ON ROBUST SEQUENTIAL ESTIMATOR BASED ON T-DISTRIBUTION WITH FORGETTING FACTOR FOR SPEECH ANALYSIS

Joohun Lee and Ki Yong Lee***

* Dept. of Information and Telecommunication, Dong-Ah Broadcasting College, Ansung, Korea

** School of Electronic Engineering, Soongsill University, Seoul, Korea

ABSTRACT

In this paper, to estimate the time-varying speech parameters having non-Gaussian excitation source, we use the robust sequential estimator(RSE) based on t-distribution and introduce the forgetting factor. By using the RSE based on t-distribution with small degree of freedom, we can alleviate efficiently the effects of outliers to obtain the better performance of parameter estimation. Moreover, by the forgetting factor, the proposed algorithm can estimate the accurate parameters under the rapid variation of speech signal.

1. INTRODUCTION

The estimation and tracking of speech parameters have long been recognized as important adjuncts to speech signal processing and the several methods based on linear predictive coding(LPC) have been developed as a useful method. However, those frame-based analysis methods are known to have problems for certain type of speech signals, including source-track interaction when periodic pulse trains are the excitation, as in voiced sounds and the fast transition between vowels and consonants. To overcome the drawbacks of those methods, the Kalman filter was proposed. However, in the presence of outliers, the Kalman filter is known to show very poor performance since it is optimal only for Gaussian noise. Also, when the speech signal varies rapidly, the parameter-tracking performance of the Kalman filter is diminished by the weight which the filter gives to the history of the signal.

In this paper, to estimate the parameters of speech signal having non-Gaussian excitation source, we use the robust sequential estimator(RSE)[2] based on t-distribution. Also, to cope with the rapid variation of speech signal, we introduce the forgetting factor. We use a loss function which assigns large weighting factor for small amplitude residuals and small weighting factor for large amplitude residuals which is for instance caused by the pitch excitations. The loss function is based on the assumption that the residual signal has an independent and identical t-distribution with α degrees of freedom. When α goes to infinite, we get the conventional LP method. Since the t-distribution with small α has more probability on its tail than that with large α, we assume that α=3. Therefore, by using the RSE based on t-distribution with small degree of freedom, we can alleviate efficiently the effects of outliers to obtain the better performance of parameter estimation. Moreover, in order to cope with the rapid variation of speech signal, we introduce the forgetting factor to this RSE. By the forgetting factor, the proposed algorithm can estimate the accurate parameters under the rapid variation of speech signal to base on estimation on only the most recent portion of the data.

Some experimental results performed on real speech signals, Korean sentences lasting about one second, show that the proposed algorithm achieves more accurate estimation and provides improved tracking performance with smaller variance and bias, compared to the robust Kalman filter[1] based on Huber's M-estimate for both Gaussian and heavy-tailed processes.

2. THE PROPOSED ALGORITHM

The residual signal ε_i can be expressed as a function of the linear prediction(LP) vector as

$$\varepsilon_i(\boldsymbol{a}) = s_i + \sum_{j=1}^{p} a_j \cdot s_{i-j} \qquad (2\text{-}1)$$

,where $a = [a_1 a_2 \mathrm{K}\ a_p]^T$ and a_j are LP coefficients. In Eq.

(2-1), s_i is a time-varying autoregressive model of order p, AR(p). The excitation source ε_i is considered a non-Gaussian process which is a combination of two Gaussian processes with different variances: one Gaussian process has a small variance which accounts for the modeling error caused by fitting the vocal tract structure with improper model parameters, and the other with a relatively much larger variance represents the error due to the spiky excitation.

In the conventional LP(CLP) speech analysis, the predictor coefficients a_j, $1 \le j \le p$, are determined to minimize the sum of the squares of the prediction residuals. Since the result is least square fit, the obtained estimate is very much affected by the strong signal parts and results in difficulties for the LP analysis of high-pitched voices. Also, In the CLP method, the structure of the source excitation is not taken into account. As mentioned above, when the excitation source is modeled by δ-contaminated normal mixture model, the least square method is biased and inefficient.

Many robust procedures can be viewed as a modified least squares(LS). Robust estimators are more efficient (lower variance) than LS when the errors are not normally distributed., and slightly less so when they are. The key step of robust procedure is to replace the square by another symmetric cost function of the residuals or to model the noise by a nonnormal, heavy-tailed distribution to account for outliers. We can then use maximum likelihood analysis to obtain robust estimates of the parameter vector. Unfortunately, the direct evaluation of maximum likelihood estimates from nonnormal distributions becomes quite complicated. But an effective means of obtaining maximum likelihood estimates for a wide class of nonnormal distributions is weighted least squares.

Let $f(\varepsilon_i)$ be any differentiable error density function which can be written in the form

$$f(\varepsilon_i) \propto \sigma^{-1} g\{(\frac{\varepsilon_i}{\sigma})^2\} \qquad (2\text{-}2)$$

where σ is a scale parameter, $g\{\cdot\}$ denotes a functional form, and $\varepsilon_i = s_i - \sum_{j=1}^{p} a_j s_{i-j}$ is the i-th actual error. Given a sample $\boldsymbol{s}$ of k observations, the logarithmic likelihood function for a_i and σ^2 is given by

$$l(\boldsymbol{a}, \sigma | \boldsymbol{\varepsilon}) = K - \sum_{i=1}^{k} [\log \sigma^{-1} + \log g\{(\frac{\varepsilon_i}{\sigma})^2\}] \qquad (2\text{-}3)$$

,where K is some constant.

From this, we can define the error criterion function as

$$J_k(\boldsymbol{a}) = \sum_{i=1}^{k} [\log \sigma^{-1} + h\{(\frac{\varepsilon_i}{\sigma})^2\}] \qquad (2\text{-}4)$$

,where $h(\cdot) = \log g(\cdot)$. For heavy-tailed Gaussian process, $h(\cdot)$ is replaced by the Huber's score function $\rho_H(\cdot)$ defined as

$$\rho_H(x) = \begin{cases} x^2/2 & |x| \le c \\ c|x| - c^2/2 & |x| \le c \end{cases} \qquad (2\text{-}5)$$

In this paper, heavy tailed error distribution is reasonably represented by a t-distribution defined by

$$f_\alpha(x) = \frac{1}{\sqrt{\alpha x}} \frac{\Gamma(\frac{\alpha+1}{2})}{\Gamma(\frac{\alpha}{2})} \frac{1}{(1+\frac{x^2}{\alpha})^{(\alpha+1)/2}} \qquad (2\text{-}6)$$

having small degree of freedom α and scaled by a parameter σ. Therefore, $g(\cdot)$ is given by

$$g\{(\frac{\varepsilon_i}{\sigma})^2\} = \{1 + \frac{\varepsilon_i^2}{(\alpha\sigma^2)}\}^{(-1/2)(1+\alpha)}. \qquad (2\text{-}7)$$

The degree of freedom is given by $\alpha = 3$, since this induces to the most accurate and the efficient estimation[3].

In addition to that, the forgetting factor λ, $0 < \lambda \le 1$, is employed to weight the most recent data more heavily to allow for tracking of varying parameters. Progressively smaller λ result in parameter being computed with effectively smaller windows of data that are beneficial in nonstationary situations. Then, the Eq. (3-4) is rewritten as

$$J_k(\boldsymbol{a}) = \sum_{i=1}^{k} \lambda^{k-i} \cdot [\log \sigma^{-1} + \log g\{(\frac{\varepsilon_i}{\sigma})^2\}]. \qquad (2\text{-}8)$$

To obtain the M-estimator, by differentiating the Eq. (2-8) with respect to $\boldsymbol{a}$ and σ^2 and equating these two equations to zero, we get the maximum likelihood estimates $\overline{\boldsymbol{a}}$ and $\overline{\sigma}^2$ as the solution of the nonlinear equations

$$\sum_i \lambda^{k-i} w_i (s_i - \sum_j \overline{a}_j s_{i-j}) s_{i-h} = 0, \quad h = 1,..., p \qquad (2\text{-}9)$$

$$\mu^2 = \overline{\sigma}^2 = \sum_i \lambda^{k-i} w_i (s_i - \sum_j \overline{a}_j s_{i-j})^2 \Big/ n$$

,where $w_i = w_i(\boldsymbol{a}, \sigma^2) = -2\left[\frac{\partial \log g(\xi)}{\xi}\right]_{\xi=(\varepsilon_i/\sigma)^2}$. (2-10)

We can solve these nonlinear equations using iteratively reweighted least squares(IRLS). Rewriting (2-9) in matrix form yields

$$\boldsymbol{H}^T \boldsymbol{W}(\boldsymbol{s} - \boldsymbol{H}\overline{\boldsymbol{A}}) = \boldsymbol{0},$$

where $\boldsymbol{s}$ is an n-vector of speech signals on the dependent variable, $\boldsymbol{H}$ is an $n \times p$ matrix of observed speech signals having rank p, $\boldsymbol{A}$ is a p-vector of parameters to be estimated, and $\boldsymbol{W}$ is a diagonal matrix defined such $\lambda^{n-i} w_i$ as

$$\boldsymbol{W} = \begin{bmatrix} \lambda^{n-1} w_1 & \mathrm{L} & 0 \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ 0 & \mathrm{L} & \lambda^0 w_n \end{bmatrix}.$$

Substituting Eq. (2-7) for g in (2-10) gives the individual weights:

$$w_i = 1 + \alpha \Big/ \alpha + (r_i / \mu)^2 \qquad (2\text{-}11)$$

,where the residual $r_i = s_i - \sum_j a_j s_{i-j}$. We use μ in this expression to distinguish the (unknown) true value of the scale parameter, σ, from an estimated value, μ, used in computing the weights.

The RSE starts from an initial robust estimate of the speech parameters and these parameters are computed by IRLS with the errors assumed to be t-distributed. It then adds the remaining data sequentially, assigning weights to each new observation based on the previous estimates.

Suppose that the robust parameter estimate for the first m observations is

$$\overline{\boldsymbol{A}}_m = (\boldsymbol{H}_m^T \boldsymbol{W}_m \boldsymbol{H}_m)^{-1} \boldsymbol{H}_m^T \boldsymbol{W}_m \boldsymbol{s}_m,$$

where $\boldsymbol{W}$ is the appropriate diagonal weighting matrix computed by the maximum likelihood analysis. After expanding $\overline{\boldsymbol{A}}_{m+1}$ as

$$\overline{\boldsymbol{A}}_{m+1} = [\lambda \boldsymbol{H}_m^T \boldsymbol{W}_m \boldsymbol{H}_m + w_{m+1} \boldsymbol{h}_{m+1} \boldsymbol{h}_{m+1}^T]^{-1} \times [\lambda \boldsymbol{H}_m^T \boldsymbol{W}_m \boldsymbol{s}_m + w_{m+1} s_{m+1} \boldsymbol{h}_{m+1}]^{-1}$$

,where $\boldsymbol{P}_m = (\boldsymbol{H}_m^T \boldsymbol{W}_m \boldsymbol{H}_m)^{-1}$ and rearranging the terms, we obtain the following recursive equations for the robust sequential algorithm:

$$\overline{\boldsymbol{A}}_{m+1} = \overline{\boldsymbol{A}}_m + \gamma_{m+1} \boldsymbol{P}_m \boldsymbol{h}_{m+1} \left[s_{m+1} - \boldsymbol{h}_{m+1}^T \overline{\boldsymbol{A}}_m \right],$$

$$\lambda \boldsymbol{P}_{m+1} = \boldsymbol{P}_m - \gamma_{m+1} \boldsymbol{P}_m \boldsymbol{h}_{m+1} \boldsymbol{h}_{m+1}^T \boldsymbol{P}_m$$

,where $\gamma_{m+1} = w_{m+1} \Big/ \lambda + w_{m+1} \boldsymbol{h}_{m+1}^T \boldsymbol{P}_m \boldsymbol{h}_{m+1}$.

The crucial problem in robust estimators is the estimation of the scale parameter. As an efficient, robust and simple approach to simultaneous scale and parameter estimation for a wide class of nonnormal distributions is obtained using maximum likelihood analysis: $\mu_m^2 = \sum_i^m w_i (s_i - \sum_j \overline{a}_j s_{i-j})^2 \Big/ m$.

3. SIMULATION RESULTS

The proposed algorithm has been tested on both synthetic speech and natural Korean speech, 'sa'. The results have been compared to those by the conventional Kalman filter and the robust Kalman filter proposed in [1]. The results are shown in Figure 1 and Figure 2. In Figure 1, a_1 trajetories for the synthetic speech, obtained by each method. In Figure 2 (b), from the top, a_1 trajetories of the conventional Kalman, the robust Kalman and the proposed algorithm are presented, respectively. They are given a little bias to show their differences apparently. As shown in these figures, the proposed algorithm can estimate the trajectory of the parameter more accurately, while the others are much affected by outliers, the pitch excitations. In this simulation.

4. CONCLUSIONS

We proposed the Robust Sequential Estimator based on t-distribution with forgetting factor. The proposed algorithm can alleviate efficiently the effects of outliers to obtain the better

parameter estimation by introducing t-distribution to sequential estimation. Also, by introducing the forgetting factor, it can estimate the rapid varying parameters which brings some drawbacks to the Kalman filter. The simulation results show that, by the proposed method, the better estimation performance can obtained.

5. REFERENCES

1.T. Yang, J. Lee, K. Lee and K. Sung, "On Robust Kalman Filtering with Forgetting Factor for Sequential Speech Analysis", *Signal Processing*, vol.63, no.2, pp.1151-1156, 1997.

2. K. Boyer, J. Mizra and G. Ganguly, "The Robust Sequential Estimator: A General Approach and its Application to Surface Organization in Range Data", *IEEE Trans. Pattern Anal. Machine Intell.*, vol.l.16, no.10, 1994.

3. J. Sanubari, K. Tokuda and M. Onoda, "Speech Analysis Based on AR Model Driven by t-distribution Process", *IEICE Trans. Fundamentals*, vol.E75-A. no.9, pp.1159-1169,. 1992

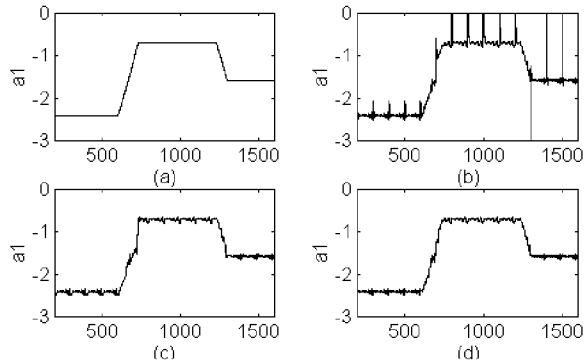

Figure 1: a_1 coefficient trajectories for synthetic speech. (a) Original a_1 trajectory, (b) Estimated a_1 trajectory obtained by the conventional Kalman filter(λ =0.85), (c) Estimated a_1 trajectory obtained by the robust Kalman filter(λ =0.85, c =0.1), (d) Estimated a_1 trajectory obtained by the proposed robust sequential algorithm(λ =0.85, α =3).

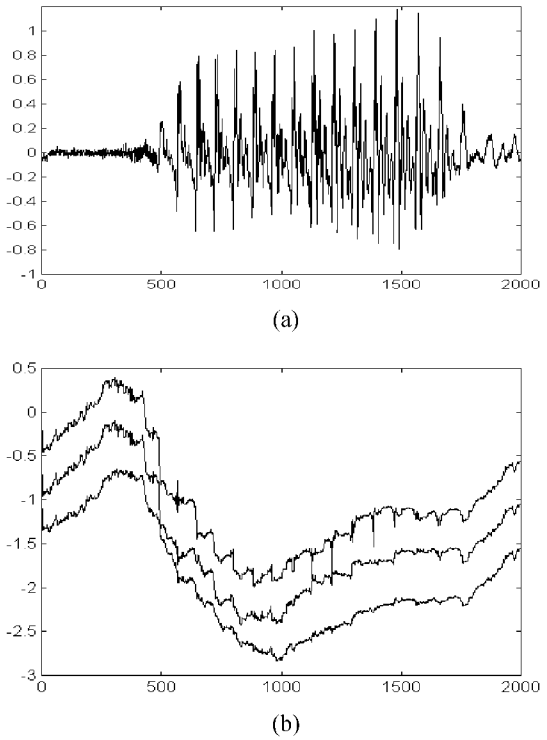

Figure 2: a_1 coefficient trajectories for real speech 'sa'. (a) Real speech 'sa', (b) From the top, each trajectory represents the estimated one of a_1 coefficients obtained by the conventional Kalman filter(λ =0.98), the robust Kalman filter(λ =0.98, c =1.5) and the proposed robust sequential algorithm(λ =0.98, α =3), respectively.