2015 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) November 9-12, 2015

Transient Noise Reduction for Hearing Aid

Yasuhito Tademoto, Naoto Sasaoka and Yoshio Itoh
Graduate School of Engineering
Tottori University
4-101 Koyama-minami, 680-8552 Tottori, Japan
tademoto@dacom2.ele.tottori-u.ac.jp, sasaoka@eecs.tottori-u.ac.jp

Abstract— **This paper proposes speech enhancement to reduce transient noise. The transient noise is composed of a wideband component which is attack of the noise, and a narrowband component due to a natural oscillation of an object. The duration of a narrowband component is long. Unfortunately, it affects a conversation more than the wideband component. Therefore, we proposed the transient noise reduction for speech. It uses a backward linear predictor using the adaptive algorithm with 4th order cumulant. However, there is a problem that the convergence rate is slow. We propose a RLFC (Recursive least 4th order cumulant) algorithm that convergence rate is faster than the previous one.**

Keywords—speeech enhancement; transient noise; Linear prediction; 4th order cumulant;

I. INTRODUCTION

When we use communication devices such as a hearing aid and smart phone, background noise prevents us from having a conversation. The short time spectral amplitude analysis (STSA) is known for speech enhancement, for example spectral subtraction [1] and minimum mean square estimation (MMSE)-STSA [2]. The conventional speech enhancement is effective in stationary noise. On the other hand, transient noise is generated by hitting an object. The transient noise represents a wideband component at the attack of noise, and then moves onto a narrowband component due to the natural oscillation when the transient noise is generated. Since the duration of the narrowband component is too long, the narrowband noise influences speech conversation critically. Unfortunately, the conventional STSA-based methods cannot reduce the transient noise in real-time processing effectively.

Speech enhancement for transient noise has been proposed in recent years. The literature [3] proposed speech enhancement based on zero phase signals. Sugiyama has proposed speech enhancement with noise detection based on phase information [4]. These methods are effective in the transient noise whose duration is shorter than the duration of speech. The speech enhancement based on GMM (Gaussian mixture models) and database was proposed for a speech recognition system [5]. Unfortunately, its computational load is too high for a hearing aid, and the method is required to be off-line processing.

We have proposed a normalized least 4th order cumulant (NLFC) algorithm with BLP (Backward linear prediction) [6] for reducing the narrowband component which influences a conversation critically. This method makes use of the kurtosis characteristic. In general, Although speech is non-Gaussian [7] with low kurtosis, the transient noise is high kurtosis than speech [8]. The NLFC algorithm converges such that the output and prediction error respectively come out the high and low kurtosis. Therefore, a linear predictor with the NLFC algorithm can estimate only the transient noise. The NLFC takes advantages of a steepest descent method. Steepest descent method can reduce the amount of operations, but cannot avoid that its convergence rate becomes slow. For improving the convergence rate, a linear predictor with a backward structure is needed. However the BLP causes the delay. The delay should be avoided for a hearing aid.

In this paper, we propose a recursive least 4th order cumulant (RLFC) algorithm for improving the convergence rate. Proposed algorithm can utilize information contained in the all input data when tap coefficients are updated. RLFC algorithm can converge faster and improve the noise reduce ability without delay.

II. PRINCIPLE OF SPEECH ENHANCEMENT

A. 4th order cumulant of transient noise

The transient noise is suddenly occurred due to a crash. Its duration is very short, and its frequency characteristic changes fast. For these reasons, the conventional STSA-based methods cannot reduce the transient noise in real-time processing effectively.

The waveform and histogram of the transient noise are shown in Fig. 1. From Fig. 1(a), the transient noise is attenuated with time. Its histogram is shown in Fig. 1(b). The histogram shows that the transient noise is high kurtosis non-Gaussian. Thus, this paper assumes that the transient noise is excited by a linear filter whose an input signal is non-Gaussian white noise. The waveform and histogram of speech are respectively illustrated in Fig. 2(a) and (b). It can be seen that the kurtosis of speech is lower than that of transient noise. Since the kurtosis is represented as 4th order cumulant with no delay[9], the adaptive algorithm for reducing the transient noise makes use of 4th order cumulant.

978-1-4673-6499-7/15/$31.00 ©2015 IEEE

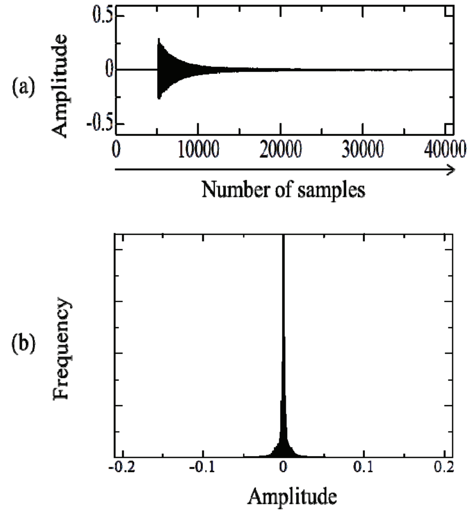

Fig. 1. Histogram of transient noise. (a)Waveform (b)Histogram of amplitude

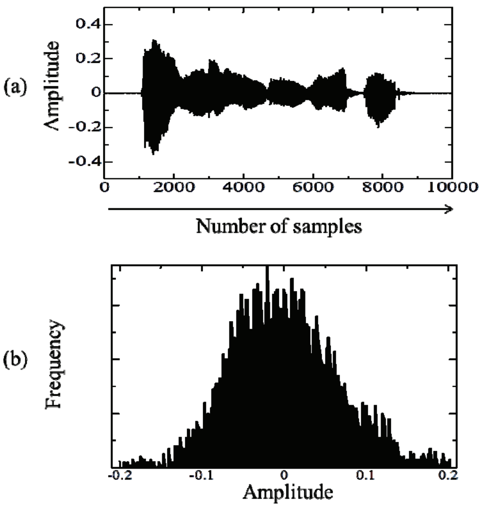

Fig. 2. Histogram of speech. (a)Waveform (b)Histogram of amplitude

B. Linear predictior with NLFC algorithm

Noisy speech $x(n)$ at time n which is composed of clean speech $s(n)$ and transient noise $\xi(n)$ is represented as

$$x(n) = s(n) + \xi(n) \tag{1}$$

The transient noise is estimated by

$$\hat{\xi}(n) = \mathbf{x}^T(n)\mathbf{h}(n) \tag{2}$$

$$\mathbf{x}(n) = [x(n-1),\ x(n-2), \cdots, x(n-M)]^T$$

$$\mathbf{h}(n) = [h_1(n),\ h_2(n), \cdots, h_M(n)]^T$$

where $h_j(n)\,(j = 1, 2, \cdots, M)$ is the j-th tap coefficient and M is the number of tap coefficients. T denotes transposition.

The enhanced speech is obtained by

$$\hat{s}(n) = x(n) - \hat{\xi}(n)\,. \tag{3}$$

The linear predictor estimates the transient noise only. Thus, an NLFC algorithm is required to converge such that the output and prediction error respectively come out the high and low kurtosis. Assuming that $E[\hat{s}(n)]$ is zero, the cost function is defined by

$$\left|Cum_4^{\hat{s}}\right| = \left|E[\hat{s}^4(n)] - 3E[\hat{s}^2(n)]^2\right| \tag{4}$$

where $E[\cdot]$ and $|\cdot|$ respectively denote an expectation operator and an absolute value. The first order partial derivative of (4) is obtained by

$$\begin{aligned} \left.\frac{\partial\,|Cum_4^{\hat{s}}|}{\partial \mathbf{h}(n)}\right|_{\mathbf{h}(n)=\mathbf{h}_0} &\approx 8\,\mathrm{sgn}(Cum_4^{\hat{s}})E[\hat{s}^3(n)\mathbf{x}(n)] \\ &= 8\,\mathrm{sgn}(Cum_4^{\hat{s}})\{\mathbf{P} - \mathbf{R}\} \\ &= \mathbf{0} \end{aligned} \tag{5}$$

where sgn $(\cdot)$ denotes the sign of $(\cdot)$. $\mathbf{p}$ and $\mathbf{R}$ are defined by

$$\mathbf{p} = E[\hat{s}^2(n)x(n)\mathbf{x}(n)] \tag{6}$$

$$\mathbf{R} = E[\hat{s}^2(n)\mathbf{x}(n)\mathbf{x}^T(n)]\,. \tag{7}$$

Thus, the normal equation is represented as

$$\mathbf{R}\mathbf{h}_0 = \mathbf{p}\,. \tag{8}$$

In order to obtain the tap coefficient vector from (8), [6] proposes a NLFC adaptive algorithm. However, its convergence rate decreases. Therefore, the backward linear predictor (BLP) is required to enhance the speech even though the delay occurs [6].

III. Recursive Least 4th Order Cumulant - based Adaptive Algorithm

The normalized least 4th order cumulant algorithm uses steepest descent method for searching the optimal tap coefficient. A steepest descent method can reduce the computational load, but becomes low convergence rate. Thus, we adapt the recursive implementation as the least 4th order cumulant based algorithm. Proposed algorithm can utilize information contained in the all input data for updating the algorithm. Therefore, RLFC algorithm can realize higher convergence rate than NLFC algorithm. From (8),

$$\mathbf{h}_0 = \mathbf{R}^{-1}\mathbf{p}\,. \tag{9}$$

is given. By using deterministic autocorrelation function to (7)

$$R(n;k,m)=E[\hat{s}^2(n)\sum_{i=1}^{n}x(i-m)x(i-k)+c\delta_{mk}] \quad (10)$$

where c and δ_{mk} are a positive constant and kronecker delta as follows:

$$\delta_{\mathrm{mk}}=\begin{cases}1, & m=k\\ 0, & m\neq k\end{cases} \quad (11)$$

Separating the term of $x(n-m)x(n-k)$ from (10), we represents

$$R(n;k,m)=R(n-1;k,m)+\hat{s}^2(n)x(n-m)x(n-k)\,. \quad (12)$$

From (12), a recursive equation to update $\mathbf{R}(n)$ is defined by

$$\mathbf{R}(n)=\mathbf{R}(n-1)+\hat{s}^2(n)\mathbf{x}(n)\mathbf{x}^T(n)\,. \quad (13)$$

Using the formula of Sherman-Morrison-Woodbury, (13) is represented as

$$\mathbf{R}^{-1}(n)=\mathbf{R}^{-1}(n-1)-\frac{\mathbf{R}^{-1}(n-1)\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{R}^{-1}(n-1)}{\hat{s}^{-2}(n)+\mathbf{x}^T(n)\mathbf{R}^{-1}(n-1)\mathbf{x}(n)}\,. \quad (14)$$

The gain vector $\mathbf{k}(n)$ and $\mathbf{T}(n)$ are defined by

$$\mathbf{T}(n)=\mathbf{R}^{-1}(n) \quad (15)$$

$$\mathbf{k}(n)=\frac{\mathbf{T}(n-1)\mathbf{x}(n)}{\hat{s}^{-2}(n)+\mathbf{x}^T(n)\mathbf{T}(n-1)\mathbf{x}(n)}\,. \quad (16)$$

Using these definitions, we rewrite (14) by

$$\mathbf{T}(n)=\mathbf{T}(n-1)-\mathbf{k}(n)\mathbf{x}^T(n)\mathbf{T}(n-1)\,. \quad (17)$$

Multiplying by $\mathbf{x}(n)$ on both sides of (17),

$$\mathbf{k}(n)=\hat{s}^2(n)\mathbf{T}(n)\mathbf{x}(n) \quad (18)$$

is obtained. For the data length n, the element of $\mathbf{p}(n)$ is

$$p(n;k)=E[\hat{s}(n)x(n)\sum_{i=1}^{n}\hat{s}(i-M)x(i-k)]\,. \quad (k=0,1,\ldots,M-1) \quad (19)$$

Isolating the term of $\hat{s}(n)x(n-k)$ from (19),

$$p(n;k)=p(n-1;k)+\hat{s}(n)x(n)\,x(n-k) \quad (20)$$

Thus, recursive equation for updating $\mathbf{p}(n)$ is obtained by

$$\mathbf{p}(n)=\mathbf{p}(n-1)+\hat{s}^2(n)x(n)\mathbf{x}(n)\,. \quad (21)$$

Substituting (15) and (21) in the right-hand side of (9),

$$\begin{aligned}\hat{\mathbf{h}}_0(n)&=\mathbf{R}^{-1}(n)\mathbf{p}(n)\\ &=\mathbf{T}(n)\{\mathbf{p}(n-1)+\hat{s}^2(n)x(n)\mathbf{x}(n)\}\\ &=\mathbf{T}(n)\mathbf{p}(n-1)+\hat{s}^2(n)x(n)\mathbf{k}(n)\hat{s}^{-2}(n)\\ &=\mathbf{T}(n)\mathbf{p}(n-1)+\mathbf{k}(n)x(n)\end{aligned} \quad (22)$$

is obtained. Substituting (17) for $\mathbf{T}(n)$ to (22),

$$\begin{aligned}\hat{\mathbf{h}}_0(n)=&\mathbf{T}(n-1)\mathbf{p}(n-1)\\ &+\mathbf{k}(n)\{x(n)-\mathbf{x}^T(n)\mathbf{T}(n-1)\mathbf{p}(n-1)\}\end{aligned} \quad (23)$$

Substituting (9) and (15) to (23), tap-weight vector $\hat{\mathbf{h}}_0(n)$ is defined by

$$\hat{\mathbf{h}}_0(n)=\hat{\mathbf{h}}_0(n-1)+\mathbf{k}(n)\eta(n) \quad (24)$$

where $\eta(n)$ is

$$\eta(n)=x(n)-\mathbf{x}^T(n)\hat{\mathbf{h}}_0(n-1)\,. \quad (25)$$

RLFC algorithm is summarized in Table Ⅰ.

TABLE Ⅰ. Summary of RLFC algorithm

Initialize the RLFC algorithm by setting
$\mathbf{R}(0)=c\mathbf{I}$, $\mathbf{T}(0)=c^{-1}\mathbf{I}$, c = small positive constant
$\mathbf{h}(0)=\mathbf{0}$
For each instant of time, $n=1,2,\ldots$, compute
$\mathbf{k}(n)=\frac{\mathbf{T}(n-1)\mathbf{x}(n)}{\lambda\hat{s}^{-2}(n)+\mathbf{x}^T(n-1)\mathbf{x}(n)}$
$\eta(n)=x(n)-\mathbf{x}^T(n)\mathbf{h}_0(n-1)$
$\mathbf{h}_0(n)=\mathbf{h}_0(n-1)+\mathbf{k}(n)\eta(n)$
$\mathbf{T}(n)=\lambda^{-1}\{\mathbf{T}(n-1)-\mathbf{k}(n)\mathbf{x}^T(n)\mathbf{T}(n-1)\}$
Where, λ = forgetting coefficien t. If $k(n)>\beta$, updating $\mathbf{T}(n)$

IV. Simulation Experiment

A. Simulation conditions

We confirmed the speech enhancement ability of the FLP using RLFC algorithm. All speech and noise that are used in simulations were sampled by 8 kHz with 16 bit resolution. The female speech was used for this simulation. The computer simulations use the speech of Acoustic Society of Japan-Japanese Newspaper Article Sentence (ASJ-JNAS) [10]. Real World Computing Partnership (RWCP) Sound Scene Database in Real Acoustical Environments [10] was used for database of transient noise. The waveforms of the female speech and transient noise are illustrated in Fig. 3. The input signal to noise ratio (SNR) is expressed by

$$\mathrm{SNR}_{\mathrm{in}}=10\log_{10}\frac{\sum_{j=1}^{N}s^2(j)}{\sum_{j=1}^{N}\xi^2(j)} \quad (26)$$

where N is the number of samples. The $\mathrm{SNR}_{\mathrm{in}}$ was set to -5.0 dB. Table II shows each parameter of the adaptive algorithms.

B. Simulation result

The spectrograms of speech enhancement are illustrated in Fig. 2. The clean speech and noisy speech are respectively illustrated in Fig. 2 (a) and (b). The enhanced speech by BLP with NLFC algorithm and FLP with proposed algorithm

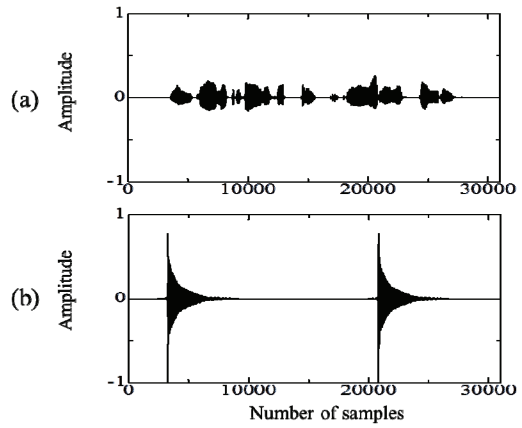

Fig. 3. Waveforms. (a)Speech (b)Transient noise

TABLE II. Parameter of each filter

NLFC algorithm	Number of tap : M	128
	Step size : μ'	0.1
	Threshold : α	0.1
RLFC algorithm	Number of tap : M	128
	Forgetting coefficient : λ	0.99
	A positive constant : c	4.0
	Threshold: β	1.0

are shown in Fig. 2 (c) and (d). Comparing the spectrograms of the conventional NLFC algorithm and the RLFC algorithm, the RLFC algorithm can reduce the reverberation component of the transient noise than the NLFC algorithm. RLFC algorithm's convergence rate is faster than NLFC algorithm. On the other hand, the speech component is maintained in the frequency band where the narrowband noise component exists although the transient noise can be reduced. In addition, the proposed system using a forward linear prediction obtains about the same value of perceptual evaluation of speech quality (PESQ) as the previous system which uses a BLP, although the delay does not occur.

V. Conclusions

The recursive least adaptive algorithm based on 4th order cumulant is proposed for improving the convergence rate. We have proposed the normalized least 4th order cumulant (NLFC) algorithm for estimating the transient noise. However, the adaptive algorithm converges slowly and then the backward linear predictor is adopted for estimating the transient noise faster with tolerance of delay. Therefore, the RLFC adaptive algorithm is proposed. It can be seen that the proposed adaptive algorithm can reduce the noise effectively according to the simulation results. We will research the reduction against the wideband component of transient noise in a future work.

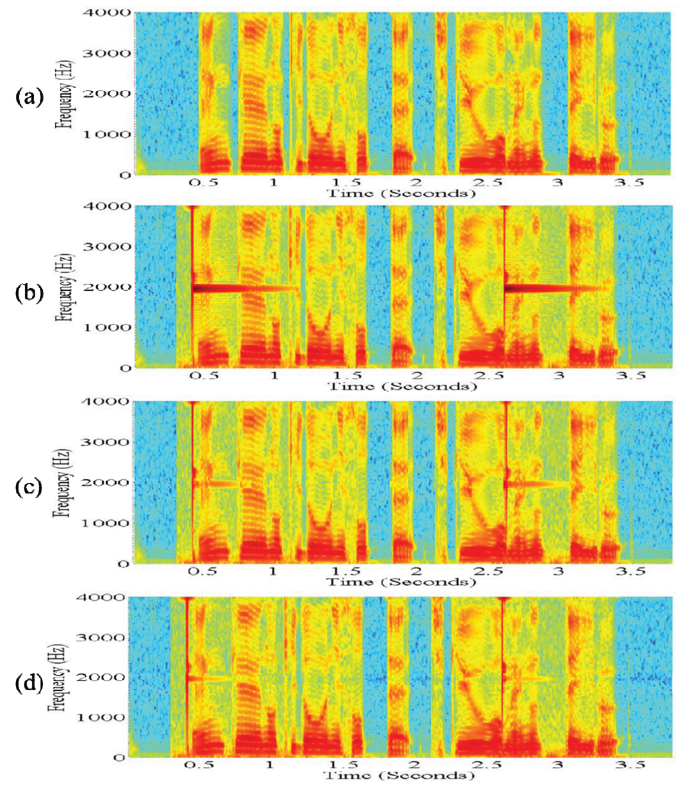

Fig. 4. Spectrograms of enhanced speech.
(a)Clean speech (b)Noisy speech(PESQ=1.94) (c)Enhanced speech by BLP with NLFC algorithm(PESQ=2.67) (d)Enhanced speech by FLP using RLFC (PESQ=2.68)

References

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transl. Acoustic, Speech, Signal Processing, vol. ASSP-27, no.2, pp. 113-120, April 1979.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," IEEE Transl. Acoust, Speech Signal Processing, vol. ASSP-32, no.6, pp.1109-1121, December 1979.

[3] W. Thanhikam, A. Kawamura and Y. Iiguni,"Noise suppression based on replacement of zero phase signal, "Proc. 2011 Int. Symp. Intelligent Signal Prpcess. Comm. Syst., December 2011.

[4] A. Sugiyama, R. Miyahara and P. Kwangsoo, "Impact-noise suppression with phase-based detection," Proc. EUSIPCO2013, September 2013.

[5] N. Miyake, T. Takiguchi and Y. Ariki, "Sudden noise reduction based on GMM with noise power estimation," J. Software Engineering & Applications, Vol.3, No.4, pp.2248-2251, April 2010.

[6] N. Sasaoka, K. Ono and Y. Itoh, "Speech Enhancement Using Normalized Cumulant-Based Adaptive Algorithm for Transient Noise," Proc. 2014 IEEE ISCAS 2014, pp.2313-2316, June 2014.

[7] T.W. Lee, Independent Component Analysis, Kluwer Academic Publishers, Boston, 1998.

[8] T. Yoshimura, F. Asano, H. Asoh and N. Kitawaki, "Investigation of voise/sound activity classifier using distribution models of fourth-order statistics," IPSJ SIG Tec. Report, HI-109, pp.137-142, July 2004.

[9] C.L. Nikias and A.P. Petropulu, Higher-Order Spectra Analysis, Prentice Hall, 1993.

[10] http://research.nii.ac.jp/src/eng/index.html.