SPEECH ENHANCEMENT USING KALMAN FILTER

Vishnu Vardhan Dhanabalan

Electrical Engineering Graduate,

Syracuse University.

LIST OF TOPICS

OVERVIEW

WHAT IS A SPEECH SIGNAL?

To give a definition of speech signal in a sentence, it is created at the vocal chords, travels through the vocal tracts and produced at speaker's mouth [1]. Speech signals are different from other audio or music signals because their frequency properties do not stay constant through out the time and changes often in a span of milliseconds. Acoustically, it reaches our ears as pressure waves [1].

NEED FOR SPEECH ENHANCEMENT

In this digital world, it is really hard for any signal in real-time environment to escape from noise. This hits us really hard when it comes to deliver a message from one place to another (that one place, sometimes, would be 20000ft from the ground level), and there is a need for cleaning up or enhancing the message signal but at the same time, not giving up any intelligibility of the message (content, not just clarity). Since speech messages has been the mode of communication everywhere, need for speech enhancement is required whenever the signal comes in contact with the real-time environment. In the upcoming sections, we are going to see how to model a speech signal digitally and methods to enhance it.

SPEECH SIGNAL MODEL

Speech signal can be regarded as Auto-Regressive model, which is essentially the output of an all-pole linear system driven by white noise sequence [2]. Thus, any K^{th} sample of a speech signal is given by:

$$S_{(k)} = \{a_1 S_{(k-1)} + \ldots + a_p S_{(k-p)}\} + u_{(k)}. \quad (1)$$

Here, 'p' represents the autoregressive filter order. 'u' is the excitation signal which is a zero-mean white Gaussian noise and S is the actual speech signal. There are totally three type of speech signals are available.

1) Voiced speech signal – Where we have data or information that we can use.
2) Un-Voiced speech signal – Where we have only noise.
3) Silence – Absence of both data and noise.

We are mainly concerned about the part of speech where data is available, that is, voiced speech signal.

METHODS OF SPEECH ENHANCEMENT

To discuss about methods of speech enhancement, we have to rollback to the past a little bit…

WIENER FILTER (frequency domain)

Speech enhancement in communication domain was an interesting topic back in 1970s. So, Norbert Wiener, who was a child prodigy, came up with a filter by himself. Named as Wiener filter, it produced an estimate of a desired or target random process by linear time-invariant filtering of an observed noisy process, assuming known signal and noise spectra, and additive noise [4].

Weiner filters approaches the problem by estimating the signal as linear sum of previous observed samples. To obtain the coefficients, we need a-priori knowledge of the joint spectral density of the actual signal and the noisy signal, and finally an estimate of noise power spectral density. But in reality, assumptions like noise and the signal are uncorrelated have to be made because it is hard to estimate the joint spectral densities. Secondly, the noise power spectral density is hard to compute if the noise is non-stationary. So assumption has to be made that it is an additive white Gaussian noise (whose PSD is easy to estimate). In real world, noise does not stay stationary so these assumptions may fail.

KALMAN FILTER

Rudolf Kalman invented Kalman filter and he published his work through his famous 1960 paper, “A new approach to linear filtering and prediction problems”. One big advantage of using Kalman filter over using Wiener filter is that, Kalman filter works on non-stationary speech models. The state and observation equations of Kalman filter models that dynamics of the speech signal generation and the

noise and observed signal respectively. *Complete explanation of conventional Kalman filter implementation is available in the next section.*

There are other speech enhancement techniques available which has their own advantages and disadvantages. Spectral subtraction is one of the commonly used technique in which frequency property of noise samples are subtracted from the contaminated signal in frequency domain. This is one of the simplest methods for speech enhancement but noise must be stationary throughout the signal and noise samples are required before starting the process to learn its frequency behavior.

APPROACH

In this project, Kalman filter was used for enhancing speech signal contaminated with zero mean Gaussian noise. The algorithm was referenced from the paper, *'A Speech Enhancement method based on Kalman Filtering*' by K.K. Paliwal and Anjan Basu. This paper, dates back to 1987, is considered widely as the very first approach on Kalman filter based speech enhancing and many papers that followed this one, were heavily based on this paper.

Modeling of the speech signal is essential for Kalman filter and the model has been explained clearly in the first section. So through out this project, speech signal has been considered as Autoregressive model with excitation signal being zero mean Gaussian noise.

Equation (1) can be rewritten as state space model as

$$X_{(k)} = \mathbf{A}\, X_{(k-l)} + G\, U_{(k)} \quad (2)$$

X, A, and G are state vector, state transition matrix and Input matrix respectively. These matrices are defined individually as,

$$X^{T}_{(k)} = [S_{(k-p+l)}, \ldots, S_{(k-l)}, S_{(k)}] \quad (3)$$

$$A = [0, I; -a_p, \ldots, -a_1] \quad (4)$$

$$G = [0, \ldots, 1]_{[1xp]}. \quad (5)$$

When only the noise corrupted signal $Y_{(k)}$ is available, it can be written as,

$$Y_{(k)} = S_{(k)} + n_{(k)} \quad (6)$$

Matrices in eqn. 6 can be rewritten as state matrix as,

$$Y_{(k)} = H\, X_{(k)} + n_{(k)} \quad (7)$$

H and G matrices are similar to each other.

Eqn. 7 and 3 are very important because it clearly suggests that Kalman filter can readily be applied for estimating state vector $X_{(k)}$.

Initializations are required for $X_{(k)}$ and P (error covariance matrix). The first 'p' amounts of input samples are used to initialize first 'p' samples of the output and P matrix is diagonally initialized with the variance value of stationary noise.

The heart of the Kalman filter follows the initializations, as explained in equations from (15) to (19) in the above-mentioned paper.

ATTEMPTS

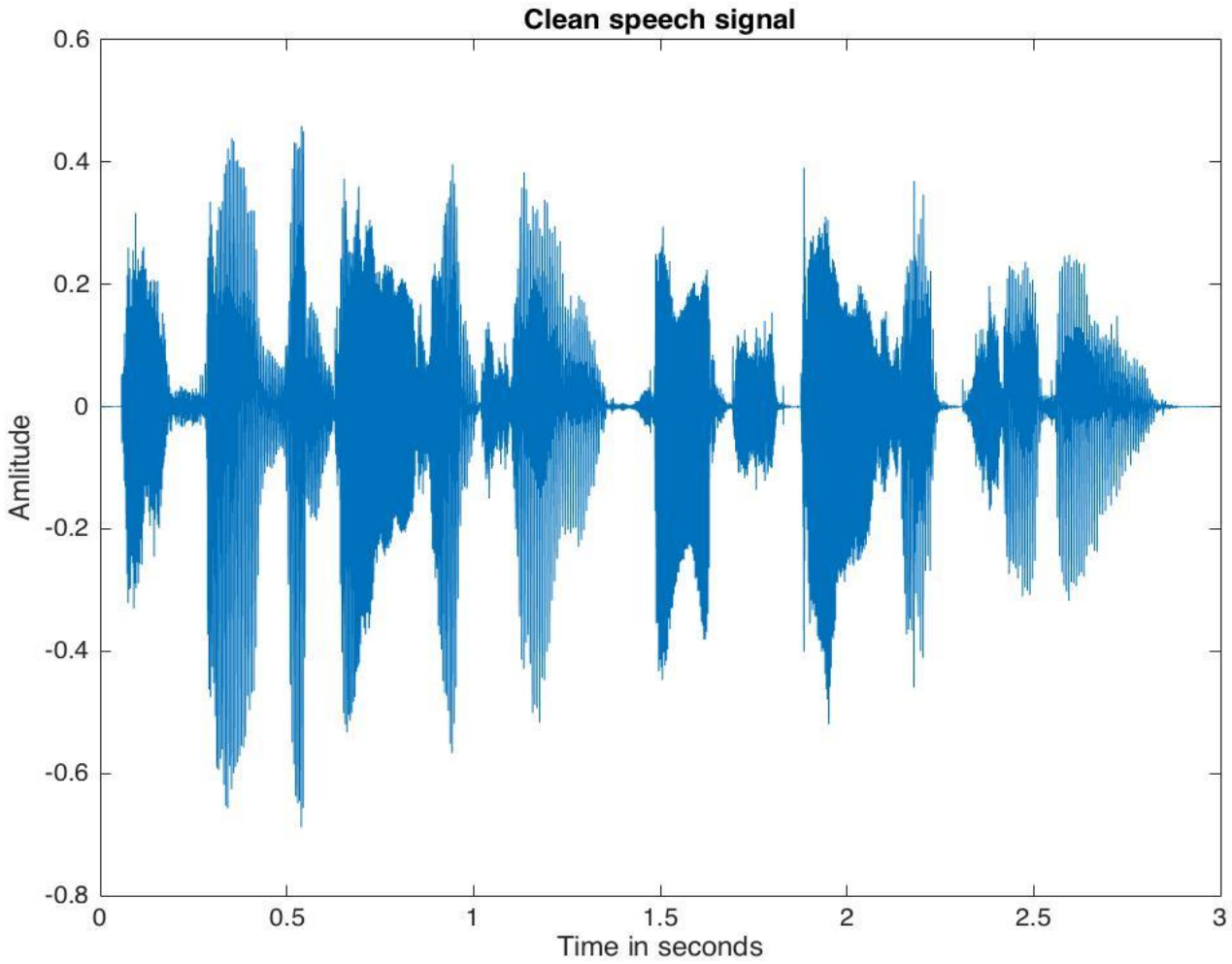
Initially, I tried filtering the entire signal with the Kalman filter but the results did not meet the requirement. So I windowed the signal using rectangular windowing, (tried Hamming windowing as well and the results are in the simulation section) and iterated the process for few times by updating the autoregressive filter coefficients for every iteration. Even though the process takes long time for a tiny speech signal data, the output can be compared with input for its similarity. In the following section, I have attached the simulation results.

SIMULATION RESULTS

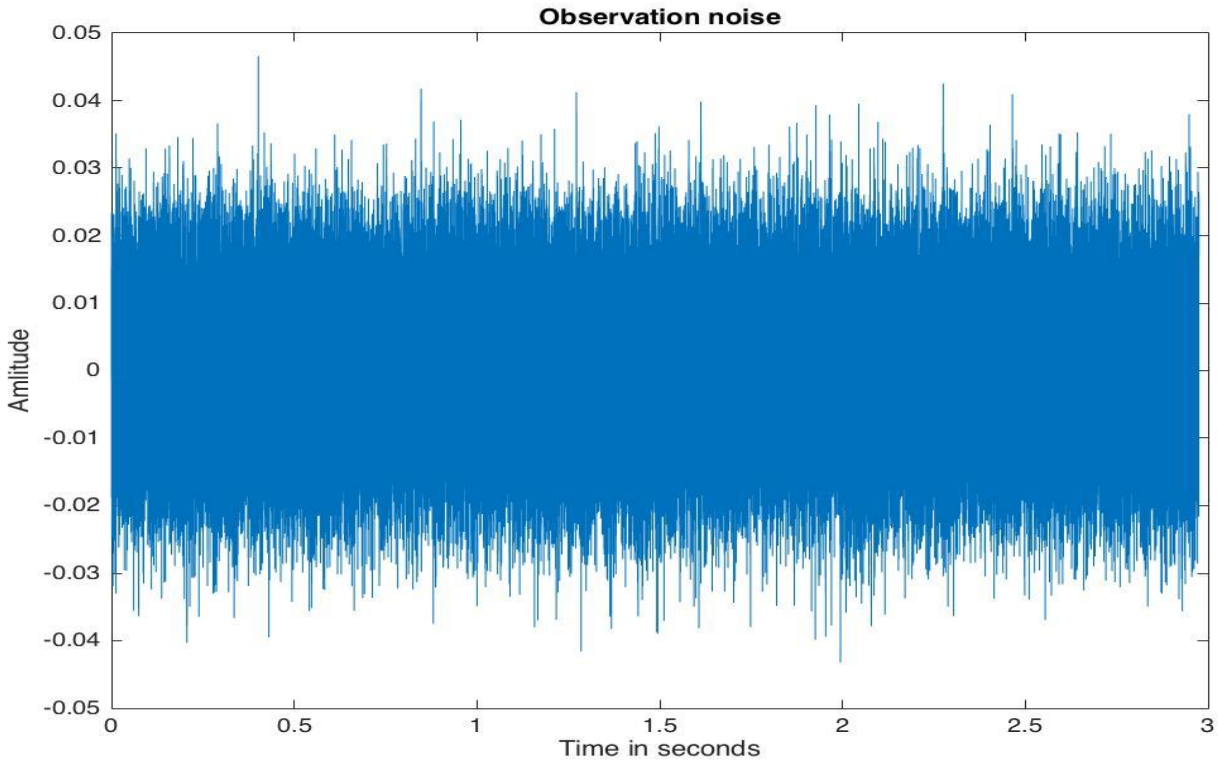
In this section, the simulation results under various situations like non-windowed processing, windowed processing, rectangular vs. hamming windowing can be seen. X-axis of all the plots denotes Time vector in seconds. Y-axis of all the plots denotes the amplitude of the respective signal in time domain.

1) Windowed processing – rectangular windowed processing.

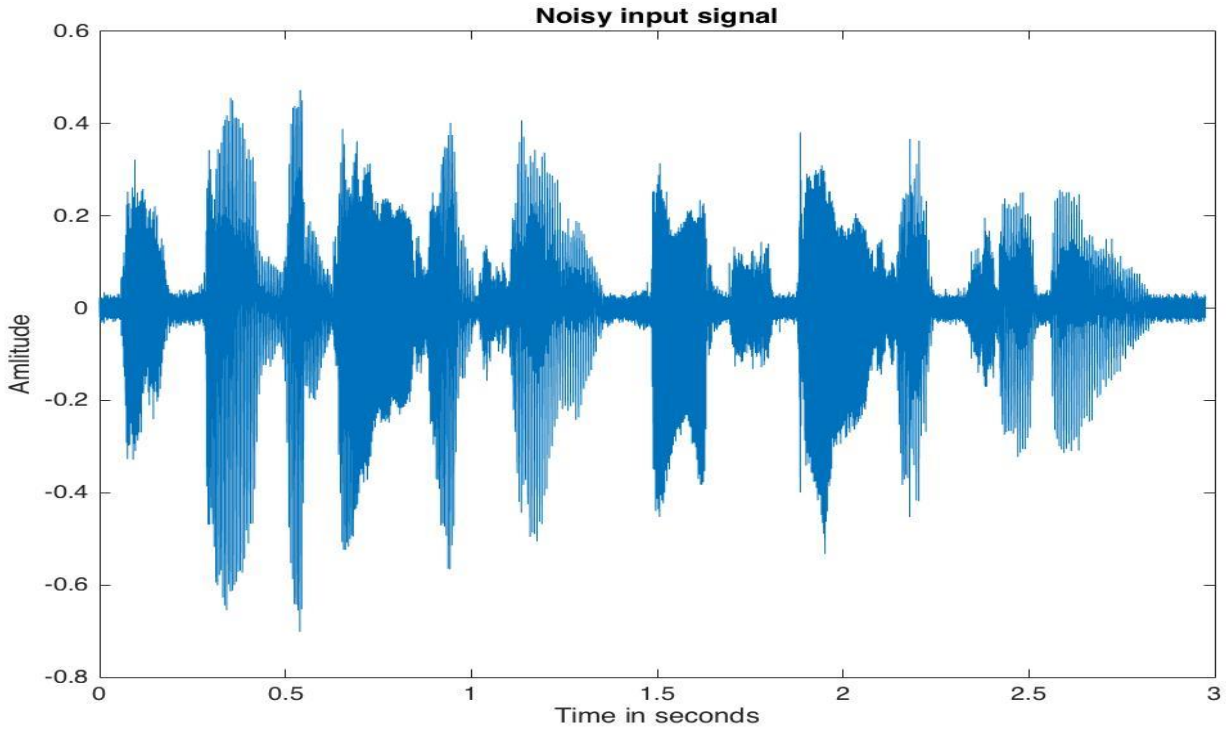
- Time vs. Clean signal.



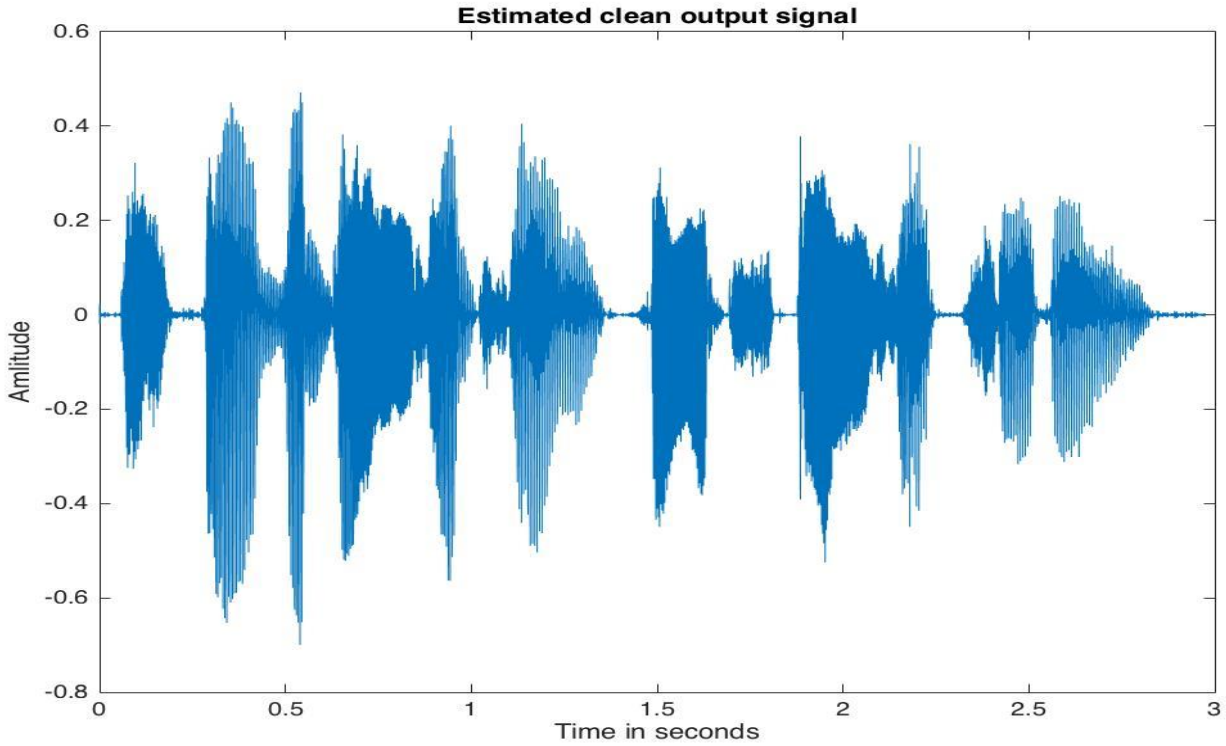
- Generated noise



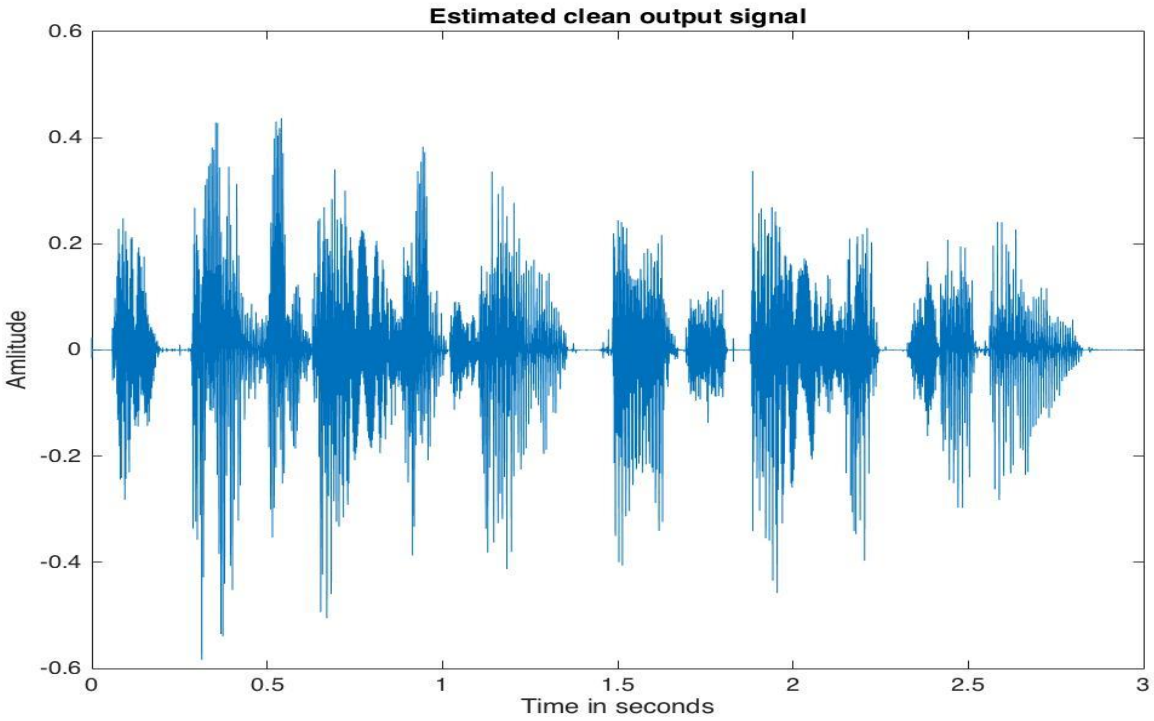
- Input + noise signal



- Estimated output



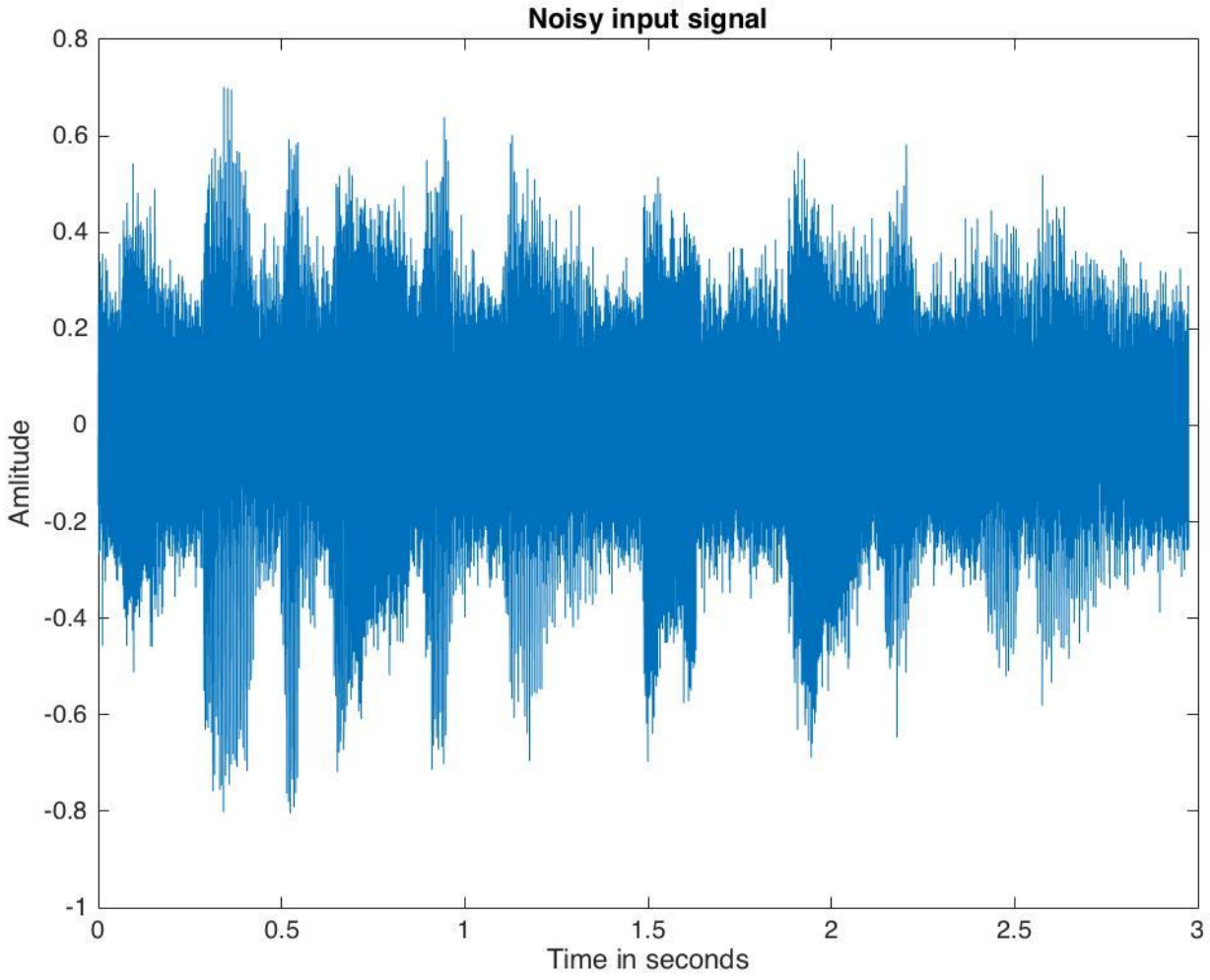
2) Hamming windowed process. (Only output is shown as input, input+noise plots are the same)



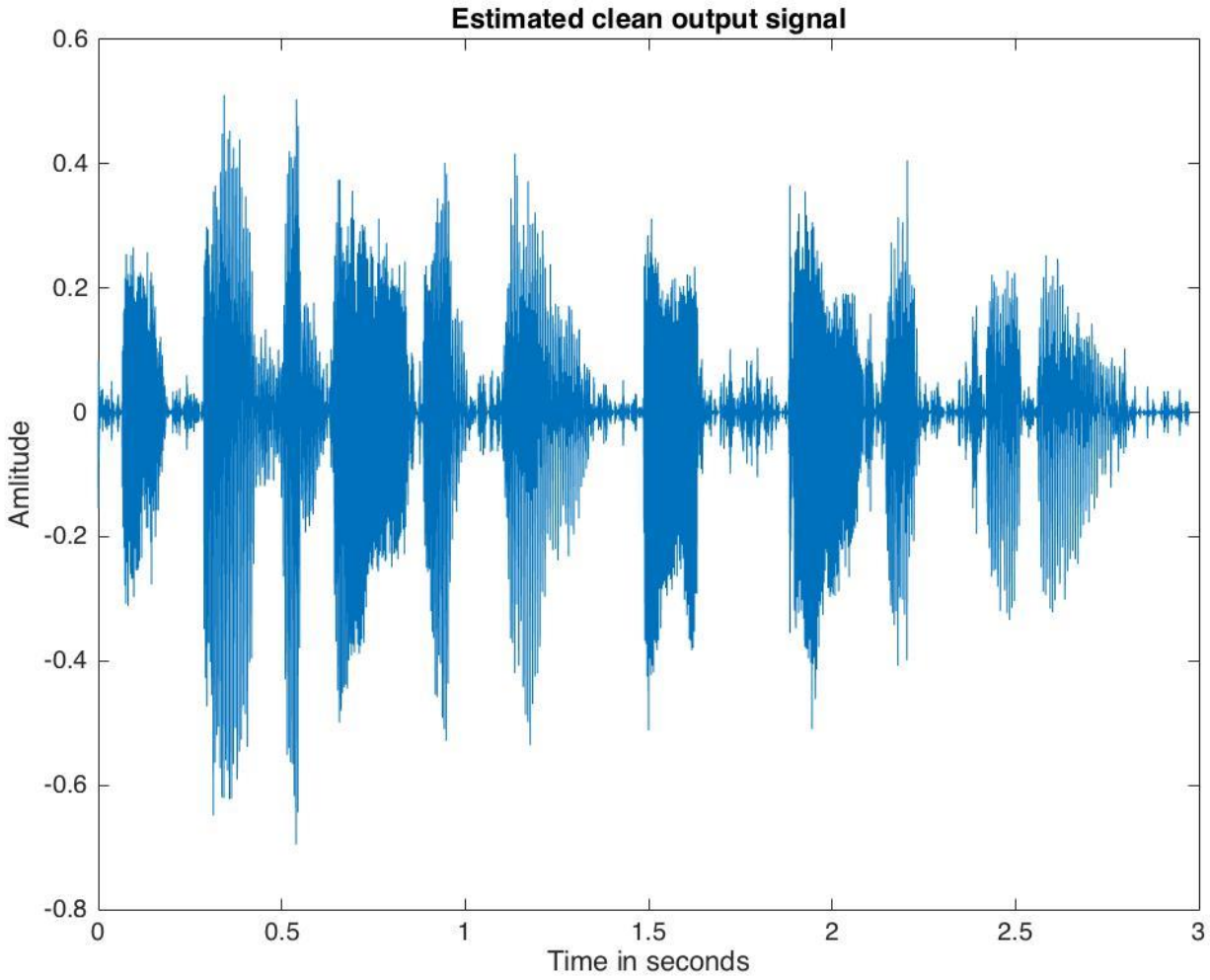
While reconstructing the hamming windowed signal output, the intelligibility was poor when compared to the rectangular window signal output. This is evident from the plot as well.

Now let's add more noise and see how well the Kalman filter estimates the clean signal…

- Input + Noise signal



- And, this is the output.



CONCLUSION AND IMPROVED ALGORITHMS

This project demonstrates how a Kalman filter can be used for purposes like speech signal enhancement. The above-mentioned idea has some big matrices (P and A for instance), whose sizes are determined by choosing appropriate autoregressive filter order. The process is slow and it is not surprising given the number of matrix multiplications it has to do for every samples. In the search for improved algorithms, both in results as well as in reducing computational overhead, I have gone through some papers, which uses Kalman filters in modulation domain [5] and another paper [3] focuses on fastening the matrix operations. But the fundamental idea behind this technique has hardly gone through improvisation and that shows the robustness of the Kalman filter algorithm in speech enhancement techniques.

REFERENCE

[1] – "*Speech Signal Basics"*, Nimvod Peleg.

[2] – "*A Speech Enhancement method based on Kalman Filtering"*, K. K. Paliwal and Anjan Basu.

[3] – "*Kalman Filtering Speech Enhancement Method Based On A Voiced-Unvoiced Speech Model*", Senton Goh, Kah Chye Tan, B. T. G. Tan.

[4] – Wikipedia – Wiener Filter.

[5] – "*Modulation-domain Kalman filtering for single-channel speech enhancement*", Stephen So, K. K. Paliwal.