AIP Conference Proceedings

Transient noise suppression algorithm in speech system

Cite as: AIP Conference Proceedings **1864**, 020006 (2017); https://doi.org/10.1063/1.4992823
Published Online: 03 August 2017

Keyu Cao, and Mingjiang Wang

View Online

Export Citation

ARTICLES YOU MAY BE INTERESTED IN

A novel noise-reduction algorithm for real-time speech processing
Proceedings of Meetings on Acoustics **19**, 055036 (2013); https://doi.org/10.1121/1.4800870

Impulse-noise suppression in speech using the stationary wavelet transform
The Journal of the Acoustical Society of America **133**, 866 (2013); https://doi.org/10.1121/1.4773264

Research on adaptive acoustic echo cancellation algorithm in digital hearing aids
AIP Conference Proceedings **1864**, 020173 (2017); https://doi.org/10.1063/1.4992990

© 2017 Author(s).

AIP Publishing

Transient Noise Suppression Algorithm in Speech System

Keyu Cao[a)] and Mingjiang Wang

School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen, China

[a)]caokeyu@stmail.hitsz.edu.cn

Abstract. In this paper, I mainly introduce the algorithm of transient noise suppression in speech system. Firstly, it divides into impulsive noise and other types of transient noise according to the characteristics of transient noise. In the impulse noise suppression algorithm, I mainly use the averaging energy threshold method to detect the impulse noise, and then I use the amplitude threshold method to reduce the impulse noise which was detected. In the other types of transient noise suppression algorithm, I mainly use the Optimally Modified-Log Spectral Amplitude estimation (OM-LSA) algorithm and the Minimum Control Recursive Average (MCRA) algorithm to suppress the transient noise.

INTRODUCTION

Noise has always been a very important issue in our work and life. Noise pollution causes many problems, it will seriously affect our health. High-intensity noise endangers our body, make people feel tired, produce negative emotions, and even cause disease. However, the duration of transient noise is very short but the amplitude is very strong, and a wide spread over the frequency domain with respect to speech phonemes. Such as knocking, keyboard typing, hammering, mouse click, etc. Transient noise is widely found in many occasions, such as hearing aids, mobile phones, interphone and other voice communication terminal equipment. It will affect the speech quality and reduce the intelligibility in the speech system. So it is very important to remove the transient noise.

In general, it is more difficult to eliminate the transient noise. Because most of the transient noise of speech signal is completely overlapping in the time domain. And they have the characteristics of discontinuity. At present, most of the speech noise suppression algorithms are for steady-state noise and continuous noise, such as spectral subtraction, adaptive filtering, Wiener filter and so on. However, these algorithms have little effect on transient noise suppression. Now the algorithm of transient noise suppression mainly adopts the methods of median filtering, model-based impulse noise suppression, and wavelet analysis. Median filtering[1] is a traditional method for impulse noise cancellation. The method is simple, but the median filtering can cause the signal distortion for the noisy signal. Vaseghi[2,3] proposed a method of analyzing and detecting impulse noise based on a linear prediction model. Nongpiur[4] proposed a method of eliminating transient noise in a wavelet transform domain. Jian Li, Shiwei Wang and Renhua Peng[5] proposed a method of Transient Noise Reduction Based on Speech Reconstruction. Although these methods have a certain degree of inhibition on transient noise, but the effect is not very satisfactory, sometimes it can cause speech distortion.

In this paper, I mainly use the method of averaging energy threshold to detect the impulse noise. As for other types of transient noise, I use the Optimally Modified-Log Spectral Amplitude (OM-LSA) algorithm. By considering the prior signal-to-noise ratio and the uncertainty of speech presence, the noise spectrum estimation is performed using Minima Controlled Recursive Averaging (MCRA) algorithm. And we will compute the spectral optimal gain function, then we can find the optimal logarithmic spectral estimation of the clean speech in the frequency domain, finally, we can get the transient noise suppression speech through IFFT.

Green Energy and Sustainable Development I
AIP Conf. Proc. 1864, 020006-1–020006-7; doi: 10.1063/1.4992823
Published by AIP Publishing. 978-0-7354-1542-3/$30.00

PROBLEM FORMULATION

Let $x(n)$ denote a clean speech signal, let $d(n)$ and $t(n)$ be the additive stationary and transient noise signals, respectively. The measured signal $y(n)$ can be given by[6] :

$$y(n) = x(n) + t(n) + d(n) \tag{1}$$

And then we make short-time Fourier transform for $y(n)$, $Y(k,l)$ is obtained by the measured signal $y(n)$ in the time-frequency domain after applying the short-time Fourier transform(STFT) :

$$Y(k,l) = \sum_{n=0}^{N-1} y(n+lM)h(n)e^{-j\frac{2\pi}{N}nk} \tag{2}$$

Where k is the frequency index number, l is the index number of the time domain frame, $h(n)$ is the window function, and M is the frame shift. We denote the formula (1) in the frequency domain as:

$$Y(K,l) = X(k,l) + T(k,l) + D(k,l) \tag{3}$$

Where $X(k,l)$, $T(k,l)$ and $D(k,l)$ are the STFTs of $x(n)$, $t(n)$ and $d(n)$ respectively.

Combined with the above derivation we can get the spectrum of this clean speech signal as shown in equation (4):

$$\hat{X}(k,l) = G(k,l)Y(k,l) \tag{4}$$

Where $\hat{X}(k,l)$ is the estimated value of the speech signal spectrum, $G(k,l)$ is the spectral optimal gain function.

If you can find an optimal spectral gain function $G(k,l)$, we can get the best estimate of the clean speech signal $\hat{X}(k,l)$, then we can get the clean speech signal $\hat{x}(n)$ which we want through the Short - time inverse Fourier transform. The formula can be written:

$$\hat{x}(n) = \sum_{l} \sum_{k=0}^{N-1} \hat{X}(k,l)\tilde{h}(n-lM)e^{j(2\pi/N)K(n-lM)} \tag{5}$$

Where $\hat{x}(n)$ is estimated value of clean speech signal in time domain, $\tilde{h}$ is the double orthogonal window of the Han Ming window h, it is designed for carrying out short-time inverse Fourier transform in advance.

PROPOSED ALGORITHM

The proposed algorithm for transient noise suppression is divided into two steps. First, in Subsection 3.1, we propose an average energy method to solve the impulse noise. Then, in Subsection 3.2, we will introduce the Optimally Modified-Log Spectral Amplitude(OM-LSA) algorithm and Minima Controlled Recursive Averaging(MCRA) algorithm.

Average Energy Algorithm for Impulse Noise

The definition of the Average Energy Algorithm is given by:

$$\text{ave}\_ene = \frac{\sum_{l=1}^{m}\sum_{i=1}^{n}(X_{\mathrm{i}}^{l})^2}{\mathrm{m}} \tag{6}$$

where m represents the number of frames and n represents the number of points within each frame. The X represents the amplitude of the speech signal.

We need to select a threshold V_{th}, and then we need to judge their size between V_{th} and the energy of each frame $E(l)$. In this paper, we let $V_{\mathrm{th}} = 3.2 \cdot ave\_ene$, it is measured by a large number of experiments. When $E(l) > V_{th}$, we think that there is impulse noise in the l-th frame. When $E(l) < V_{th}$, we think that there is no impulse noise in the l-th frame.

Through the above method, we successfully detected the impulse noise, and then we use the method of limiting the amplitude to deal with impulse noise. The formula can be written as:

$$\tilde{X}^{1} = A \cdot X^{l} \tag{7}$$

We know that because of the different selection of coefficient A, it will affect the speech processing effect. In order to avoid this situation, we use an influence function in the selection of the coefficient A. The definition of the function can be written as:

$$A = \begin{cases} \text{ave\_ene}\left(0 < \text{ave\_en} < 1\right) \\ 0.5 \cdot ave\_ene\left(ave\_ene > 1\right) \end{cases} \tag{8}$$

The OM-LSA Algorithm for Transient Noise

The Optimally Modified Log Spectral Amplitude (OM-LSA)Estimation Algorithm

The core idea of the OM-LSA algorithm is to find the minimum variance of the logarithmic spectrum[7], we need to find the minimum value of $E\{(\log A(k,l) - \log \hat{A}(k,l))^2\}$. And then deriving the optimal spectral gain function. Where $A(k,l) = |X(k,l)|$ is the spectrum of clean speech amplitude, then we need to estimate the $\hat{A}(k,l)$. Where $\hat{A}(k,l)$ is defined as:

$$\hat{A}(k,l) = \exp\{E[\log A(k,l) \,|\, Y(k,l)]\} \tag{9}$$

Then we need to estimate the noise spectrum of transient component, and calculate the spectral gain function. Figure 1 depicts a block diagram of the proposed algorithm.

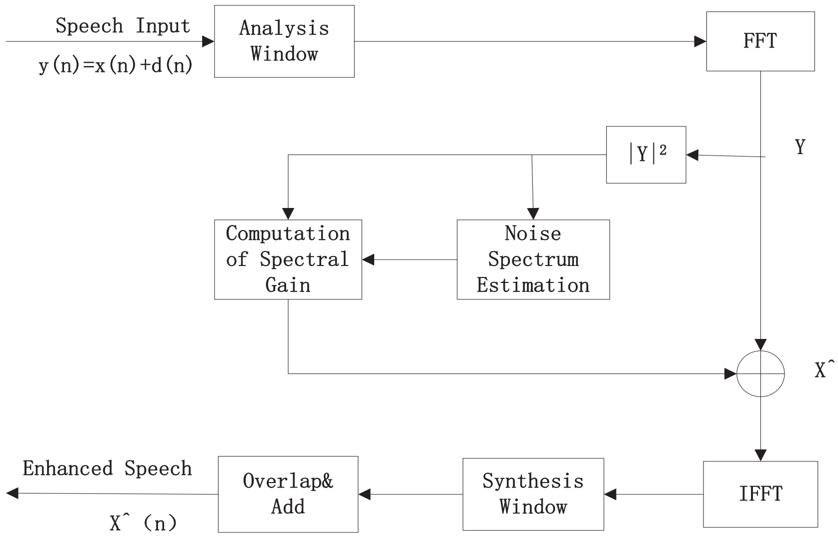

FIGURE 1. The block diagram of OM-LSA algorithm

Transient Noise Spectrum Estimation

We mainly use the MCRA algorithm[8] to estimate the noise spectrum, this algorithm is recursively averaging the previous signal frame, and then to search the minimum values for the previous frame of the signal in a causal window, next to find the noisy frequency point. The PSD estimation based on a smoothed periodogram obtained by a temporal recursive averaging of the spectral amplitude as follows:

$$S(k,l)=\alpha_s S(k,l-1)+(1-\alpha_s)\,|\,Y(k,l)\,|^2 \tag{10}$$

Where $S(k,l)$ is the time-domain smoothing for noise-free speech power spectrum, α_s is the time-domain smoothing coefficient for noise-free speech power spectrum. The value of α_s is generally in the range of 0.7to 0.9.

Next, we search for all the previous frames through a causal window, here the length of the search window L is roughly at 125. The formula is as follows:

$$S_{min}^{L}(k,l)=\min\left\{S(k,l),S(k,l-1),\cdots,(k,l-L+1)\right\} \tag{11}$$

Whether the transient noise point is captured depends on the following equation:

$$S_r(k,l)\equiv\frac{S(k,l)}{S_{min}^{L}(k,l)}\gtrless\delta \tag{12}$$

Where δ is the empirical threshold with noise. When $S_r(k,l)>\delta$, the current position is considered as contain transient noise, otherwise it is considered as a clean speech or background noise. Then we can estimate the instantaneous noise spectrum $\hat{\lambda}_t$.

In this section,we use the output of subsection 3.2.2(the transient PSD estimate $\hat{\lambda}_t$)as an additional input for a second application of the OM-LSA filter as presented in Figure 1.

We let $\xi(k,l)=\dfrac{\lambda_x(k,l)}{\lambda_t(k,l)}$ is a priori signal-to-noise ratio, $\gamma(k,l)=\dfrac{|Y(k,l)|^2}{\lambda_t(k,l)}$ is posteriori SNR. Then $v(k,l)=\dfrac{\gamma(k,l)\xi(k,l)}{1-\xi(k,l)}$, according to the conditional voice existence probability $p(k,l)$, the speech loss probability $q(k,l)$, the priori signal-to-noise ratio $\xi(k,l)$, the posteriori SNR $\gamma(k,l)$, we can derived the spectral gain[9] as follows:

$$G(k,l)=\left\{G_{H_1}(k,l)\right\}^{p(k,l)}G_{min}^{1-p(k,l)} \tag{13}$$

Where $G_{\min}$ is a constant low gain used when speech is absent, $G_{H_1}(k,l)$ is the gain function when the speech is present.[9].

$$G_{H_1}(k,l)=\frac{\xi(k,l)}{1+\xi(k,l)}\exp\left(\frac{1}{2}\int_{v(k,l)}^{\infty}\frac{e^{-t}}{t}dt\right) \tag{14}$$

EXPERIMENTS

We use the Matlab to simulate the above two algorithms, the simulation results are shown in Figure 2 and Figure 3.

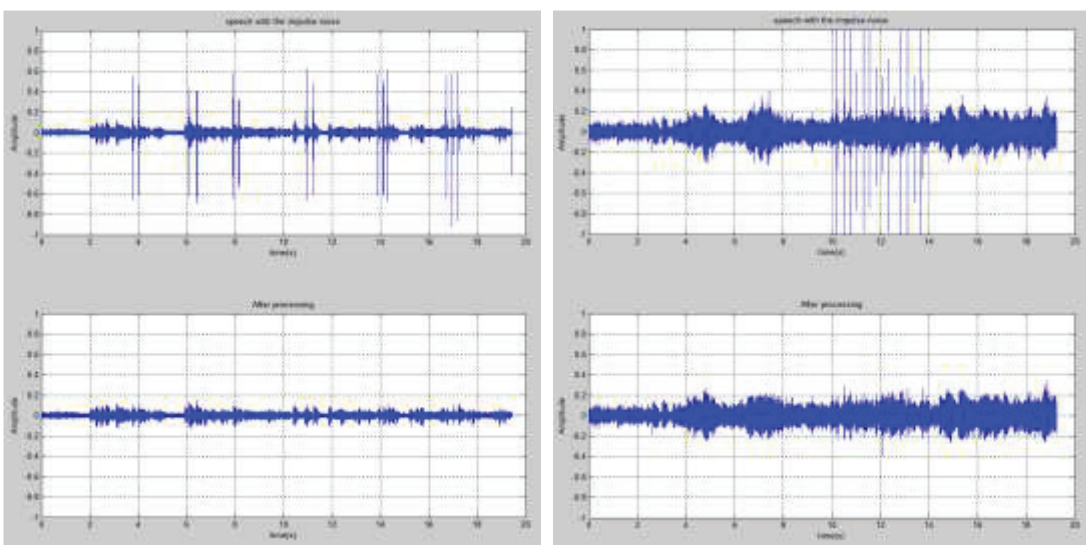

FIGURE 2. The simulation of Average Energy algorithm

In this experiment of impulse noise, we recorded two pieces of audio, one is the music with the impulse noise, another is the speech with impulse noise. We use the Averaging Energy algorithm to deal with the two pieces of audio separately. We can see the experimental results shown in Figure 2. In the left of the chart shows the speech with the impulse noise, In the right of the chart shows the music with the impulse noise, at the top of the chart, it is the waveform before processing, and at the bottom of the chart, it is the waveform after processing. In the figure, the spike is the impulse noise, we can see that it has been effectively suppressed by the Average Energy Algorithm.

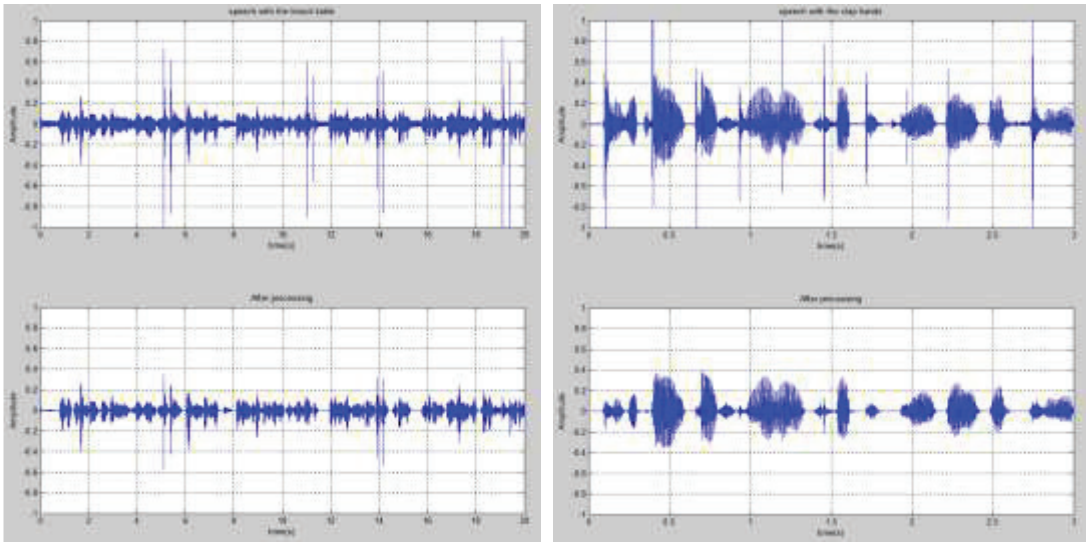

FIGURE 3. The simulation of OM-LSA algorithm

In other types of transient noise, we also recorded two pieces of audio, one is the speech with knock table, and another is the speech with clap hands strongly. Then we use OM-LSA algorithm to deal with these audio. We can see the experimental results shown in Figure 3. In the left of the chart shows the speech with the knock table, in the right of the chart shows the speech with the clap hands, at the top of the chart, it is the waveform before processing, and at the bottom of the chart, it is the waveform after processing. In the figure, we can see that it has been effectively suppressed by the OM-LSA Algorithm.

CONCLUSION

In the experiment of impulse noise treatment, we found that the algorithm of average energy has a great effect on short-term, repetitive, and large impulsive noise. And its calculation and complexity is relatively simple, but the speech quality is not very well. If we cannot select a suitable threshold or limiting factor, then it will affect the speech distortion.

In the OM-LSA algorithm, we recorded the experimental audio is not only a large amplitude of the clapping sound, but also the knocking cup of sound that there is little difference between noise amplitude and speech signal amplitude. Through the simulation results we can see that the OM-LSA algorithm have a great inhibitory effect. When we listen carefully, we will find that the speech quality is also better, at the same time it also greatly improve the intelligibility of speech. And through a lot of experiments, we can find that the algorithm has a better inhibitory effect on all kinds of transient noise, whether single or multiple transient noise can be suppressed. The performance of the algorithm is well verified.

ACKNOWLEDGMENTS

This research was supported by Basic layout of Shenzhen City (no. JCYJ20150827165024088), Supporting platform project in Guangdong Province (no. 2014B0909B001).

REFERENCES

1. Huang Zan. A Median Filter Based on Judging Impulse Noise by Statistic and Adaptive Threshold[J]. Image and Signal Processing, 2008(1): 207-210.
2. S.V.Vaseghi and P.J.W.Rayner. Detection and suppression of impulsive noise in speech communication

system[C]. IEEE Proc: Commun, Speech and Vision, 137(1): 38-46, 1900.
3. S.V.Vaseghi. Advanced Digital Signal Processing and Noise Reduction [M], New York: Wiley, 2006.
4. R.C.Nongpiur. Impulse noise removal in speech using wavelets[C]. IEEE International Conference on Acoustices, Speech and Signal Processing, Las Vegas, March 2008: 1593-1596.
5. Jian Li, Shiwei Wang and Renhua Peng. Transient Noise Reduction Based on Speech Reconstruction[C]. The 21st International Congress on Sound and Vibration, Beijing/China, July 2014: 1-8.
6. Chengshi Zheng,Shiwei Wang,Renhua Peng and Xiaodong Li.Delayless Method to Suppress Transient Noise Using Speech Properties and Spectral Coherence[C]. Audio Engineering Society, New York, USA, October 2013: 1-7.
7. Zhaowei Zhang.Research on Suppression Algorithm of Transient Noise in Speech [D]. Da Lian: Dalian University of Technology, 2013.
8. I Cohen. Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging[C]. IEEE Transactions on Speech and Audio Processing, VOL, 11, September 2003: 466-474.
9. I Cohen and B.Berdugo. Speech enhancement for non-stationary noise environments. Signal Process, vol. 81, no, 11, pp. 2403-2418, Nov. 2001.