LABORATORY PROJECT ON SPEECH ENHANCEMENT

Aakash Tripathi, Alexander Yu
Rowan University
Date: May 7, 2021

Table of Contents

Introduction

For this lab, students were tasked to use signal processing to understand speech enhancement. Speech enhancement is the process of improving speech signals which were not of good quality using techniques of signal processing. Speech signals can be considered "poor quality" if there is noise or interference from the surroundings of the recording area. However, to improve the quality, there are quite a bit of obstacles, thus making the process difficult. The main reason is that corrupted speech signals vary in different situations. This makes the speech enhancement algorithm vary in performance, as well. However, there are multiple types of noise reduction algorithms. These include adaptive filtering, spectral subtractive, Wiener filtering, statistical model based and subspace algorithm methods. For this lab, the spectral subtractive and wiener filtering methods were the focus. Although there are many types of signal transformations. the two signal transformations that were used were the discrete Fourier transform (DFT) and the short time Fourier transform (STFT). For speech enhancement, the most important signal transformation is the DFT. This is because DFT has a lower computational complexity. In addition to the lower computational complexity, DFT's are easier to implement, while also having a natural resemblance to the auditory processes that occur in the human ear. The DFT method modifies the speech signal and then transforms it back into the time domain. This is where STFT comes in, as there are times where a speech signal is not stationary. This leads to said speech signals to only be modified in certain time intervals (10-40ms). STFT looks to take out coefficients from the DFT so the speech processed noise can be eliminated, as it is an additive effect. To do so, an amplitude-phase component is used. Generally, in a DFT, the phase component and the amplitude component are unknown. But to estimate both together is more difficult than to estimate them separately. When estimated separately, they can combine to produce the speech coefficients. STFT on the other hand focus more on the two noise reduction algorithms. To end, the motivation to do this lab is to improve and prevent errors in devices that rely heavily on speech enhancement. Speech enhancement is used in hearing aids, mobile devices, teleconferencing systems, and speech recognition. The more important ones are speech recognition and hearing aids can always be improved on and are vital to the daily lives of those who use them.

Objectives

1. Add noise to a speech signal.
2. Learn and Understand PESQ evaluation
3. Use a Wiener filter on a noisy signal.
4. Analyze and understand the effects filtering has on PESQ values.
5. Calculate and plot the 95% confidence interval of the PESQ values.

Background – Speech Enhancement

For this lab students were told to look and understand two types of noise reduction algorithms when it comes to speech enhancement. The first one that was looked at was the Spectral Subtractive Methods. Spectral subtraction was one of the first proposed category of algorithms for noise reduction in the frequency domain. Spectral subtraction is heavily reliant of the beliefs of noise spectrums. So clear speech comes from eliminating an estimate of noise on a noisy speech spectrum. Another way to look at it is by looking at the similarity of phases between the noisy and clean speech. This differs from the Wiener Filtering method. This is because the Wiener filtering methods do not assume that subtracting will make a noisy signal clean. It seems that the only similarities between the two are that they both use STFT's and are calculations on sound spectrum. The spectral subtractive method is more based on intuition and is easier to compute compared to the Wiener filter.

Adding Noise to Speech

The signal to noise ratio or SNR is defined as the ratio of the power of the clean input signal to the power of unwanted noise and controls the level at which noise affects the sound:

$$SNR = \frac{P_{clean\ signal}}{P_{noisy\ signal}}$$

Where SNR can be found by taking the log base 10 of the L2 norm of clean signal divided by the L2 norm of the noisy signal.

$$SNR_{dB} = 10 \log_{10} SNR$$

For this section, white noise was added to a clean speech file, at various signal to noise ratio (SNR) equal to 30 dB, 20 dB, 10 dB and 0 dB. To test the behavior of speech as the SNR changed, the audio was played with the different levels of white noise SNR added to the clean speech file. From this it was found that, as the SNR decreases it became more difficult to hear the clean speech file. This result confirms the relationship computed earlier. As more noise is added, the value of the fraction approaches 1 and since the log of 1 is equal to 0, the speech becomes harder to hear and understand.

Next, different noise signals suppled such as, exhibition, train, and street were listened to and compared against the white noise used in the procedure. There was no perceptual difference between the different noisy signals. Though they were difficult to hear, there was still enough information to portray the signal properly.

Perceptual Evaluation of Speech Quality (PESQ)

Perceptual Evaluation of Speech Quality or PESQ is an industry standard for automating the assessment and testing of speech quality. Typically used in the telecommunication industry by phone manufactures to ensure that their products reach a certain quality standard for speech.

PESQ as an evaluation tool improves upon the non-automated the opinion based Mean Opinion Score (MOS) and the algorithm ranges from -0.5 to 4.5. There are two benefits of using PESQ over MOS. Firstly, the process is automated and does not require opinion-based surveys to take place, such as the case for MOS. Secondly, the process is repeatable and will return consistent results, which can be useful for tuning the filter to a certain specification.

There are two algorithms used for testing using PESQ, full reference and no reference. For full reference (FR) algorithm, the original signal without the added noise is used as a reference for the signal with the added noise. The benefit of this is the higher accuracy but can only be used for live networks. In comparison, no reference (NR) algorithm, only available for transport stream analysis needs just the noisy signal [2]. It comes at a cost of accuracy but can proof very useful in scenarios where the original clean signal is not available.

Wiener Filter

Unlike PESQ, Wiener Filters are used for enhancing noisy signals by filtering the noise out and returning the original cleaner speech signal. This section outlines how the Wiener Filter was utilized to remove the added train noise from the input signal and is compared to the signals generated previously in the “Adding Noise to Speech” section.

For the SNR equal to 30 there no perceptual difference between the noisy signal and the enhanced signal, this behavior is expected as at a SNR value of 30, the train noise is not strong enough for the enhanced signal to make a difference. For values of SNR 20 and below, there is a noticeable degradation in the original sound added by the train noise, hence a more noticeable difference is expected in the enhanced versus the noisy version. SNR of 20 provided the best noise reduction and while SNR 10 was useful in reducing the noise, it was not able to eliminate it completely. Finally, for SNR of 0 there was too much noise for the enhancement to make a difference.

Once the enhanced signals were compared with their noisy signal equivalent at each SNR values. For each of those the signals the PESQ values were found. PESQ is the industry standard for analyzing the quality of a signal, more details on this can be found in the PESQ

section above. The figure below shows the PESQ results for the SNR values ranging from 0, 10, 20, and 30.

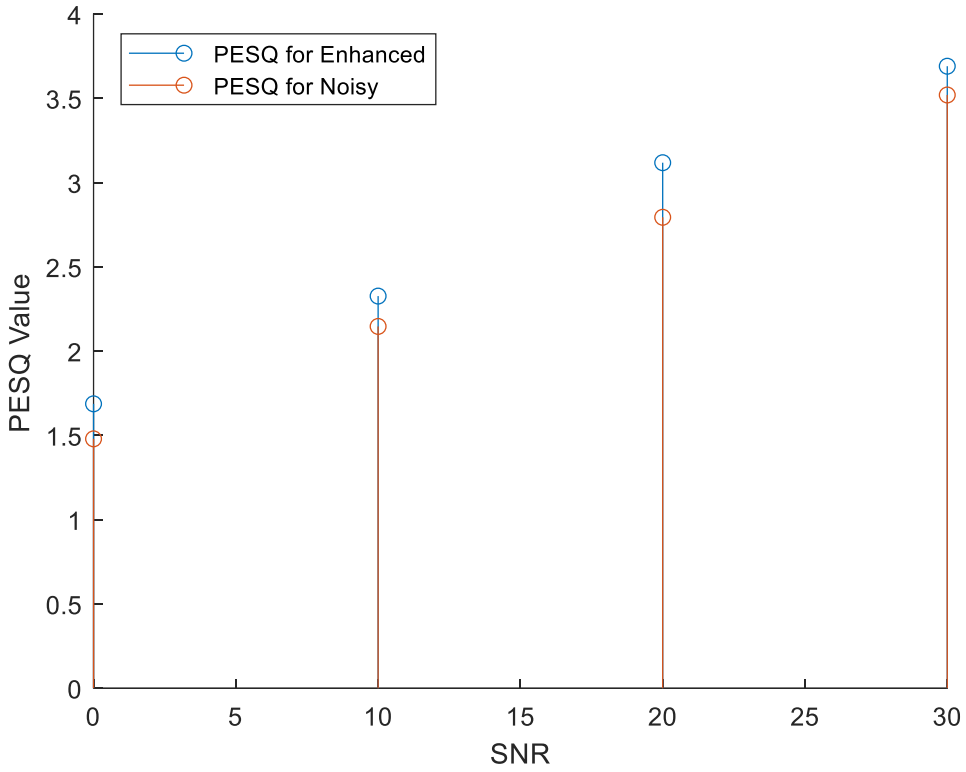

Figure 1: PESQ for Enhanced and Noisy Signals

The figure 1 given above shows the PESQ of the enhanced and the noisy signal for the following SNR values; 0, 10, 20, 30. As the results show, most of the significant differences appeared for the SNR values of 10 and 20. Confirming our previous claims, as SNR values of 10 and 20 had the best ratios of noise to signal and allowed for the enhanced to perform well as compared to the rest.

95% Confidence Interval of PESQ

Figures 2 through 5 given below show the PESQ value for a noisy and an enhanced signal for every clean signal given with each of the four noise signals, totally a 120 noisy and 120 enhanced signals. This process was then repeated at different SNR values, starting from 30 and going down to 0 at an interval of 10.

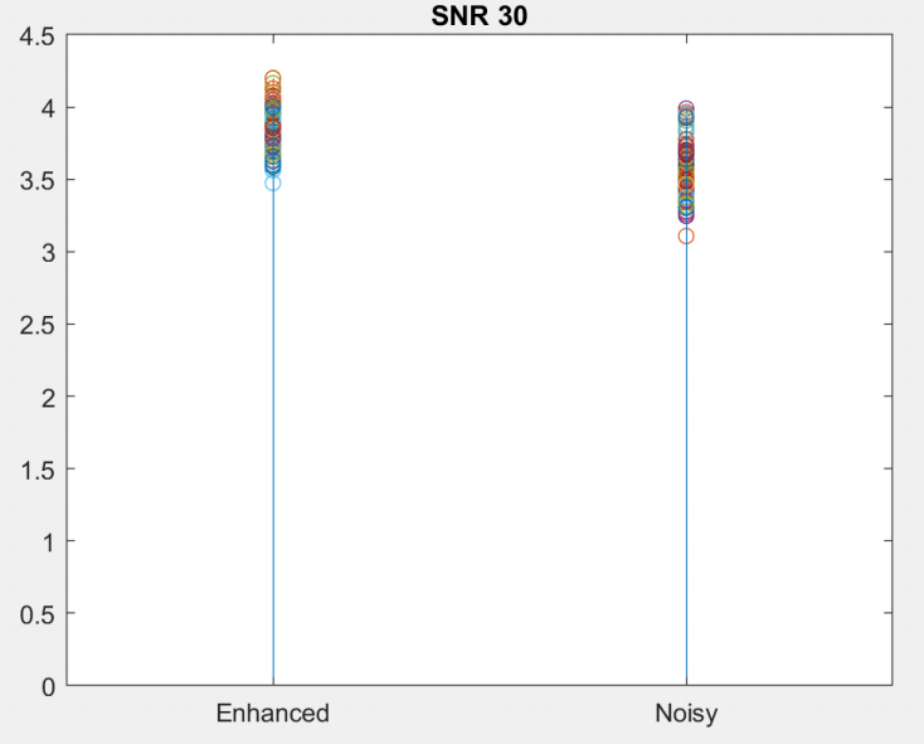

Figure 2: PESQ for Every Enhanced and Noisy Signals at 30 SNR

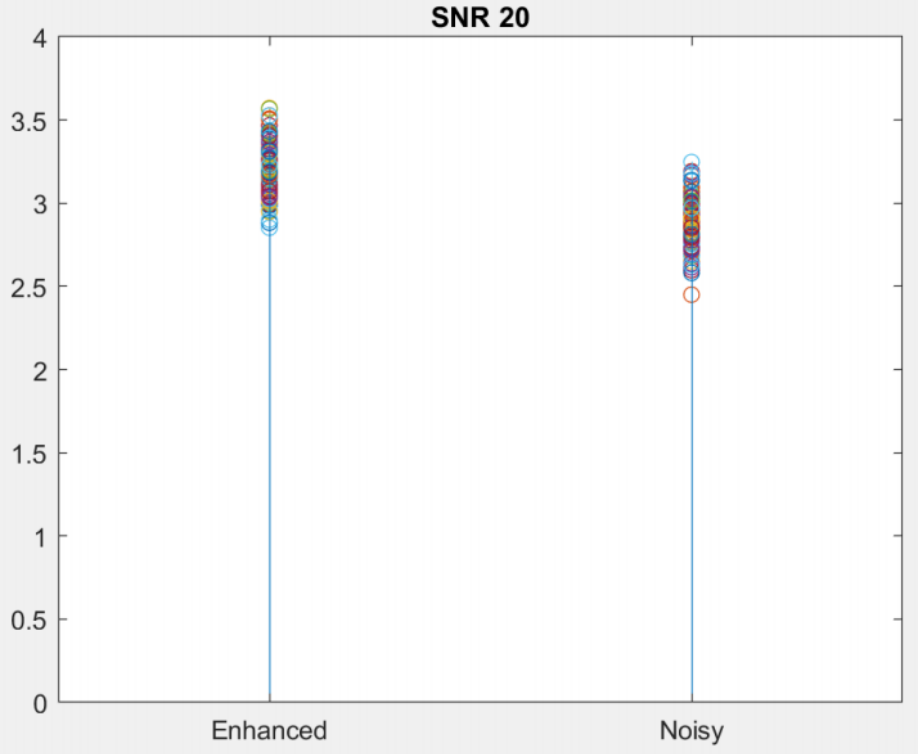

Figure 3: PESQ for Every Enhanced and Noisy Signals at 20 SNR

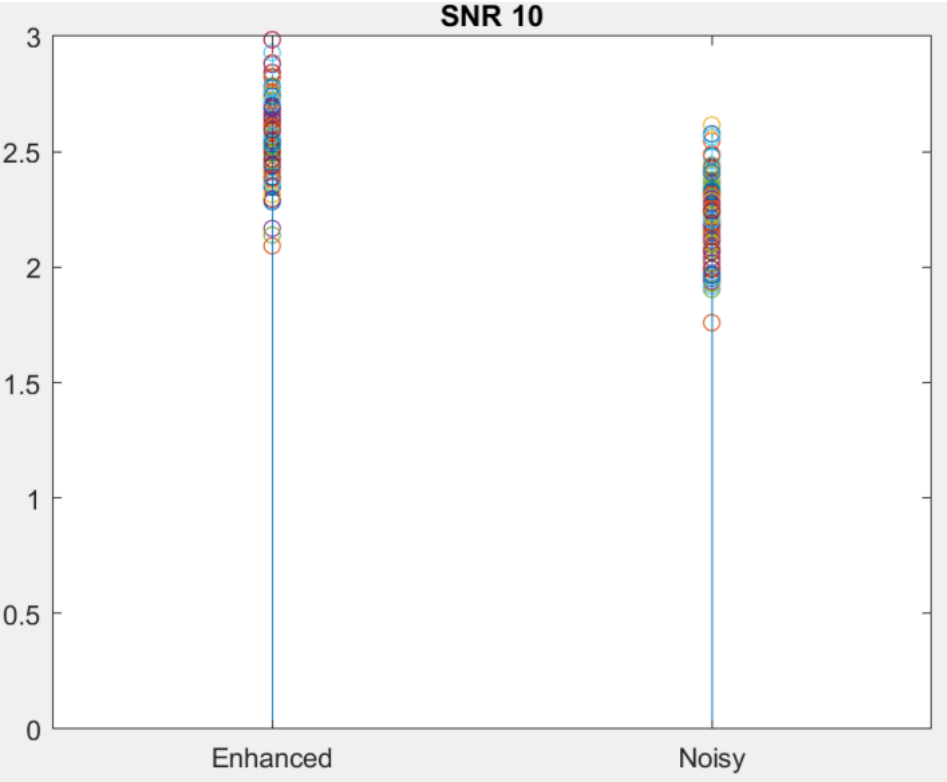

Figure 4: PESQ for Every Enhanced and Noisy Signals at 10 SNR

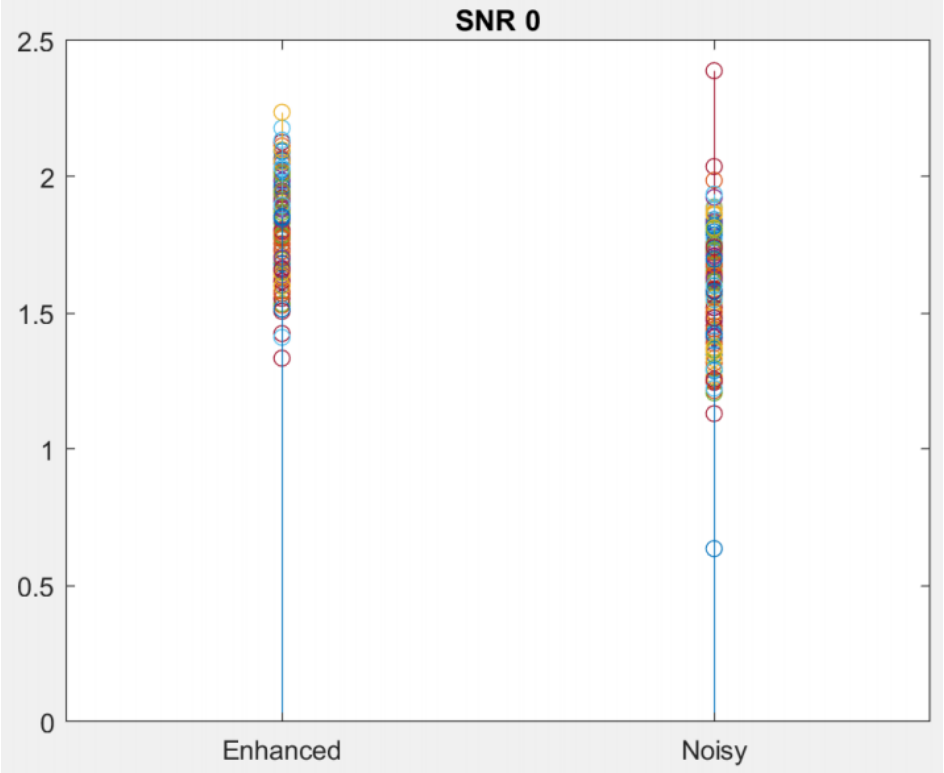

Figure 5: PESQ for Every Enhanced and Noisy Signals at 0 SNR

Once the four plots shown above were made their respective 95% confidence intervals were computed for their given data points. This was done to check if there is a statistical difference between the noisy and the enhanced signals.

In the case, the confidence intervals of the noisy and the enhanced signals overlap, then the signal is said to not be significantly distinguishable from one another. Else if there is no overlap present, then the two signals are set to be distinguishable.
Using the source given in [4], the confidence interval can be defined as a range of in which 95% of the values of the dataset are distributed within.

Confidence intervals are usually utilized when there is a lot of variation in the given data. To compute the confidence, a few key pieces of information are needed such as, the mean, the Z value, the standard deviation, and the total number of observations. For a 95% confidence interval the z value is 1.96. The equation below gives the outline of the confidence interval equation.

$$mean \pm Z\left(\frac{standard\ deviation}{\sqrt{number\ of\ oberservations}}\right)$$

Using this equation, the following confidence interval plot were developed for four different SNR values of 30, 20, 10, and 0. Each of the plots contain the enhanced and the noisy signals confidence interval.

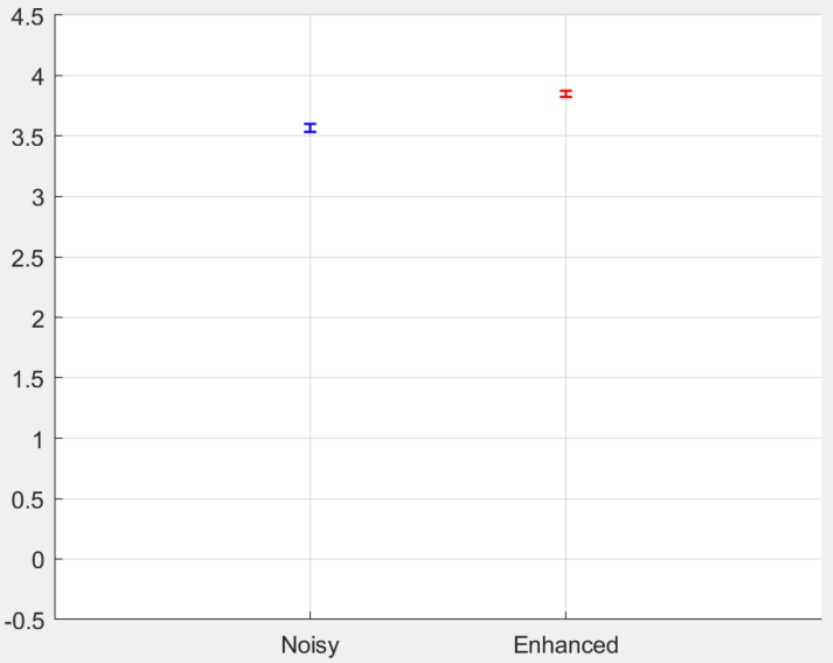

Figure 6: 95% Confidence Interval for PESQ for Every Enhanced and Noisy Signals at 30 SNR

The figure above shows the confidence interval for PESQ for a SNR value of 30. Here there is no overlap between the confidence intervals, hence, the signals are said to be statistically different.

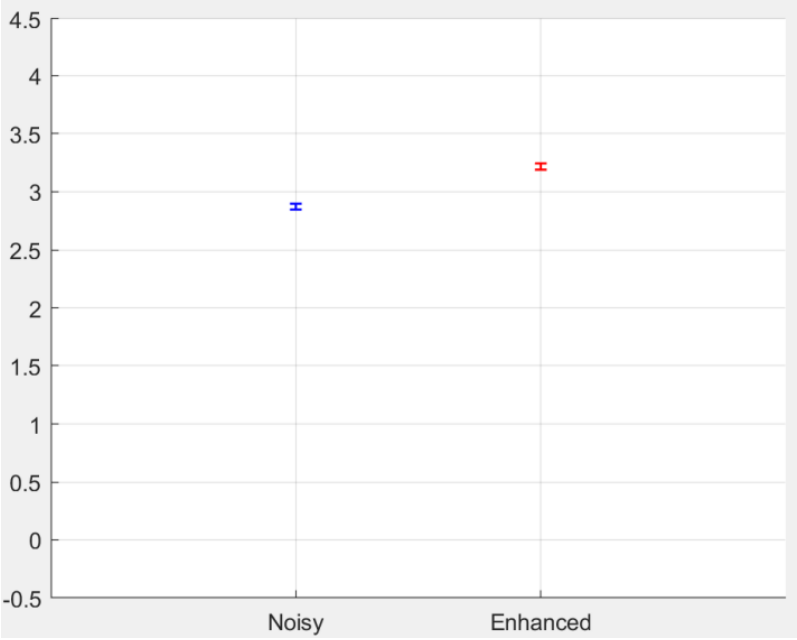

Figure 7: 95% Confidence Interval for PESQ for Every Enhanced and Noisy Signals at 20 SNR

Next, the figure above shows the confidence interval for PESQ for a SNR value of 20. Here there is no overlap between the confidence intervals, hence, the signals are said to be statistically different.

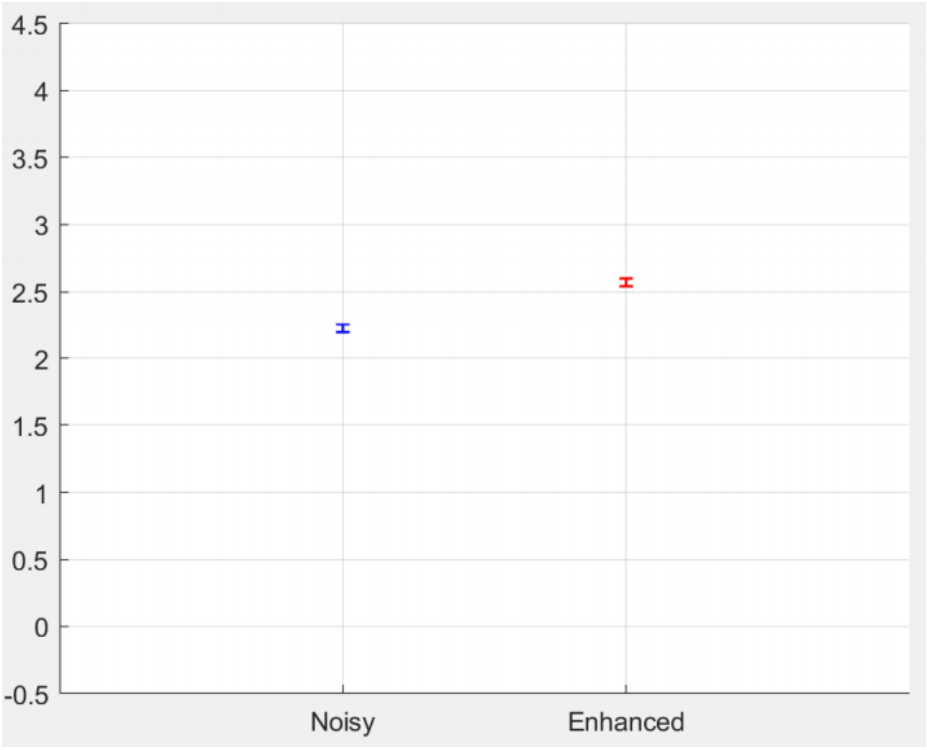

Figure 8: 95% Confidence Interval for PESQ for Every Enhanced and Noisy Signals at 10 SNR

Next, the figure above shows the confidence interval for PESQ for a SNR value of 10. Here there is no overlap between the confidence intervals, hence, the signals are said to be statistically different.

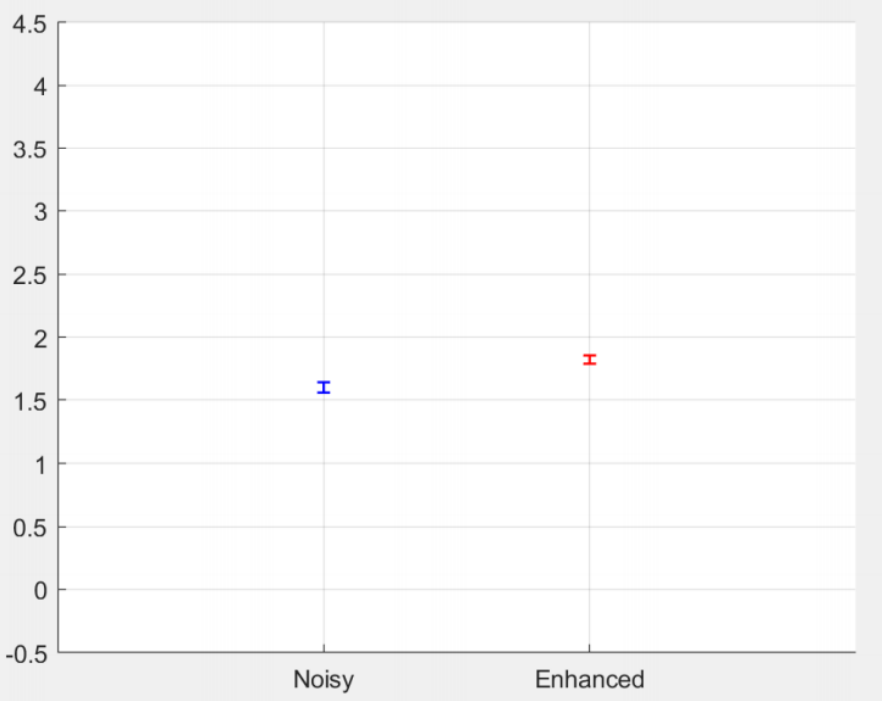

Figure 9: 95% Confidence Interval for PESQ for Every Enhanced and Noisy Signals at 0 SNR

Lastly, the figure above shows the confidence interval for PESQ for a SNR value of 0. Here there is no overlap between the confidence intervals, hence, the signals are said to be statistically different.

Summary and Conclusion

For the lab students learned a lot about speech enhancements. To do so, first understanding signal to noise ratio was crucial. With an SNR of 0, 10, 20, and 30 dB, the four types, exhibition, street, train, and white noises were added to 120 clean speech signals. After that was done, the noisy and the clean signals were compared. Next on the agenda was to learn and understand PESQ evaluation. This was done by using a Wiener filter on a noisy signal to then use PESQ to test the qualities of the enhanced and noisy signals at the different SNR values. Finally, a 95% confidence interval was plotted and calculated to compare the PESQ values. From that step in was concluded that the noisy and enhanced signals did not overlap at any SNR values. This means that the noisy and enhanced signals were significantly different from each other. This lab was very helpful as it helped with understanding signal processing in speech enhancement.

Acknowledgement

The MATLAB code for the Wiener filter implementation and the PESQ calculation was taken from [5].

References

1. M. Parchami, W.-P. Zhu, B. Champagne, and E. Plourde, "Recent Developments in Speech Enhancement in the Short Time Fourier Transform Domain", *IEEE Circuits and Systems Magazine*, pp. 45—77, September 2016.
2. https://en.wikipedia.org/wiki/PESQ
3. https://blog.empirix.com/take-a-closer-look-what-is-pesq/
4. https://www.mathsisfun.com/data/confidence-interval.html
5. P. C Loizou, Speech Enhancement: Theory and Practice, CRC Press, 2013.
6. https://www.spearline.com/blog/post/what-is--pesq-/
7. https://en.wikipedia.org/wiki/Signal-to-noise_ratio
8. https://en.wikipedia.org/wiki/Wiener_filter