F_0 ESTIMATION BASED ON ROBUST ELS COMPLEX SPEECH ANALYSIS

Keiichi Funaki

Computing & Networking Center, Univ. of the Ryukyus
Senbaru 1, Nishihara, Okinawa, 903-0213, Japan
phone: +(81)98-895-8946, fax: +(81)98-895-8963, email: funaki@cc.u-ryukyu.ac.jp

ABSTRACT

A robust fundamental frequency (F_0) estimation algorithm based on robust ELS (Extended Least Square) complex-valued speech analysis for an analytic speech signal is proposed in this paper. Speech spectrum can be accurately estimated in low frequencies since the analytic signal provides spectrum only over positive frequencies. The remarkable feature makes it possible to realize more accurate F_0 estimation using complex residual extracted by the speech analysis. We have already proposed F_0 estimation using complex LPC residual, in which the autocorrelation function weighted by AMDF was adopted for the criterion. The method adopted an MMSE-based complex LPC analysis and it has been reported that it can estimate more accurate F_0 for IRS filtered speech corrupted by white Gauss noise although it cannot perform better for the IRS filtered speech corrupted by pink noise. In this paper, robust complex speech analysis based on ELS method is introduced in order to solve the problem and evaluated with larger number of speech data. The experimental results for additive white Gauss or pink noise demonstrate that the proposed algorithm based on robust complex residual can perform better than other methods.

1. INTRODUCTION

The F_0 estimation plays an important role in speech processing such as speech coding, tonal speech recognition, speaker recognition, and speech enhancement. For example, in CELP(Code Excited Linear Prediction) speech coding, F_0 is estimated by adaptive codebook search as a fractional delay in open-loop pre-search as well as closed-loop final search [1],[2]. Recently, free IP-based voice communication system called Skype is commonly used by many people to enjoy free long distance call. In Skype, F_0 is also estimated by first stage of open-loop adaptive codebook in the residual domain [3]. Needless to say, the F_0 estimation error causes degradation of speech quality. However, speech processing is commonly applied in realistic noisy environment, hence, the performance is degraded seriously. Accordingly, more robust F_0 estimation algorithm is desired. Many F_0 estimation algorithms have already been proposed [4], for example, autocorrelation method (AUTOC), AMDF (Amplitude Magnitude Difference Function) [5], and cepstrum analysis based methods [6]. In general, AUTOC is robust against additive noise. However, it suffers from error estimation of halving or doubling F_0. In order to cope with the drawback, Prof. T.Shimamura has proposed robust F_0 estimation method against additive noise, whose criterion is the autocorrelation function weighted by a reciprocal of the AMDF(AUTOC/AMDF)[7]. Indeed, it can improve the estimation accuracy, however, the performance is not sufficient. The reason why the F_0 estimation invokes the estimation error is that F_0 is closed to the first formant (F_1) frequency. One of the solutions to alleviate such formant effects is to calculate the criteria such as AUTOC using the LPC residual that is called "modified autocorrelation method". On the other hand, complex LPC speech analysis methods for analytic speech signal have already been proposed [8],[9]. Analytic signal is a complex-valued signal in which its real part is speech signal and its imaginary part is Hilbert transform of the real part. Since the analytic signal provides the spectrum only on positive frequencies, the signals can be decimated by a factor of two with no degradation. As a result, the complex analysis offers attractive features, for example, more accurate spectral estimation in low frequencies. We have already proposed time-varying complex AR (TV-CAR) speech analysis methods [10],[11] in which the AR model parameter is represented by complex basis expansion. It is well known that the complex speech analysis for analytic speech can estimate more accurate spectrum in low frequencies [9]. Consequently, if the complex-valued speech analysis is applied before calculation of the criterion, vocal tract spectrum is removed more effectively from speech signal and then more appropriate criterion can be obtained by using the corresponding complex residual. For this reason, we have already proposed robust F_0 estimation algorithms [12],[13] based on complex LPC residual that is extracted by the MMSE(Minimizing Mean Squared Error)-based analysis[10] or the ELS(Extended Least Square)-based analysis[11], respectively. It has been reported in [12] that the method can estimate more accurate F_0 for IRS (Intermediate Reference System) filtered speech corrupted by white Gauss noise although it cannot estimate more accurate F_0 for the speech corrupted by pink noise. The IRS filter is band pass filter whose frequency response corresponds to that for analog part of the transmitter of telephone equipment. In order to evaluate the F_0 estimation method for the speech data processed by speech coding, the IRS filter was introduced. The reason why the performance is down for additive pink noise is as follows. Pink noise provides spectral tilt and the tilt compensates the HPF (High Pass Filter) characteristics of the IRS filter, thus, F_0 is more accurately estimated by complex LPC speech analysis, as a result the performance is down. In order to solve the problem, robust complex speech analysis based on ELS method [11] was introduced and complex residual signal extracted by the robust analysis was used [13]. The ELS analysis employs a whiten filter that whitens equation error so that it allows for unbiased estimation and is feasible to estimate more accurate and less flat formant peaks from noisy speech whereas the MMSE analysis can only estimate flat speech spectrum[11]. By introducing the ELS analysis, more accurate formant peaks can be estimated, thus, the formant effects can be removed more effectively on the corresponding AR residual. For this reason, it assigns more robust F_0 estimation against the additive noise. It has been reported in [13] that more accurate F_0 can be estimated not only for additive white Gauss noise but also for additive pink noise. However, number of the evaluated speech data was only 6. The threshold of GPE was $10[Hz]$. The number of speech sample is too small and the threshold is not appropriate for high-pitched speech. In this paper, 20 number of sentences are used and 20% of threshold is adopted for appropriate evaluation.

2. TV-CAR SPEECH ANALYSIS

2.1 Analytic speech signal

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

$$y^c(t) = \frac{y(2t) + j \cdot y_H(2t)}{\sqrt{2}} \tag{1}$$

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, respectively. Notice that superscript c denotes complex value in this paper. Since analytic signals provide the spectra only over the range of $(0, \pi)$, analytic signals can be decimated by a factor of two. $2t$ means the decimation. The term of $1/\sqrt{2}$ is multiplied in order to adjust the power of an analytic signal with that of the observed one.

2.2 Time-varying complex AR (TV-CAR) model

Conventional LPC model is defined by

$$Y_{LPC}(z^{-1}) = \frac{1}{1 + \sum_{i=1}^{I} a_i z^{-i}} \tag{2}$$

where a_i and I are i-th order LPC coefficient and LPC order, respectively. Since the conventional LPC model cannot express the time-varying spectrum, LPC analysis cannot extract the time-varying spectral features from speech signal. In order to represent the time-varying features, the TV-CAR model employs a complex basis expansion shown as

$$a_i^c(t) = \sum_{l=0}^{L-1} g_{i,l}^c f_l^c(t) \tag{3}$$

where $a_i^c(t)$,I,L,$g_{i,l}^c$ and $f_l^c(t)$ are taken to be i-th complex AR coefficient at time t, AR order, finite order of complex basis expansion, complex parameter, and a complex-valued basis function, respectively. By substituting Eq.(3) into Eq.(2), one can obtain the following transfer function.

$$Y_{TVCAR}(z^{-1}) = \frac{1}{1 + \sum_{i=1}^{I} \sum_{l=0}^{L-1} g_{i,l}^c f_l^c(t) z^{-i}} \tag{4}$$

The input-output relation is defined as

$$\begin{aligned} y^c(t) &= -\sum_{i=1}^{I} a_i^c(t) y^c(t-i) + u^c(t) \\ &= -\sum_{i=1}^{I} \sum_{l=0}^{L-1} g_{i,l}^c f_l^c(t) y^c(t-i) + u^c(t) \end{aligned} \tag{5}$$

where $u^c(t)$ and $y^c(t)$ are taken to be complex-valued input and analytic speech signal, respectively. In the TV-CAR model, the complex AR coefficient is modeled by a finite number of arbitrary complex basis. Note that Eq.(3) parameterizes the AR coefficient trajectories that continuously change as a function of time so that the time-varying analysis is feasible to estimate continuous time-varying speech spectrum. In addition, as mentioned above, the complex-valued analysis facilitates accurate spectral estimation in the low frequencies, as a result, this feature allows for more accurate F_0 estimation if formant structure is removed by the inverse filtering. Eq.(5) can be represented by vector-matrix notation as

$$\begin{aligned} \bar{y}_f &= -\bar{\Phi}_f \bar{\theta} + \bar{u}_f \\ \bar{\theta}^T &= [\bar{g}_0^T, \bar{g}_1^T, \cdots, \bar{g}_l^T, \cdots, \bar{g}_{L-1}^T] \\ \bar{g}_l^T &= [g_{1,l}^c, g_{2,l}^c, \cdots, g_{i,l}^c, \cdots, g_{I,l}^c] \\ \bar{y}_f^T &= [y^c(I), y^c(I+1), y^c(I+2), \cdots, y^c(N-1)] \\ \bar{u}_f^T &= [u^c(I), u^c(I+1), u^c(I+2), \cdots, u^c(N-1)] \\ \bar{\Phi}_f &= [\bar{D}_0^f, \bar{D}_1^f, \cdots, \bar{D}_l^f, \cdots, \bar{D}_{L-1}^f] \\ \bar{D}_l^f &= [\bar{d}_{1,l}^f, \cdots, \bar{d}_{i,l}^f, \cdots, \bar{d}_{I,l}^f] \\ \bar{d}_{i,l}^f &= [y^c(I-i) f_l^c(I), y^c(I+1-i) f_l^c(I+1), \\ &\qquad \cdots, y^c(N-1-i) f_l^c(N-1)]^T \end{aligned} \tag{6}$$

where N is analysis interval, $\bar{y}_f$ is $(N-I,1)$ column vector whose elements are analytic speech signal, $\bar{\theta}$ is $(L \cdot I, 1)$ column vector whose elements are complex parameters, $\bar{\Phi}_f$ is $(N-I, L \cdot I)$ matrix whose elements are weighted analytic speech signal by the complex basis. Superscript T denotes transposition.

2.3 MMSE-based algorithm[10]

MSE criterion is defined by

$$\begin{aligned} \bar{r}_f &= [r^c(I), r^c(I+1), \cdots, r^c(N-1)]^T \\ &= \bar{y}_f + \bar{\Phi}_f \hat{\theta} \end{aligned} \tag{7}$$

$$r^c(t) = y^c(t) + \sum_{i=1}^{I} \sum_{l=0}^{L-1} \hat{g}_{i,l}^c f_l^c(t) y^c(t-i) \tag{8}$$

$$E = \bar{r}_f^H \bar{r}_f = (\bar{y}_f + \bar{\Phi}\hat{\theta})^H (\bar{y}_f + \bar{\Phi}\hat{\theta}) \tag{9}$$

where $\hat{g}_{i,l}^c$ is the estimated complex parameter, $r^c(t)$ is an equation error, or complex AR residual and E is Mean Squared Error (MSE) for the equation error. To obtain optimal complex AR coefficients, we minimize the MSE criterion. Minimizing the MSE criterion of Eq.(9) with respect to the complex parameter leads to the following MMSE algorithm.

$$(\bar{\Phi}_f^H \bar{\Phi}_f)\hat{\theta} = -\bar{\Phi}_f^H \bar{y}_f \tag{10}$$

Superscript H denotes Hermitian transposition. After solving the linear equation of Eq.(10), we can get the complex AR parameter $(a_i^c(t))$ at time t by calculating the Eq.(3) with the estimated complex parameter $\hat{g}_{i,l}^c$.

2.4 ELS-based algorithm[11]

If the equation error shown as in Eq.(8) is white Gaussian, the MMSE estimation is optimal, however, it is rare case. As a result, MMSE estimation suffers from biased estimation. In the ELS method, an AR filter is adopted to whiten the equation error as follows.

$$r^c(t) = -\sum_{k=1}^{K} b_k^c r^c(t-k) + e^c(t) \tag{11}$$

where b_k^c is k-th parameter of the AR filter whose order is K and $e^c(t)$ is 0-mean white Gaussian of equation error at time t. The inverse filter of Eq.(11) is called a whiten filter. The TV-CAR model can be represented using Eq.(5) and Eq.(11) as follows.

$$\begin{aligned} y^c(t) &= -\sum_{i=1}^{I}\sum_{l=0}^{L-1} g_{i,l}^c f_l^c(t) y^c(t-i) \\ &\quad -\sum_{k=1}^{K} b_k^c r^c(t-k) + e^c(t) \end{aligned} \tag{12}$$

Eq.(12) is the ELS model. The parameter is estimated so as minimize the MSE for the whitened equation error in the ELS algorithm whereas the parameter is estimated so as minimize the MSE for the equation error in the MMSE algorithm.

Eq.(12) can be expressed by the following vector-matrix notation.

$$\begin{aligned} \bar{y}_f &= -\bar{\Phi}_f \bar{\theta} - \bar{R}_f \bar{b} + \bar{e}_f \\ &= -(\bar{\Phi}_f \quad \bar{R}_f) \begin{pmatrix} \bar{\theta} \\ \bar{b} \end{pmatrix} + \bar{e}_f \end{aligned} \tag{13}$$

where

$$\begin{aligned} \bar{R}_f &= \begin{pmatrix} r^c(I-1) & r^c(I-2) & \cdots & r^c(I-K) \\ r^c(I) & r^c(I-1) & \cdots & r^c(I+1-K) \\ \vdots & \vdots & \ddots & \vdots \\ r^c(t) & r^c(t-1) & \cdots & r^c(t-K) \\ \vdots & \vdots & \ddots & \vdots \\ r^c(N-2) & r^c(N-3) & \cdots & r^c(N-1-K) \end{pmatrix} \\ \bar{b} &= [b_1^c, b_2^c, \cdots, b_K^c]^T \\ \bar{e}_f &= [e^c(I), e^c(I+1), e^c(I+2), \cdots, e^c(N-1)]^T \end{aligned} \tag{14}$$

By minimizing the MSE for Eq.(13), one can get the following equation.

$$\begin{pmatrix} \bar{\Phi}_f^H \bar{\Phi}_f & \bar{\Phi}_f^H \bar{R}_f \\ \bar{R}_f^H \bar{\Phi}_f & \bar{R}_f^H \bar{R}_f \end{pmatrix} \begin{pmatrix} \hat{\theta} \\ \hat{b} \end{pmatrix} = -\begin{pmatrix} \bar{\Phi}_f^H \bar{y}_f \\ \bar{R}_f^H \bar{y}_f \end{pmatrix} \tag{15}$$

By applying the well-known inversion Matrix lemma to Eq.(15), one can obtain the following equation.

$$\left(\bar{\Phi}_f^H \bar{\Phi}_f\right) \hat{\theta}_{bias} = \bar{\Phi}_f^H \bar{R}_f \hat{b} \tag{16}$$

$$\hat{\theta} = \hat{\theta}_0 - \hat{\theta}_{bias} \tag{17}$$

The MMSE estimated parameter $\hat{\theta}_0$ contains the biased element $\hat{\theta}_{bias}$. The unbiased estimation of $\hat{\theta}$ is calculated by $\hat{\theta}_0 - \hat{\theta}_{bias}$. The ELS algorithm is equivalent to the GLS (Generalized Least Square) algorithm and more sophisticated algorithm. Since the equation error $r^c(t)$ cannot be observed, the iteration algorithm is required by estimating the $A(z)$ and $B(z)$ simultaneously. The iteration procedure is shown as follows.

(1)	Initial $\hat{\theta}_0$ is estimated by MMSE (Eq.(10)).
(2)	The equation error is calculated by Eq.(8).
(3)	$\hat{b}$ is estimated so as to minimize Eq.(18) using $r^c(t)$.
(4)	The bias parameter $\hat{\theta}_{bias}$ is calculated by Eq.(16).
(5)	The unbiased parameter $\hat{\theta}$ is calculated by Eq.(17).
(6)	Go to (2).

$$\frac{1}{2\pi j}\oint_{|z|=1} |R(z)B(z)|^2 \frac{dz}{z} = 0 \tag{18}$$

In Eq.(18), $R(z)$ is z-transform of $r^c(t)$ and $B(z)$ is the transfer function of the whiten filter. The procedures from (2) to (5) are iterated with the pre-determined number. The ELS algorithm estimates two kinds of AR filters, $A(z)$ and $B(z)$, iteratively. If L is 1, the ELS model is only series connection of two AR filters. Thus, there is a suspicion that ELS algorithm is equivalent to high-order MMSE algorithm. ELS method can effectively whiten the MMSE equation error using whiten filter $B(z)$ whereas high-order MMSE algorithm cannot force the equation error closed to white Gaussian effectively since the MMSE estimation is carried out so as to minimize the mean squared equation error. Accordingly, the MMSE algorithm suffers from bias estimation. The ELS method can realize more robust spectral estimation than high-order MMSE algorithm due to the whiten filter, $B(z)$.

3. F_0 ESTIMATION METHOD

3.1 Shimamura Method[7]

Autocorrelation function (AUTOC) is defined by

$$f(\tau) = \frac{1}{N}\sum_{t=0}^{N-1} x(t)x(t+\tau) \tag{19}$$

where $x(t)$ is target signal such as speech signal, LPC residual or so on, N is frame length and τ means delay. F_0 is selected as peak frequency for Eq.(19) within certain range of F_0.

AMDF is defined as follows.

$$p(\tau) = \frac{1}{N}\sum_{t=0}^{N-1} |x(t) - x(t+\tau)| \tag{20}$$

F_0 is selected as notch frequency for Eq.(20) within certain range of F_0.

In Shimamura method [7], the AUTOC is weighted by a reciprocal of the AMDF shown as Eq.(21). Since the weighting makes it possible to suppress other peaks, the method can estimate more accurate F_0 than AUTOC or AMDF. The value of m is set to be 1 in order to avoid the value of 0 at the denominator.

$$G(\tau) = \frac{f(\tau)}{p(\tau) + m} \tag{21}$$

where $f(\tau)$ and $p(\tau)$ are AUTOC shown as in Eq.(19) and AMDF shown as in Eq.(20), respectively.

3.2 Proposed Method

In this paper, Shimamura criterion shown as Eq.(21) is applied to complex AR residual extracted by the robust TV-CAR speech analysis. The complex parameter is estimated and complex AR residual is calculated with the estimated complex parameter. Note that pre-emphasis is operated for speech analysis such as real-valued AR or TV-CAR speech analysis, and inverse filtering is applied for the non pre-emphasized speech signal so as not to eliminate F_0 spectrum on the residual signal. Real part of AUTOC is used to calculate the AUTOC for complex-valued signal.

4. EXPERIMENTS

Speech signals used in the experiment are 10 sentences uttered by male speaker and 10 sentences uttered by female speaker of ATR database. Speech signals are filtered by an IRS filter [14]. The IRS filter is band pass FIR filter whose frequency response corresponds to that for analog part of the transmitter of telephone equipment. In order to evaluate the proposed method for the speech data processed by speech coding, the IRS filter has to be introduced. It provides the characteristics of HPF in low frequencies and LPF in high frequencies shown as in [12]. Since the IRS filter is designed for 16 KHz sampled data, speech data are corrupted by additive noise in 16KHz sampling and the IRS filtering is carried out and then down sampling from 16KHz to 10KHz is operated. Frame length is 25.6[msec] and frame shift length is 10[msec]. Analysis orders are 14 and 7 for real-valued analysis and complex-valued analysis, respectively. Note that the basis expansion order L is 1, consequently, non time-varying analysis is adopted in the experiments. Hence, no basis function is adopted in speech analysis. White Gauss noise or pink noise [15] is adopted for additive noise and the levels are 30, 20, 10, 5, 0, and -5 [dB]. In order to extract more accurate F_0, 3-point Lagrange's interpolation is adopted. The experimental conditions are summarized in Table 1.

Commonly used criterion for F_0 estimation, Gross Pitch Error (GPE) [6], is adopted for objective evaluation. F_0 estimation error is defined as

$$e_p(n) = F_e(n) - F_t(n) \tag{22}$$

where $F_t(n)$ is true F_0 value and $F_e(n)$ is the estimated one. The true F_0 values are derived by pit file (*.pit) of ATR database, in which F_0 was estimated by cepstral method and the estimation error was modified by hand. In Eq.(22), if $|e_p(n)| \geq F_t(n) \times THR/100$ then the estimation error is regarded as ERROR and GPE is the probability of the error frames. Figures 1 and 2 shows the experimental results setting the THR as $20[\%]$. Figure 1 shows the results for the MMSE-based method[12]. Figure 2 shows the results for the proposed ELS-based method. In the figures, (1) denotes the GPEs for additive white Gauss noise, (2) denotes the GPEs for additive pink noise, While the MMSE method is implemented by Eq.(10) with L=1, the ELS method is implemented by Eq.(17) in which L=1, K=5, the iteration number is 6 and initial parameter θ_0 is estimated by Eq.(10). **SP** (solid line), **AN** (dotted line), **LPC** (dashed line) and **CLPC** (dash-dotted line) mean the results for speech, analytic signal, real AR residual and complex AR residual, respectively. Note that **SP** means the Shimamura method [7], viz., Shimamura criterion for speech signal and **CLPC** means the proposed method. In all figures, X-axis means noise level of $30, 20, 10, 5, 0, -5[dB]$. Y-axis means GPE$[\%]$.

Table 1: Experimental Conditions

Speech data	ATR database set B Male speaker 1, 10 sentences Female speaker 1, 10 sentences
IRS filter	64-th FIR[14]
Target signal	(1)speech signal (2)real AR residual (3)analytic speech signal (4)complex AR residual
Sampling	10kHz/16bit
Analysis window	Window Length: 25.6[ms] Shift Length: 10.0[ms]
F_0 search range	50 to 400 [Hz]
Real-valued AR Pre-emphasis	I=14, L=1 (time-invariant) $1-z^{-1}$
Complex-valued AR Pre-emphasis	I=7, L=1 (time-invariant) $1-z^{-1}$
Criterion	AUTOC/AMDF[7]
Noise Noise Level	(1)white Gauss noise (2)pink noise[15] 30,20,10,5,0,-5[dB]
Interpolation	3 point Lagrange's

Figure 1 (2) demonstrates that the MMSE-based complex residual cannot estimate more robust F_0 estimation for additive pink noise. It results from the spectral compensation in low frequencies due to the spectral tilt of pink noise. Figures 1 and 2 demonstrate that the proposed ELS-based method can perform better not only for additive white Gauss noise but also for Pink noise in terms of GPE. The reason why the ELS-based complex method can perform better for pink noise is as follows. The ELS method can estimate more accurate first or second formants of speech and resonance frequency of additive pink noise than MMSE method. For this reason, the estimated complex residual contains less first or second formants and the pink noise resonance. As a result, the proposed F_0 estimation based on the ELS method can realize more robust F_0 estimation.

5. CONCLUSIONS

This paper has proposed robust fundamental frequency estimation algorithm for noisy speech signal. In this method complex AR residual extracted from analytic speech signal by means of the ELS-based complex AR analysis is applied to calculate the criterion for F_0 estimation. The experiments using IRS filtered speech corrupted by white Gauss noise or pink noise demonstrate that the proposed method outperforms for noisy speech than the other methods in terms of GPE. The proposed F_0 estimation can be introduced as a open-loop adaptive codebook in CELP speech coding [1],[2] or iLBC[3]. Since closed-loop final adaptive codebook search is carried out within the neighboring range for the lag pre-selected by open-loop search, it can realize more accurate adaptive codebook search for noisy speech, as a result, the speech quality can be improved in realistic environment. In order to improve the performance more, forward and backward linear prediction [16], low pass filtering and decimation and optimal pre-emphasis will be introduced and examined. This paper evaluates only non-time invariant, frame-based F_0 estimation, thus, time-varying analysis will be introduced. When noise level is low such as 30 or 20$[dB]$ of ordinary environment, the performance is not sufficient. In order to improve the estimation accuracy for the environment, time-varying F_0 contour estimation[17] will be introduced as a pre-selection of the proposed frame-based F_0 estimation.

REFERENCES

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

[2] ITU-T Recommendation G.722.2, "Wideband coding of speech at around 16kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)," 2003.

[3] IETF RFC 3951 "Internet Low Bit Rate Codec (iLBC)," http://www.ietf.org/rfc/rfc3951.txt, 2004.

[4] W.J.Hess, "Pitch and voicing determination," in Advances in Speech Signal Processing, ed. S.Furui and M.sondhi, Marcel Dekker, 1992.

[5] M. J. Ross, H. L. ShaiTer, A.Cahen, R. Freudberg, and. H.J.Manley, "Average magnitude difference function pitch extractor," IEEE Trans. ASSP-22, No.5, pp. 353-362, 1974.

[6] L.R.Rabiner, M.Cheng, A.Rosenberg, C.McGonegal, "A comparative performance study of several pitch detection algorithms," IEEE Trans. ASSP-24, no.5, pp.399-418, 1976.

[7] T.Shimamura and H.Kobayashi,"Weighted Autocorrelation for Pitch Extraction of Noisy Speech," IEEE Trans. Speech and Audio Processing, vol. 9, no. 7, pp. 727-730, 2001.

[8] S.M.Kay,"Maximum entropy spectral estimation using the analytic signal," IEEE Trans. ASSP-26, pp.467-469, 1980.

[9] T.Shimamura and S.Takahashi,"Complex linear prediction method based on positive frequency domain," IEICE Trans., Vol.J72-A, pp.1755-1763, 1989. (in Japanese)

[10] K.Funaki, Y.Miyanaga, and K.Tochinai,"On a time-varying complex speech analysis," Proc. EUSIPCO-98,Rodes,Greece, Sep. 1998.

[11] K.Funaki,"A time-varying complex AR speech analysis based on GLS and ELS method," Proc.EUROSPEECH2001, Aalborg Denmark, Sep., 2001.

[12] K.Funaki and T.Kinjo,"Robust F0 Estimation Based on Complex LPC Analysis for IRS Filtered Noisy Speech," IEICE Trans. on Fundamentals, Vol. E90-A, No.8.,1579-1586, Aug. 2007.

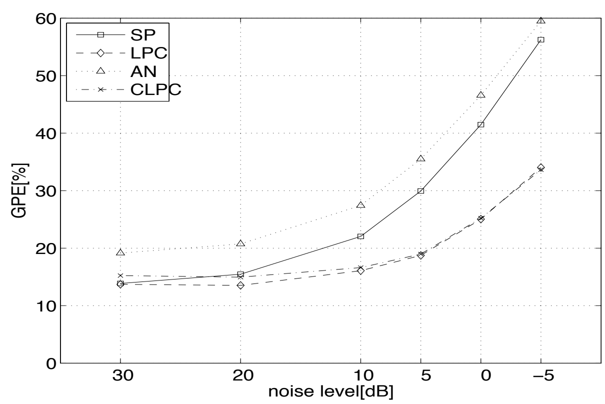
[13] K.Funaki and T.Kinjo,"Robust F0 Estimation using ELS-based Complex Speech Analysis," IEICE Trans. on Fundamentals, Vol. E91-A, No.3.

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

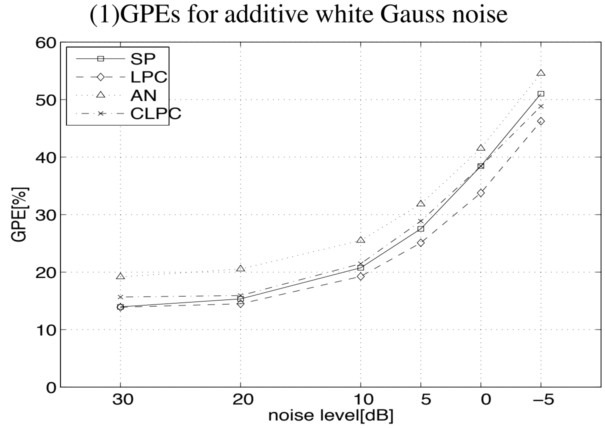
[15] NOISE-X92, http://spib.rice.edu/spib/select_noise.html

[16] K.Funaki,"Improvement of the ELS-based time-varying complex speech analysis," Proc.ICSLP-2002, Denver, USA, Sep.17-20, 2002.

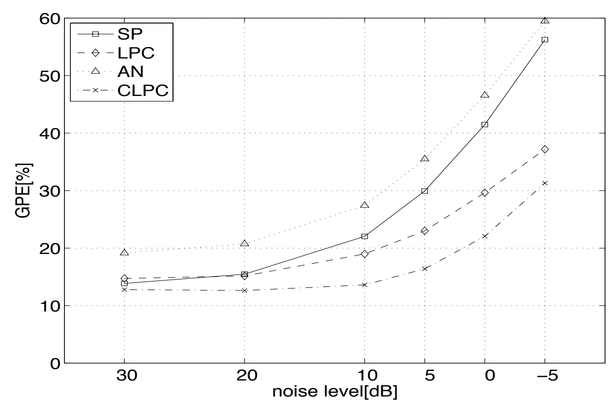
[17] K.Funaki,"F_0 contour estimation based on time-varying complex speech analysis," Proc.Acoustics'08, Jun.2008.



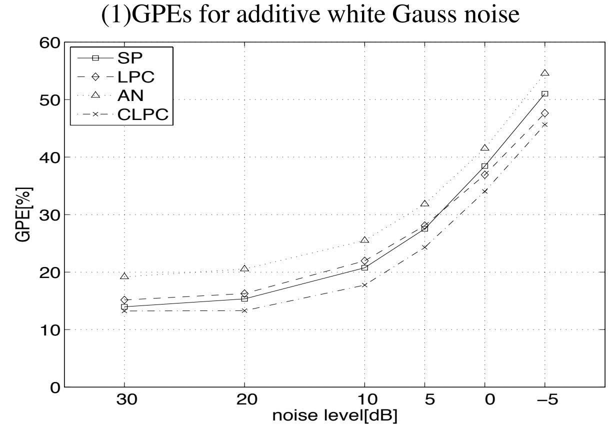
(1)GPEs for additive white Gauss noise



(2)GPEs for additive pink noise
Figure 1 MMSE-based method



(1)GPEs for additive white Gauss noise



(2)GPEs for additive pink noise
Figure 2 ELS-based method