



Correlation based Transient Removal Method for Speech Signal Enhancement

Pushpraj Tanwar
ECE Department
MANIT Bhopal, India

Ajay Somkuwar
ECE Department
MANIT Bhopal, India

ABSTRACT

In this article, unwanted transients are specified by correlating the previous shade power standards and eliminated the detected transients. Various studies have been done on the autocorrelation based method to perform the noise reduction in speech signals. The speech signal is a one dimensional signal, for that the correlation may be with its delayed function. The proposed method uses recursive approach and the autocorrelation coefficient as a constraint or stopping criterion. The algorithm solves the transient problem of threshold for transient reduction and provides the alternative. The final simulation modelling shows the results.

Keywords

Transient noise, signal assessment, autocorrelation, correlation coefficient, transient noise assessment

1. INTRODUCTION

The signal generation and its storage for further processing, is done by using different types of sensors. These sensor placed at any location may have some noise. For the real time application it is also affected by the various types of noise. Here we are considering the transient noise and take the step to enhance the quality of signal. Many researchers have work in this field based on the autocorrelation based methodologies. In 1981 Grenier et al. [9] presented a linear prediction technique (LPT) for broadcast, analysis and recognition. Whenever the noise degrades the signal, the autoregressive model is suitable to be biased. This LPT is reduces the bias. The modified Yule Walker equations are used to find the variation of noise. The noise is reduced recursively as a nonlinear approach. In 1991, the evaluation of the speech signal through the cepstral function [7] of autocorrelation is used with the central clipping algorithm. Hu et al. [8] in 2001 offered a spectral subtraction method using the correlation between noise and speech. To enhance the speech quality and minimize the noise distortion, a perceptual weighting function based on correlation was used.

The autocorrelation derived features and phase variations are noise robust. The “Phase Auto Correlation” (PAC) features consist of PAC spectrum is the autocorrelation feature [1]. The phase correlation features are comparatively more noise robust. In another way the sum of squared autocorrelation (SAM) minimization based adaptive channel shortening is used for noise reduction. The single lag autocorrelation minimization (SLAM) algorithm [2] is used to decline noise with shortening signal to noise ratio (SSNR) as a stopping criterion. An advanced spectral subtraction is used to minimise noise, based on cross-correlation between speech and noise signal for the normalization of mean and variance of energy and cepstral parameter [4]. Another correlation

coefficient that is Pearson correlation coefficient based techniques are also introduced. Benesty et al. [10] in 2010 presented the squared Pearson correlation coefficient (SPCC) based cost function for noise reduction technique. The noise decline concerned with the SPCC-based cost function provides a more appropriate criterion for optimization and introduces another parameter compared to the MSE.

The various experiments are done by the researchers on the autocorrelation function. The short term auto-correlation property of speech signal is implemented. The summation of weighted short-term auto-correlation (WSAC) [3] is used by applying the auto-correlation property of current speech frame and frames nearby. Another experiment and evaluation by Abe et al. in 2009, [11] is done with correlation coefficient. It employs correlation coefficient of the filter output and the difference between the input and the filter output as the evaluation function of the parameter setting. The adjustment in the optimal parameter of ϵ -filter is done for the optimization algorithm. The minimum variance distortion less response (MVDR) beam former for multichannel noise reduction [12] is implemented. It uses the noise correlation matrix that computed by noisy input and a noise reference which can be obtained in the absence of speech. The inter-frame correlation of the clean speech signal is used for the optimal FIR filters for speech enhancement in DFT subbands [13]. In 2014 the single-channel speech is estimated in short-time frequency domain. It uses the minimum variance distortion less response (MVDR) [14] approach. These above models drive properly the inter-frame correlation of the speech and noise signals. It requires the investigation of their statistics on a large dataset is computed.

The application of one domain signals in the Global Positioning System (GPS) is very common. The auto (cross) correlation properties of codes used in GPS [5] and effect of noise on code correlation have the great impact on the communication. The navigational satellites allocation and its number increase/decrease are based on code correlation. A comparison of peaks of code correlated signal is done based on the signal to noise ratio. The inter-frame and inter-band correlations are computed with conditional minimum mean squared error (MMSE) for filters. These filters [15] take into account the speech presence uncertainty for each time-frequency unit is used to estimate the clean speech.

The proposed method uses recursive approach and the autocorrelation coefficient as a constraint or stopping criterion. The algorithm solves the transient problem of threshold for transient reduction. The further paper is organized as. The second Sector covers problem formulation with sector III the methodology based on autocorrelation. The

Sector IV contains the trial results and finally segment V concludes precisely.

2. PROBLEM FORMULATION

The speech signal is general a one dimensional signal, for that the correlation may be with its delayed function. The transient noise problem is reduced through the autocorrelation function. The noise reduction is performed by considering constant mean to recover the signal of interest $s(k)$ (clean speech) from the noisy signal observation (microphone signal)

$$S(k) = s(k) + n(k) \quad (1)$$

Where $n(k)$ is the unwanted additive transient noise which is assumed to be a zero-mean random process (white or colored) and uncorrelated with $s(k)$. An estimate of $s(k)$ can be obtained by passing $S(k)$ through a linear filter, i.e.

$$\bar{S}(k) = H^{-1}S(k) \quad (2)$$

$$\bar{S}(k) = H^{-1}(s(k) + n(k)) \quad (3)$$

$$H = [H_0, H_1, \dots, H_{L-1}]^T \quad (4)$$

Representing the finite impulse response (FIR) of length L , superscript T represents the transpose of the vector. With this formulation, the objective of noise reduction is to find an optimal filter that would attenuate the noise as much as possible while keeping the distortion of the clean speech low.

3. METHODOLOGY BASED ON AUTOCORRELATION

Autocorrelation is defined as a measure of how similar a considered signal is to itself with its delayed version. All signals are 100% correlated with their self-version. Hence we leave the peak of autocorrelation function occurring at origin point and find the peak next point. The general informative speech signal is not repetitive in nature. That why the autocorrelation is giving single peak corresponding the zero time lag. The position of second and other peaks are one fourth of the highest peak and other peaks are one tenth of the highest peak for the speech signal. The coefficient for the musical signal is half and one fourth respectively. The frequency of the periodic signal is finding out by considering the first peak on which the autocorrelation have maximum value. This maximum value corresponds to the number of samples (or time duration) after which signal is repeats itself.

$$R(k) = E[s(m).s(m+k)]$$

$$(5) R(k) = \frac{1}{2N+1} \sum_{m=-N}^N s(m).s(m+k) \quad (6)$$

Where $E[.]$ is the statistical expectation which is progressive average, whenever the signal length is large enough or of infinite samples. For speech signal takes on an obvious character of short term stability that it can be regarded to be stationary within 30ms. So in the technique of speech processing, it is general to frame speech with a short window in time domain and regard it as stationary. The autocorrelation

function for the speech signals, using the finite window of length $N-1$, at time index "n", the "k"th lag is:

$$R_n(k) = \sum_{m=0}^{N-1-k} s(n+m)\bar{w}(m).s(n+m+k)\bar{w}(m+k) \quad (7)$$

Where $\bar{w}(m) = w(-m)$ and $\bar{w}(m+k) = w(-m-k)$ the energy of the signal can also be found by taking $k=0$. The frame length considered is 512 and sampling is 8 KHz.

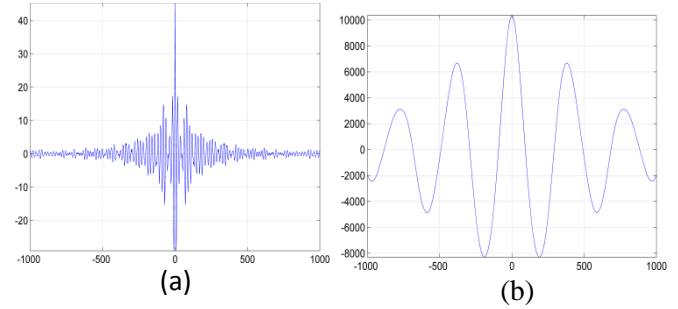


Fig. 1 (a) Autocorrelation of sp09 speech sample signal (b) Autocorrelation of Finger-Bass musical sample signal

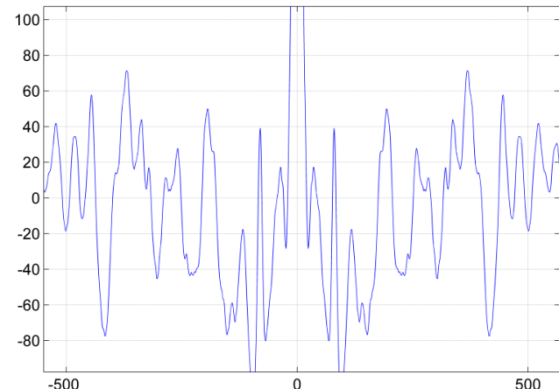


Fig. 2 Signal with cannon transient noise

The analyses from figure 1 is that the speech signals have the second peaks are approximately one fourth of the highest peak. And for the musical signals the second peaks are approximately half of the highest peak. The autocorrelation coefficient is maintained up to 25-35 percent for the speech signal from its second peak at zero delay and up to 45-50 percent for the musical signal except for the zero time delay lag. The iterative median filtering is used for the autocorrelation as constraint. The median filtering smoothing is applied to eliminate speech like noise components of the speech signal. Smoothing can be non-linear as well as linear.

The noise effect on the speech signal is illustrated in figure 2. The autocorrelation spectrum is completely disturbed from its original shape as the peaks other than highest and second peaks are following the second peaks. This property is used to reduce the effect of noise in the speech signal. The complete flowchart of the proposed algorithm is shown in figure 3. In this flowchart the first step is to make a noisy signal by the addition of clean speech signal with transient noise signal. The second step applies the median filtering iteratively. The median filtering may by linear or nonlinear. Generally the

nonlinear filtering or smoothing is preferred for this implementation. The nonlinear smoothing is median smoothing in which L samples of speech are taken at every time instant and the smoothed signal is the median of those L samples.

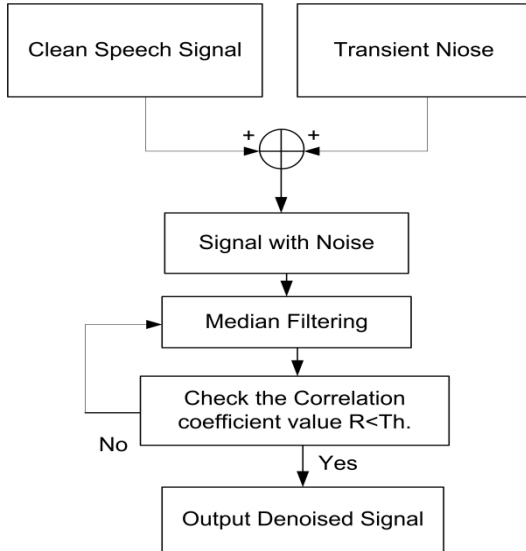


Fig. 3 Flowchart of the autocorrelation based proposed algorithm

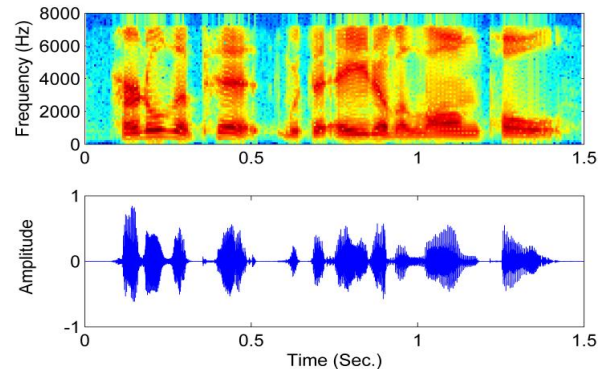
The empirical relation for the threshold (TH.) is given as follows:

$$TH. = \begin{cases} \alpha_1 A_m, & \text{for speech signal} \\ \alpha_2 A_m, & \text{for the musical signal} \end{cases} \quad (8)$$

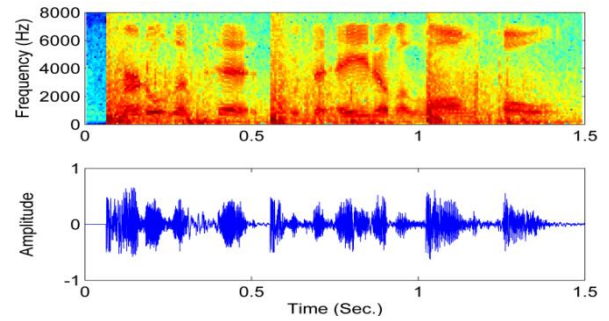
Where α_1 and α_2 are the parameters that are in the range 0 to 1. The appropriate values of $\alpha_1 = 0.35$ and $\alpha_1 = 0.45$ is used for the simulation of the result. A_m represents the second highest peak amplitude of the autocorrelation function. The threshold value from autocorrelation coefficient is used as a constraint for the iterative median filtering.

4. RESULT ANALYSIS

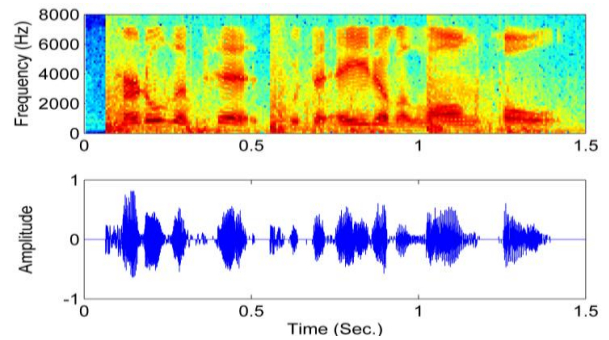
The result obtained by simulation is illustrated through the following mentioned figures. This method is simulated in MATLAB 2012a Version 8.1a with assuming the data is previously recorded for verification of the result. The processor used for this simulation is i3 2040M, 2.3GHz. The original speech signal waveform in time domain and frequency domain is illustrated in figure 4 (a). The spectrum is having no significant component above the 7 KHZ frequency.



(a) Original signal

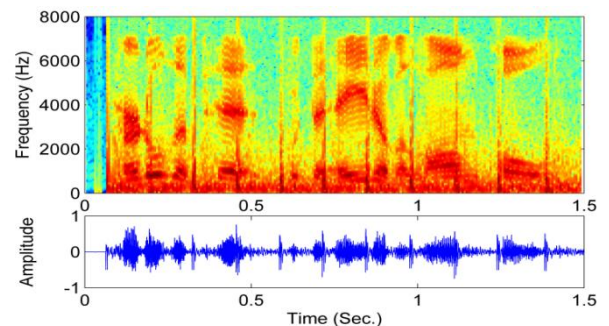


(b) Signal with noise



(c) De-noised signal

Fig. 4 The frequency spectrum and the time domain waveform of the (a). Original clean speech signal, (b). The noisy signal with combination of both clean speech signal and transient noise of cannon, (c) The denoised signal



(a) Noisy Signal

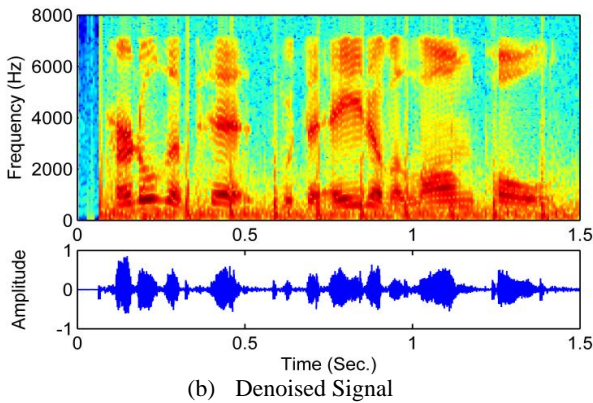


Fig. 5 The frequency spectrum and the time domain waveform of the (a). Noisy signal with combination of both clean speech signal and transient noise of AK-47, (b) the denoised signal

The figure 4 (b) shows the signal with noise and from the spectrum of this combined signal shows that the signal components lies in all the frequency range. Figure 4 (c) clearly illustrates that after filtering the signal, noise is reduced above 7 KHz. The time domain waveform also clears the view to understand the effect of noise removal method. The de-noised signals spectrum is like the original signals spectrum.

The Segmented SNR is computed for both the noisy and denoised signals. The computed Segmented SNR for speech signal is mentioned in table 1.

Table 1 Segmental SNR in dB for the sp09 speech signal with two types of transient noise

Noise component	Input SSSNR(dB)	Output SSSNR(dB)
Cannon noise	-0.1696	2.2069
AK-47 noise	1.0279	3.7341

The original speech signal is same for another noise analysis whose waveform in time domain and spectrum in frequency domain is illustrated in figure 4 (a). The figure 5 (a) shows the signal with AK-47 transient noise and from the spectrum of this combined signal, shows that the signal components lies in all the frequency range above 7KHz also. Figure 5 (b) shows clearly that after filtering the signal, noise is reduced. The denoised signals spectrum is like the original signals spectrum.

5. CONCLUSION

Iterative procedure mean is an ordinarily jumble-sale manner for assessment of unwanted transient power spectrum. The proposed method uses recursive approach and the autocorrelation coefficient as a constraint or stopping criterion. The algorithm solves the transient problem of threshold for transient reduction. The median filtering may be linear or nonlinear. Generally the nonlinear filtering or smoothing is preferred for this implementation. The Implementation of the algorithm is considering 50 samples of speech and musical. Generally this method is more suitable for the speech signals. The Segmented SNR is quite stable for the speech signals. It can be enhanced in the future for up gradation of the method. The further improvements can be done through the various parameters computations.

6. REFERENCES

- [1] S. Ikbul, H. Misra, H. Bourlurd, “Phase Autocorrelation (PAC) Derived Robust Speech Features” IEEE, ICASSP 2003, pp. 134-136.
- [2] R. Nawaz and J. A. Chambers, “A novel single lag auto-correlation minimization (slam) algorithm for blind adaptive channel shortening” IEEE, ICASSP 2005, pp. III- 885-888.
- [3] Z. Shuyin, G. Ying, W. Buhong “Auto-correlation Property of Speech and Its Application in Voice Activity Detection” International Workshop on Education Technology and Computer Science 2009, pp. 265-268.
- [4] G. Faraham, S. M. Ahadi, and M. M. Homayounpour, “Improved autocorrelation based noise robust speech recognition using kernel based cross correlation and overestimation parameters” EUSIPCO 2007, pp. 2355-2359.
- [5] A. Unnisa and A. K. Dwivedi, “Noise Analysis on Auto Correlation of GPS Coarse Acquisition code” ICCNT 2014, pp.178-181.
- [6] C. Nadeu , J. Pascual and J. Hernando “pitch determination using the cepstrum Of the one-sided autocorrelation sequence” IEEE 1991, pp. 3677-3680.
- [7] Y. Hu , M. Bhatnagar and P. C. Loizou, “A cross-correlation technique for enhancing speech corrupted with correlated noise” IEEE 2001, pp. 673-676.
- [8] V. Grenier, K. Bry, J. L. Roux, M. Sulpis, “Autoregressive models for noisy speech signals” IEEE 1981, pp. 1093-1096.
- [9] J. Benesty, J. Chen, and Y. (Arden) Huang, “On the Importance of the Pearson Correlation Coefficient in Noise Reduction” IEEE Trans. on Audio, Speech, and Language Process., vol. 16, no. 4, May 2008, pp. 757-765.
- [10] T. Abe, M. Matsumoto, S. Hashimoto, “Parameter optimization in ϵ -filter for acoustical signal based on correlation coefficient” IEEE 2009, pp. 1417-1420.
- [11] R. C. Hendriks and T. Gerkmann, “Noise Correlation Matrix Estimation for Multi-Microphone Speech Enhancement’ IEEE Trans. on Audio, Speech, and Language Process., vol. 20, no. 1, Jan. 2012, pp. 223-233.
- [12] A. Schasse and R. Martin, “Online inter-frame correlation estimation methods for speech Enhancement in frequency subbands” IEEE, ICASSP 2013, pp. 7482-7486.
- [13] A. Schasse and R. Martin, “Estimation of Subband Speech Correlations for Noise Reduction via MVDR Processing” IEEE/ACM Trans. on Audio, Speech, and Language Processing, vol. 22, no. 9, Sept., 2014, 1355-1365.
- [14] H. Momeni, H. R. Abutalebi and E.A. P. Habets, “conditional mmse-based single-channel speech enhancement using inter-frame and inter-band correlations” IEEE, ICASSP 2016, pp. 5215-5219.