid pitch related bias that is an overestimation of first formant (F_1) in the case of high F_0 [39]. In the RLP, l_2-norm for the spectral changes in the

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

frequencies is introduced as the l_2-norm regularization term and AR coefficients can be estimated by solving a linear equation with no iteration. Furthermore, P.Alku et al. have proposed time regularized LP (TRLP) analysis whose l_2-norm regularization term is the l_2-norm of the difference between current and previous frame parameters and have shown that the TRLP performs better than the RLP in terms of spectral distance and phoneme distinguish performance [40].

In this paper, we propose l_2-norm regularized RLP-based TV-CAR speech analysis and evaluate the performance on F_0 estimation using IRAPT (Instantaneous RAPT) [41] that is the improved version of well-known and commonly used RAPT (Robust Algorithm for Pitch Tracking) [42].

II. Regularized LP

A. LP Analysis

LP analysis is the l_2-norm optimization method estimating an i-th AR coefficient $a_i (i = 1, 2, 3, ..., I)$ so as to minimize a mean squared error (MSE) for an AR (Auto-Regressive) model shown in Eq.(1).

$$\frac{1}{A(z^{-1})} = \frac{1}{1 + \sum_{i=1}^{I} a_i z^{-i}} \tag{1}$$

The power spectrum of the AR model is represented by Eq.(2).

$$S(\omega, \mathbf{a}) = \frac{1}{|A(e^{j\omega})|^2} \tag{2}$$

In the LP analysis, the l_2-norm criterion is shown in Eq.(3).

$$\mathcal{D} = E\left[e^2(t)\right] = \mathbf{a}^T \mathbf{R} \mathbf{a} + 2\mathbf{a}^T \mathbf{r} + r_0 \tag{3}$$

where $e(t)$ is the residual signal at time t, $\mathbf{R}$ is the symmetric Toeplitz matrix whose elements are the auto-correlation function $r_i (i = 0, 1, ..., I-1)$, $\mathbf{a}$ is $[a_1, a_2, ..., a_I]^T$, $\mathbf{r}$ is $[r_1, r_2 ..., r_I]^T$ and T means Transpose. Minimizing Eq.(3) viz., $d\mathcal{D}/d\mathbf{a}^T = 0$ results in the following linear equation.

$$\mathbf{R}\hat{\mathbf{a}} = -\mathbf{r} \tag{4}$$

This Yule-Walker equation can be solved efficiently by using Levinson recursion. Eq.(4) is also named as the auto-correlation LP analysis commonly used in mobile phone, smartphone, or Skype nowadays.

B. Regularized LP (RLP) analysis [39]

It is well-known that LP analysis suffers from pitch related bias that is to estimate the unnaturally sharp peak of the F_1 for high pitch speech. In order to cope with the pitch related bias, the RLP analysis introduces an l_2-norm regularization term shown in Eq.(5) that means l_2-norm of the AR spectral changes in the frequency domain.

$$\mathcal{R}(S(\omega, \mathbf{a})) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{d}{d\omega} \log S(\omega, \mathbf{a}) \right]^2 d\omega \tag{5}$$

The criterion of the RLP is $\mathcal{D} + \lambda \mathcal{R}$. λ is called the Regularized coefficient that controls the contribution for the regularized term. In order to estimate the parameter, $\mathbf{a}$, with no iteration, Eq.(5) is approximated to be Eq.(6).

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{d}{d\omega} \log\left(A\left(e^{j\omega}\right)\right) \right|^2 d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{A'\left(e^{j\omega}\right)}{A\left(e^{j\omega}\right)} \right|^2 d\omega \tag{6}$$

By using Eq.(6), Eq.(5) turns to be Eq.(7)

$$\widehat{\mathcal{R}}(S(\omega, \mathbf{a})) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{A'\left(e^{j\omega}\right)}{W(\omega)} \right|^2 d\omega \tag{7}$$

where $|W(\omega)|^2$ is a rough estimation of $|A(\omega)|^2$.

$$A'\left(e^{j\omega}\right) = -\sum_{k=0}^{M} jk a_k e^{jk\omega} \tag{8}$$

Thus, Eq.(7) turns to be Eq.(9).

$$\sum_{k=0}^{I} \sum_{m=0}^{I} k a_k m a_m \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{-j\omega(k-m)}}{|W(\omega)|^2} d\omega \tag{9}$$

Since the integral in Eq.(9) is an inverse discrete transform of $|1/W(\omega)|^2$, Eq.(7) turns to be Eq.(10).

$$\widehat{\mathcal{R}}(S(\omega, \mathbf{a})) = \sum_{k=0}^{I} \sum_{m=0}^{I} k a_k m a_m h(m-k) \tag{10}$$

where

$$h(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{j\omega x}}{|W(\omega)|^2} d\omega \tag{11}$$

that is the inverse Fourier transform of the power spectrum so that it is the auto-correlation function. As a result, Eq.(10) turns to be Eq.(12).

$$\widehat{\mathcal{R}}(S(\omega, \mathbf{a})) = \mathbf{a}^T \mathbf{D}^T \mathbf{F} \mathbf{D} \mathbf{a} \tag{12}$$

where $\mathbf{D}$ is a diagonal matrix whose element is $d(m, m) = m$, $\mathbf{F}$ is Toeplitz auto-covariance matrix. From Eq.(3) and Eq.(12), the criterion of RLP, $\mathcal{D} + \lambda \mathcal{R}$ is as follows.

$$\mathbf{a}^T (\mathbf{R} + \lambda \mathbf{D}^T \mathbf{F} \mathbf{D}) \mathbf{a} + 2\mathbf{a}^T \mathbf{r} + r_0 \tag{13}$$

Minimizing Eq.(13), $d(\mathcal{D} + \lambda \mathcal{R})/d\mathbf{a}^T = 0$ results in the following linear equation.

$$(\mathbf{R} + \lambda \mathbf{D}^T \mathbf{F} \mathbf{D}) \hat{\mathbf{a}} = -\mathbf{r} \tag{14}$$

The RLP analysis can be realized by solving Eq.(14). Note that if λ is 0, the RLP analysis is the same as the LP analysis.

III. Regularized TV-CAR analysis

A. TV-CAR model

The TV-CAR model can be defined by Eq.(15).

$$\begin{aligned} Y_{TVCAR}(z^{-1}) &= \frac{1}{A(z^{-1})} = \frac{1}{1 + \sum_{i=1}^{I} a_i^c(t) z^{-i}} \\ &= \frac{1}{1 + \sum_{i=1}^{I} \sum_{l=0}^{L-1} g_{i,l}^c f_l^c(t) z^{-i}} \end{aligned} \tag{15}$$

where $a_i^c(t)$, L, $g_{i,l}^c$ and $f_l^c(t)$ is i-th complex AR coefficient at time t, an order of complex basis expansion, complex parameter and complex basis function, respectively. The input-output relationship for Eq.(15) is shown as in Eq.(16).

$$\begin{aligned} y^c(t)= & -\sum_{i=1}^{I} a_i^c(t) y^c(t-i) + u^c(t) \\ = & -\sum_{i=1}^{I}\sum_{l=0}^{L-1} g_{i,l}^c f_l^c(t) y^c(t-i) + u^c(t) \end{aligned} \tag{16}$$

where $y^c(t)$ is the target analytic signal at time t and $u^c(t)$ is a complex input signal at time t. Analytic signal is complex-valued signal whose real part is speech signal and the imaginary part is the Hilbert transformed signal of the real one. Since the analytic signal yields the spectrum only over positive frequencies, the signal can be decimated by a factor of two, consequently, the complex analysis can estimate more accurate spectrum in low frequencies. Moreover, the TV-CAR analysis is a time-varying analysis that introduces complex basis expansion of the AR parameter to represent the parameter as a function of time.

Eq.(16) can be formulated by the following vector-matrix representation.

$$\begin{aligned} \mathbf{y}_f =& -\mathbf{\Phi}_f \theta + \mathbf{u}_f \\ \theta^T =& [\mathbf{g}_0^T, \mathbf{g}_1^T, \cdots, \mathbf{g}_l^T, \cdots, \mathbf{g}_{L-1}^T] \\ \mathbf{g}_l^T =& [g_{1,l}^c, g_{2,l}^c, \cdots, g_{i,l}^c, \cdots, g_{I,l}^c] \\ \mathbf{y}_f^T =& [y^c(I), y^c(I+1), y^c(I+2), \cdots, y^c(N-1)] \\ \mathbf{u}_f^T =& [u^c(I), u^c(I+1), u^c(I+2), \cdots, u^c(N-1)] \\ \mathbf{\Phi}_f =& [\mathbf{S}_0^f, \mathbf{S}_1^f, \cdots, \mathbf{S}_l^f, \cdots, \mathbf{S}_{L-1}^f] \\ \mathbf{S}_l^f =& [\mathbf{s}_{1,l}^f, \mathbf{s}_{2,l}^f, \cdots, \mathbf{s}_{i,l}^f, \cdots, \mathbf{s}_{I,l}^f] \\ \mathbf{s}_{i,l}^f =& [y^c(I-i) f_l^c(I), y^c(I+1-i) f_l^c(I+1), \\ & \cdots, y^c(N-1-i) f_l^c(N-1)]^T \end{aligned} \tag{17}$$

where N is analysis length, $\mathbf{y}_f$ is $(N-I, 1)$ column vector whose element is the analytic signal, θ is $(L \cdot I, 1)$ column vector whose element is the complex parameter, $\mathbf{\Phi}_f$ is $(N-I, L \cdot I)$ matrix whose element is the weighted analytic signal by the complex basis.

B. Proposed RLP-based TV-CAR analysis

Since the TV-CAR analysis is the complex, time-varying and covariance type of LP analysis, Eq.(14) turns to be Eq.(18) with integrating the RLP onto the TV-CAR analysis. As the l_2-norm regularized term, the power spectrum at the center sample of the frame, $N/2$, is applied.

$$(\mathbf{\Phi}_f^H \mathbf{\Phi}_f + \lambda \mathbf{D}_{tv}^H \mathbf{F} \mathbf{D}_{tv}) \hat{\theta} = -\mathbf{\Phi}_f^H \mathbf{y}_f \tag{18}$$

where H is an Hermite operator and $\mathbf{D}_{tv}$ is as follows.

$$\mathbf{D}_{tv} = [\mathbf{d_0}, \mathbf{d_1}, ..., \mathbf{d_l}, ..., \mathbf{d_{L-1}}] \tag{19}$$

$$\mathbf{d_l} = \begin{pmatrix} f_l^c(N/2) & 0 & \cdots & 0 \\ 0 & 2f_l^c(N/2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & If_l^c(N/2) \end{pmatrix} \tag{20}$$

$\mathbf{d_l}$ is (I, I) matrix and $\mathbf{D_{tv}}$ is $(I, L \cdot I)$ matrix that is generated by aligning L number of $\mathbf{d_l}(l = 0, 1, ..., L-1)$.

IV. F_0 ESTIMATION

In this paper, IRAPT [41] is used to implement the F_0 estimation. Fig.1 depicts the flow of the IRAPT algorithm. In the IRAPT, an instantaneous frequency is used to estimate F_0 instead of NCCF (Normalized Cross Correlation Function) in RAPT. The instantaneous frequency is estimated by using the analytic signal. For this reason, the IRAPT can be used for complex residual signal estimated by complex analysis. Moreover, a time-warping is operated to the input speech signal by using the estimated F_0, and then F_0 is estimated again by using the time-warped signal. The experiments with real-valued speech signal exhibit that the IRAPT leads to improved performance in comparison to the RAPT [41]. In this paper, the method using the time-warped signal is called IRAPT2.

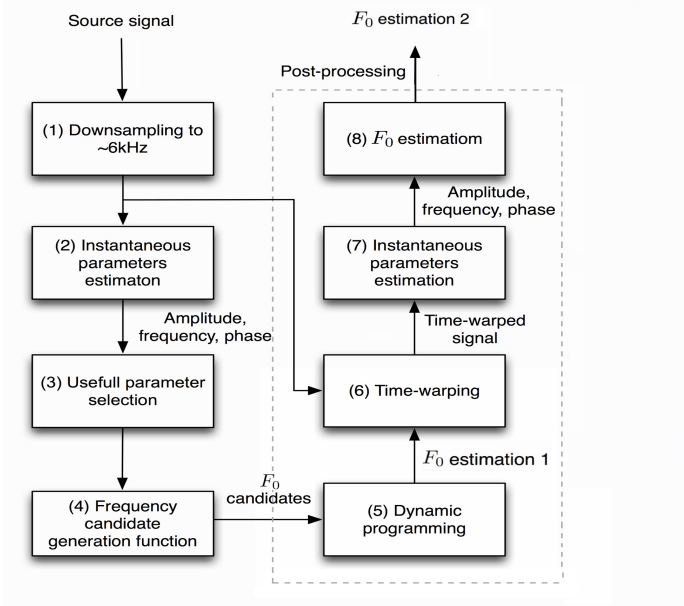


Fig.1: Blockdiagram of IPART

As mentioned above, the residual signal contains much less formant components since the residual signals are computed by the inverse AR filtering. Accordingly, the residual is much more suitable for F_0 estimation than a speech signal. Furthermore, the F_1 is more removed in the complex residual than real residual since a complex speech analysis for an analytic signal can estimate more accurate speech spectrum in low frequencies due to the nature of the analytic signal. Thus, we can take into account that the complex residual is more appropriate. Time-varying analysis can estimate the parameter in any sample, thus, it can remove the formant frequencies

from speech. The half value of the sampling rate of an input speech signal is set the sampling frequency to estimate F_0.

V. Experiments

The performance of the proposed TV-CAR analysis is evaluated by using the F_0 estimation with the IRAPT2. Speech signal, residual signal estimated by the LP analysis, complex residual signal estimated by the RLP-based TV-CAR analysis, complex residual signal estimated by the MMSE-based TV-CAR analysis are compared as the input of the IRAPT2 [41]. The experimental conditions are shown in Table 1. The IRS (Intermediate Reference System) [43] filtered noise-corrupted speech signals are used in the experiments for speech coding application. The corrupted speech is generated by adding white Gaussian noise or Pink noise [44] to speech in Keele Pitch database [45]. F_0 estimation performance is evaluated by means of GPE (Gross Pitch Error) and FPE (Fine Pitch Error). If the estimation error is less than p-percent of the true F_0, the estimation is regarded as SUCCEED. Otherwise, the estimation is regarded as FAILURE. The GPE is percentage of FAILURE frames and the FPE is a variance of the estimation error at the SUCCEED frames.

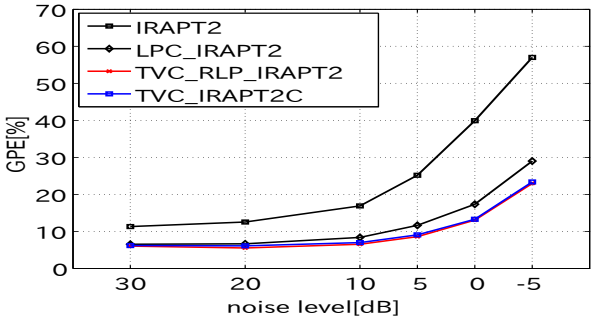
Table 1: Experimental Conditions

Speech data	Keele Pitch Database [45] 5 long Male sentence 5 long Female sentence
Sampling	10kHz/16bit
Analysis window	Window Length: 25.6[ms] Shift Length: 10.0[ms]
TV-CAR	$I = 7, L = 2$ (Time-Varying)
Basis	$f_l^c(t) = t^l/l!$
Pre-emphasis	$1 - z^{-1}$
RLP	$\lambda = 0.0001$
Noise	White Gauss or Pink noise [43]
Noise Level	30,20,10,5,0,-5[dB]

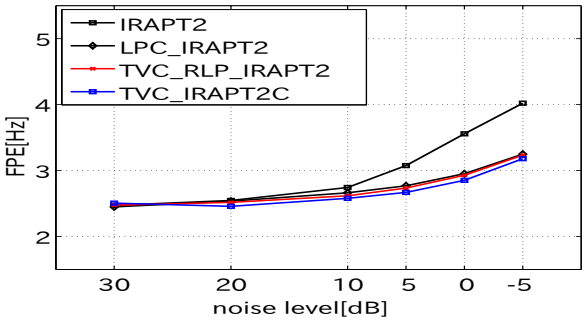
Fig.2 shows 10[%] GPE and FPE for additive white Gauss noise while Fig.3 shows 10[%] GPE and FPE for additive pink noise. X-axis means noise level$[dB]$ and Y-axis means GPE$[\%]$ or FPE$[Hz]$. In figures 2 and 3, four lines indicate as follows.

(a)IRAPT2 (■black line) is the results of the IRAPT2 for speech [41]
(b)LPC_IRAPT2 (◆black line) is the results of the IRAPT2 for the LP residual.
(c)TVC_RLP_IRAPT2 (red line) is the results of the IRAPT2 for complex residual computed by the proposed RLP-based TV-CAR analysis.
(d)TVC_IRAPT2C (blue line) is the results of the IRAPT2 for complex residual computed by the MMSE-based TV-CAR analysis.

Needless to say, (a) and (b) are the conventional methods. Figures 2 and 3 demonstrate that the proposed RLP-based TV-CAR method performs slightly better than the MMSE-based method and performs better than the conventional methods, the original IRAPT2 and LP analysis in terms of GPE that means large estimation error resulting in making the fatal performance down. The value of the GPEs are large since the IRS filtered noise corrupted telephone speech data are used.

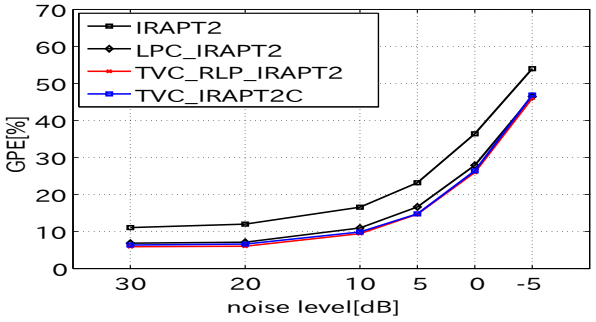


(1)GPE (10%)

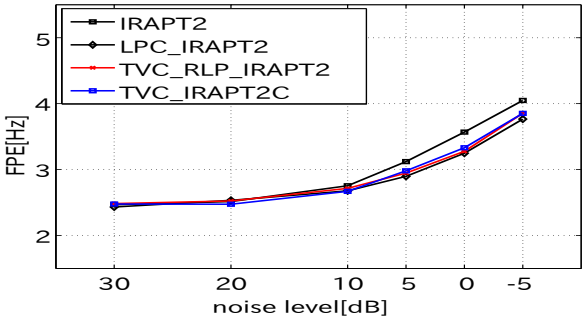


(2)FPE (10%)

Fig.2: F_0 estimation performance (Gauss)



(1)GPE (10%)



(2)FPE (10%)

Fig.3: F_0 estimation performance (Pink)

VI. Conclusions

In this paper, we have explained the l_2-norm regularized LP analysis (RLP) that penalizes the rate of spectral changes

in the frequency domain, and then we have proposed l_2-norm regularized TV-CAR speech analysis integrated by the RLP framework and have evaluated by using F_0 estimation with the IRAPT2 under noisy conditions. The experimental results demonstrate that the integration of the RLP leads to improved performance compared to the conventional MMSE-based TV-CAR analysis, the LP and the original IRAPT. The reason why the proposed RLP-based method improves the performance slightly than the MMSE method is as follows. The complex analysis is able to suppress the pitch related bias since the bandwidth turns to be twice due to the decimated analytic signal by a factor of two, thus the complex analysis can separate F_0 from F_1. Pre-emphasis is operated to remain the F_0 in the residual, consequently, the complex analysis can separate F_0 from F_1 more. Furthermore, the time-varying analysis suppresses the rapid spectral changes since the basis expansion offers constraint to be slow spectral changes within the frame. In this paper, the regularized coefficient λ is fixed. As a future work, the coefficients will be flexed. In addition, we are going to propose the TRLP-based TV-CAR analysis.

REFERENCES

[1] J.Makhoul, "Linear prediction: A tutorial review," Proc. IEEE, vol. 63, no. 4, pp. 561-580, Apr. 1975.

[2] ITU-T G.729: "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)," Mar., 1996.

[3] "Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB) Service Option 62 for Spread Spectrum Systems," 3GPP2. C.S0052-0 Version 1.0, pp.73-85, 3GPP2, June, 2004.

[4] T.Liebchen, T.Moriya, N.Harada, Y.Kamamoto, and Y.A.Reznik, "The MPEG-4 Audio Lossless Coding (ALS) Standard - Technology and Applications," Audio Engineering Society, Convention Paper 119th Convention, Oct., New York, NY, USA, 2005. http://elvera.nue.tu-berlin.de/files/0737Liebchen2005.pdf

[5] T.Kinjo and K.Funaki, "Robust F0 Estimation Based on Complex LPC Analysis for IRS Filtered Noisy Speech," IEICE Trans., E90-A, No.8, 2007.

[6] J.S.Lim and A.Oppenheim, "All-pole Modeling of Degraded Speech," IEEE Tran. ASSP, 1978.

[7] ETSI Advanced Front-End, ES 202 050 v1.1.5(2007-01), Jan.2007.

[8] 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.

[9] A.van den Oord, S.Dieleman, H.Zen, K.Simonyan, O.Vinyals, A.Graves, N.Kalchbrenner, A.Senior, K.Kavukcuoglu, "WaveNet: A generative model for raw audio," arXiv:1609.03499, 2016.

[10] L.Juvela, V.Tsiaras, B.Bollepalli, M.Airaksinen, J.Yamagishi, P.Alku, "Speaker-independent raw waveform model for glottal excitation," Proc. Interspeech-2018, 2018.

[11] Y.Miyanaga, N.Miki, N.Nagai, "Adaptive identification of a time-varying ARMA speech model," IEEE Trans. ASSP-34, 423-433, 1986.

[12] M.G.Hall, A.V.Oppenheim, A.S.Willsky, "Time-varying parametric modeling of speech," Signal Processing, Vol.5, Dec. 1977.

[13] Y.Grenier, "Time-dependent ARMA modeling of nonstationary signals," IEEE Trans. on ASSP., vol.31, no.4, 1983.

[14] S.Kay, "Maximum entropy spectral estimation using the analytic signal," IEEE Trans. ASSP-26, 1980.

[15] T.Shimamura and S.Takahashi, "Complex linear prediction method based on positive frequency domain," Trans. IEICE A-72, 1989. (in Japanese)

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

[17] M. Airaksinen, T. Raitio, B. Story, and P. Alku, "Quasi closed phase glottal inverse filtering analysis with weighted linear prediction," IEEE/ACM TASL., vol.22, no.3, Mar. 2014.

[18] H.Fujisaki and M.Ljungqvist, "Estimation of Voice Source and Vocal Tract Parameters Based on ARMA Analysis and a Model for the Glottal Source Waveform," Proc. ICASSP-87, 1987.

[19] T.Ohtsuka and H.Kasuya, "Robust ARX-based speech analysis method taking voicing source pulse train into account," The Journal of the Acoustical Society of Japan, Vol.58, 2002.(in Japanese)

[20] K.Funaki, Y.Miyanaga, K.Tochinai, "Recursive ARMAX speech analysis based on a glottal source model with phase compensation," Signal Processing, Vol. 74, 1999.

[21] K.Funaki, Y.Miyanaga, K.Tochinai, "On Subband analysis based on Glottal-ARMAX speech model," ESCA 3RD INTERNATIONAL WORKSHOP ON SPEECH SYNTHESIS, Bluemountain, Australia, Nov., 1998.

[22] Y.Li, K.Sakakibara and M.Akagi, "Estimation of glottal source waveforms and vocal tract shapes from speech signals based on ARX-LF model," Proc. ISCSLP-2018, Taipei, Nov., 2018.

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

[24] K.Funaki, Y.Miyanaga, K.Tochinai, "On Robust speech analysis based on time-varying complex AR model,"ICSLP-98, Sydney, Australia, Dec., 1998.

[25] K.Funaki, "A time-varying complex speech analysis based on IV method," Proc. ICSLP-2000, Beijing, China, Oct.,2000.

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

[27] K.Funaki,"F0 estimation based on robust ELS complex speech analysis," EUSIPCO-2008, Lausanne, Switzerland, Aug.2008.

[28] K.Funaki,"Speech Enhancement based on Iterative Wiener Filter using Complex Speech Analysis," EUSIPCO-2008, Lausanne, Switzerland, Aug.2008.

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

[30] K.Hotta and K.Funaki "On a robust F0 estimation of speech based on IRAPT using robust TV-CAR analysis," Proc. APSIPA-2014, Dec. 2014.

[31] K.Higa and K.Funaki, "Improved ETSI advanced front-end for ASR based on robust complex speech analysis," Proc. APSIPA-2016, Jeju, Korea, Dec. 2016.

[32] M.Elad,"Sparse and Redundant Representations From Theory to Applications in Signal and Image Processing," Springer; 2010

[33] Y.Sugiura and T.Shimamura, "Speech Enhancement Based on Sparse Representation in Logarithmic Frequency Scale," Proc. ISPACS-2018, Nov. 2018.

[34] E. Denoel and J-P.Solvay, "Linear Prediction of Speech with a Least Absolute Error Criterion," IEEE Trans. ASSP., Vol.33, No.6, 1985.

[35] T.Jensen, D.Giacobello, M.G.Christensen, S.H.Jensen, M.Moonen, "Real-Time Implementations of Sparse Linear Prediction for Speech Processing," Proc. ICASSP-2013, 2013.

[36] D.Giacobello, M.G.Christensen, M.N.Murthi, S.H.Jensen, M.Moonen, "Sparse Linear Prediction and its Applications to Speech Processing," IEEE Trans. ASLP., Vol.20, No.5, 2012.

[37] R.Tibshirani, "Regression shrinkage and selection via the lasso," J. Royal. Statist. Soc B., Vol. 58, No. 1, pages 267-288, 1996.

[38] H.Zou, "The Adaptive Lasso and Its Oracle Properties," Journal of the American Statistical Association, Vol.101, 2006.

[39] L. A. Ekman, W. B. Kleijn, and M. N. Murthi,"Regularized linear prediction of speech," IEEE Trans. ASLP., Vol.16, No.1, 2008.

[40] M.Airaksinen, L.Juvela, O.J.Rsnen, P.Alku, "Time-regularized Linear Prediction for Noise-robust Extraction of the Spectral Envelope of Speech," Proc. Interspeech-2018, India, 2018.

[41] E.Azarov, M.Vashkevich, A.Petrovsky, "Instantaneous pitch estimation based on RAPT framework," Proc. EUSIPCO-2012, Bucharest, Romania, Aug., 2012.

[42] D.Talkin, "A Robust Algorithm for Pitch Tracking (RAPT)," in Speech Coding and Synthesis, W.B.Kleijn and K. K.Palatal (eds), pp.497-518, Elsevier Science B.V., 1995.

[43] ITU-T Recommendation G.191, Software tools for speech and audio coding standardization, Nov. 2000.

[44] NOISE-X92, http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html

[45] F.Plante, G.F.Meyer, W.A.Ainsworth, "A Pitch Extraction Reference Database," Proc.EUROSPEECH-95, 1995.