



FINAL PROJECT
SPEECH
DENOISING/ENHANCEMENT

09.05.2023

Digital Signal Processing

GROUP 9

BI12-076 Mai Hải Đăng

BI12-074 Đoàn Đình Đăng

BI12-073 Trần Hải Đăng

BI12-040 Trần Ngọc Ánh




BI12-099 Nguyễn Thanh Đức

1. Introduction	3
2. Materials	3
3. Methods/ Procedures	3
3.1. General terminology	3
3.2. Spectral Subtraction Method	4
3.3. Spectral Filtering	6
3.3. Wiener Filtering	6
3.4. Signal Subspace Approach (TNA)	9
4. Comparison	10

1. Introduction

- The field of Digital Signal Processing (DSP) plays a crucial role in improving the quality of audio signals by mitigating unwanted noise and enhancing the overall listening experience. In various real-world scenarios, such as audio recordings in noisy environments or degraded audio transmissions, the presence of noise can significantly degrade the quality and intelligibility of the desired audio signal.
- In this report, we focus on evaluating four commonly used methods: Spectral Subtraction Method, Spectral Restoration Method, Wiener Filtering, and Signal Subspace Approach.
- The purpose of this study is to investigate the performance of these methods under different noisy environments. By applying each method to audio recordings corrupted with various types and levels of noise, we can analyze their effectiveness in suppressing noise while preserving the desired audio content. Additionally, we aim to compare the results obtained from each method and evaluate their respective strengths and limitations in terms of noise reduction and audio quality improvement.

2. Materials

- Audio files: .wav
- All the codes, images, sound files can be found in the following github repository:
<https://github.com/Incomprehensibilitative/Speech-Denoising-Enhancement>
- Noise: Only taken the first few seconds, < 10s
 -  RESTAURANT AMBIENCE • 10H Busy Coffee Shop Background Noise
 -  Airport Sounds - One Hour!!! The Most Complete Airport Ambience!
 -  EPIC THUNDER & RAIN | Rainstorm Sounds For Relaxing, Focus or Sleep | Whi...

3. Methods/ Procedures

3.1. General terminology

3.1.1. The Difference between frequency and spectral when analyzing digital signal

- Frequency refers to the rate at which a periodic waveform repeats per unit of time. It represents the number of complete cycles of a waveform that occur in one second and is typically measured in Hertz (Hz)
- Spectral refers to the frequency content or distribution of a signal. It involves analyzing a signal in the frequency domain to understand the different frequencies that make up the signal and their corresponding amplitudes or power levels. The spectral analysis provides information about how the signal's energy or power is distributed across various frequencies.

- Frequency tells us how fast a signal oscillates, while the spectral content tells us which specific frequencies are present in the signal and how much energy they contain.

3.1.2. What are the spectral properties

- **Power Spectral Density (PSD):** The power spectral density describes the power distribution of a signal over its frequency range. It provides a more detailed analysis of the signal's frequency content by quantifying the power or energy present in different frequency bins.
- **Spectral Shape:** Spectral shape refers to the overall pattern or shape of the frequency spectrum. It includes characteristics such as the slope, peaks, valleys, and bandwidth of the spectrum. Spectral shape can provide insights into the timbre or tonal qualities of an audio signal.
- **Spectral Envelope:** The spectral envelope represents the general outline or contour of the frequency spectrum. It captures the amplitude variations across different frequency bands and helps identify the fundamental frequency and harmonic structure of the signal.
- **Spectral Peaks and Valleys:** Peaks and valleys in the frequency spectrum indicate regions of high or low energy concentration, respectively. Peaks correspond to strong frequency components, such as fundamental frequencies of harmonics, while valleys represent areas with less energy or spectral gaps.
- **Spectral Flatness:** Spectral flatness measures the ratio between the peak and average values in the spectrum. It indicates whether the energy is evenly distributed across the frequency bins (flat spectrum) or concentrated in a few dominant components (peaked spectrum).
- **Spectral Centroid:** The spectral centroid represents the center of mass or average frequency of the spectrum. It provides information about the spectral balance of the signal, indicating whether the energy is concentrated towards low or high frequencies.
- **Spectral Roll-off:** The spectral roll-off is the frequency at which a certain percentage of the total signal energy resides below it. It describes the rate at which the energy decreases beyond a particular frequency and can provide insights into the overall brightness or dullness of the signal.

3.2. Spectral Subtraction Method

3.2.1. Definitions:

- A method for restoration of the power spectrum, or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum.
- Spectral subtraction is a widely used algorithm in audio denoising applications. It operates by estimating the long-term average background noise and subtracting it from the audio signal. If the estimated average background noise has peaks in it that do not correspond to the current audio signal, these peaks have nothing to cancel and instead generate musical noise¹. The proposed architecture uses a novel approach to estimate environmental noise from speech

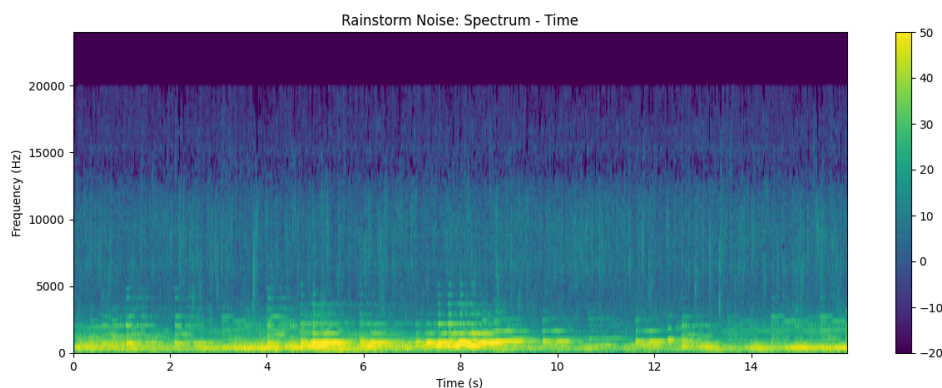
adaptively. After estimating the noise from the input speech the noise samples are subtracted, making it noise-free.

3.2.2. Method: (How Spectral Subtraction works)

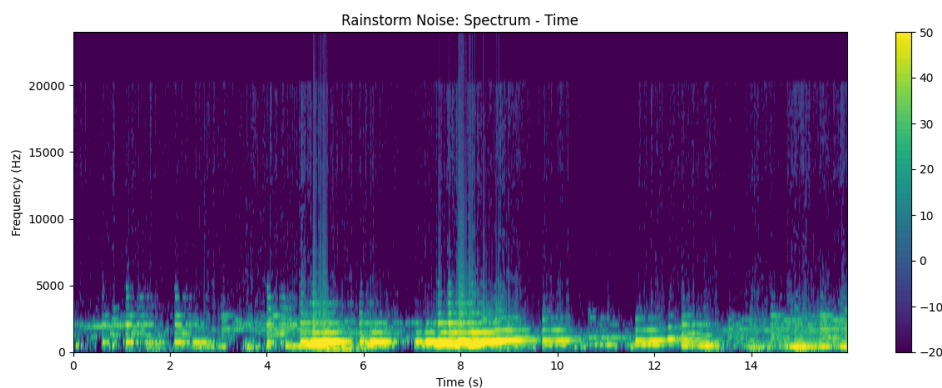
1. **Analysis:** The input signal, which consists of both the desired signal and the background noise, is divided into small frames of a fixed duration. These frames are typically overlapping to capture the time-varying characteristics of the signal.
2. **Fourier Transform:** Each frame is then subjected to the Fourier transform to convert the time-domain signal into the frequency domain. This transformation yields a representation of the signal's spectral content.
3. **Noise Estimation:** In this step, an estimate of the noise spectrum is obtained. This can be done in various ways, such as using a noise-only segment of the signal or assuming that the lowest energy spectral components represent the noise.
4. **Spectral Subtraction:** The estimated noise spectrum is subtracted from the spectrum of the observed signal on a frequency-by-frequency basis. This subtraction operation attenuates the noise components in the spectrum.
5. **Inverse Fourier Transform:** After the spectral subtraction, the modified spectrum is converted back to the time domain using the inverse Fourier transform.
6. **Overlap and Add:** The frames are recombined by overlapping and adding them together to reconstruct the filtered signal. Overlapping frames help to ensure smooth transitions between adjacent frames.

3.2.3. Outputs

- Before and After picture in time domain and frequency domain for each audio file, comment
- There are 3 denoised examples, here is one of them: The Rainstorm noise in the background of a piano song:
Noisy Audio:



Denoised Audio:



3.3. Spectral Filtering

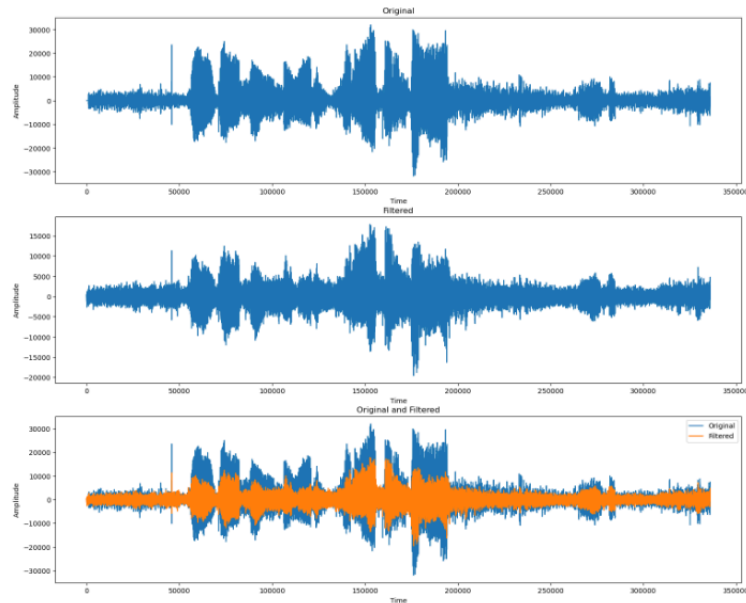
3.3.1. Definition:

- The basic idea: Estimate the power spectrum density of the noise. The noise corresponds to the low variance of the PSD, while the signal corresponds to the maximum variance. Due to the nature of speech mostly relies on specific frequency bands leading to higher power in those regions of the PSD. Noise, on the other hand, is often random and spread across a wide frequency range hence their power is distributed across the entire spectrum. This leads to a relatively flat or more uniform power distribution in the noise component of the PSD.
- Then we select the frequency region where the PSD is at max to be our's clean audio.

3.3.2. The detail outline of the steps involved in spectral filtering:

- **Analysis:** Transform the degraded signal from the time domain to the frequency domain using techniques (the Fourier transform or the Short-Time Fourier Transform (STFT)) to provide a representation of the signal's spectral components.
- **Noise estimation:** Estimate the noise or unwanted components in the degraded signal by analyzing portions of the signal that are considered noise-only (lower PSD) by calculating the power spectral density.
- **Filtering:** Once the noise is estimated, we create a specific threshold that only allows the range where the power spectrum is highest (the desired signal) to pass through.
- **Synthesis:** Transform the restored spectral components back to the time domain using the inverse Fourier transforms. This yields the enhanced signal with improved quality and reduced unwanted artifacts.

3.3.3. Outputs



3.3. Wiener Filtering

3.3.1. Definition:

- The Wiener filter looks at both the noisy signal and the noise itself, assuming that they have certain properties. It then calculates a filter that tries to estimate the original, clean signal by taking into account the characteristics of the desired signal and the noise. The goal is to minimize the difference between the filtered signal and the original clean signal.
- The Wiener filter uses the concept of power spectral density, which describes the power of a signal at different frequencies. By analyzing the power of the desired signal and the noise at different frequencies, the filter can determine how to adjust the noisy signal to make it clearer.
- The filter operates by multiplying the Fourier transform of the observed signal by a transfer function that is the ratio of the desired signal's PSD to the sum of the desired signal's PSD and the noise PSD. This transfer function is known as the Wiener transfer function.
- However, as you can see above, the Wiener filter assumes certain conditions that may not always hold in practical scenarios:
 - + Stationarity: The Wiener filter assumes that both the desired signal and the noise are stationary. Stationarity means that their statistical properties, such as mean and variance, do not change over time. In practical terms, this assumption implies that the characteristics of the signal and noise remain constant throughout the duration of the filtering process.
 - + Known statistical properties: The Wiener filter assumes that the statistical properties of both the desired signal and the noise are known. This includes their probability distributions and power spectral densities. In practice, however, it is often challenging to

have complete knowledge of these properties, and approximations or estimations may be used instead.

- + Linearity: The Wiener filter assumes linearity of the system. This means that the relationship between the desired signal and the observed signal can be represented by linear equations. If the relationship is highly nonlinear, the performance of the Wiener filter may be compromised.
- + Additive noise: The Wiener filter assumes that the noise is additive, meaning it is combined with the desired signal in a straightforward manner. It assumes that the noise is not correlated with the desired signal. If the noise has other characteristics, such as multiplicative or non-additive properties, the Wiener filter may not be appropriate.

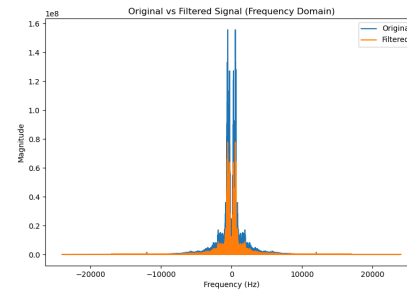
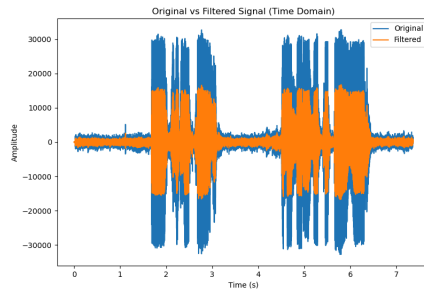
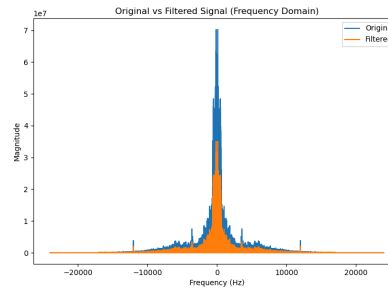
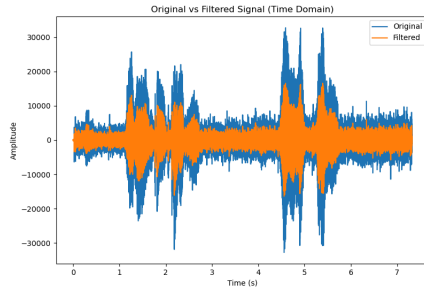
3.3.2. Step by step explanation:

- Compute the cross-power spectral density (CPSD):
 - + The cross-power spectral density measures the correlation between two signals, (`channel_1` and `channel_2`) in the frequency domain. More specifically, it measures the strength of the relationship between the frequency components of the two signals.
 - + The Fast Fourier Transform (FFT) transforms the input signals, `channel_1` and `channel_2`, from the time domain to the frequency domain. Then, the FFT of `channel_1` is multiplied by the complex conjugate of the FFT of `channel_2` to calculate the CPSD
- Compute the auto-power spectral density (APSD) of the reference channel:
 - + The auto-power spectral density measures the power distribution of a single signal in the frequency domain. It provides information about the power content at different frequencies in the signal.
 - + The reference channel is `channel_1` and the FFT of `channel_1` is then squared to calculate the APSD.
- Compute the Wiener filter:
 - + The Wiener filter utilizes the CPSD and APSD to estimate the signal-to-noise ratio (SNR) at each frequency component. The SNR is estimated as the ratio of the CPSD to the APSD
 - + The Wiener filter then calculates a weighting factor for each frequency component. This weighting factor is the ratio of the CPSD to the sum of the CPSD and APSD.
 - + This ratio represents the filter gain at each frequency interval. Higher values indicate more noise suppression, while lower values preserve more of the noisy signal.
- Apply the Wiener filter to the input signal:
 - + The Wiener filter is multiplied with the FFT of `channel_2` to obtain the filtered frequency-domain representation of `channel_2`.
 - + The inverse FFT is then applied to the filtered frequency-domain signal to convert it back to the time domain.
 - + The real component of the complex-valued inverse FFT result is extracted, as the filtered signal is expected to be real-valued.
- Combine the filtered channel with the reference channel:

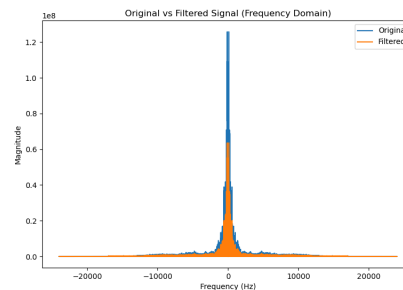
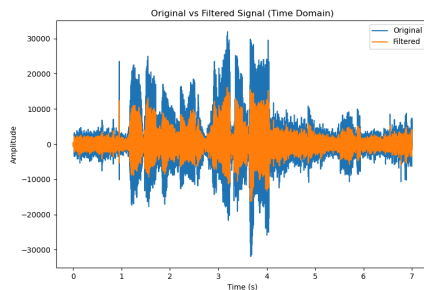
- + The filtered channel, `filtered_channel_2`, is combined with the reference channel, `channel_1`, to form the filtered data.
- + Then, the two channels are stacked horizontally, creating a two-column array representing the filtered data.

3.3.3. Outputs:

- Airport



- Cafe
- Rainstorm



3.4. Signal Subspace Approach

3.4.1. Definition:

The subspace-based approach exploits the fact that the covariance matrix of a noisy speech signal frame can be decomposed into two mutually orthogonal vector spaces: a signal (+noise) subspace and a noise subspace. Noise reduction is obtained by discarding the noise subspace completely, while modifying the noisy speech components in the signal (+noise) subspace.

3.4.2. Method:

- **Frame Segmentation:** Divide the noisy speech signal into short frames.
- **Noise Estimation:** Select a portion of the signal that contains only background noise and estimate the statistical properties of the noise. This can be done by considering frames or segments of the signal where there is minimal or no speech activity.
- **Covariance Matrix Calculation:** Calculate the covariance matrix of each frame or segment of the noisy speech signal. This matrix represents the statistical relationships between different elements of the signal.
- **Eigenvalue Decomposition:** Perform eigenvalue decomposition or Singular Value Decomposition (SVD) on the covariance matrix. This decomposition yields the eigenvectors and eigenvalues of the matrix.

$$A = U \Sigma V^T$$

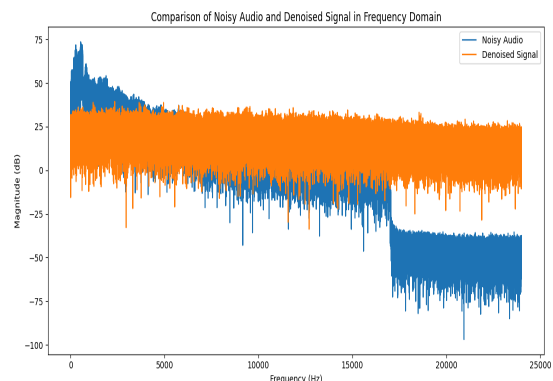
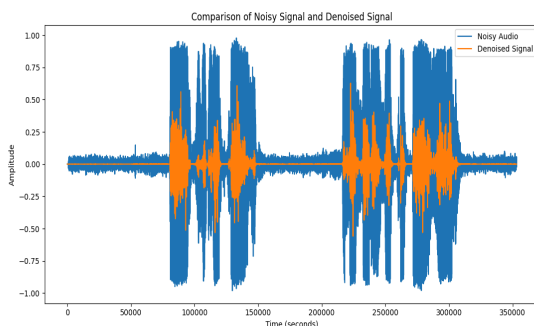
Formula of SVD:

The matrix sigma is rectangular diagonal, and has the same dimension as matrix A. The numbers on the diagonal are the singular values of matrix A arranged in descending order, every other entry is 0. The singular values of matrix A are $\sqrt{\text{eigenvalues of A}}$

- **Eigenvalue Sorting:** The eigenvalues obtained from the decomposition represent the variance of the signal along the corresponding eigenvectors. Sorting the eigenvalues in descending order allows us to identify the eigenvectors associated with the dominant signal components and those associated with the noise components.
- **Noise Subspace Identification:** Based on the sorted eigenvalues, a threshold or cutoff point is determined to distinguish the noise subspace from the signal (+noise) subspace. The eigenvectors corresponding to eigenvalues below the threshold are considered to belong to the noise subspace.
- **Noise Subspace Projection:** The eigenvectors associated with the noise subspace are used to construct a projection matrix that projects the signal onto the noise subspace. This projection matrix effectively removes the noise components from the signal, as the noise subspace is orthogonal to the signal (+noise) subspace. (the project operation)
- **Signal Subspace Modification:** Modify the remaining eigenvectors or eigenvector coefficients, which correspond to the signal (+noise) subspace.

3.4.3: Output:

Cafe:



4. Comparison

- If a row is shared between different columns, the characteristics are shared among the techniques/ methods. Otherwise it just belong to one technique
- These comparisons are both general characteristics and also specific implementations and the noise condition we chose.
- The audio for Signal Subspace Method wasn't the best to decide it advantage and disadvantage

	Spectral Filtering	Spectral Subtraction (SS)	Wiener Filter	Subspace Method
Concept	- Simple, utilize PSD to it best	- Very basic and simple method.	- Simple to understand, intuitive	- The most complex method - Required understanding of linear algebra subspace, covariance, ...
Dependencies	- Dependent on the distribution of the signal power	- For the best performance, it assumes that the noise is stationary within each frame and that the noise spectrum estimation is accurate. In reality, most noises are non-stationary, and their characteristics may change over time.		Estimating accurately the threshold or cutoff point determined to distinguish the noise subspace from the signal (+noise) subspace.
		- Dependent on the accuracy of the noise and signal models - Requires a reliable estimate of the noise and signal spectra		
Implementation	- Require good knowledge of spectral properties and how to manipulate - Can use many different method to estimate the noise		- Easiest to implement and test	Not much resource, documentation to follow Because of lacking information, I followed techniques and implemented but output attenuated both the signal and noise
Advantages	- Better than (SS) when removing more dynamic noise, while without distorting the original too much	- Best when it come to removing noise from the original audio - Can provide not effective noise reduction in stationary noise	- Provides an optimal solution. - Achieve good noise reduction while preserving the signal quality - Adaptive	

		scenarios		
Disadvantages	<ul style="list-style-type: none"> - Introduce musical noise or residual noise artifacts - Signal distortion can occur if the estimated noise power is inaccurate 	<ul style="list-style-type: none"> - Requires prior knowledge about the signal(s) - Linear: poor handling of nonlinear phenomena such as harmonics, intermodulation, or clipping - Can introduce some undesirable effects such as ringing, overshooting, or blurring, due to its frequency-domain smoothing. 		