Speech Enhancement Final Project

Nick Schooley, Fowad Sohail
Rowan University

– ECE 09.351: Final Project –

April 25, 2020

Table of Contents

1 Introduction

Speech enhancement is the process of improving the intelligibility and quality of degraded speech signals. Speech signals can be degraded due to a variety of reasons. For example, the quality of instrumentation and the environment that the speech signal is recorded play large roles in the degradation of the original signal. Reverberations, echoes, interference and ambient noise in an environment all contribute to noise that is added to the speech signal. Because these factors are incredibly variable, the field of speech enhancement becomes increasingly difficult - as any potential solution has to generalize to many different scenarios.

The applications of speech enhancement are very broad, including everything from mobile phones, voice over internet protocols (VOIP), teleconferencing systems, speech recognition and even hearing aids [1]. Because there are so many applications, speech enhancement is a highly researched area. As such, there are several fundamental techniques to speech enhancement, including Spectral Subtractive Algorithms, Wiener Filtering, statistical methods and subspace algorithms. Recently, there have even been Deep Learning approaches to speech enhancement [6]. In practical applications, however, the Discrete Fourier Transform (DFT) is typically preferred due to a variety of reasons including lower computational complexity and easier implementation. Nevertheless, as technology that makes use of speech signals becomes more commonplace in everyday life, the need for speech enhancement grows.

2 Objectives

- Examining clean sentence signals and noise signals.
- Mix the clean signals and noise signals at different signal to noise ratios.
- Enhance the mixed signals using a Wiener Filter to clean up the sentence signals.
- Use the Perceptual Evaluation of Speech Quality (PESQ) to determine the quality of the enhanced signals and compare it to the mixed signals before enhancement.
- Plot the 95$\%$ confidence interval of each signal to noise ratio and determine if there is a statistical difference between the enhanced signals and the original mixed signals.
- Determine overall usefulness of the Wiener Filter in clearing up noisy signals at different signal to noise ratios

3 Background - Speech Enhancement

The Spectral Subtractive Method to speech enhancement relies on the basic principle that an estimate of the clean speech signal $\hat{S}(k,l)$ can be obtained by the difference between an estimate of the noisy speech signal $X(k,l)$ and an estimate of the noise spectrum $\hat{V}(k,l)$. This relationship is seen in the equation below:

$$\hat{S}(k,l) = \mid X(k,l) - \hat{V}(k,l) \mid e^{j\theta X(k,l)} \quad (1)$$

The main advantage of Spectrical Subtractive Methods is that they are computationally simple and are quick enough for real world use. There are, however, mathematical assumptions made in Spectral Substractive Methods which degrade their performance and, as a result, they can not be claimed as the optimal speech enhancement solution. There is, however, active research ongoing to combat these issues.

Wiener Filtering based methods of speech enhancement estimate the speech signal $\hat{S}(k,l)$ as equal to the product of the gain function $W(k,l)$ and the noisy speech signal $X(k,l)$. This relationship is seen in the equation below:

$$\hat{S}(k,l) = W(k,l)X(k,l) \quad (2)$$

The Wiener Filter is a more complex method to speech enhancement, mainly as a result of the noise gain function $W(k,l)$. There are several ways to calculate Wiener Filter gain, as explored in the literature. However, the conventional way relies on a priori signal to noise ratios, which in and of itself presents a challenge.

As a whole, both Spectral Subtractive and Wiener Filtering methods offer promising results of enhanced speech. Spectral Subtractive Methods are intuitive and computationally simple, however may not always be optimal. Wiener Filtering methods, on the other hand, are typically more accurate at the expense of more complex calculations.

4 Adding Noise to Speech

In this part of the project, pure sentences have noise added to them. This is done by controlling the sound to noise ratio or SNR, which controls the amount of noise that is effecting the sound. Several white noise clips were tested at different SNR levels of 30db, 20dB, 10dB and 0dB. It was observed that as the SNR got lower, the static and noise over the clear white noise got higher. This made the signals harder to hear. Therefore, at 0dBs the signal was hardest to hear.

The different noise signals were also listened to individually and compared to the white noise. These noise signals are not as clear and there is not as much of an obvious difference between them, compared to the different white noises. Even though they are hard to hear clearly, it is still clear enough to portray the signal properly.

The SNR is found by taking the log, with base 10, of the L2 norm of the speech divided by the L2 norm of the noise. This relationship shows that with more noise, the fraction approaches 1. The log of 1 is 0, meaning the speech is harder to hear when the noise approaches the value of the speech. When there is more speech than the noise the value in the log will be greater and therefore the SNR value will be greater.

5 Perceptual Evaluation of Speech Quality (PESQ)

PESQ is an industry standard testing methodology that automates the assessment of speech quality. Phone manufacturers and telecommunications operators use PESQ to ensure a high quality of speech is experienced by a user of their systems. PESQ improves upon the Mean Opinion Score (MOS) method of speech quality control, which is not automated because people must be asked for their opinion. The greatest benefits of PESQ are that it gives an automated and accurate representation of quality and that it is traceable, meaning that a change in PESQ score means something has changed in the pipeline. There are two testing algorithms used in PESQ, full reference (FR) and no reference (NR). An FR algorithm uses the original signal as a reference for comparison. FR algorithms have the highest accuracy but can only be conducted for live networks. An NR algorithm only uses the degraded signal that has been effected by noise and does not have access to the original signal. NR algorithms are not as accurate as FR algorithms and is limited to transport stream analysis [6] Regardless of the algorithm used, PESQ readings range from -0.5 to 4.5.

6 Wiener Filter

The Wiener Filter is used to enhance a noisy signal and make it sound more clear. In this lab a Wiener Filter was used to clear up one of the input signals that has been mixed with the train noise and different SNR values. These enhanced signals were compared to the signals generated in the ”Adding Noise to Speech” section. When the SNR was 30, there wasn't much of a difference between the enhanced and noisy signals, considering that the train noise is not that strong at a SNR of 30 before being enhanced. At a SNR of 20, the train noise starts to become more obvious for the noisy signal, but it is still very hard to hear the train noises. Once the SNR gets down to 10 and 0 the noisy signal becomes really tough to make out what the original signal is saying. Although, the enhanced signal is still extremely easy to make out. There is still some noise associated with these signals, but the enhanced signal allows for the original signal to still be heard and understood.

After using the Wiener Filter to find the enhanced signals, the PESQ of each signal was found. The PESQ is discussed in the section ”Perceptual Evaluation of Speech Quality (PESQ)” and is basically a industry standard to determine the quality of a signal. The PESQ was found for both the enhanced signals at each SNR and the noisy signal at each SNR. This was done to compare if there is actually a difference between the two signals at each SNR mathematically. These values were then plotted to make it easy to see which signals produced a higher PESQ and therefore a higher quality signal. This plot can be seen in 1.

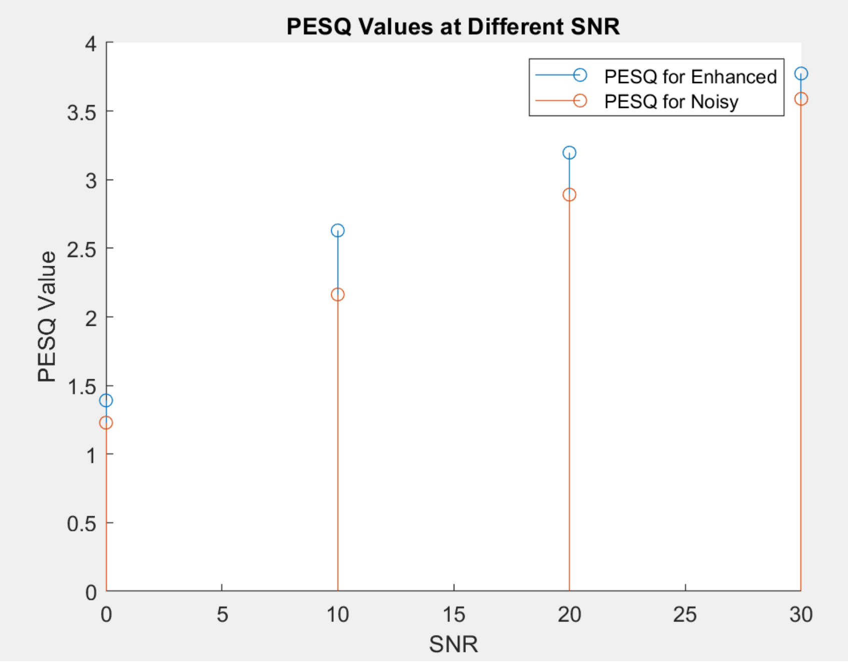

Figure 1: PESQ Calculations for Enhanced and Noisy Signals

This figure shows that the PESQ for the enhanced signal was larger than the noisy signal for every SNR value. This backs up what was observed earlier on when the signals were listened to, especially how it is seen that the biggest difference in the two signals in terms of PESQ came for the SNR values of 10 and 20.

7 95% Confidence Interval of PESQ

In the 4 figures below, an enhanced signal and a noisy signal was found for every clean sentence signal at each of the 4 noise signals for a total of 120 enhanced signals and 120 noisy signals. This was done at each of the 4 SNR values, 0, 10, 20, 30. The figures show the PESQ value that was calculated for each of these signals and the plots show the variation in the enhanced and noisy signals internally and the difference between the two types of signals.

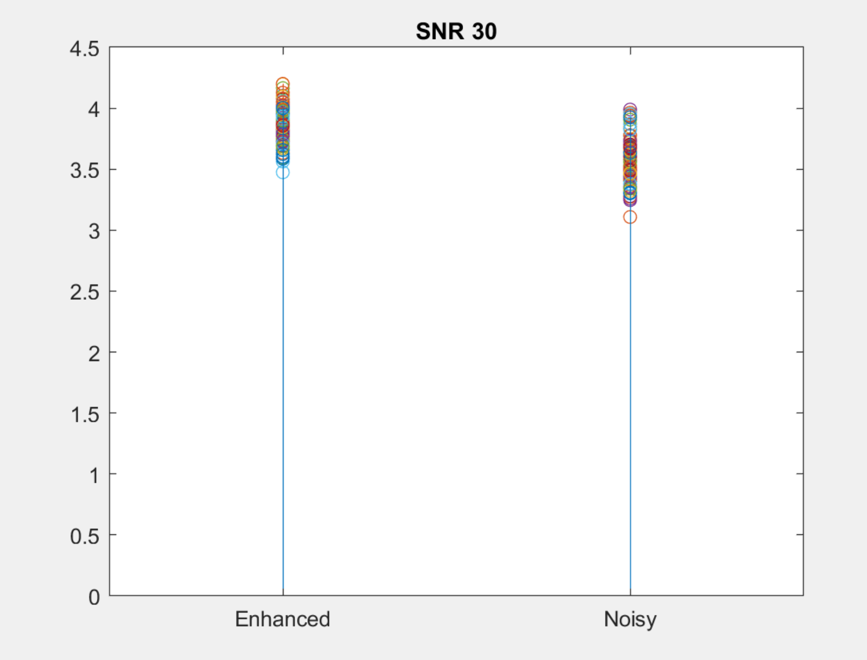

Figure 2: PESQ Calculations for Every Enhanced and Noisy Signals at 30 SNR

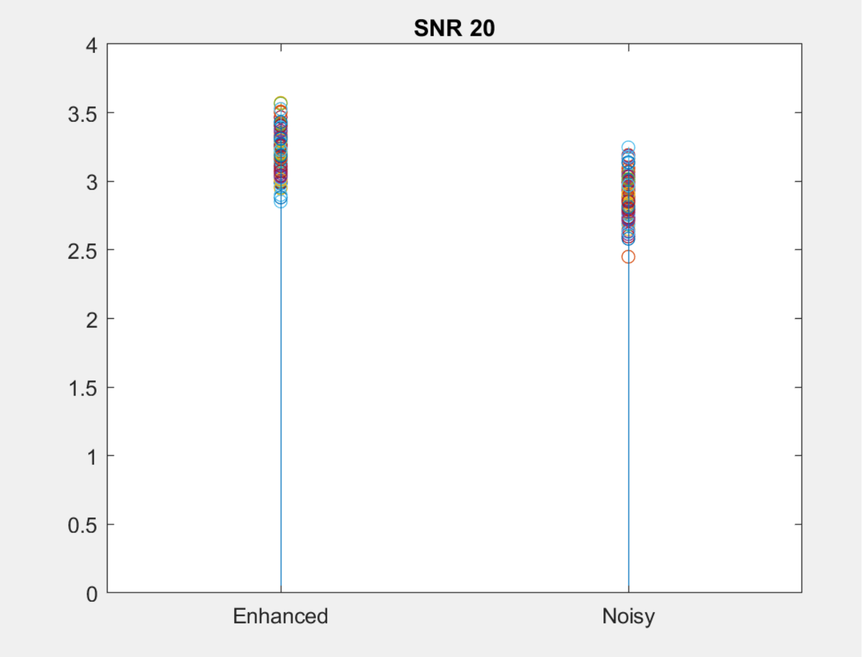

Figure 3: PESQ Calculations for Every Enhanced and Noisy Signals at 20 SNR

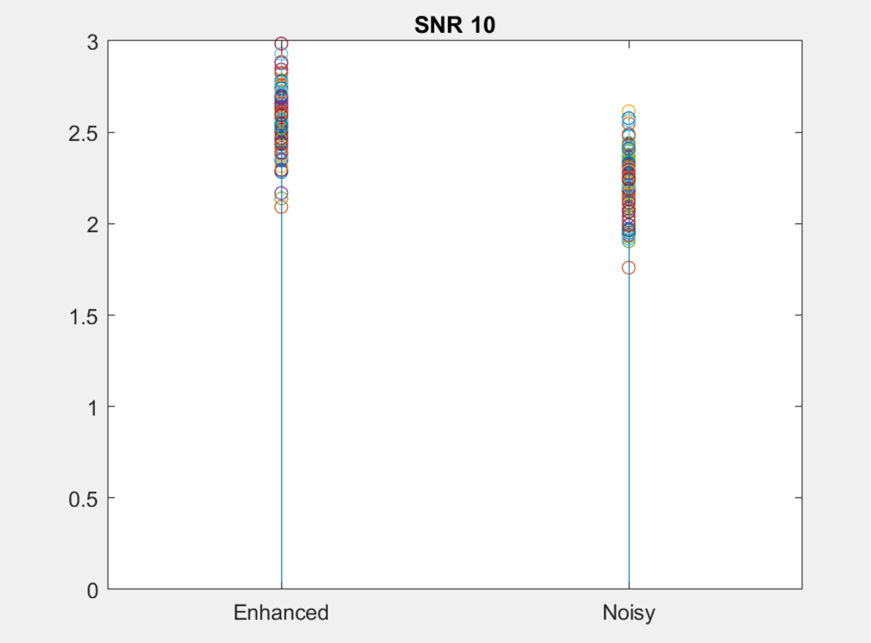

Figure 4: PESQ Calculations for Every Enhanced and Noisy Signals at 10 SNR

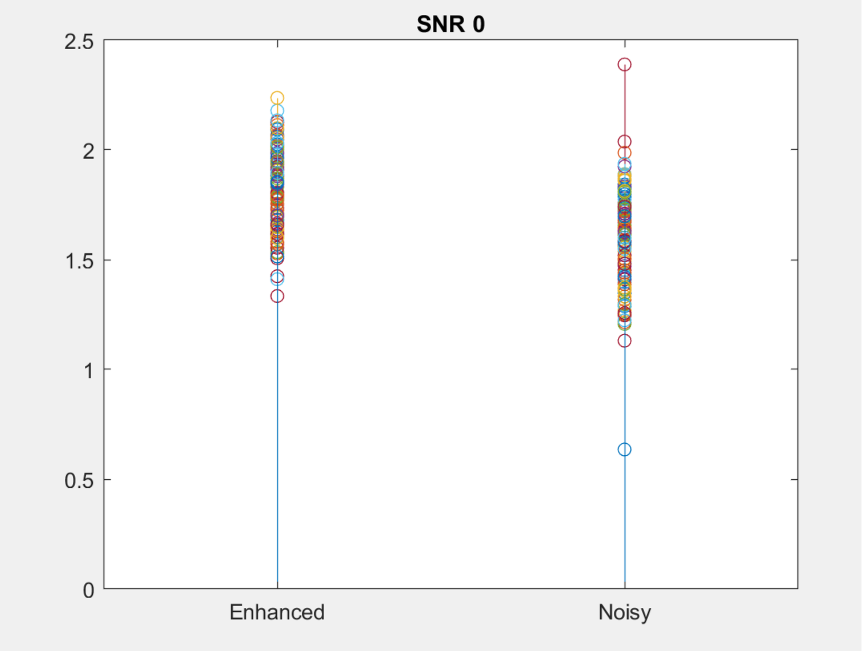

Figure 5: PESQ Calculations for Every Enhanced and Noisy Signals at 0 SNR

After these four plots were made, the 95% confidence interval was determined using all of the data points. In doing this, the enhanced and noisy signal will be able to be assessed to see if they are statistically distinguishable or not. If the intervals overlap the two signals would be indistinguishable, but if they don't the two signals are statistically distinguishable. Although, in order to do this, it must first be understood what confidence intervals are and how they can be calculated. All of the information that will be discussed was found in [4]. A confidence interval is a set of values that the user is sure their true answer is in, therefore the 95% percent confidence interval says that the user is confident the true value will fall in the confidence range 95 percent of the experiments. Confidence intervals are used when data has a lot of variation and is helpful with limiting this variation. The confidence interval can be calculated with a few simple items such as the the mean, the standard deviation and the number of observations. The z value is also needed, which is a value that is associated with different confidence levels and can be found on a z value table. For a 95% confidence the z value would be 1.96. Once these values are all done, a true 95% confidence can be represented in the equation below.

$$Mean \pm Z\frac{standarddeviation}{\sqrt{numberofobservations}} \tag{3}$$

With this knowledge on the confidence intervals, the confidence interval for each of the 4 SNR plots above was found and plotted below. There is a different confidence interval for the enhanced and noisy signals and the two intervals were compared to find if they were statistically distinguishable or not.

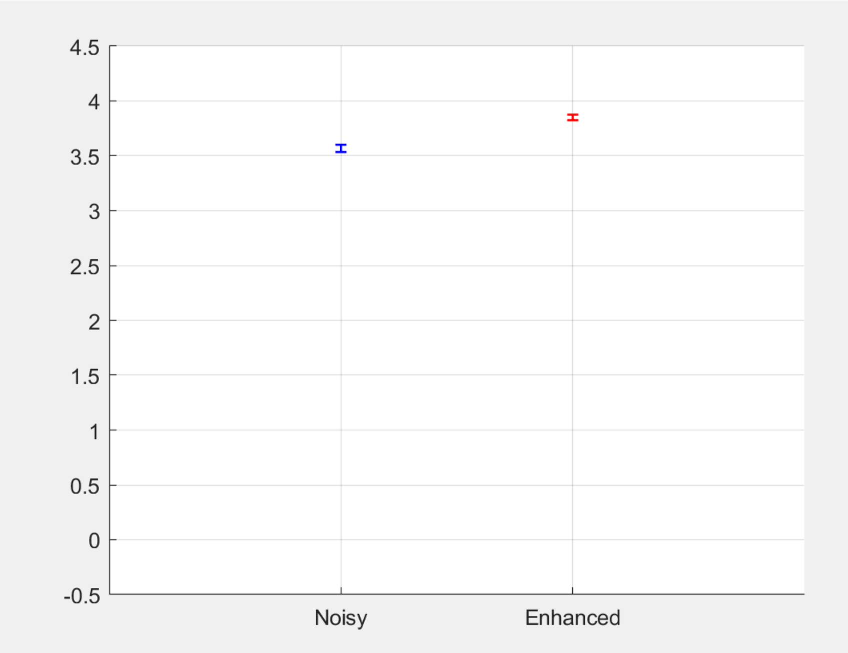

Figure 6: 95 % Confidence Interval for PESQ Calculations for Every Enhanced and Noisy Signals at 30 SNR

The first confidence interval plot, seen in Figure 6, is for the SNR value of 30. It is seen that there is a confidence interval for both the noisy and the enhanced signals, but, these two intervals do not over lap. This shows that at a SNR of 30, the enhanced and the noisy signals are distinguishable statistically.

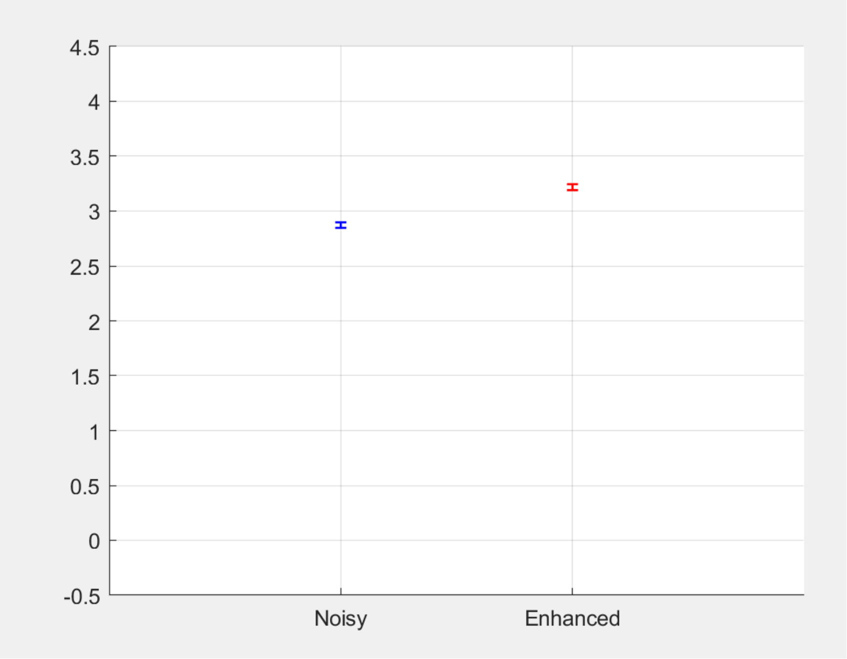

Figure 7: 95 % Confidence Interval for PESQ Calculations for Every Enhanced and Noisy Signals at 20 SNR

The second confidence interval plot, seen in Figure 7, is for the SNR value of 20. With the same set up as the 30 SNR plot, it is again seen that there is no overlap between the confidence intervals and therefore the signals are statistically distinguishable at a SNR of 20.

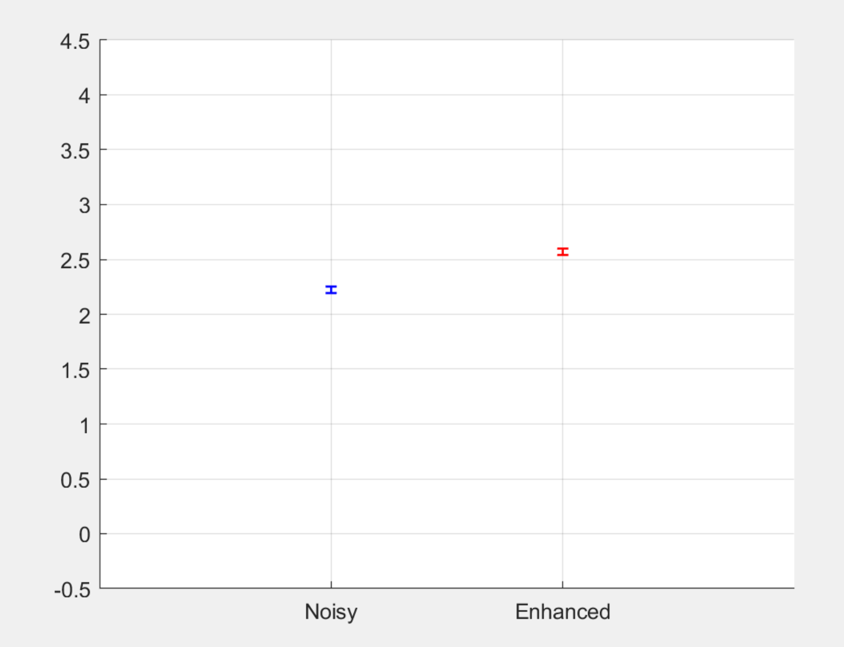

Figure 8: 95 % Confidence Interval for PESQ Calculations for Every Enhanced and Noisy Signals at 10 SNR

The third plot seen in Figure 8 coincides with the SNR of 10. Again, there is no overlap in the two intervals so the two groups of signals can be seen as statistically distinguishable.

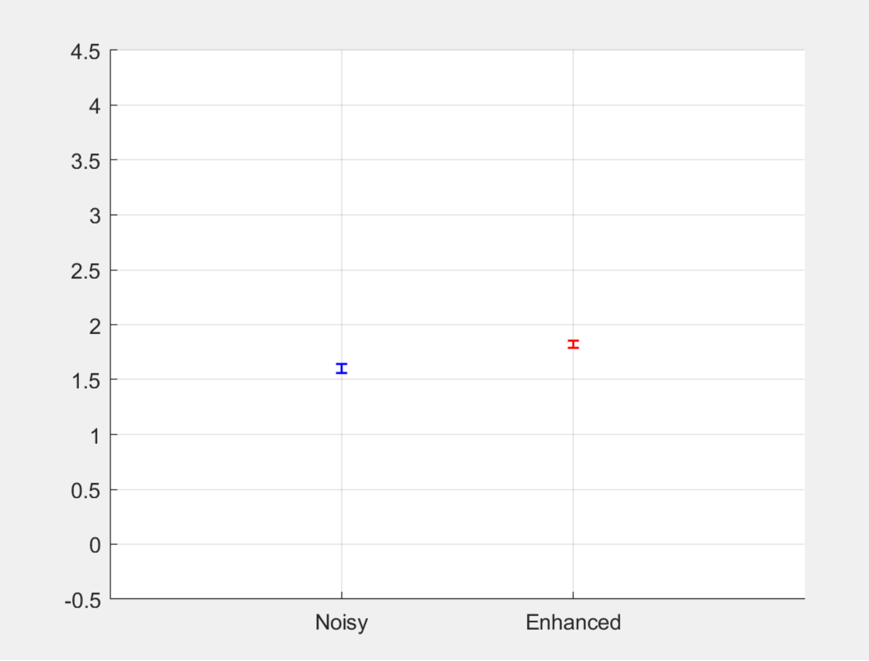

Figure 9: 95 % Confidence Interval for PESQ Calculations for Every Enhanced and Noisy Signals at 0 SNR

Lastly, the SNR of 0 plot is seen in Figure 9. There is again no overlap of the confidence intervals, so just like the the three other SNR values, the signals are statistically distinguishable.

8 Summary and Conclusion

Overall, all of the goals for this project were accomplished using MATLAB and other resources found online and in the literature. At first, clean sentence signals were taken and had different noise signals added to them at different signal to noise ratios. After this, a Wiener Filter was used to enhance the quality of the noisy signals, and the noisy signals were compared to enhanced signals by finding the PESQ. Lastly, the 95% confidence interval was used to determine if the enhanced signals and the noisy signals were statistically distinguishable from each other at each of the four SNR values. It turns out that the enhanced signals were statistically distinguishable from the noisy signals at each of the different SNR values. However it is worth mentioning that the confidence intervals were closer to overlapping for some SNR values than they were for others. For the SNR values of 10 and 20, there was about a 0.5 PESQ difference between the two intervals, but for SNRs of 0 and 30 there was not much of a difference at all. This is likely due to signal to noise ratio itself, since at a ratio of 30 there is not much noise for the Wiener Filter to limit anyway, so it makes sense that the noisy signal and enhanced signal would be similar. The opposite idea would be applied to an SNR of 0 because for this SNR the signal is barely able to be heard due to all of the noise, so there should be a large difference between the enhanced signal and the noisy signal. The reason this may not be the case is due to the amount of noise present for the SNR, the Wiener Filter may not be able to make much of a difference. Although, it was clearly seen for the SNR value of 10 and 20 that the Wiener Filter really does make a difference in the PESQ of the different signals. As a whole, the properties and applications of speech enhancement metrics and techniques were observed in this project.

9 Acknowledgements

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] “Perceptual Evaluation of Speech Quality,” Wikipedia, 13-Apr-2020. [Online]. Available: https://en.wikipedia.org/wiki/PerceptualEvaluationofSpeechQuality. [Accessed: 25-Apr-2020].

[7] Y. Xu, J. Du, L. Dai and C. Lee, ”A Regression Approach to Speech Enhancement Based on Deep Neural Networks,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 1, pp. 7-19, Jan. 2015.

[8] PESQ – Perceptual Evaluation Speech Quality. (2018, January 07). Retrieved April 25, 2020, from https://www.opalesystems.com/pesq/