Problem set 2

Author: Cheng-Liang Lu | s220034

You are missing some of the rows. The matrix should be of size N+L-1 X L, since we we have the situation were we only have one measurement of x in both the beginning (x only contains = x_0, applied to the first filter coefficient) and the end (x only contains x_N, but applied to the last filter coefficient. Try e.g. to write up the convolution asked in problem set 1 on matrix form. See eq. (2.126) in the book.

2.1

- Constructing Matrix X, which is constructed such that each row corresponds to different timestamps of the input signal. Matrix X has N rows accounting for the convolution overlap, with each row containing L consecutive values from the input signal. The construction proceeds as follows:
  - The first row contains $x_0, 0, 0, \ldots, 0$.
  - The second row contains $x_1, x_0, 0, \ldots, 0$.

    Try writting the matrix explicitly!
  - The third row contains $x_2, x_1, x_0, 0, \ldots, 0$.
  - This continues until the last row, which contains $x_{N-1}, x_{N-2}, x_{N-3}, \ldots, x_{N-L}$.

  This results in an N rows by L column matrix X.
- Constructing Vector θ, which contains the coefficients of the filter, i.e., $\theta = [\theta_0, \theta_1, \theta_2, \ldots, \theta_{L-1}]$.
- Computing Output y. It can be computed through the matrix product $y = X\theta$, where X represents the matrix representation of the input signal and θ is the vector of filter coefficients. The result of the matrix product will be a vector y containing the filtered signal.

2.2

1.

missing square

too fast! why do you add and substract that quantity? Its rigth, but why not explaining why you are doing it?

$$MSE = \mathbb{E}[(\hat{\theta} - \theta_0)] = \mathbb{E}[((\hat{\theta} - \mathbb{E}[\hat{\theta}]) + (\mathbb{E}[\hat{\theta}] - \theta_0))^2]$$

incorrect, if you expand the expression it would have been more clear. You jumped too quickly into the answer

The cross-term vanish since the estimated values and the true parameters are uncorrelated, we left with the following result,

$$MSE = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] + (\mathbb{E}[\hat{\theta}] - \theta_0)^2$$

So,

$$MSE = Variance + Bias^2$$

2.

correct

First we could define the biased estimator as a scaled estimate of the unbiased estimator,

$$\hat{\theta}_b = (1+\alpha)\hat{\theta}_u$$

$\hat{\theta}_u$ is the unbiased estimator and $\hat{\theta}_b$ is biased estimator. According that, $\mathbb{E}[\hat{\theta}_u] = \theta_0$.
We assume $MSE(\hat{\theta}_b) > 0$ and $\theta_0 > 0$.

We calculate Biased MSE:

$$MSE(\hat{\theta}_b) = \mathbb{E}[(\hat{\theta}_b - \hat{\theta}_0)^2] = \mathbb{E}[((1+\alpha)\hat{\theta}_u - \theta_0)^2]$$

We add $\alpha\theta_0$:

$$\begin{aligned} MSE(\hat{\theta}_b) &= \mathbb{E}[((1+\alpha)\hat{\theta}_u + \alpha\theta_0 - \alpha\theta_0 - \theta_0)^2] \\ &= \mathbb{E}[((1+\alpha)\hat{\theta}_u + \alpha\theta_0 - \theta_0(1+\alpha))^2] \\ &= \mathbb{E}[((1+\alpha)(\hat{\theta}_u - \theta_0) + \alpha\theta_0)^2] \\ &= \mathbb{E}[(1+\alpha)^2(\hat{\theta}_u - \theta_0)^2 + \alpha^2\theta_0^2 + 2(1+\alpha)(\hat{\theta}_u - \theta_0)\alpha\theta_0] \end{aligned}$$

$$MSE(\hat{\theta}_b) = (1+\alpha)^2\mathbb{E}[(\hat{\theta}_u - \theta_0)^2] + \alpha^2\theta_0^2 + 2\alpha(1+\alpha)(\mathbb{E}[\hat{\theta}_u] - \theta_0)\theta_0$$

Since $MSE(\hat{\theta}_u) = \mathbb{E}[(\hat{\theta}_u - \hat{\theta}_0)^2]$ and $\mathbb{E}[\hat{\theta}_u] = \theta_0$, so

$$\begin{aligned} MSE(\hat{\theta}_b) &= (1+\alpha)^2 MSE(\hat{\theta}_u) + \alpha^2\theta_0^2 + 2\alpha(1+\alpha)(\theta_0 - \theta_0)\theta_0 \\ &= (1+\alpha)^2 MSE(\hat{\theta}_u) + \alpha^2\theta_0^2 \end{aligned}$$

We assume that that

$$MSE(\hat{\theta}_b) < MSE(\hat{\theta}_u)$$

we need to know the solution of α.

By subsitution,

$$\begin{aligned}
(1+\alpha)^2 MSE(\hat{\theta}_0) + \alpha^2\theta_0^2 < MSE(\hat{\theta}_u) \\
=> MSE(\hat{\theta}_0) + \alpha^2 MSE(\hat{\theta}_0) + 2\alpha MSE(\hat{\theta}_0) + \alpha^2\theta_0^2 < MSE(\hat{\theta}_u) \\
=> \alpha(\alpha MSE(\hat{\theta}_0) + 2MSE(\hat{\theta}_0) + \alpha\theta_0^2) < 0
\end{aligned}$$

We multiply both sides by $\frac{1}{\theta_0^2 + MSE(\hat{\theta}_u)}$ (positive), we get

$$\alpha(\frac{2MSE(\hat{\theta}_u)}{\theta_0^2 + MSE(\hat{\theta}_u)} + \alpha) < 0$$

Now, we need to consider the $\alpha < 0$ and $\alpha > 0$. ($\alpha = 0$ don't make sence).

Consider $\alpha > 0$,

$$\begin{aligned}
\alpha(\frac{2MSE(\hat{\theta}_u)}{\theta_0^2 + MSE(\hat{\theta}_u)} + \alpha) < 0 \\
=> \frac{2MSE(\hat{\theta}_u)}{\theta_0^2 + MSE(\hat{\theta}_u)} + \alpha < 0
\end{aligned}$$

As $MSE(\hat{\theta}_u)$ and θ_0^2 are both positive, so the inequality cannot hold.

Then, we consider $\alpha < 0$,

$$\begin{aligned}
\alpha(\frac{2MSE(\hat{\theta}_u)}{\theta_0^2 + MSE(\hat{\theta}_u)} + \alpha) < 0 \\
=> \frac{2MSE(\hat{\theta}_u)}{\theta_0^2 + MSE(\hat{\theta}_u)} + \alpha > 0 \\
=> -\frac{2MSE(\hat{\theta}_u)}{\theta_0^2 + MSE(\hat{\theta}_u)} < \alpha
\end{aligned}$$

As $\frac{MSE(\hat{\theta}_u)}{\theta_0^2 + MSE(\hat{\theta}_u)}$ is positive, so

$$0 < \frac{MSE(\hat{\theta}_u)}{\theta_0^2 + MSE(\hat{\theta}_u)} < 1$$

Which means

$$-2 < \alpha < 0 => -1 < 1 + \alpha < 1 => |1 + \alpha| < 1$$
$$=> |\hat{\theta}_b| = |(1 + \alpha)\hat{\theta}_u| = |(1 + \alpha)||\hat{\theta}_u| < |\hat{\theta}_u|$$

So

$$|MSE(\hat{\theta}_b)| < |MSE(\hat{\theta}_u)|$$

2.3

As we know the convex definition:

correct

A function

$$f : \mathcal{X} \subseteq \mathbb{R}^l \mapsto \mathbb{R}$$

is called $convex$ if $\mathcal{X}$ is convex and $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$ the following holds true:

$$f(\lambda \boldsymbol{x}_1 + (1 - \lambda)\boldsymbol{x}_2) \leq \lambda f(\boldsymbol{x}_1) + (1 - \lambda) f(\boldsymbol{x}_2), \quad \lambda \in [0, 1].$$

So using the Triangle inequality and the fact that the norm is absolutely scalable, you can see that every Norm is convex:

$$||\lambda \boldsymbol{x}_1 + (1 - \lambda)\boldsymbol{x}_2|| \leq ||\lambda \boldsymbol{x}_1|| + ||(1 - \lambda)\boldsymbol{x}_2|| = \lambda ||\boldsymbol{x}_1|| + (1 - \lambda)||\boldsymbol{x}_2||$$

So by definition every norm is convex.

2.4

1.

As

$$y_n = x_n + b$$

$r_y(k)$ could be turned to

$$\begin{aligned} r_y(k) &= \mathbb{E}[y_n y_{n-k}] \\ &= \mathbb{E}[(x_n + b)(x_{n-k} + b)] \\ &= \mathbb{E}[x_n x_{n-k}] + b\mathbb{E}[x_n] + b\mathbb{E}[x_{n-k}] + b^2 \\ &= r_x(k) + b\mathbb{E}[x_n] + b\mathbb{E}[x_{n-k}] + b^2 \end{aligned}$$

As we already know that $\mathbb{E}[x_n] = 0$ and also $\mathbb{E}[x_{n-k}] = 0$.

Hence, we can simplify as

$$r_y(k) = r_x(k) + b^2$$

2.

As we know

$$r_x(k) = \mathbb{E}[x_n x_{n-k}]$$

We calculate $r_x(k)$:

$$\begin{aligned} r_x(k) &= \mathbb{E}[x_n x_{n-k}] \\ &= \mathbb{E}[(a x_{n-1} + v_n)(a x_{n-k-1} + v_{n-k})] \\ &= a^2\mathbb{E}[x_{n-1} x_{n-k-1}] + a\mathbb{E}[x_{n-1} v_{n-k}] + a\mathbb{E}[v_n x_{n-k-1}] + E[v_n v_{n-k}] \end{aligned}$$

Since $\mathbb{E}[x_n] = 0$, so

$$\mathbb{E}[x_{n-1} x_{n-k-1}] = r_x(k-1)$$

and

$$\mathbb{E}[v_n v_{n-k}] = \sigma_v^2 \delta(k)$$

As x_n and v_n are uncorrelated, so $\mathbb{E}[x_{n-1} v_{n-k}] = \mathbb{E}[v_n x_{n-k-1}] = 0$

Hence,

$$r_x(k) = a^2 r_x(k-1) + \sigma_v^2 \delta(k)$$

Only $k = 0$, the last term will have value σ_v^2

$$
\begin{aligned}
r_x(0) &= \mathbb{E}[x_n x_n] \\
&= \mathbb{E}[(a x_{n-1} + v_n)(a x_{n-1} + v_n)] \\
&= a\mathbb{E}[x_n x_n] + \mathbb{E}[v_n^2] \\
&= a^2 r_x(0) + \sigma^2 \\
&= \frac{\sigma^2}{1 - a^2}
\end{aligned}
$$

Since that and $r_x(-k) = r_x(k)$

Incorrect, your final derivation is wrong or you introduced the wrong numbers(3.1.4)

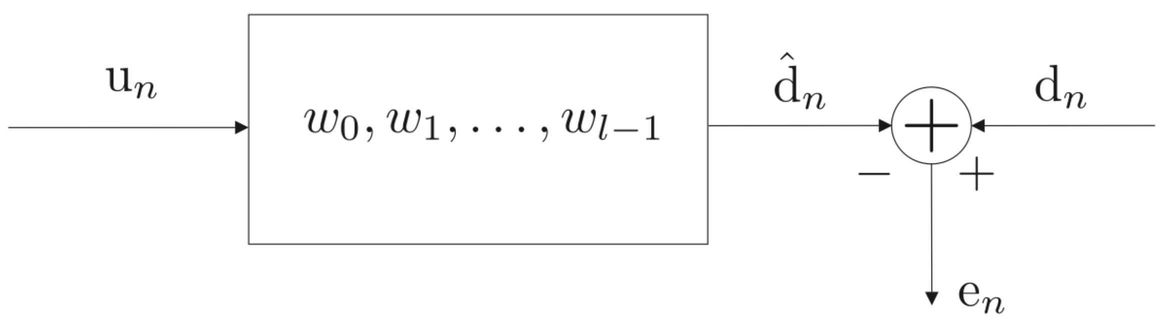
$$r_x(k) = \frac{a^{2|k|}}{1 - a^2}\sigma_v^2$$

So when we see $r_y(-2)$ for $a = 0.8$, $b = 0.5$ and $\sigma_v^2 = 1$.

$$r_y(-2) = r_x(-2) + b^2$$

$$r_y(-2) = \frac{0.8^{2\cdot\cancel{2}\,1}}{1 - 0.8^2} \cdot \overset{2}{1} + 0.25 = \cancel{1.387}$$

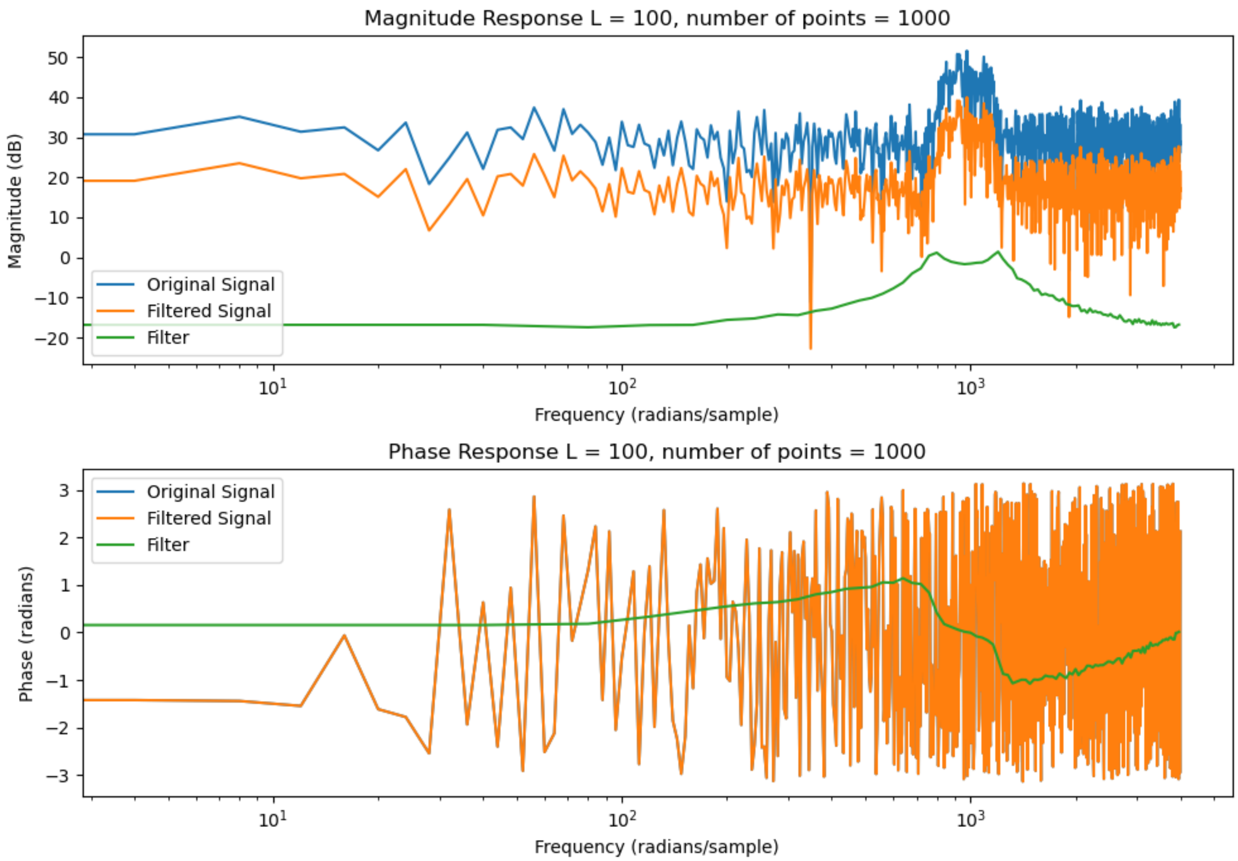
2.5

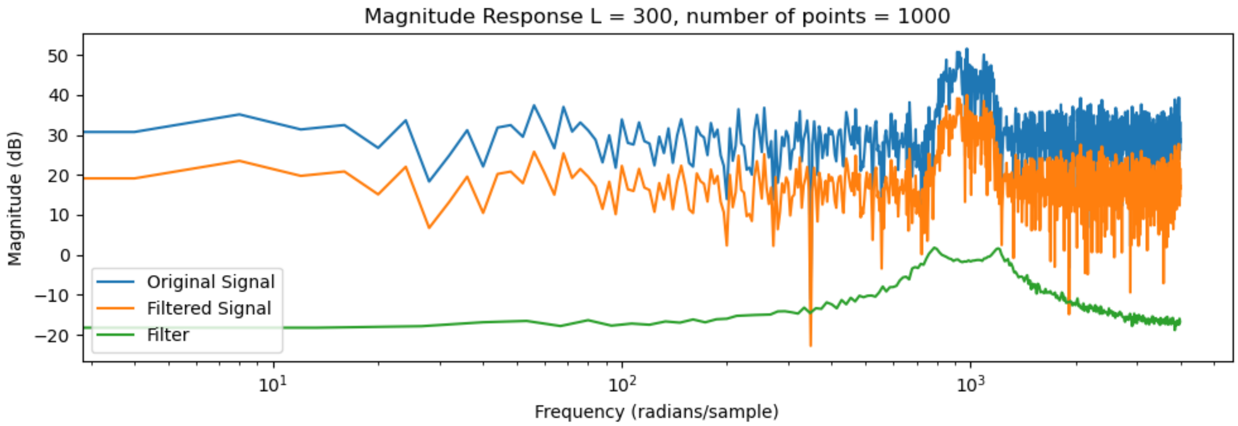


The desired signal, d_n, represents the clean version of the signal that you want to obtain after noise removal. It's essentially the component of the input signal that you're interested in isolating or enhancing.

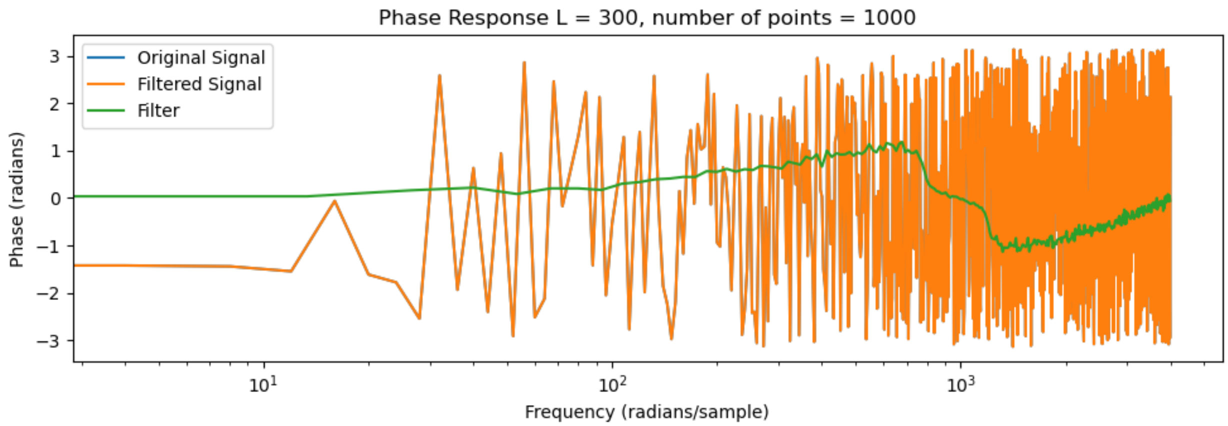
and

The error signal, e_n, is the difference between the desired signal and the filter's output. It represents the residual noise that remains in the filtered signal.

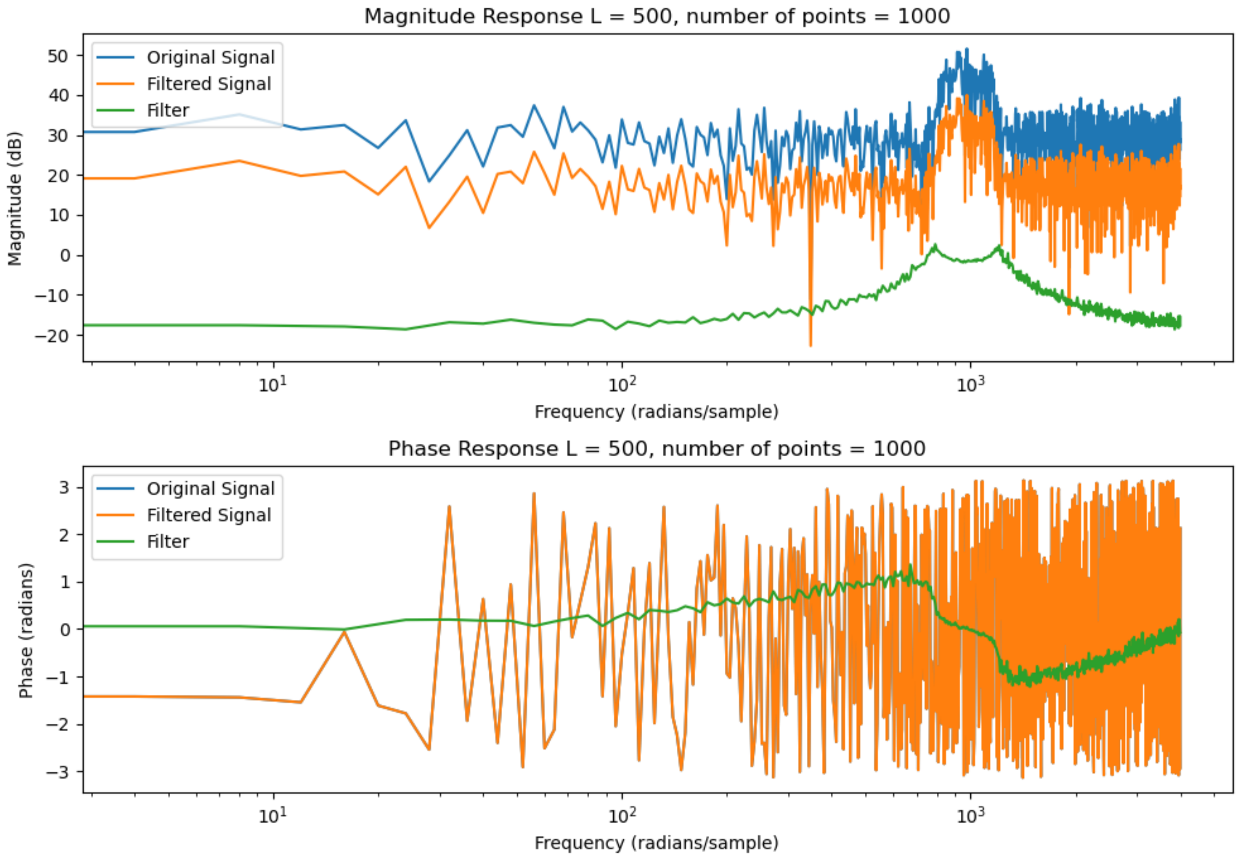




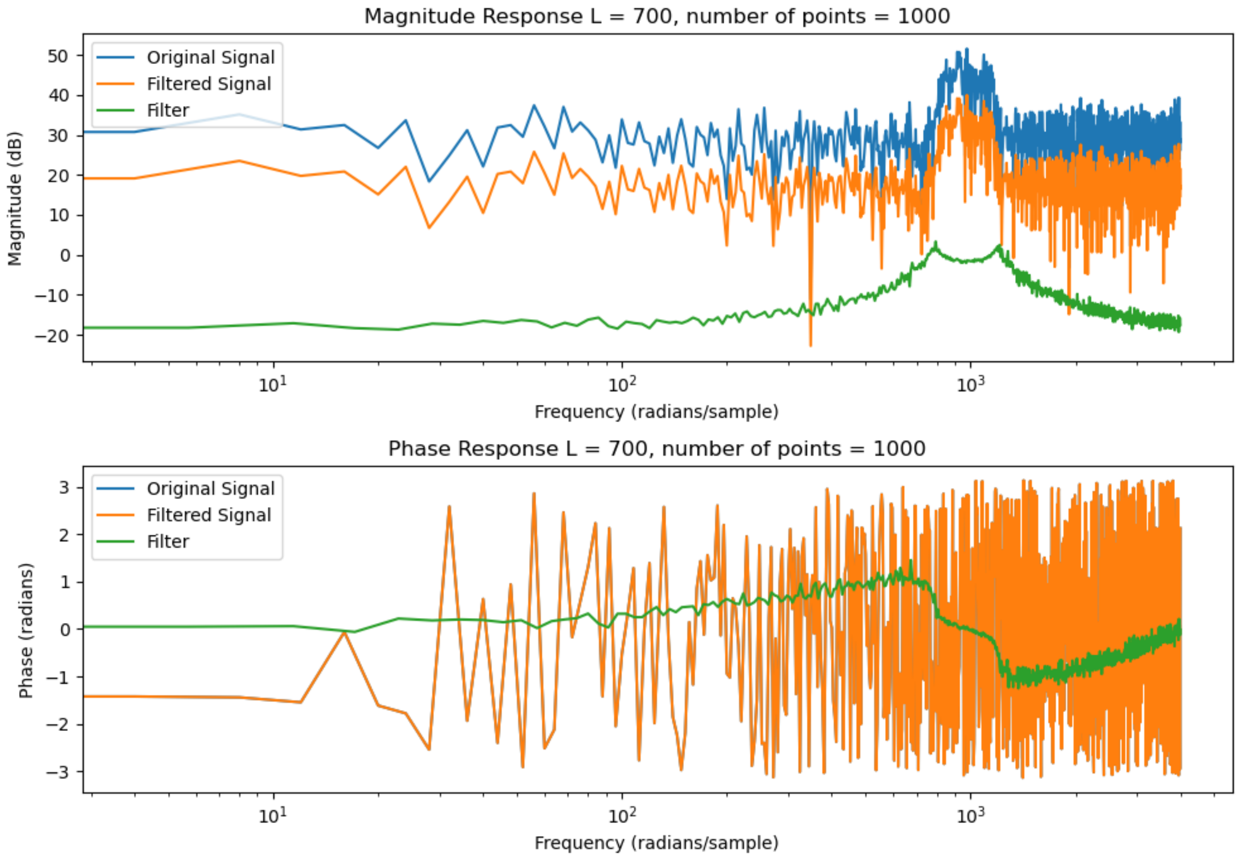
Magnitude Response L = 300, number of points = 1000
50
40
30
20
10
0
−10
−20
Magnitude (dB)
Original Signal
Filtered Signal
Filter
10^1
10^2
10^3
Frequency (radians/sample)



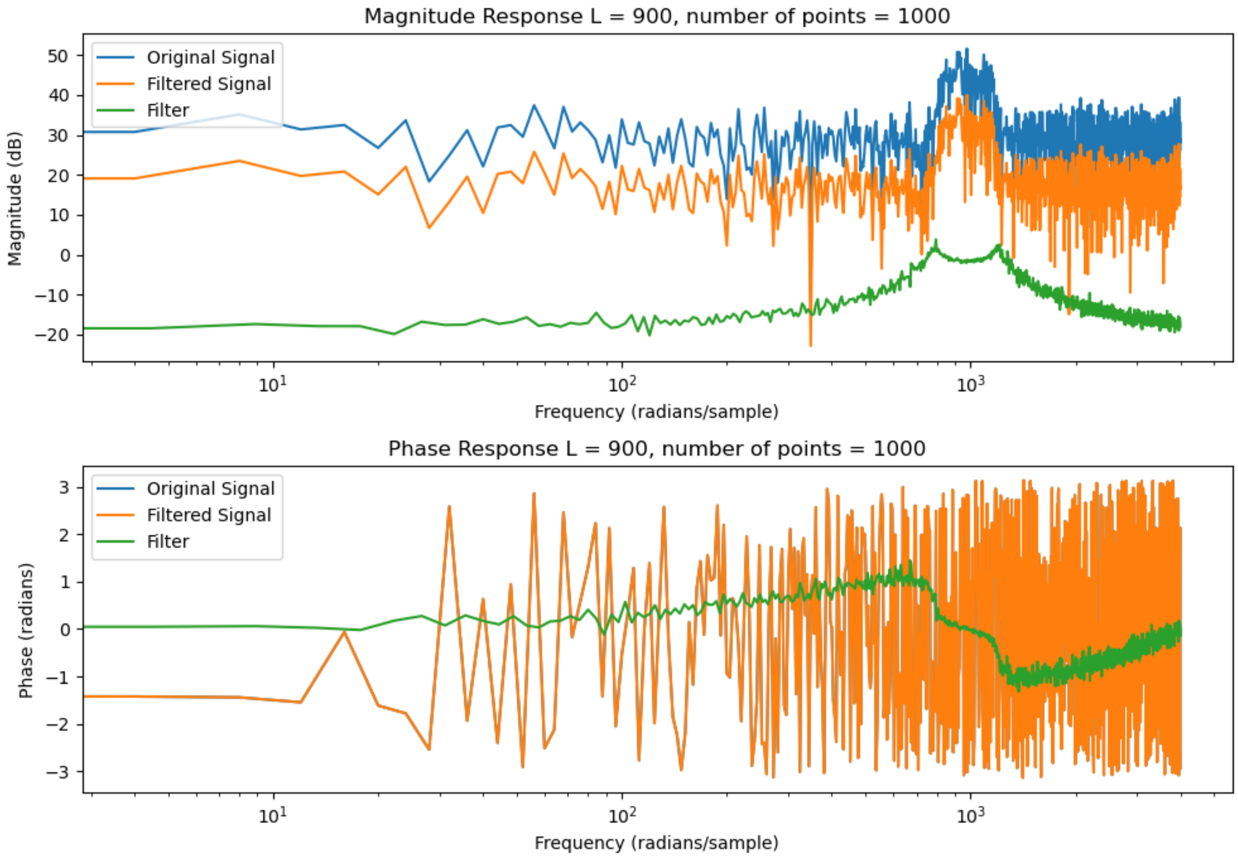
Phase Response L = 300, number of points = 1000
3
2
1
0
−1
−2
−3
Phase (radians)
Original Signal
Filtered Signal
Filter
10^1
10^2
10^3
Frequency (radians/sample)



Magnitude Response L = 500, number of points = 1000
Original Signal
Filtered Signal
Filter
Magnitude (dB)
50
40
30
20
10
0
−10
−20
10^1
10^2
10^3
Frequency (radians/sample)
Phase Response L = 500, number of points = 1000
Original Signal
Filtered Signal
Filter
Phase (radians)
3
2
1
0
−1
−2
−3
10^1
10^2
10^3
Frequency (radians/sample)



Magnitude Response L = 700, number of points = 1000
Original Signal
Filtered Signal
Filter
Magnitude (dB)
50
40
30
20
10
0
−10
−20
10^1
10^2
10^3
Frequency (radians/sample)
Phase Response L = 700, number of points = 1000
Original Signal
Filtered Signal
Filter
Phase (radians)
3
2
1
0
−1
−2
−3
10^1
10^2
10^3
Frequency (radians/sample)



As you could see I used different filter length to compare in the Frequency Response's graphs. As we know simple signals with slowly varying characteristics may require shorter filters and noisy signals with more complex noise patterns may benefit from longer filters. Also, Longer filters provide better frequency resolution, which is important, but longer filters require more computational resources to implement.
Because we need to choose the trade-off one, I will choose the length 300 since we could see the length 100 still a little bit not obvious for frequency resolution. However, the filter length surpass 300 may waste more computational resource with the same pattern and vanish the slowly varying characteristics.

---

100 is huge, 15 works out. I think it is better to compute the MSE between the reconstructed signal and the original signal for differen filter lengths and then plot MSE vs filter length. The from the plot you can choose the best point according to the trade off.

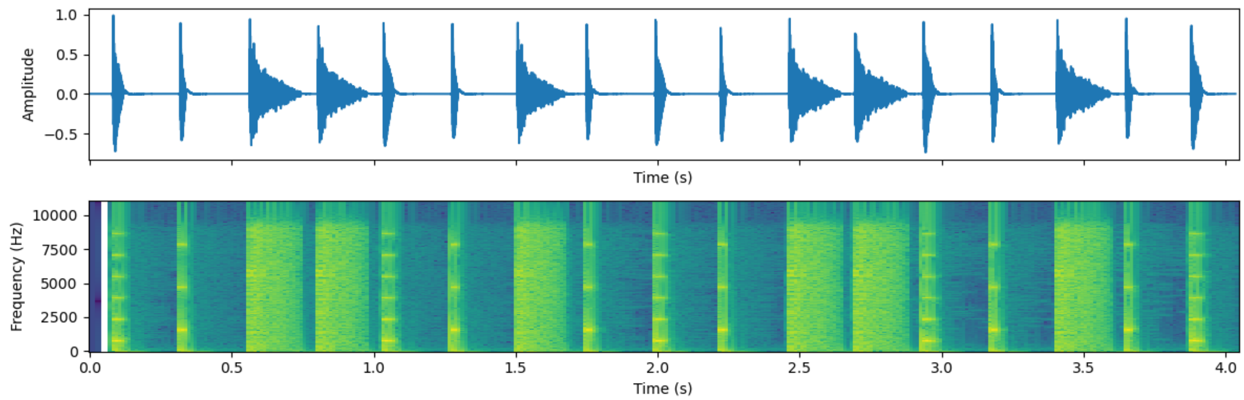
2.6

I eventually chose window size = 516 and the overlap is 256 since the higher perfomance effect in the spectrogram of this recording. The reason to choose this window size is the analysis of the results since we need to choose the best trade-off between fine-grained time analysis, using a smaller window size, and capturing detailed

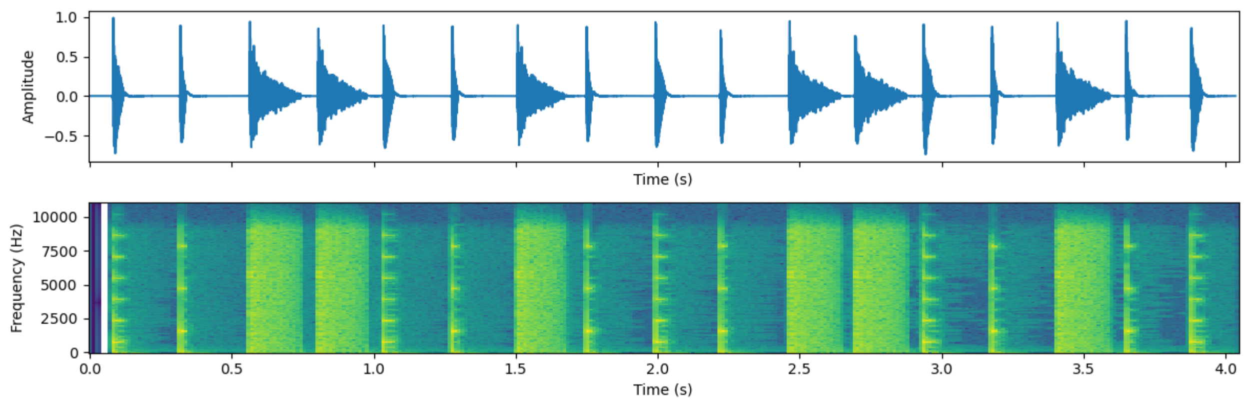
correct

frequency information, use a larger window size. Furthermore, for the overlap, we chose the typical choice is 50 percent, helping in reducing artifacts in the spectrogram. It is based on the balance between smoothness and computational complexity.
As the window function, I chose the 4 types of the categories to compare, which are Rectangular, Bartlett, Hann, and Hamming.

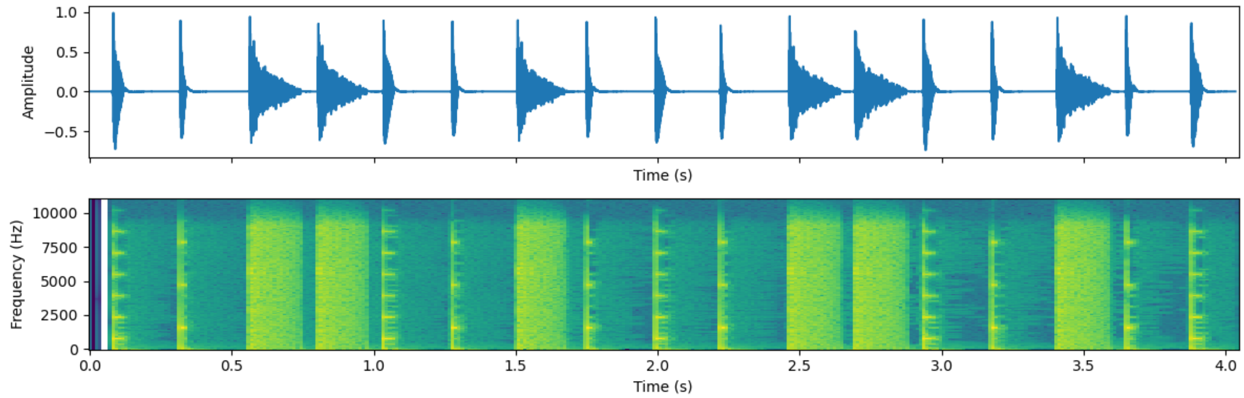
Rectangular:



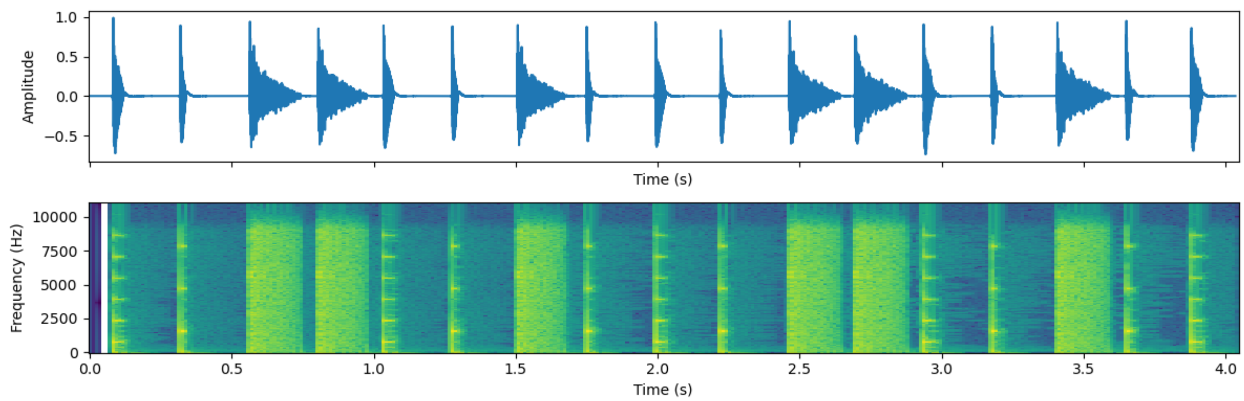
Barlett:



Hann:



Hamming:



I will select Hamming and Hanning windows since they have lower side lobe levels but wider main lobes, making them suitable for most applications. Compare to Barlett(Triangular) with pretty low side lobes and Rectangular too high side lobes(too smooth), I think the best choice in this situation is Hamming and Hanning windows.

2.7

correct

$$a : wd$$
$$b : s + h * n - a$$
$$c : w^i = w^{i-1} - \mu \nabla (J(w^{i-1}))$$
$$the\ last\ term\ this\ will\ change\ depending\ on\ the\ algorithm$$
$$d : noise\ signal$$
$$h : the\ filter\ we\ add\ on\ noise\ signal$$
$$w : the\ metrix\ that\ turn\ noise\ signal\ to\ the\ prediction\ of\ h * n.$$

Functional principle of the noise canceling setup:

- **Microphones:** The setup typically uses at least two microphones: one for capturing the desired signal (e.g., speech or music) and another for capturing the noise. These microphones are often referred to as the "primary" and "reference" microphones, respectively.
- **Signal Acquisition:** The primary microphone captures both the desired signal and the unwanted noise, resulting in a mixed signal that contains both components. The reference microphone captures primarily the noise component.
- **Adaptive Filter:** An adaptive filter is used to process the signal from the reference microphone. The filter estimates the characteristics of the noise based on the reference signal and adapts its coefficients to minimize the correlation between the filtered reference signal and the noise component in the mixed signal.
- **Cancellation:** The output of the adaptive filter is subtracted from the mixed signal (captured by the primary microphone). The result is an enhanced signal that ideally contains the desired signal with reduced or canceled noise.
- **Adaptive Algorithm:** The adaptive filter uses an adaptive algorithm (e.g., LMS, NLMS, RLS) to continuously adjust its coefficients to minimize the residual noise in the output.

The assumption of uncorrelated source signal and noise is important in a noise canceling setup for several reasons:

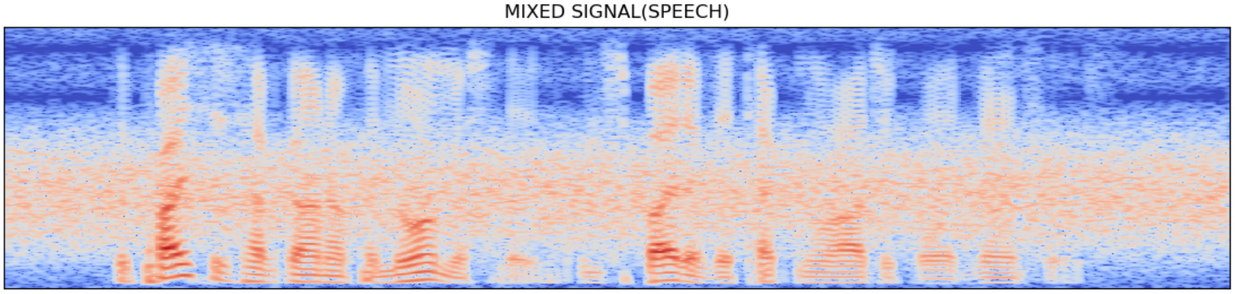
- **Adaptation Stability:** Adaptive filters rely on the assumption that the reference signal (capturing the noise) and the mixed signal (containing noise and source) are uncorrelated. This assumption ensures the stability of the adaptation process and prevents the filter from over-adapting to the source signal.

- **Minimization of Residual Noise:** The primary goal of noise canceling is to minimize residual noise in the output. When source and noise are uncorrelated, the adaptive filter can effectively estimate and cancel the noise, resulting in a cleaner output.
- **Efficient Use of Resources:** Assuming uncorrelated source and noise simplifies the filtering process and allows the system to focus on noise reduction. Correlated source and noise may introduce complexity and make noise cancellation less effective.

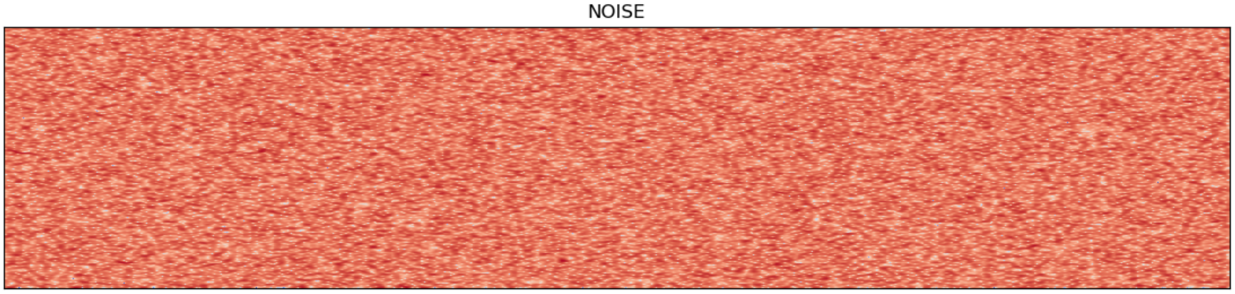
LMS

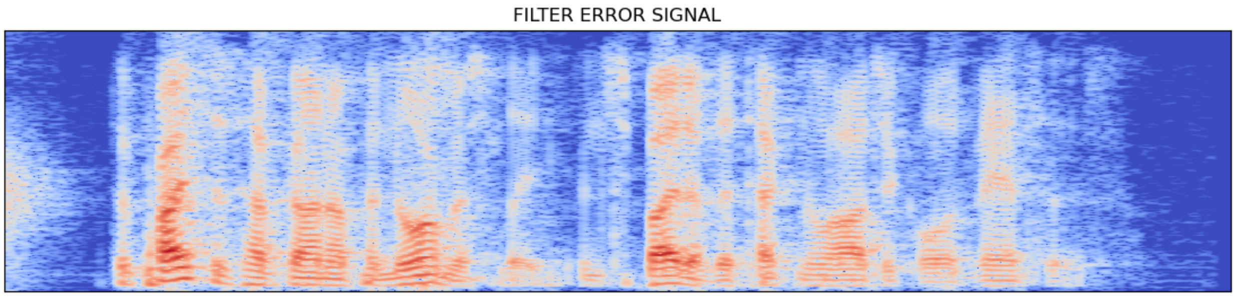
- Mixed Signal = Speech + Filer Noise Signal(bpir) [Mean squared error = $3.21 \cdot 10^{-4}$]

MIXED SIGNAL(SPEECH)

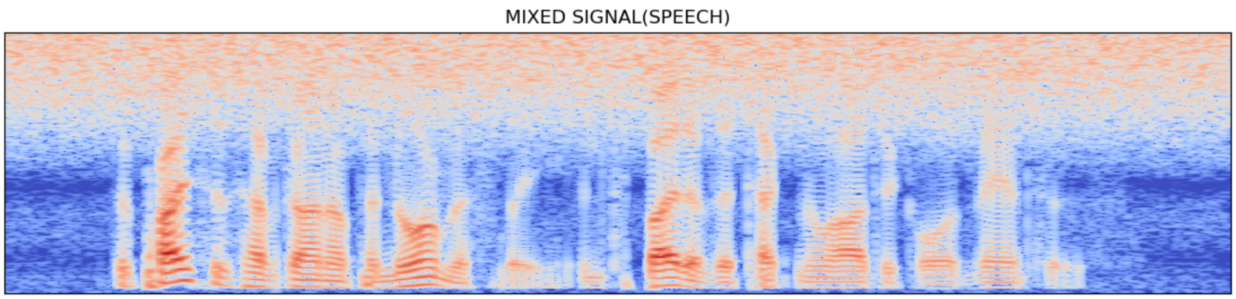


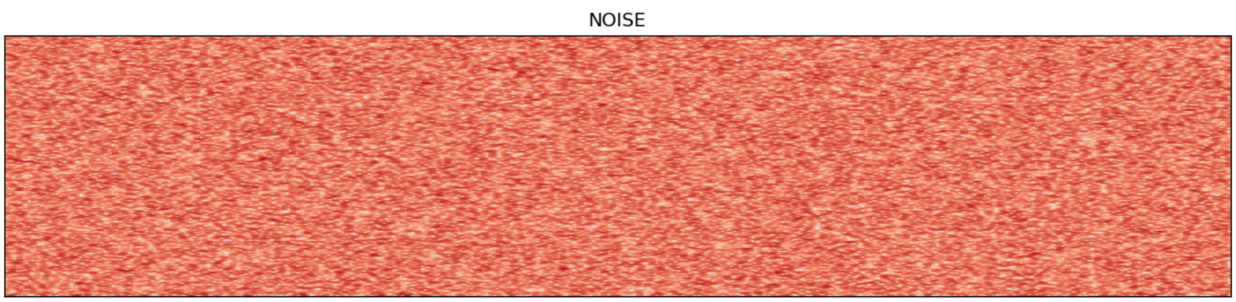
NOISE

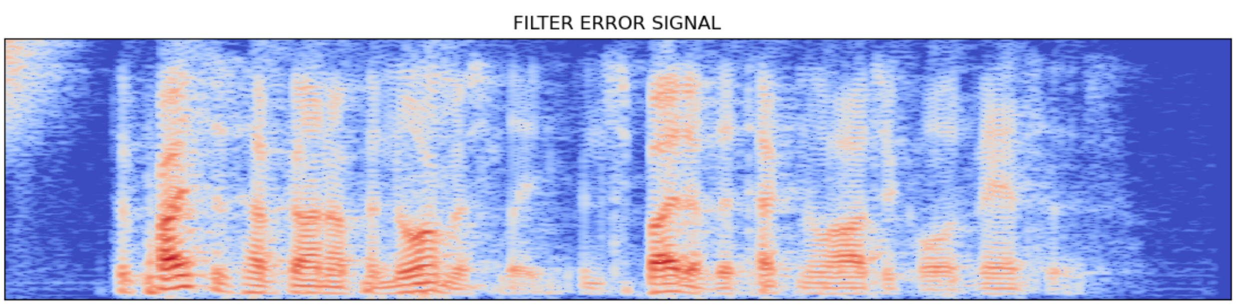




- Mixed Signal = Speech + Filer Noise Signal(hpir)[Mean squared error = $3.20 \cdot 10^{-4}$]

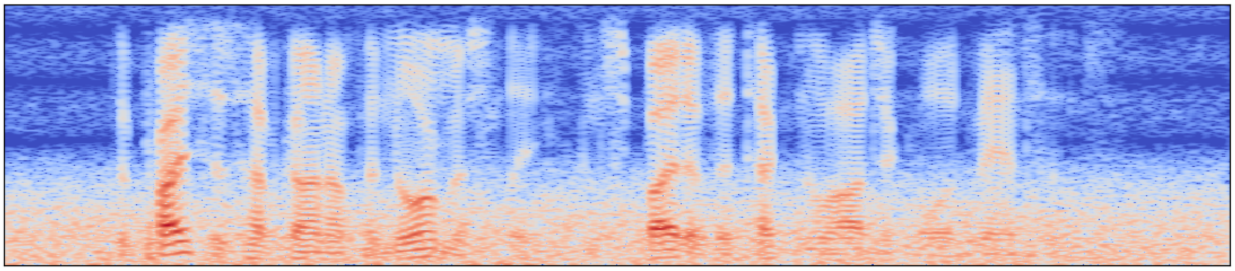




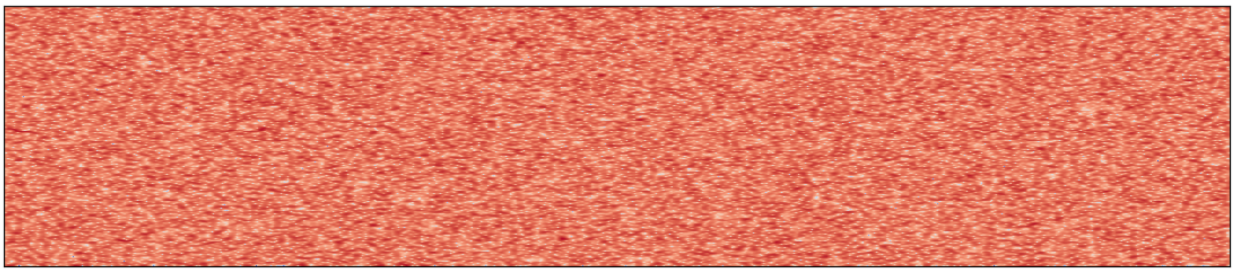


- Mixed Signal = Speech + Filer Noise Signal(lpir)[Mean squared error = $3.21 \cdot 10^{-4}$]

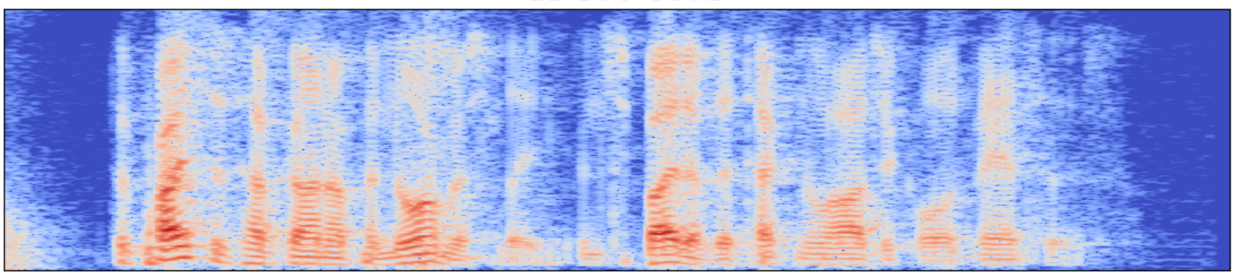
MIXED SIGNAL(SPEECH)



NOISE



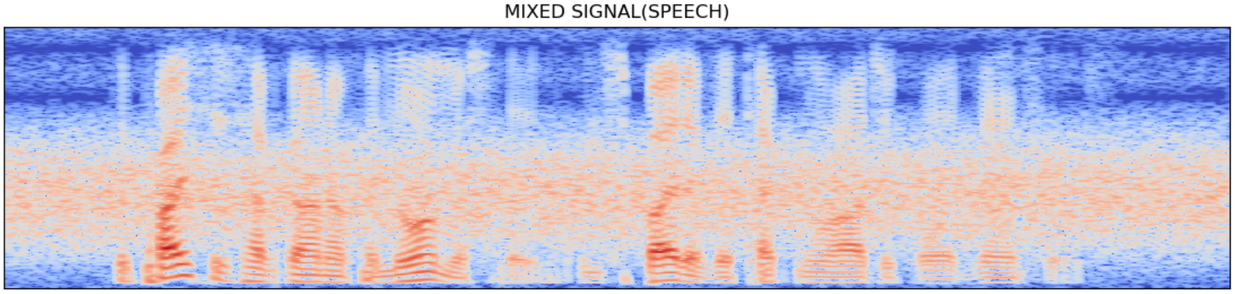
FILTER ERROR SIGNAL



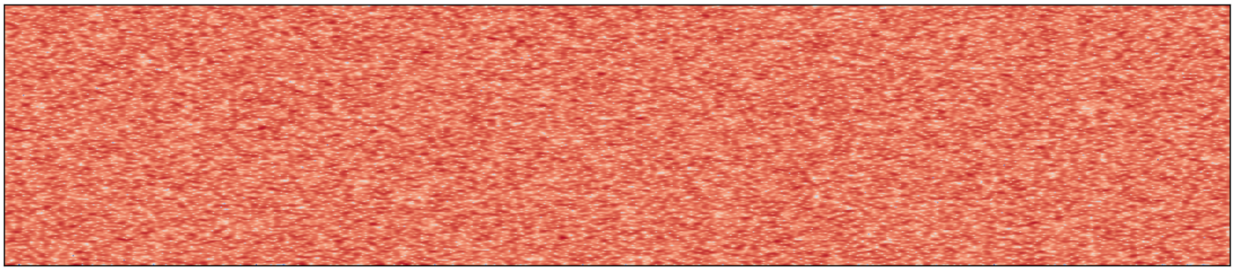
NLMS

- Mixed Signal = Speech + Filer Noise Signal(bpir) [Mean squared error = $1.85 \cdot 10^{-4}$]

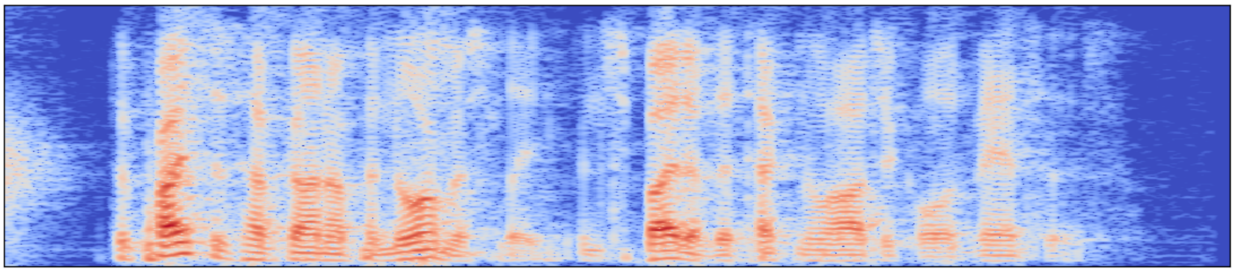
MIXED SIGNAL(SPEECH)



NOISE

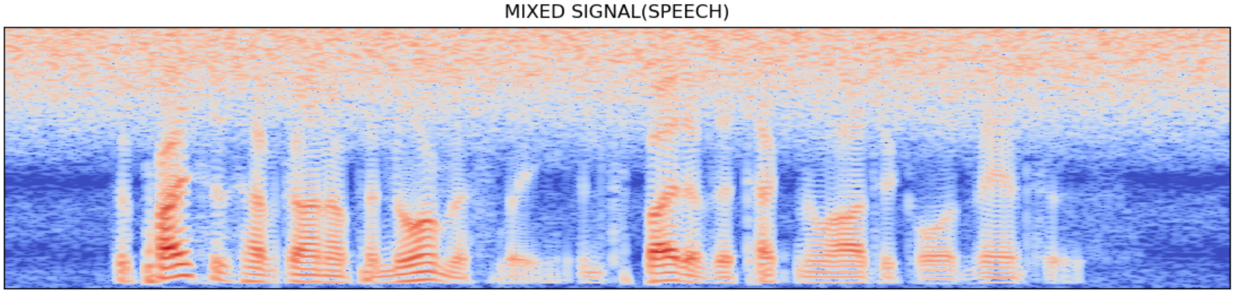


FILTER ERROR SIGNAL

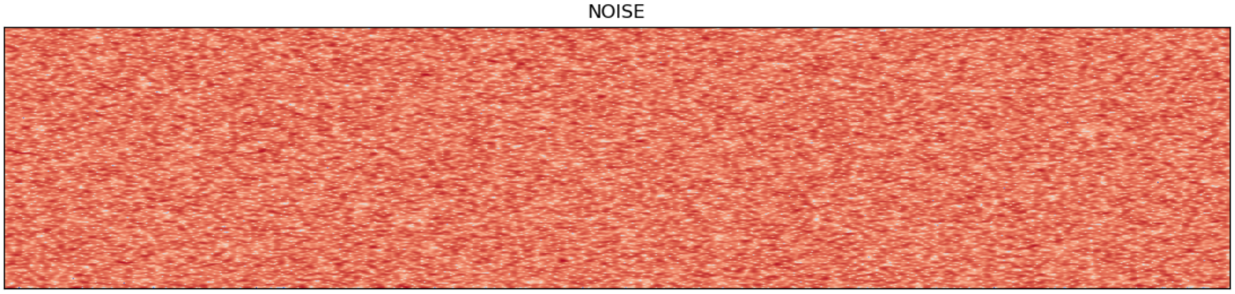


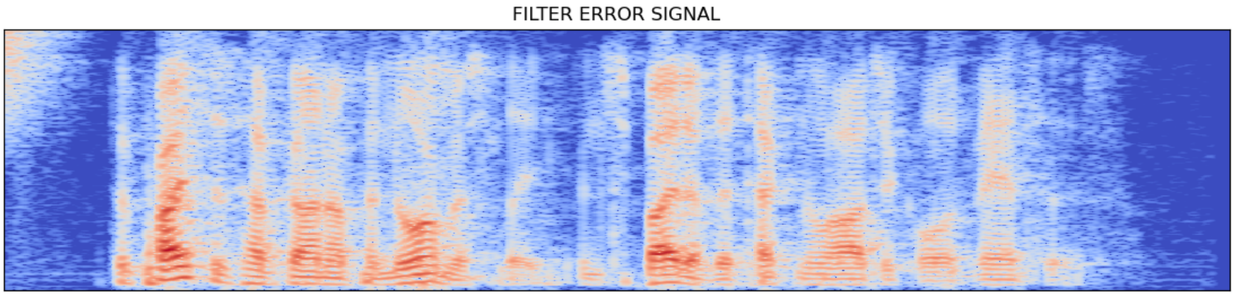
- Mixed Signal = Speech + Filer Noise Signal(hpir)[Mean squared error =$1.86 \cdot 10^{-4}$]

MIXED SIGNAL(SPEECH)

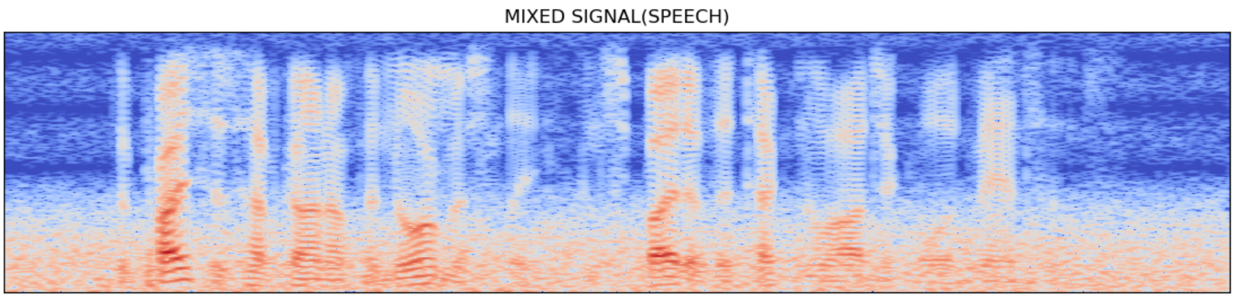


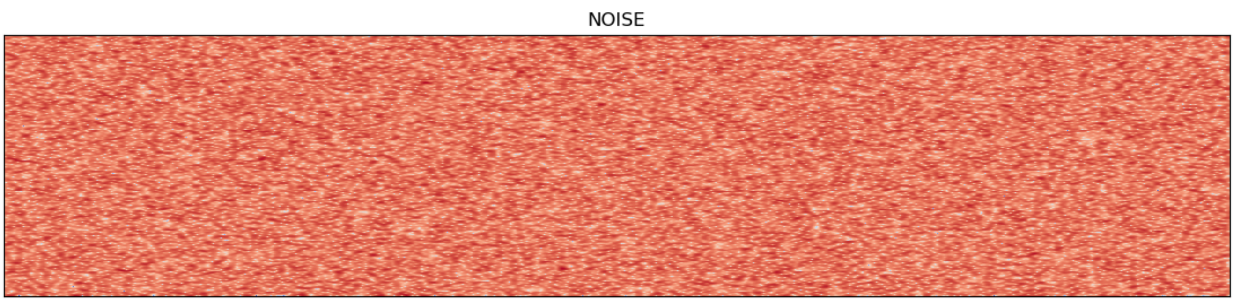
NOISE

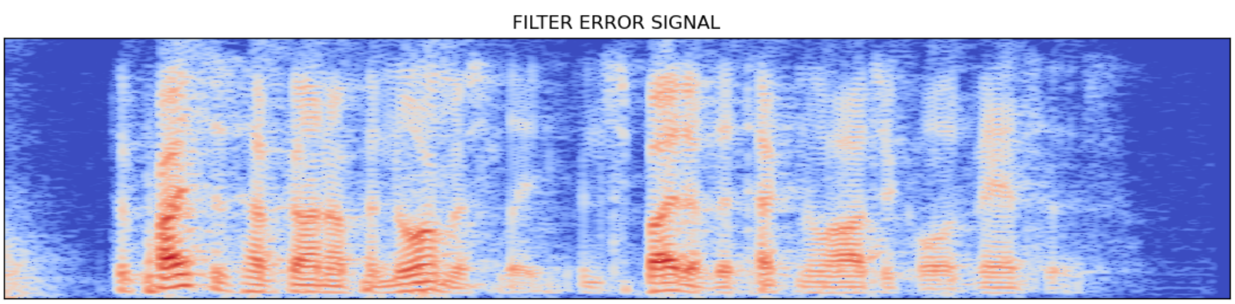




- Mixed Signal = Speech + Filer Noise Signal(lpir)[Mean squared error = $1.85 \cdot 10^{-4}$]



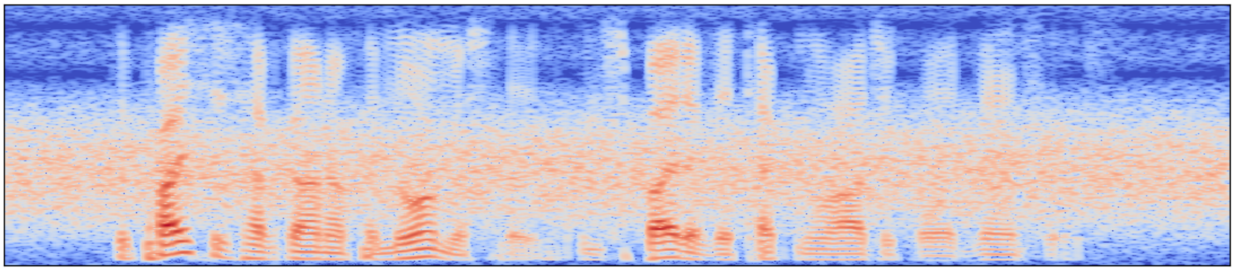




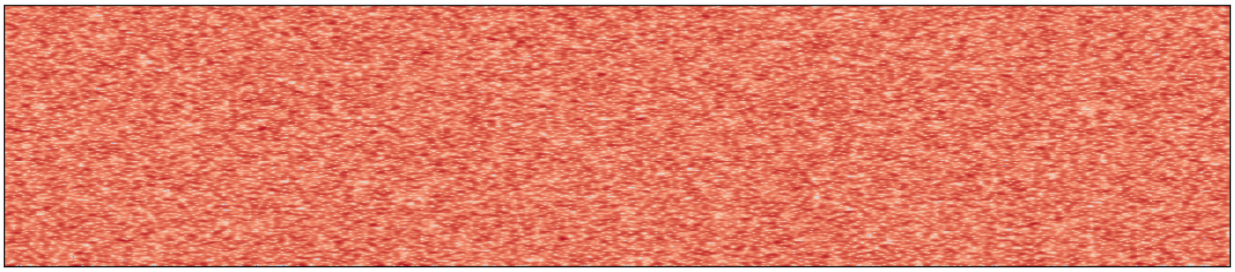
RLS

- Mixed Signal = Speech + Filer Noise Signal(bpir)[Mean squared error = $2.49 \cdot 10^{-4}$]

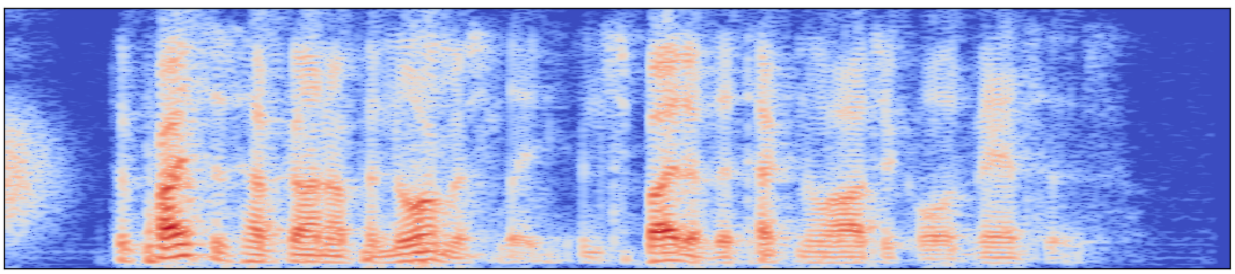
MIXED SIGNAL(SPEECH)



NOISE

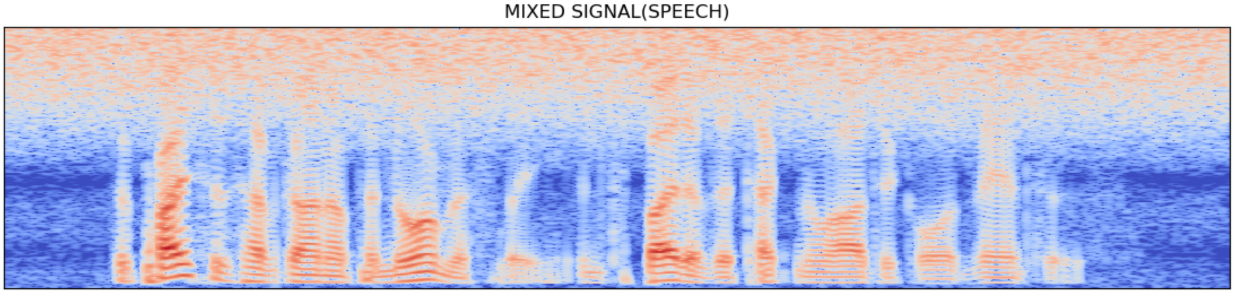


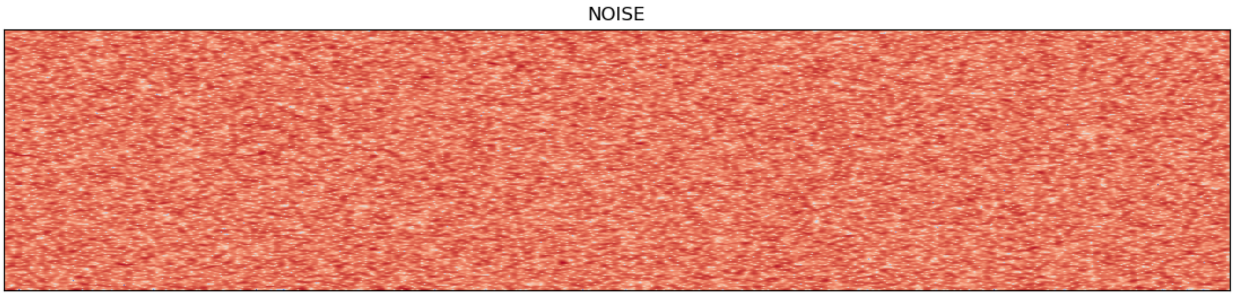
FILTER ERROR SIGNAL

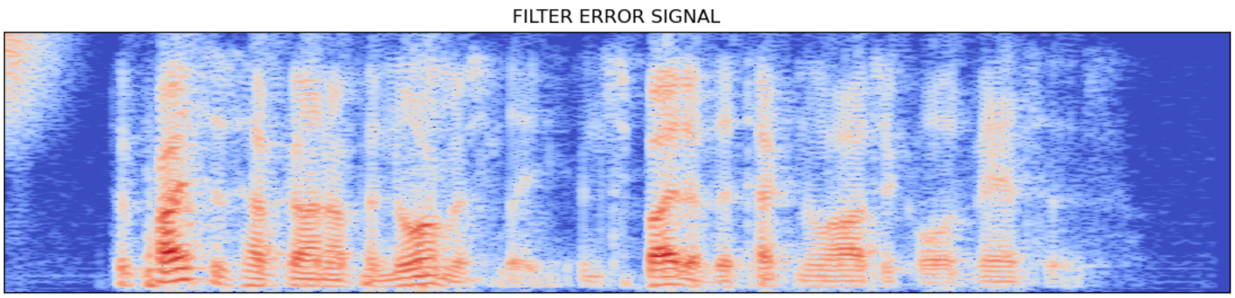


- Mixed Signal = Speech + Filer Noise Signal(hpir)[Mean squared error = $2.49 \cdot 10^{-4}$]

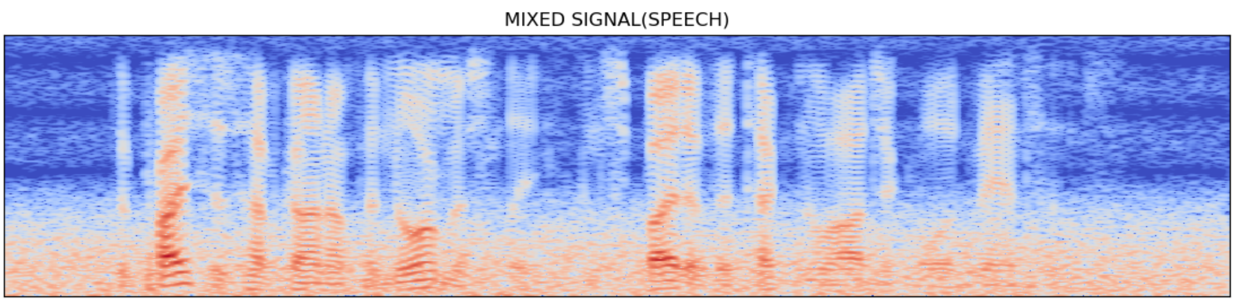
MIXED SIGNAL(SPEECH)

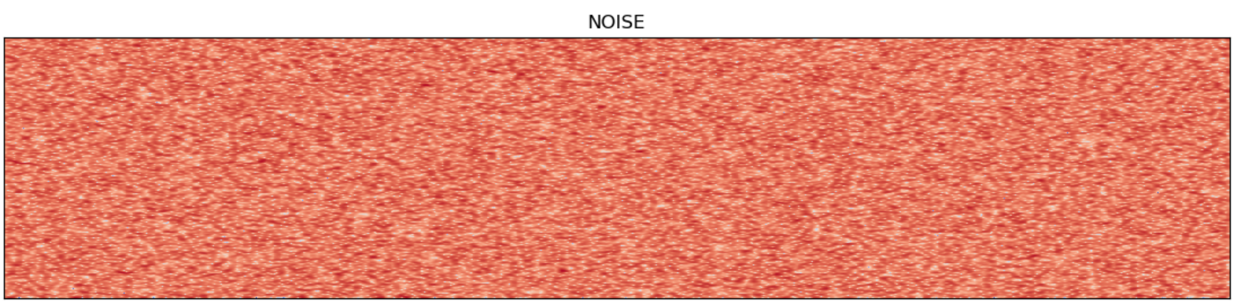




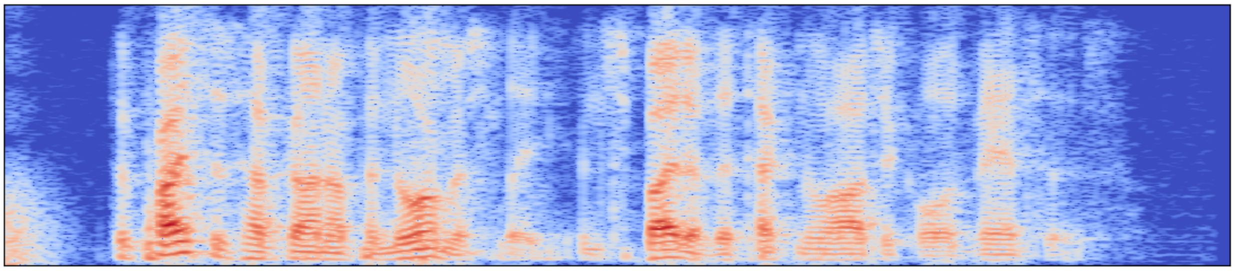


- Mixed Signal = Speech + Filer Noise Signal(lpir)[Mean squared error = $2.49 \cdot 10^{-4}$]





FILTER ERROR SIGNAL



The performance is surprising that NLMS will be better than the RLS. For LMS cannot compete with those two. Furthermore, in the Mixed Signal graphs, we could easily recognized the true characteristics of these filters: bpir represents “Bandpass Filter”, hpir represents “Highpass Filter”, lpir represents “Lowpass Filter”. These could turn the noise to specific frequency noise, high frequency noise, and low frequency noise respectively.