On Cause and Cancellation Techniques of Acoustic Echo in Voice Communication

CAI, Ning [a]   DENG, Chun-Lin* [b]

College of Electrical Engineering, Northwest University for Nationalities, Lanzhou, China

[a]caining91@tsinghua.org.cn   [b]876142595@qq.com

Keywords: Acoustic Echo; Voice communication; Cancellation; Adaptive filter; NLMS.

Abstract. Acoustic echo cancellation is common in voice communication. The interference signal caused by acoustic echo will reduce the quality of voice transmission. The current paper discusses the cause and technique for cancellation of acoustic echo, mostly based on the use of adaptive filter. Adaptive filter iteratively alters its parameters to minimize the difference between its output and the actual signal. Morever, this paper makes a comparison of various adaptive filtering algorithms and elaborates the detailed operational process of NLMS filtering algorithm.

Introduction

It has many problems in the technology of voice communication, with the most significant being the echo. Echo is from voice speaker which is transferred by microphone again after being reflected many times in voice communication process. It will seriously reduce the quality of the reproduced voice signal. The acoustic echo cancellation is a key technology to eliminate echo so that the quality of voice in the voice signal transmission process could be improved [1]. This paper mainly addresses on introducing the overall generating mechanism and cancellation principle of echo cancellation. Futhermore, this paper draws a conclusion that NLMS filtering algorithm is more effective and uncomplicated than others after making a comparison of various adaptive filtering algorithms.

Generating Mechanism of Acoustic Echo

The acoustic echo is a phenomenon that the original signal is delayed and anamorphic, and then collected again by microphone. The echo can be divided to two types: the direct echo and the indirect echo. The direct echo is the sound from speaker without any reflection, being directly transmitted into the microphone. The delay time of direct echo is short. The indirect echo is the mixed sounds from the voice speaker which has been reflected many times before transmitted into the microphone together with the original sound. The impulse response of indirect echo will vary quickly and wildly, and the maintained time is longer than direct echo. There is no echo but voice with lower quality when the return time of reflected wave is very short or the amplitude of echo is small. But the echo will be significant when its delay time exceeds tens of milliseconds or amplitude exceeds 30dB [2]. The generating mechanism of acoustic echo is shown in figure 1.
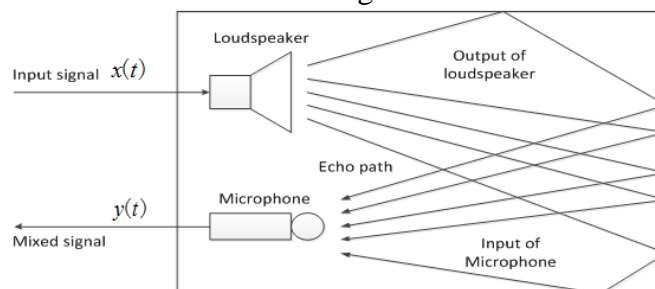

Fig. 1 Generating mechanism of acoustic echo

Principle of Acoustic Echo Cancellation

The principle of acoustic echo cancellation is to build a model which can simulate the actual echo path exactly and the output of this model is the prediction of actual signal. The echo will be eliminated when it subtracts the predicted signal. The most mature method in actual acoustic echo cancellation is to use an adaptive filter to simulate the echo path, with the adaptive filter being the core of echo cancellation. The principle of acoustic echo cancellation is shown in figure 2.
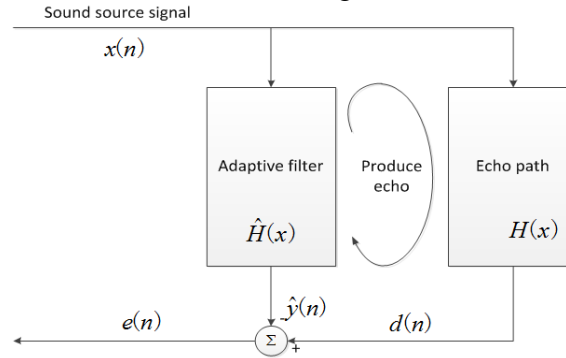

Fig. 2 Principle of acoustic echo cancellation

In figure 2, the echo path $H(x)$ can be treated as a function, and it depends on the actual environment. The adaptive filter can also be treated as a function $\hat{H}(x)$, and it should approach $H(x)$ as close as possible. They should produce the same result when their input is identical. The adaptive filer can adjust its parameters with the actual environmental changes, which means that $\hat{H}(x)$ can keep on simulating $H(x)$. The output of filter is $\hat{y}(n)$. The ideal error $e(n) = d(n) - \hat{y}(n)$ should be zero, with $d(n)$ as the actual output, which implies that the echo has been eliminated thoroughly. The utility of adaptive filter is illustrated in figure 3.
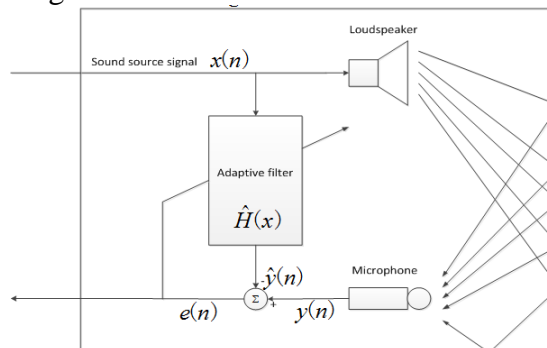

Fig. 3 Function of adaptive filter

The principle of adaptive filter is very simple. Finite impulse response filter (FIR), as a frequently-used filter, consists of multipliers, unit delay cells and adders. It usually represents the predicted value of the future output signal by the linear combination of a period of time signals:

$$\hat{y}(n) = \sum_{i=0}^{N-1} w_i(n)x(n-i)$$

n is the time predicted signal appears, N is the order of adaptive filter, $w_0, w_1, \ldots w_{N-1}$ is N adjustable parameters. The structure of FIR filter is shown in figure 4.
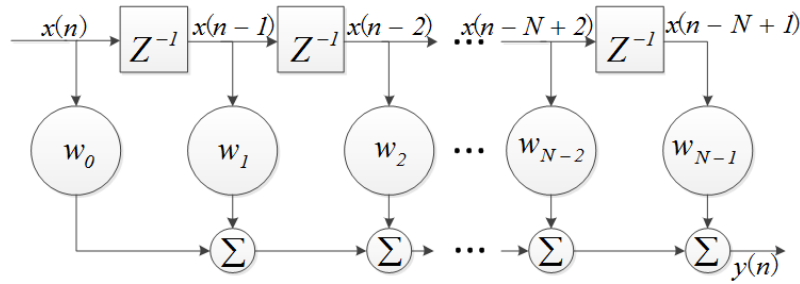
Fig. 4 Structure of FIR filter

The concrete structure of filter depends on the value of N adjustable parameters w_0, w_1, ... w_{N-1}. The adaptive filter can change the value of these parameters in real time, which are usually denoted by a filter parameter vector:

$$W(n) = [w_0(n) \; w_1(n) \; w_2(n) \; ... \; w_{N-1}(n)]^T$$

The value of these parameters depends on the difference between the actual signal and predicted signal. When the actual signal equals to the prediction signal, it means that the filter can simulate the actual echo path greatly. Therefore it is not necessary to update the value of these parameters w_0, w_1, ... w_{N-1}. The echo path will change when the environment changes. Perhaps, the filter will need an adaptive process to adjust the parameter vector until it is appropriate when the echo cancellation system starts to run [3-5].

Actual Acoustic Echo Cancellation System in Engineering

The implementation of acoustic echo cancellation system in engineering is more complex than the mere theoretical principles. An applicable system usually includes a Double Talk Detector (DTD) and a Nonlinear Processor (NLP) in addition. The structure of the actual acoustic echo cancellation system in engineering is shown in figure 5.
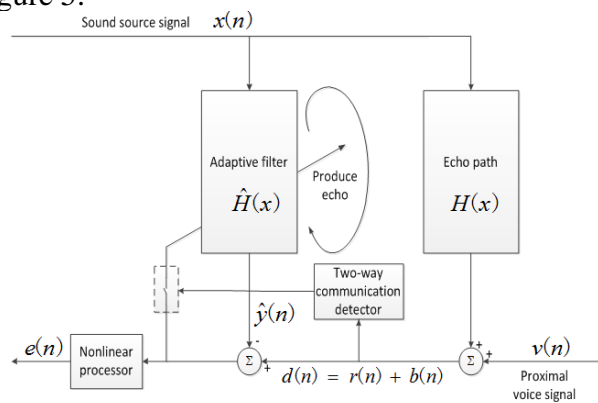

Fig. 5 Integral structure of actual acoustic echo cancellation system in engineering

The operation of adaptive filter will be interfered by the new proximal voice signal when the new proximal voice signal appears suddenly in the course of echo cancellation. That will result in abnormal adjustment of parameter vector $W(n)$. More seriously, it will make the output signal of adaptive filter be chaotic so that the filter is divergent eventually. The effective solution to solve this problem is to detect whether or not new proximal voice signal exists in microphone before updating these parameters w_0, w_1, ... w_{N-1}. The actual solution is to install a two-way communication detector in the input channel of microphone. System will stop to update the parameter vector $W(n)$ when there is a new proximal voice signal in microphone input channel [6].

Generally, the final output signal still needs a nonlinear processor for further polishing, e.g. eliminating noise clutter, eliminating residual echo and adding some background effects, etc.

Choice of Algorithm

The main difference among various algorithms is that the method for updating parameter vector $W(n)$ of each algorithm is different. The least mean square (LMS) algorithm is the most fundamental adaptive filtering algorithm, and it has a very reliable tendency of convergence when the input signal is steady. Meanwhile, it is simpler than other algorithms. It has been proved theoretically that the output signal of adaptive filter based on LMS algorithm can track the actual signal exactly and the error between the output of adaptive filter and the actual echo will gradually become zero after a period of running time when the original signal is a random and steady signal.

According to the data provided in table 1, NLMS algorithm should be the optimal algorithm. Because its average echo attenuation is close to 30dB, and complexity is small. In addition, NLMS algorithm will update all filter coefficients when it calculates a sample value.

Table 1 Comparison of various adaptive filtering algorithms

Name of algorithm	Average echo attenuation	Calculation complexity
LMS	-18.2dB	2N+2
NLMS	-27.9 dB	3N+1
VSLMS	-9.8 dB	4N+1
VSNLMS	-9.9 dB	5N+1
RLS	-34.2 dB	$4N^2$

Normalized Least Mean Square Algorithm

The normalized least mean square algorithm (NLMS) is an extension of least mean square algorithm (LMS). NLMS algorithm can avoid calculating the maximum step length existing in LMS algorithm. The execution steps of NLMS are similar to LMS. Each iteration of NLMS algorithm requires the following steps:

1) *To acquire the output of adaptive filter $y(n)$:*

$$y(n) = \sum_{i=0}^{N-1} w_i(n)x(n-i) = W^T(n)x(n)$$

$y(n)$ is the input data vector with delay time, its value:

$$x(n) = [x(n)\ x(n-1)\ x(n-2)\ \ldots\ x(n-N+1)]^T$$

2) *To calculate the error signal $e(n)$:*

$$e(n) = d(n) - y(n)$$

3) *To calculate the calculation step μ:*

$$\mu(n) = \frac{1}{x^T(n)x(n)}$$

4) *To update the filter parameter vector $W(n)$ with the recursive algorithm and to prepare for next iteration:*

$$w(n+1) = w(n) + \mu(n)e(n)x(n)$$

μ is a known step parameter, and its value should be a small positive constant, which can adjust the update of the parameter vector $W(n)$. The convergence speed depends on the value of μ. This is suitable as the value of μ is very small, and the adaptive filter will converge to the optimal solution in this case. But the adaptive filter will be unstable when the value of μ is too large. More seriously, the output of adaptive filter will be divergent.

Theoretically, the error $e(n) = d(n) - y(n)$ should be zero when there is no communication and echo, but this situation does not exist in actual environment. Therefore, the update of the tap-weights of

the adaptive filter also need input signal $x(n)$ and error signal $e(n)$ so as to achieve the minimum mean square error.

The choice of the step of NLSM algorithm follows the normalized principle:

$$w(n + 1) = w(n) + \frac{1}{x^T(n)x(n)} e(n)x(n)$$

The purpose of controlling the length of step, as a function of error signal, is to let the energy of error signal reflect the extent of convergence. Actually, the NLMS filter can be treated as an LMS filter with variant step length [7-10].

Conclusion

With the extensive application of voice signal transmission, one of the major problems in voice signal transmission is echo. The NLSM echo cancellation algorithm introduced in this paper attempts to find a set of solution for the problem of echo in the voice signal transmission environment. This technique ought to be effective for canceling residual echo without clipping of the reference voice signal.

Acknowledgements

References

[1] L. S. H. Ngia, *System Modeling Using Basis Functions and Application to Echo Cancellation*, Doctoral Dissertation, Chalmers University of Technology, 2000.

[2] S. Furui and M. M. Sondhi, *Advances in Speech Signal Processing*, Marcel Dekker Inc., 1992.

[3] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996.

[4] M. L. Honig, D. G. Messerschmitt, *Adaptive Filters: Structures, Algorithms and Applications*, Springer Verlag, 1984.

[5] J. Homer, *Adaptive Echo Cancellation in Telecommunications*, Doctoral Dissertation, Australian National University (University of Newcastle), 1994.

[6] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer Verlag, 2001.

[7] P. S. R. Diniz, *Adaptive Filtering*, Springer Verlag, 1997.

[8] C.F.N. Cowan, P.M. Grant, P.F. Adams, *Adaptive Filters*, Prentice Hall, 1985.

[9] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*, John Wiley & Sons, 2013.

[10] M. Bellanger, G. Maurice, *Adaptive Digital Filters*, Marcel Dekker Inc., 2001.