Acoustic Echo Cancellation with Double-Talk Detection
DSP Lab Course WS 2017/2018

Suraj Khan (suraj.khan@fau.de), Shrishti Saha Shetu (shrishti.shetu@fau.de) & Raza Azam (raza.ul.azam@fau.de)

Supervisor : Michael Günther

1 Introduction

The need for acoustic echo cancellation arises whenever a loud speaker and a microphone are placed together in nearby vicinity of each other and as a result the signals radiated by the loud speaker and its reflections are picked up by the microphone. As a result, the electroacoustic circuit may become unstable and produce undesirable howling. In order to mitigate this problem attenuation of the acoustic path between the loudspeaker and microphone is essential. A straight forward solution to this problem can be achieved by using directional microphones or arranging the microphones and loudspeakers in such a manner so that do not see each other.

However, as more signal processing is becoming more economical it possible to develop an adaptive filter parallel to the loudspeaker-microphone system. In an ideal case if it is possible to match the impulse response of the loudspeaker-microphone system, then the echoes, generated from the loudspeaker and its reflections, captured by the microphone can be completely suppressed and the local speech signal can be perfectly isolated.

In this project we model the impulse response as a linear system for the sake of simplicity while at the same time having sufficient performance. Since, we have to model a filter with a sufficient long impulse response, IIR filters might appear to be suitable choice at the first glance. However, due to possibility of instability in IIR filters, FIR filter implementation is preferred. In addition, FIR filters allow us to have a large number of adjustable parameters, which aids in obtaining a sufficient replica of the impulse response.

In this project we use the regularized Normalized Least Mean Square(NLMS) algorithm to generate the coefficients of the adaptive filter and to have a control over the filter adaption we couple it with a double talk detector using Magnitude Squared Coherence(MSC) to extract the local speech signal.

2 Algorithm

2.1 Normalized Least Mean Square (NLMS) algorithm

The normalized mean square algorithm[1] is a type of stochastic gradient algorithm. Hence, we can use Wiener-Hopf equations to obtain the solution by the using an iterative application of gradient method. The idea is to obtain the filter coefficients by the application of negated gradient as depicted in 2.1 iteratively on the incoming signal samples.

$$\nabla E\{e^2[k]\} = \frac{\partial E\{e^2[k]\}}{\partial h[k]} \tag{2.1}$$

The incremental update of the filter coefficients can obtained as 2.2

$$h[k+1] = h[k] - \frac{\mu}{2}\nabla E\{e^2[k]\} \tag{2.2}$$

where μ denotes the adaptation size. The advantage of using iterative approach helps us achieving lower complexity that the matrix inversion. In addition, it enables us to identify an unknown time variant system.

However, the estimation of the expectation value of $E\{e^2[k]\}$ is very difficult, hence we substitute it by $e^2[k]$, hence we can correspondingly obtain 2.3 from 2.2 as

$$\nabla e^2[k] \;= \frac{\partial e^2[k]}{\partial h[k]} = 2e[k]\frac{\partial e[k]}{\partial h[k]} = 2e[k]\frac{\partial(y[k] - h^T x[k])}{\partial h[k]} = -2e[k]x[k] \tag{2.3}$$

Now from the above basic assumption and formulation we can obtain the equation for the filter adaptation of the NLMS algorithm as 2.4

$$h[k+1] = h[k] + \frac{\alpha}{||x[k]||^2 + \delta} e[k]x[k] \tag{2.4}$$

The constant δ is important when we consider non-stationary signals, such as speech. Its effect is to limit the step size when $||x[k]||^2$ is near zero i.e. during speech or signal pauses.

2.2 Magnitude Square Coherence (MSC)

In order to detect local speech activity we use the magnitude coherence formulation. The main idea behind using Magnitude Squared Coherence[2] is that, whenever the there is a local speech signal present in the microphone signal, the two signals appear uncorrelated and the function returns a smaller value. The MSC can be calculated as 2.5:

$$|M_{xy}(e^{j\Omega})|^2 = \frac{|C_{xy}(e^{j\Omega})|^2}{C_{xx}(e^{j\Omega})C_{yy}(e^{j\Omega})} \tag{2.5}$$

The term C_{xy} refers to the cross power spectral density (PSD) of two signals X and Y while the term C_{xx} and C_{yy} refers to the PSDs of the signals X and Y respectively. It is imperative to mention that MSC is a normalized function as evident from 2.5 and the function returns a value of 1 for maximum correlation and a value of 0 for uncorrelated signals.

3 Implementation

For implementing the NLMS algorithm we use equation 3.1 with a value of $0 < \mu < 2$:

$$h[k+1] = h[k] + \mu \frac{x[k]e[k]}{x^T[k]x[k] + \delta} \tag{3.1}$$

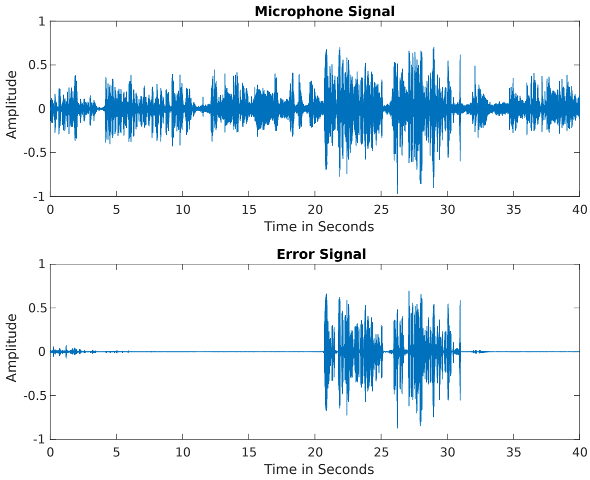
For implementing the MSC function we compute the power spectral densities using the Welch's[**?**] Method with an overlap of 75% for obtaining a smooth PSD estimate. In this project to compute the PSDs we use the pwelch function built in matlab and for computing the cross power spectral density we use the cpsd function.

It must be noted that while implementing the system together i.e the NLMS algorithm with the Double Talk Detection, we had to apply a delay of 125 ms double talk detection and between filter adaptation. The reason being stopping the filter adaptation in realtime considers some portion of the local speech signal and the obtained isolated signal has a possibility to get muffled. However, it must also be noted that if the filter response converges sufficiently we can stop the filter adaptation instantaneously without having to worry about the isolated signal getting muffled. As per our experimental observations with a filter length of 2001 coefficients, 8-10 secs. is sufficient enough for the convergence of the adaptive filter. However, presence of any local speech signal before 8 secs. after initialization requires a delay of approx 100 ms to obtain a fairly good isolated signal.

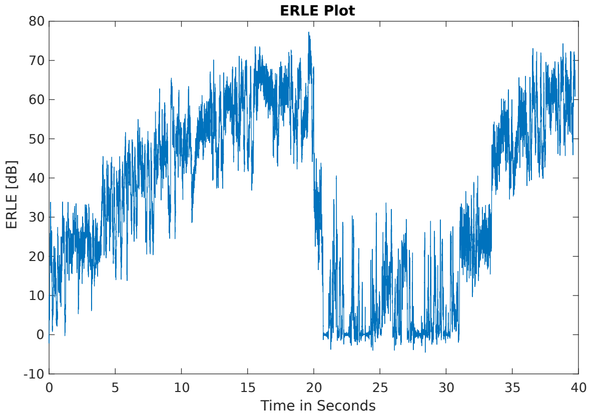
4 Evaluation and Results

For evaluating the system performance we generate several microphone signals with the occurrence of local speech signal at various instants after system initialization. For evaluating the performance we use metrics like system distance measure[3], echo return loss enhancement[4] in addition to simple auditory listening of the obtained error signal. In addition, we also compare the estimated impulse response and true response.

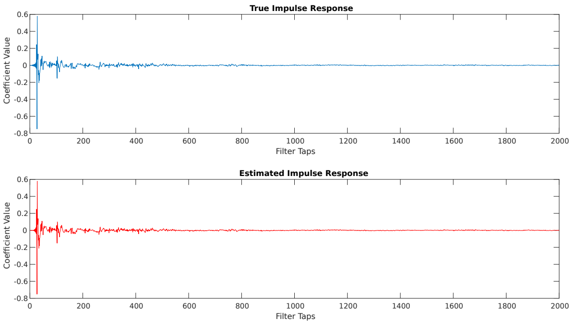
In the following graphs we have mixed the echoed signal with a local speech signal located approximately at 20 seconds.



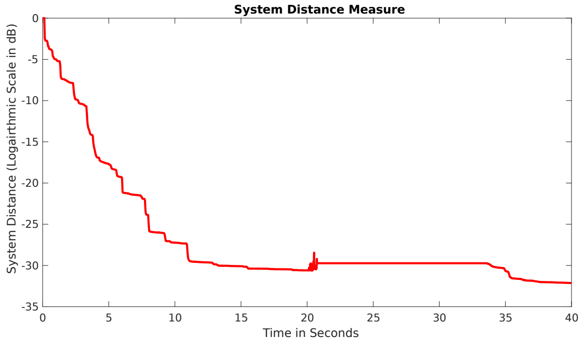
(a) Comparison between microphone signal and error signal



(b) Echo return loss enhancement in dB



(c) Comparison of True and Estimated Impulse Response



(d) System Distance Measure

From figure (a), we can observe that as the filter is adapting there is a little bit of echoes that can be observed and as time passes, the filter is able to suppress echoes to near perfection. Figure (b) depicts the echo return loss enhancement.

However, figure (d), depicts the performance of the system most accurately. It can be observed that as more number of samples are observed, the estimated filter response becomes closer to the true impulse response. In addition, it can also be observed that the system distance measure is constant, which depicts that the double talk detection filter blocks the filter adaptation during the presence of the local speech signal.

References

[1] G.-O. Glentis, K. Berberidis, and S. Theodoridis. Efficient Least Squares Adaptive Algorithms for FIR Transversal Filtering: A Unified View, *IEEE Signal Proc. Magazine*, 1999.

[2] G. Carter, C. Knapp and A. Nuttall, Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing, *IEEE Transactions on Audio and Electroacoustics* vol. 21, no. 4, pp. 337-344, Aug 1973.

[3] Kellermann W., Lecture Notes Digital Signal Processing Laboratory, *Multimedia and Communications and Signal Processing* University of Erlangen-Nuernberg.

[4] M. Rages and K. C. Ho, Limits on echo return loss enhancement on a voice coded speech signal, *The 2002 45th Midwest Symposium on Circuits and Systems*, pp. II-152-II-155 vol.2, 2002.