Adaptive Non-Linear Prediction with Order Statistics for Speech in Impulsive Noise

HIROBUMI TANAKA, YUICHIROH OHHASHI and TETSUYA SHIMAMURA
Graduate School of Science and Engineering
Saitama University
255 Shimo-Okubo, Sakura-ku, Saitama, 338-8570
JAPAN

Abstract: - In this paper, we investigate the linear prediction of speech signals in an impulsive noise environment. Both schemes of batch processing and adaptive processing are comparatively studied and it is shown that the adaptive processing scheme is basically suitable in a highly impulsive noise environment. As an extended version of the Order Statistic Least Mean Square (OSLMS) algorithm addressed by Shimamura *et al.*, an OSLMS algorithm involving ambiguous sorting is developed. The performance of the proposed algorithm is demonstrated and it is shown that the effects of impulse noise are significantly suppressed by the proposed algorithm.

Key-Words: - prediction, speech signals, order statistics, impulsive noise

1 Introduction

In recent communication systems, a transmitted speech signal very often corrupted by impulsive noise. This is caused by the fact that recently, various types of connection are used in the transmitted and received circuits and mismatching of them invokes a kind of impulsive disturbances. Furthermore, in a certain communication system, power supply is controlled by switching, which results in an another source of impulses. Impulsive type disturbances distort the speech waveform, leading to a degradation of quality and intelligibility of the transmitted speech [1, 2].

Linear Prediction (LP) is a widely used technique in speech processing systems, because it efficiently provides a compact representation of speech. The performance of LP, however, is affected by additive noise. In [3, 4], it has been reported that LP is very sensitive to white Gaussian noise. Furthermore, in [5], the performance of LP is affected by colored Gaussian noise as well. It is expected that LP is affected in an impulsive noise environment as well. Although a method to suppress an impulsive disturbance on a speech signal is proposed in [7], the performance of LP in impulsive noise is not clear.

Two commonly used methods for LP are the autocorrelation method and the covariance method. Both are batch processing in which one set of parameters of LP is obtained from a segmented data. In this case, an impulse contained in the segmented data may corrupt all parameters of LP. In many cases, impulsive type disturbances appear at a very small region of a segmented data. This means that if LP is performed sequentially, then it may be possible to reduce the effects of impulse noise, depending on the convergence characteristics of the adaptation scheme of LP.

From this point of view, we consider two adaptive algorithms; LMS [6] and OSLMS [7]. The LMS algorithm is simple and standard, while the OSLMS algorithm retains the time information the speech data has and has the potential to reduce impulsive noise. The OSLMS algorithm is based on **strict sorting**. If the OSLMS is applied directly when a frame length N is short, then even if the impulsive noise is absent, the convergence property of the OSLMS becomes worse than that of the LMS. This is because much more coefficients have to be updated for the OSLMS algorithm.

In this paper, we introduce a concept of **ambiguous sorting** instead of the strict sorting, and develop an OSLMS algorithm with ambiguous sorting (OSLMS with AS). The basic idea of the AS is to combine the LMS algorithm with the OSLMS algorithm. In the AS, when an impulse is present, the OSLMS algorithm works. If an impulse is not present, the LMS algorithm works normally. The OSLMS with AS provides a convergence speed which is similar with the LMS, but has robustness against impulsive noise, which is not shared with the LMS. These properties of the OSLMS with AS are demonstrated in comparison with the LMS and original OSLMS algorithms.

2 OSLMS Predictor Based on Ambiguous Sorting

In this section, the OSLMS algorithm with AS is described. This algorithm is derived from the OSLMS algorithm proposed by Shimamura *et al.*[7]. The difference point between the two algorithms is the sorting operation for the input vector. In the OSLMS with AS, a parameter inducing ambiguity in sorting, s_{diff}, is used.

Assume that the speech signal corrupted by impulse noise is given by

$$s(n) = s_t(n) + d(n), \tag{1}$$

where $s_t(n)$ is a noiseless speech signal and $d(n)$ is an impulsive noise. When the $s(n)$ is used as the input signal of the estimator, the OSLMS algorithm with AS is described as follows.

I Prepare a coefficient matrix $C(n)$ as

$$C(n) = \begin{bmatrix} c_{11}(n) & c_{12}(n) & \cdots & c_{1M}(n) \\ c_{21}(n) & c_{22}(n) & \cdots & c_{2M}(n) \\ \vdots & \vdots & \ddots & \vdots \\ c_{M1}(n) & c_{M2}(n) & \cdots & c_{MM}(n) \end{bmatrix}, \tag{2}$$

and a coefficient vector $\mathbf{c(n)}$ as

$$\mathbf{c(n)} = \left[\begin{array}{cccc} c_{m(1)1}(n) & c_{m(2)2}(n) & \cdots & c_{m(M)M}(n) \end{array} \right]^{\mathbf{T}}. \tag{3}$$

These are initialized as $C(1) = O$ (zero-matrix) and $\mathbf{c(1)} = \mathbf{0}$ (zero-vector).

II Ambiguously sort the elements of the input vector

$$\mathbf{s(n)} = \left[\begin{array}{cccc} s(n-1) & s(n-2) & \cdots & s(n-M) \end{array} \right]^{\mathbf{T}} \tag{4}$$

based on the rule of the AS; that is, if

$$s(n-l-1) > s(n-l) + s_{diff}(n) \tag{5}$$
$$\text{for } \left\{ \begin{array}{l} k = 1, 2, \ldots, M-1 \\ l = M, M-1, \ldots, k+1 \end{array} \right.$$

is satisfied, then $s(n-l-1)$ and $s(n-l)$ are replaced.

III Assume that the resulting sorted vector is given by

$$\mathbf{x(n)} = \left[\begin{array}{cccc} x_1(n) & x_2(n) & \cdots & x_M(n) \end{array} \right]^{\mathbf{T}}, \tag{6}$$

where $x_1(n) \leq x_2(n) \leq \cdots \leq x_M(n)$. And determine the integer in (3), $m(j)$, as

$$m(j) = i, \quad \text{if } x_i(n) = s(n-j), \tag{7}$$

and actually select the corresponding elements $c_{m(j)j}(n)$ from the coefficient matrix $C(n)$.

IV Update the coefficient vector $\mathbf{c(n)}$ by means of the normalized LMS algorithm [6] as

$$e(n) = s(n) - \mathbf{s(n)^T c(n)}, \tag{8}$$
$$\mathbf{c(n+1)} = \mathbf{c(n)} + \frac{\mu}{\mathbf{s(n)^T s(n)} + \beta} \mathbf{e(n) s(n)}. \tag{9}$$

And insert the elements of the updated coefficient vector $\mathbf{c(n+1)}$ into the coefficient matrix $C(n+1)$.

V Increase the iteration number as $n \to n+1$ (for $n = 1, 2, \ldots, N$) and go to Step II.

To implement the above algorithm, the ambiguity parameter in (5), $s_{diff}(n)$, must be determined. To do this, the following is calculated simultaneously with the above adaptation operation.

$$s_{diff}(n) = \max_{\substack{k=1,2,\ldots,n \\ \tilde{s}(k) \neq 0}} \tilde{s}(k) - \min_{\substack{k=1,2,\ldots,n \\ \tilde{s}(k) \neq 0}} \tilde{s}(k), \tag{10}$$

$$\tilde{s}(n) = \left\{ \begin{array}{ll} 0 & \text{if } |s(n)| > \rho \max_{k=1,2,\ldots,n-1} |\tilde{s}(k)| \\ s(n) & \text{otherwise} \end{array} \right. \tag{11}$$

where $\tilde{s}(n)$ is a temporary sample and ρ is a scaling parameter. When $n = 1$, $\tilde{s}(n) = s(n)$. It is expected that for the above OSLMS algorithm with AS, the sorting of the input vector elements is not implemented as long as the input signal $s(n)$ has an amplitude within that of the noiseless signal. Therefore, only the input signal corrupted by an impulsive disturbance will be sorted and as a result, effectively the effect of impulse noise will be suppressed. The above OSLMS algorithm with AS reduces to the OSLMS algorithm in the case of $s_{diff}(n) = 0$ for all n. On the other hand, in the case of $s_{diff}(n) = +\infty$ for all n, it reduces to the normalized LMS algorithm [6].

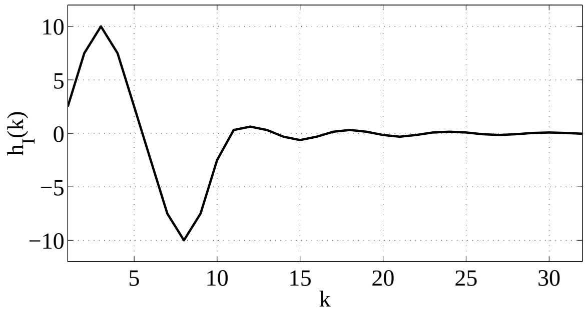


Figure 1: Impulse shaping filter.

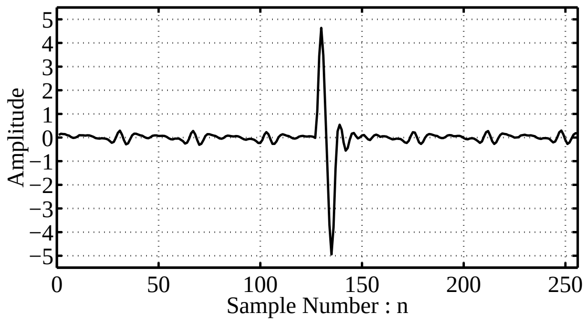


Figure 2: An example of speech signal corrupted by impulsive noise on a real female vowel /a/ (in the case where the impulse with SINR $= 10\,$[dB] appears at $n = 128$).

3 Experiments

We conducted experiments to confirm the effectiveness of the proposed OSLMS algorithm with AS.

It was assumed that the impulse noise $d(n)$ in (1) is generated based on the *Bernoulli Gaussian* model [2] as

$$d(n) = \sum_{k=0}^{L-1} h_I(k)\left\{g(n-k)b(n-k)\right\}, \quad (12)$$

where $g(n)$ is white Gaussian noise and $b(n)$ is a binary random sequence with the values of 1 or 0. The impulse response of the impulse shaping filter, $h_I(n)$, is shown in Figure 1 where L is set to $L = 32$ (this setting is commonly used in the experiments). And, we assumed that a speech signal as shown in Figure 2, for example, is analyzed. This means that an impulse appears as the disturbance *only once* (*not one sample*) at a random time n in each analysis frame.

The signal-to-impulsive-noise-ratio (SINR) is defined by

$$\text{SINR} = 10\log_{10}\frac{P_{\text{signal}}}{\alpha\cdot P_{\text{impulse}}}, \quad (13)$$

where P_{impulse} is the average power of each impulse component, and P_{signal} is the noiseless signal power. The α is defined as $\alpha = N_I/N$, where N is the frame length, and N_I is the number of the signal samples contaminated by impulsive noise in the frame.

We use a pair of male and female 5 vowels /a,i,u,e,o/, and all of the vowels are commonly sampled with 10 [kHz]. And, the predictive order M, frame length N, step size μ and stabilized parameter β employed by normalized LMS in (4) are set to $M = 10$, $N = 256$ (25.6 [ms]), $\mu = 1.0$, and $\beta = 0.05$, respectively. In addition, it is assumed $N_I = L$, then $\alpha = L/N = 0.125$.

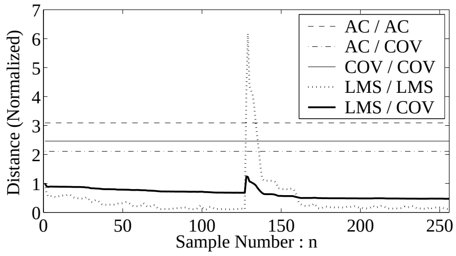


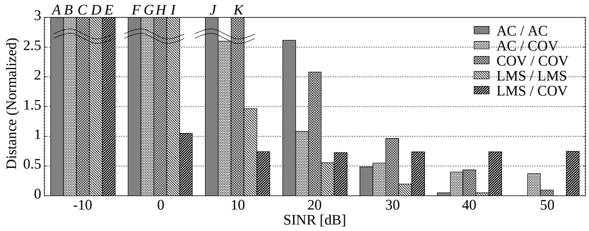
Figure 3: Comparison of the batch and adaptive methods on a real female vowel /a/ (in the case where the impulse with SINR $= 10\,$[dB] appears at $n = 128$). The AC, COV and LMS correspond to the auto-correlation, covariance and LMS methods, respectively. The mark of method$_{(\hat{\mathbf{c}})}$ / method$_{(\mathbf{c})}$, for example AC / COV, means that the prediction is conducted by the method$_{(\hat{\mathbf{c}})}$ and the true one is obtained by the method$_{(\mathbf{c})}$.

3.1 Performance Comparison of Batch and Adaptive Methods

At first, in an impulsive noise environment, we implemented the auto-correlation method and the covariance method, both of which are batch processing, and compared them with the LMS method, adaptive processing. The performance of the three methods on a real female vowel /a/ is shown in Figure 3 where the distance given by

$$d = \frac{\|\mathbf{c}-\hat{\mathbf{c}}\|^2}{\|\mathbf{c}\|^2} = \frac{(\mathbf{c}-\hat{\mathbf{c}})^T(\mathbf{c}-\hat{\mathbf{c}})}{\mathbf{c}^{\mathbf{T}}\mathbf{c}}, \quad (14)$$

where $\mathbf{c}$ and $\hat{\mathbf{c}}$ are the true (noiseless case) and estimated coefficient vectors, respectively, is measured. The impulsive disturbance in Figure 3 was generated at $n = 128$. From Figure 3, we see that both batch methods have been affected by the impulsive disturbance, while the adaptive method has behaved robustly against it. The coefficient vector obtained by the adaptive method at the end of the analysis frame is more accurate than both batch methods.



$^{\dagger}A = 23.4,\ B = 6.69,\ C = 10.6,\ D = 36.3,\ E = 4.62,\ F = 17.4,\ G = 4.95,\ H = 5.88,\ I = 4.41,\ J = 8.48,\ K = 3.51.$

Figure 4: Comparison of the batch and adaptive methods on male and female 5 vowels (average). When the value of d is larger than 3.0, each is denoted by $A, B, C, \ldots$. The exact value of each is shown below the corresponding figure.

Figure 4 shows the dependency of the three methods on SINR where each plot has been obtained by the average of 500 individual trials based on noise generation. The evaluation is the distance in (14) for the coefficient vector obtained at the end of each analysis frame. From Figure 4, we see again that the adaptive method is more accurate than the batch methods. In the case of SINR ≥ 30 [dB], both batch methods are superior. This is because the batch methods provide more accurate estimates than the adaptive method in a noiseless environment. This means that at a low SINR, the adaptive method works more effectively than the batch methods.

3.2 Performance Comparison of LMS, OSLMS and OSLMS with AS

Next, the performance of the adaptive method is further investigated. Three adaptive algorithms, LMS, OSLMS and OSLMS with AS, are compared. The prediction error by each algorithm is evaluated and a signal-to-noise-ratio (SNR) given by

$$\text{SNR [dB]} = 10\log_{10}\frac{\sum_{n=N-\nu}^{N}\{s(n)\}^2}{\sum_{n=N-\nu}^{N}\{s(n)-\hat{s}(n)\}^2} \quad (15)$$

(The N and ν are defined as $N = 256,\ \nu = 100$, respectively) is measured. In addition, the scaling parameter ρ shown in (11) is set to $\rho = \sqrt{2}$. Figure 5 shows an example of convergence in an impulsive noise environment. It is observed that the influence by the impulse disturbance is suppressed by the OSLMS algorithm with AS effectively.

Figure 6 shows the SNR obtained by the three adaptive algorithms in impulsive noise environments where

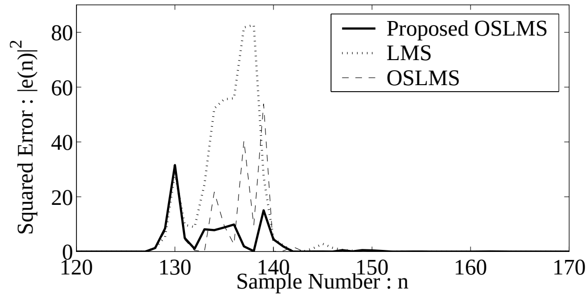


Figure 5: Convergence of the three adaptive algorithms on a real female vowel /a/ (in the case where the impulse with SINR $= 10$ [dB] appears at $n = 128$).

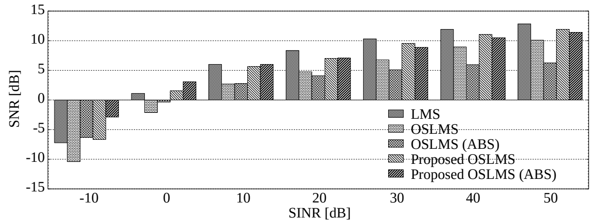


Figure 6: Comparison of the three adaptive algorithms on male and female 5 vowels (average). The ABS means that for sorting, an absolute value version of $\mathbf{s(n)}$, $\mathbf{s(n)} = [\ |\mathbf{s(n-1)}|\ \ |\mathbf{s(n-2)}|\ \cdots\ |\mathbf{s(n-M)}|\ ]^{\mathbf{T}}$, is used.

500 individual trials are averaged for each plot. From Figure 6, we see that in the region of SINR ≤ 10 [dB], the OSLMS algorithm with AS provides better SNR than the LMS and OSLMS algorithms. This result is obviously obtained by the effect of the impulsive noise suppression. Also, we see from Figure 6 that the sorting operation based on the absolute value of each element of the input vector provides an improvement commonly for the OSLMS and OSLMS with AS. Impulsive noise has plus or minus values. By taking account of absolute values of noisy speech samples, the impulsive noise components included would be moved to $x_M(n)$ or around of $x_M(n)$ in the sorted vector $x(n)$ constantly. This may be the reason why.

3.3 Discussion

Summarizing the experimental results, we can say that the OSLMS algorithm with AS provides the best performance in the region of SINR ≤ 10 [dB], while the batch processing should be used in the region of SINR ≥ 30 [dB]. A further performance improvement of the OSLMS algorithm with AS is obtained by involving the absolute value operation in sorting.

4 Conclusions

In this paper, we investigated the performance of LPC in impulse noise, and suggested that the OSLMS algorithm with AS works effectively in a highly impulsive noise environment.

References:

[1] S.Vaseghi and P.Rayner, “Detection and suppression of impulsive noise in speech communication systems”, Proc. IEE, Part-I, Communications, Speech and Vision, vol.137, no.1, pp.38-46, 1990.

[2] S.Vaseghi and B.Milner, “Speech recognition in impulsive noise”, Proc. International Conference on Acoustics, Speech, and Signal Processing, pp.437-440, 1995.

[3] M.Sambur and N.Jayant, “LPC analysis / synthesis from speech inputs containing quantizing noise or additive white noise”, IEEE Trans. Acoustics, Speech and Signal Processing, vol.ASSP-24, no.6, pp.488-494, 1976.

[4] J.Tierney, “A study of LPC analysis of speech in additive noise”, IEEE Trans. Acoustics, Speech and Signal Processing, vol.ASSP-28, no.4, pp.389-397, 1980.

[5] Shimamura T. and Arima Y., “Spectral estimation of voiced speech based on system identification”, Proc. ISCIE Int. Symp. Stochastic Systems Theory and Its Appications, pp.307-314, 1999.

[6] S.Haykin, “Adaptive Filter Theory”, 4th edition, Prentice Hall, 2001.

[7] T.Shimamura and H.Hayakawa, “Adaptive nonlinear prediction based on order statistics for speech signals”, Proc. European Conference on Speech Communication and Technology, Budapest, pp.347-350, 1999.