CHALMERS
UNIVERSITY OF TECHNOLOGY


# Speech enhancement for non-stationary noise around a machine cabin

Master's thesis in Sound and Vibration

YILIANG ZHOU

**DEPARTMENT OF Architecture and Civil Engineering**
**Division of Applied Acoustics**

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023
www.chalmers.se

Master's thesis 2023

# Speech enhancement for non-stationary noise around a machine cabin

YILIANG ZHOU


CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
*Division of Applied Acoustics*
Chalmers University of Technology
Gothenburg, Sweden 2023

Speech enhancement for non-stationary noise around a machine cabin

YILIANG ZHOU



Supervisor: Frenne Nicklas, Volvo Construction Equipment AB
Examiner: Jens Ahrens, Division of Applied Acoustics

Master's Thesis 2023
Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: The various sound sources around the testing VOLVO L60H wheel loader.
The illustration picture is from the previous thesis student Tomoya Otsuka.
Image URL: https://www.volvoce.com/europe/en/products/wheel-loaders/l60h/

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2023

Speech enhancement for non-stationary noise around a machine cabin
YILIANG ZHOU
Division of Applied Acoustics
Chalmers University of Technology

# Abstract

This thesis is mainly concerned with the solution for speech enhancement in the presence of non-stationary noise around the machine cabin. This allows outside speech to enter the cabin and reduces unwanted noise. The application scenario of this work is a signal processing system using a microphone array outside the cabin to capture signals and using different algorithms to enhance the speech signals. With the help of this system, the machine operator is able to get speech information in a noisy environment. The noise sources in this situation are more complex and non-stationary. Examples of noise sources include engine noise, traffic noise, or other construction activities. Previous work done by Tomoya chose the microphone array configuration and developed the beamforming method. From the results of the valuable work, it is found that beamforming is able to increase the signal-to-noiseratio (SNR) in the current situation, but the sound quality and SNR are still limited due to the low input SNR and non-stationary noise environment. Therefore, modified beamforming and new methods are implemented in this work. noise cancellation(NC) predicts the transfer path for the noise signal and removes it by controlling the minimum error of the output. Noise suppression(NS) uses the scheme of spectral subtraction to subtract the noise spectrum from noisy speech spectrum. A combination of beamforming and noise cancellation and a combination of beamforming and noise suppression method are developed and evaluated. The result shows a better performance for this low input SNR and non-stationary noise case.

Keywords: Speech enhancement, Non-stationary noise environment, Beamforming, Noise cancellation, Noise estimation.

# Acknowledgements

I appreciate all the valuable advice and instructions from Nicklas Frenne, my supervisor at Volvo Construction Equipment AB. He was always active and helpful during the whole thesis program. It was a pleasure to work with him and Volvo Construction Equipment. I would like to show my appreciation to Jens Arhens, my supervisor and examiner at Chalmers. He has given me a lot of support and suggestions for my thesis. Thank you Sophie Poulsen(my dear girlfriend) for patiently pushing me to finish this thesis faster. Finally, I wish to thank my family and friends for their unconditional love and encouragement throughout my life.

Yiliang Zhou, Copenhagen, October 2023

# List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

| | |
|---|---|
| SNR | Signal-to-Noise Ratio |
| STFT | short-time Fourier transform |
| NC | Noise Cancellation |
| FRF | Frequency Response Function |
| FFT | Fast Fourier Transform |
| IFFT | Inverse Fast Fourier Transform |
| LMS | Least Mean Squares |
| FXLMS | Filtered-X Least Mean Squares |
| NS | Noise Suppression |
| CNNs | Convolutional Neural Networks |
| RNN | Recurrent Neural Networks |
| LSTM | Long Short-Term Memory |
| MCRA | Minima Controlled Recursive Averaging |
| PSD | Power Spectral Density |
| LLR | Log-Likelihood Ratio |
| WSS | Weighted Spectral Slope |
| SPP | Speech Presence Probability |
| PESQ | Perceptual Evaluation of Speech Quality |

# Contents

# List of Figures

# List of Tables

# 1
# Introduction

Speech communication is essential in many environments, including vehicle cabins. However, background noise and engine noise can interfere with speech intelligibility and pose challenges to clear communication between the machine operator and other individuals. At Volvo Construction Equipment AB, some machines are equipped with cabins that are well acoustically isolated from the environment. In the cabin, normally a good noise reduction has been achieved which means the noise and sound from outside are isolated. However, there is a need to allow specific sound sources, such as speech, to be passed through.
To address this problem, an important point is to use a system to capture mixed signals and process the signal to get clear signals, ie. Speech signals. Previous work has been done using a 10-microphone array and beamforming to develop a system for speech enhancement and noise reduction. However, the results were not optimal in this low-input SNR and non-stationary noise situation. Therefore, the aim of this thesis is to optimize and complete the signal-processing system implement new methods for speech enhancement and noise reduction in the vehicle cabin try to achieve better performance, also investigate the limitations of different methods.

## 1.1 Purpose

By improving the existing system, this research will contribute to enhancing speech intelligibility and enabling clear communication in noisy vehicle cabins. This thesis will explore and compare different signal processing techniques and their limits to optimize the performance of the system, with a focus on improving speech quality and reducing noise. The results of this research will be valuable for developing effective communication systems in noisy environments, which can have important applications in various industries, including communication, construction, and transportation.

## 1.2 Related Work

Ideally, we would like a speech enhancement process to improve both quality and intelligibility. It is possible to reduce the background noise but at the expense of introducing speech distortion, which in turn may impair speech intelligibility. Hence, the main challenge in designing effective speech enhancement algorithms is to suppress noise without introducing any perceptible distortion in the signal. Thus far, most speech enhancement algorithms have been found to improve only

the quality of speech [10]. This is also a limitation of this thesis that we are getting less improvement in speech intelligibility.

The solution to the general problem of speech enhancement depends largely on the application at hand, the number of microphones or sensors available, the relationship (if any) of the noise to the clean signal, and the characteristics of the noise source or interference. The interference could be noise-like (e.g., fan noise) or speech-like such as an environment (e.g. a restaurant) with competing speakers [3].

Furthermore, the number of microphones available can influence the performance of speech enhancement algorithms. Typically, the larger the number of microphones, the easier the speech enhancement task becomes. Adaptive cancellation techniques can be used when at least one microphone is near the noise source. The noise may also be statistically correlated or uncorrelated with the clean speech signal, like the minimum mean-square error information[4].

Regarding the on-site situation in this thesis and these three facts, three kinds of corresponding methods are implemented. They are Beamforming, noise cancellation, and noise suppression.

## 1.3 Previous Work and Setup

Previous work was done by Tomoya by exploring the optimal arrangement for beamforming microphone positions and recording relevant noise and speech signals. Comparison between different microphone arrangements was evaluated and the final choice had the best performance among all. The optimal setup is shown on the left side of Figure. 1.1, and 6 locations are selected around the cabin for measuring. There are 10 microphones in all in the beamforming setup. The microphone setup is symmetric along the Y-axis. The names are marked on the right side of Figure. 1.1. (Front Right 1 and 2, Rear Right 1 and 2, and Corner West 1, on the other side they are Front Left 1 and 2, Rear Left 1 and 2, and Corner East 1).

After choosing the optimal position, three measurements were conducted: frequency response function measurements, speech, and machine recordings. There are 6 different locations where the source was placed and measurements of the frequency response functions and recordings of noise speech were taken. The measurement plan is shown in the Appendix A.

After getting all the measurements, the delay and sum beamforming method was used to implement the signal processing process. SNR improvement is shown in the Result Chapter 4.


**Figure 1.1:** Optimal microphone positions from the previous work, drawing, and the real picture.
Image: Nicklas Frenne, March 22, 2022

## 1.4 Structure of the thesis

Following the introduction, chapter 2 illustrates the theory that is needed in this thesis. Theory of different methods and theory for speech quality evaluation. Chapter 3 gives the methodology that was applied in conducting this thesis. Beamforming, noise cancellation, and noise suppression are used individually and combined. Chapter 4 will display the results of different methods, both objective and subjective results will be shown. Also, the robustness test result is included in this chapter. Further discussions on the results and limitations include elements that could have been handled differently, and suggestions of real applications are included in Chapter 5.

# 2
# Theory

In most applications, the aim of speech enhancement is to improve the quality and intelligibility of degraded speech. The improvement in quality is highly desirable as it can reduce listener fatigue, particularly in situations where the listener is exposed to high levels of noise for long periods of time (e.g., manufacturing).[10]

## 2.1 Stationary noise and Non-Stationary noise

Noise is an unwanted random variation that can corrupt or degrade a signal. Understanding the characteristics of noise is crucial for developing effective noise reduction techniques.
Noise can be broadly categorized into two types: stationary noise and non-stationary noise.

### 2.1.1 Stationary Noise

Stationary noise is a type of noise that exhibits a constant statistical property over time. In other words, the statistical parameters of stationary noise, such as mean and variance, remain constant or change very slowly over short time intervals. Stationary noise is often characterized by its spectral properties, which can be represented by a power spectral density (PSD) that remains relatively constant across time. Common examples of stationary noise include white noise, machine noise, and certain types of background hum.
Mathematically, if we denote the speech signal corrupted by stationary noise as $s[t]$ and the stationary noise component as $v[t]$, we can model the observed signal $x[t]$ as the sum of the two:

$$x[t] = s[t] + v[t] \tag{2.1}$$

The stationary noise $v[t]$ is often assumed to be uncorrelated with the speech signal $s[t]$ and can be represented by a time-invariant noise PSD $N(f)$. This assumption allows for the development of various noise reduction techniques based on spectral subtraction, Wiener filtering, and other linear filtering approaches.
Examples of stationary noise include white Gaussian noise, pink noise, and many background noises such as the hum of electrical equipment or the hiss in an audio recording.
Stationary noise is often characterized by its mean and variance, which remain constant over time. This means that the average value of the noise signal does not change, and the spread or dispersion of the noise values remains the same.

### 2.1.2 Non-Stationary Noise

Non-stationary noise, in contrast, is noise whose statistical properties change significantly over time. Non-stationary noise sources are time-varying and can be more challenging to model and suppress compared to stationary noise. Examples of non-stationary noise include complex engine noise, traffic noise, babble noise, and environmental sounds with time-varying characteristics.
Modeling non-stationary noise requires more advanced techniques, as traditional stationary noise reduction methods may not be effective. Adaptive filtering algorithms and time-frequency analysis approaches are commonly employed to track and adapt to the changing properties of non-stationary noise in real time.
Mathematically, we can represent the observed signal $x[t]$ corrupted by non-stationary noise $v'[t]$ as:

$$x'[t] = s[t] + v'[t] \tag{2.2}$$

Unlike in stationary noise, the non-stationary noise $v'[t]$ is not assumed to be uncorrelated with the speech signal $s[t]$. Its characteristics may vary significantly across time and frequency, requiring more sophisticated algorithms for effective noise reduction.
Non-stationary noise poses additional challenges for noise reduction algorithms because the statistical properties of the noise change rapidly. As a result, traditional noise reduction techniques that assume stationary noise may not be effective in reducing non-stationary noise.
To deal with non-stationary noise, advanced techniques such as time-frequency analysis and adaptive filtering are often employed [13]. These techniques aim to track and adapt to the changing characteristics of the noise over time, allowing for more effective noise reduction.

## 2.2 Beamforming

Beamforming is a versatile approach to spatial filtering that has found applications in diverse fields such as radar, sonar, wireless communication, and medical imaging. It is a technique that focuses a transmitted or received signal in a specific direction, effectively enhancing the signal-to-noise ratio and improving system performance.
The concept of beamforming dates back to the early 20th century, with its roots in antenna design and radar systems. Over the years, it has evolved and adapted to various technological advancements. Beamforming relies on constructive and destructive interference to steer a beam of electromagnetic waves in a desired direction. This can be achieved through various algorithms and techniques, including delay-and-sum, minimum variance, and adaptive beamforming[2].
Beamforming is a technique used to enhance the desired signal by selectively combining the signals from an array of microphones. It works by focusing the array response towards the desired direction while suppressing the interference from other directions. The beamformer weights are typically determined based on the spatial properties of the signal and the noise, as well as the geometry of the microphone array.

In Figure. 2.1, the upper diagram displays a target location with differing delays ($t_2$, $t_1$, and 0) applied to the microphone signal. These delays are deliberately chosen such that the signal constructively interferes when summed. While signals from other directions remain at their original amplitude because the delays applied do not align the incoming signals. Thus increasing the signal-to-noise ratio (SNR). However, depending on the signal frequency and spacing of the microphones, a signal from an undesired direction may be amplified.
In the process of creating delay and virtually moving the source location, an arrival time difference $\Delta t$ is calculated as follows:

$$\Delta t = \frac{(x_A - x_B)^2 + (y_A - y_B)^2}{c} \tag{2.3}$$

where $x_A$, $y_A$ and $x_B$, $y_B$ are the coordinates of the evaluation points.
Since the signals in this thesis are mainly speech signals with noise, a broadband beamforming process was implemented. The method process is shown in section 3.1 where a frequency domain beamforming is performed. The inverse frequency response function is used to store the phase and amplitude information for the delay-and-sum beamforming.

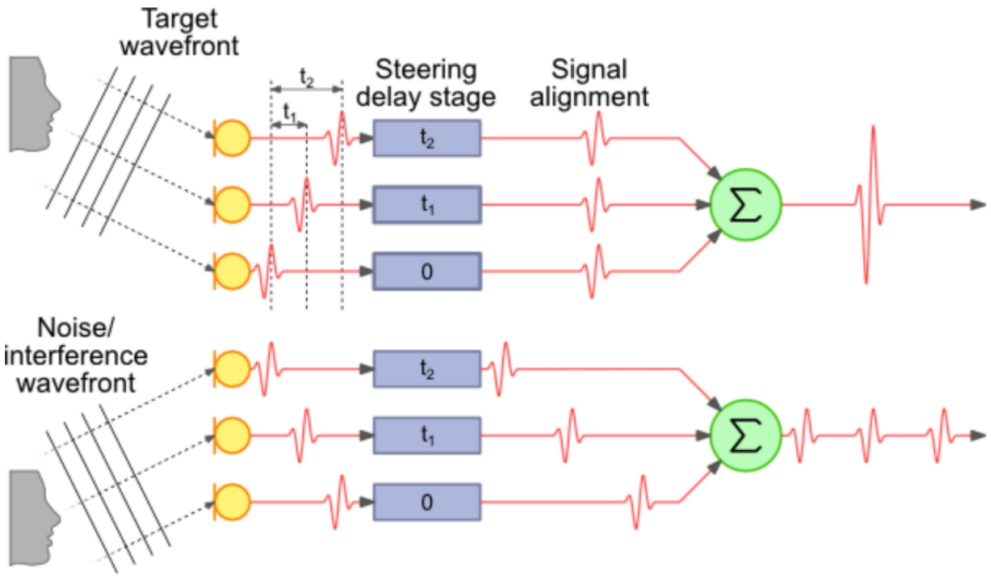


**Figure 2.1:** Delay-sum beamforming diagram. Image from[11]

## 2.3 Noise cancellation

Noise cancellation algorithms are a class of digital signal processing techniques used to suppress unwanted noise from a corrupted signal, enhancing the quality of the desired signal. These algorithms find extensive applications in speech processing, audio enhancement, and communication systems. This chapter provides an overview of noise cancellation principles, and explains the assumptions made in this thesis in order to perform noise cancellation.

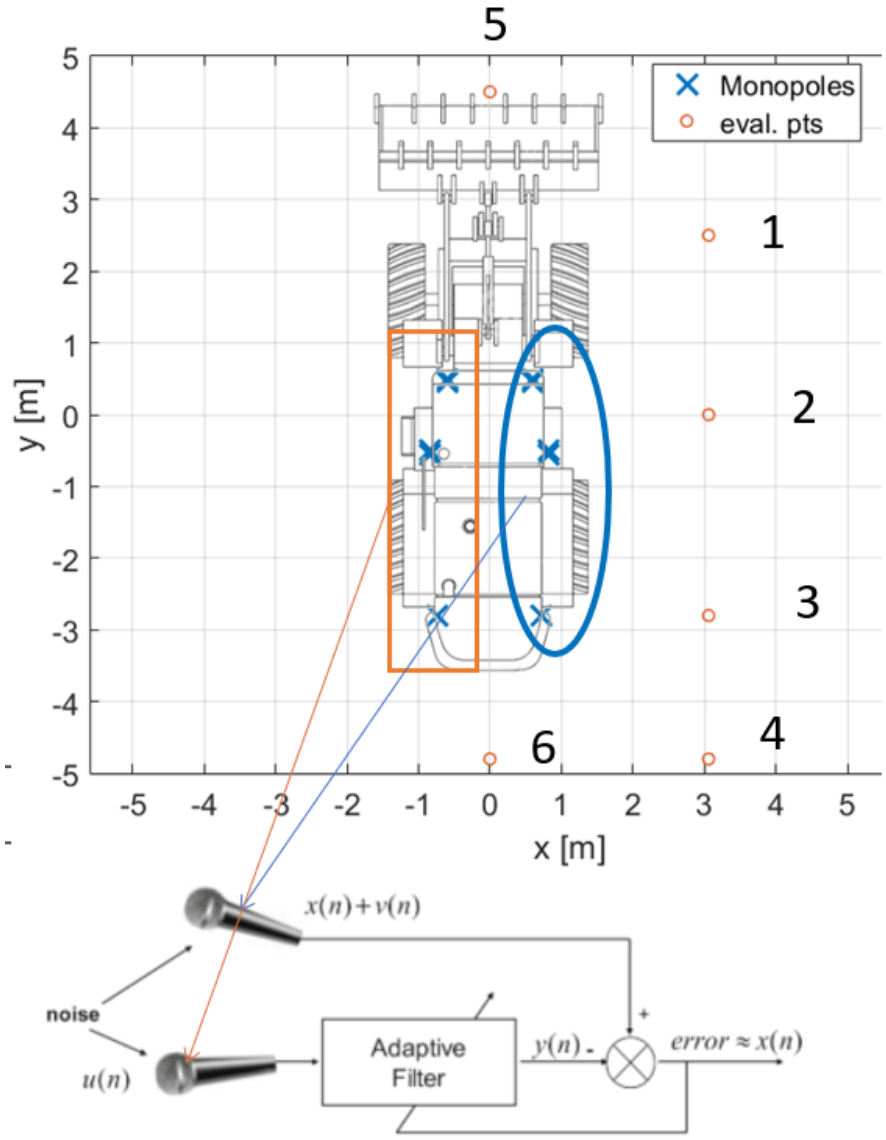


**Figure 2.2:** Noise Cancellation theory in current case[12]

Noise cancellation is based on separating the desired signal from the background noise, assuming that the noise can be modeled or estimated. From Figure. 2.2, the corrupted signal received by the speech and noise microphone can be represented as the same as Equation 2.1:

$$x[t] = s[t] + v[t]$$

where $x[t]$ is the observed signal, $s[t]$ is the desired signal, and $v[t]$ is the noise component. The goal of noise cancellation algorithms is to estimate or approximate the noise $v[t]$ and then subtract it from $x[t]$ to obtain an enhanced version of the desired signal $s[t]$[12].
In this case, suppose the speech comes from position 2 and the noise mainly around the cabin. We assume the left side 5 microphones primarily receive the noise signals and the right side 5 microphones receive the noise and speech signals. $u[n]$ is the

noise microphone signal which mainly contains noise. An adaptive filter is applied to predict the noise signal $y[n]$in speech and noise microphone position and make the error the least. In this situation, if the prediction is close to the actual noise, a minimum error will be addressed which is the desired signal $x[n]$.
Adaptive noise cancellation is a widely used technique that employs adaptive filtering to estimate and remove the noise component from the observed signal. The algorithm uses an adaptive filter to approximate the characteristics of the noise and adaptively update its coefficients to minimize the error between the desired signal and the filtered signal.
The most common adaptive noise cancellation algorithm is the *Filtered-X Least Mean Squares (FXLMS)* algorithm. In this approach, the adaptive filter's coefficients are updated iteratively based on the error between the reference signal (estimated noise) and the observed signal. It extends the classical Least Mean Squares (LMS) algorithm to efficiently adapt filters for the purpose of reducing unwanted components in signals.
The core equation of the FXLMS algorithm can be expressed as follows:

$$\theta(k+1) = \theta(k) + \mu \cdot \mathbf{e}(k) \cdot \mathbf{x}(k) \tag{2.4}$$

Where:

$$\begin{aligned} \theta(k+1) &: \text{Updated filter coefficients at iteration } (k+1) \\ \theta(k) &: \text{Current filter coefficients at iteration } k \\ \mu &: \text{Adaptation step size (learning rate)} \\ \mathbf{e}(k) &: \text{Error signal at iteration } k \\ \mathbf{x}(k) &: \text{Reference input at iteration } k \end{aligned}$$

The goal of the FXLMS algorithm is to minimize the error signal $\mathbf{e}(k)$, which is the difference between the desired output and the actual output. By iteratively adjusting the filter coefficients using this formula, the algorithm seeks to converge to a set of coefficients that effectively cancels or reduces the unwanted components in the signal.
To better understand how the FXLMS algorithm works, let's consider an illustrative example. Suppose you have a reference input $\mathbf{x}(k)$ that contains noise, and your goal is to cancel this noise from the output signal. The FXLMS algorithm adapts a filter with coefficients $\theta(k)$ to generate an estimate of the noise, denoted as $\hat{n}(k)$. The filtered output $\hat{y}(k)$ can then be computed as:

$$\hat{y}(k) = \mathbf{x}(k) * \theta(k) \tag{2.5}$$

Where:

$$\begin{aligned} \hat{y}(k) &: \text{Filtered output at iteration } k \\ \mathbf{x}(k) &: \text{Reference input at iteration } k \\ \theta(k) &: \text{Filter coefficients at iteration } k \\ * &: \text{Convolution operation} \end{aligned}$$

The error signal $\mathbf{e}(k)$ is calculated as the difference between the desired output and the filtered output:

$$\mathbf{e}(k) = d(k) - \hat{y}(k) \tag{2.6}$$

Where:

$$\mathbf{e}(k) : \text{Error signal at iteration } k$$
$$d(k) : \text{Desired output at iteration } k$$

The FXLMS algorithm uses this error signal to update the filter coefficients according to the formula mentioned in Equation 2.4. Through this iterative process, the filter adapts to minimize the error, ultimately resulting in the effective reduction of noise in the output signal.

## 2.4 Noise suppression

### 2.4.1 Overview

Noise suppression is a crucial aspect of various applications, including audio signal processing, speech recognition, and telecommunications. Over the years, researchers have explored and developed a wide range of techniques to reduce or eliminate unwanted noise from signals. This literature overview provides a summary of some key methods and studies related to noise suppression. One thing to be mentioned is that noise suppression is a term with a wide definition. The methods introduced in this section expect no pre-knowledge about the noise which is different from the beamforming and noise cancellation.
Popular approaches include spectral subtraction, wiener filtering, and deep Learning-Based Approaches.
One of the earliest methods for noise reduction is the spectral subtraction technique. Initially proposed by Ephraim and Malah in 1984, this approach estimates the noise power spectral density and subtracts it from the noisy signal's spectrum. The spectral subtraction method has been widely used in various applications such as speech enhancement and audio denoising [4].
Wiener filtering, introduced by Norbert Wiener, is a classical method for signal estimation and noise reduction. It aims to minimize the mean square error between the desired signal and the estimated signal. This approach has found applications in image processing, audio, and speech signal enhancement [6].
Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have shown remarkable success in noise suppression tasks. These networks learn complex features directly from the data and can effectively reduce noise in various applications, including audio and image processing [10].
Recurrent Neural Networks(RNN), such as Long Short-Term Memory (LSTM) networks, are employed in speech enhancement tasks to capture temporal dependencies and context. RNN-based models have demonstrated state-of-the-art results in speech denoising [10].
The specific method used in this thesis is Minima Controlled Recursive Averaging (MCRA), which is a special kind of spectral-subtractive algorithm.

### 2.4.2 General spectral-subtractive speech enhancement configuration

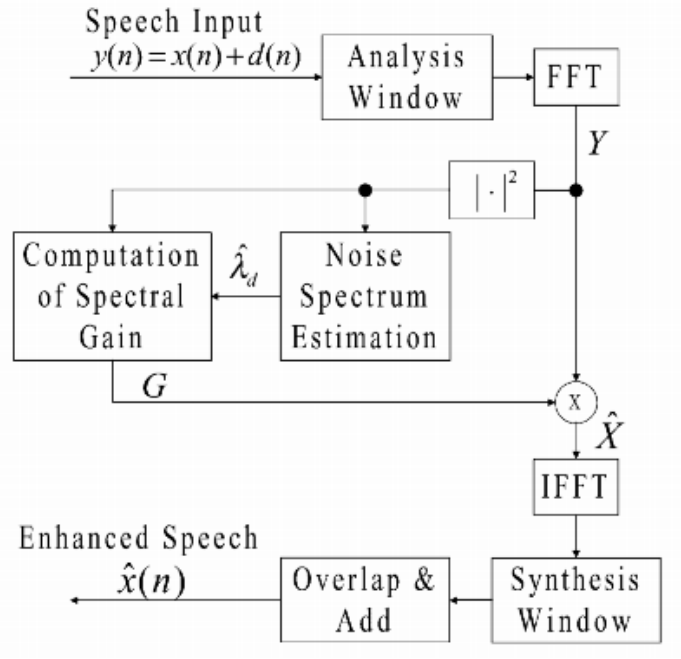


**Figure 2.3:** General form of the spectral subtraction algorithm. Flow chart from [10]

The general form of the spectral subtraction algorithm follows the process shown in Figure.2.3.
Suppose the Speech input is represented as:

$$y[t] = x[t] + v[t]$$

where the noisy speech $y[t]$ equals the sum of clean speech $x[t]$ and noise $v[t]$. This signal is divided into overlapping frames by the application of a window function and analyzed using the short-time Fourier transform (STFT). Specifically,

$$Y(k,l) = \sum_{n=0}^{N-1} y(n+lM)h(n)e^{-j(2\pi/N)nk} \tag{2.7}$$

where $k$ is the frequency bin index, $l$ is the time frame index, $h$ is an analysis window of size $N$ (e.g., Hanning window), and $M$ is the framing step (number of samples separating two successive frames). Let $X(k,l)$denote the STFT of the clean speech, then its estimate is obtained by applying a specific gain function to each spectral component of the noisy speech signal:

$$\hat{X}(k,l) = G(k,l)Y(k,l) \tag{2.8}$$

Using the inverse STFT, with a synthesis window $\hat{h}$ that is biorthogonal to the analysis window $h$, the estimate for the clean speech signal is given by

$$\hat{x}(n) = \sum_{l} \sum_{k=0}^{N-1} \hat{X}(k,l)\hat{h}(n-lM)e^{j(2\pi/N)k(n-lM)} \tag{2.9}$$

where the inverse STFT is efficiently implemented using the weighted overlap-add method.
The crucial step in this process is Noise spectrum estimation. One approach to noise estimation is to use the assumption that the noise is stationary and estimate its statistics from a segment of the audio signal where there is no speech present. Another approach is to use the fact that the noise is often present during silent intervals and estimate its statistics from these intervals.

### 2.4.3 Minima Controlled Recursive Averaging (MCRA) noise estimation

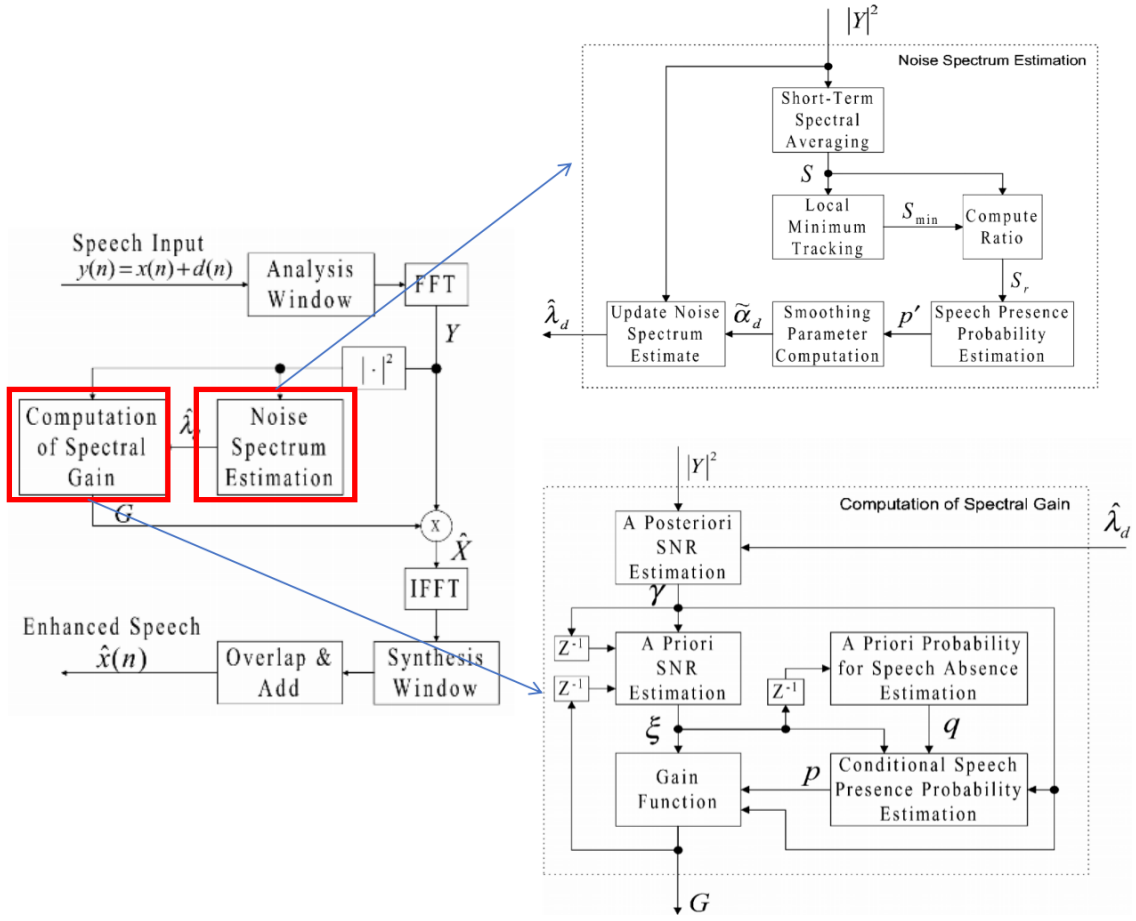


**Figure 2.4:** Theory of MCRA algorithm.Flow chart from[10]

The algorithm works by recursively averaging over a given frequency bin across adjacent processing blocks, thus smoothing out any random noise that may be present. The amount of smoothing is controlled by a parameter called the forgetting factor,

which determines how much weight is given to older versus newer samples in the averaging process.
From the flow chart created by Louzu, Figure.2.4, the process for noise estimation and Spectral gain is shown. The formula for the MCRA algorithm can be expressed as follows in

$$y_n = \alpha(x_n - min(x_{n-m}, ..., x_{n+m})) + (1 - \alpha)y_{n-1} \tag{2.10}$$

where: $y_n$ is the current output sample at time $n - x_n$ is the current input sample at time n - m is the half-length of the smoothing window $\alpha$ is the forgetting factor, typically chosen to be close to 1 to favor more recent samples $min(x_{n-m}, ..., x_{n+m})$ is the minimum value over the previous $2m + 1$ samples $y_{n-1}$ is the previous output sample.
The Minima Controlled Recursive Averaging (MCRA) noise estimation method proposed by Israel Cohen is a widely used approach for estimating the noise power spectral density (PSD) in non-stationary noise environments [9]. The method is based on the observation that the minimum value of the PSD is a good estimate of the noise PSD in the absence of speech or other signals of interest. The MCRA algorithm estimates the noise PSD by recursively averaging the minimum PSD estimates over time and frequency.
Let $x(n)$ be the noisy speech signal at time instant $n$, and let $X(k, n)$ be the discrete Fourier transform (DFT) of a frame of $x(t)$ at frequency bin $k$. The noise PSD estimate $\hat{V}(k, n)$ at frequency bin $k$ and time instant $n$ is given by:

$$\hat{V}(k, n) = \alpha_k \cdot \min(V(k, n - 1), |X(k, n)|^2) + (1 - \alpha_k) \cdot |X(k, n)|^2 \tag{2.11}$$

where $V(k, n - 1)$ is the estimated noise PSD at frequency bin $k$ and the previous time instant $n-1$, and $\alpha_k$ is a smoothing factor that controls the rate of adaptation of the noise estimate. The smoothing factor $\alpha_k$ is given by:

$$\alpha_k = \begin{cases} \alpha_{\text{low}}, & \text{if } V(k, n - 1) \leq |X(k, n)|^2 \\ \alpha_{\text{high}}, & \text{otherwise} \end{cases} \tag{2.12}$$

where $\alpha_{\text{low}}$ and $\alpha_{\text{high}}$ are small and large smoothing factors, respectively. The idea behind the choice of $\alpha_k$ is to adapt the noise estimate quickly when the input signal contains non-stationary components and to adapt slowly when the input signal is mainly noise.
The MCRA algorithm estimates the noise PSD by applying Equation (1) recursively over time and frequency. The estimated noise PSD at time instant $n$ and frequency bin $k$, denoted as $\hat{D}(k, n)$, is used to estimate the speech presence probability (SPP) at the same time and frequency, denoted as $\hat{P}(k, n)$, using the following equation:

$$\hat{P}(k, n) = \frac{|X(k, n)|^2}{|X(k, n)|^2 + \hat{V}(k, n)} \tag{2.13}$$

The estimated SPP is used in subsequent speech enhancement algorithms to suppress the noise and enhance the speech.

## 2.5 Musical noise

Apart from stationary noise and non-stationary noise, musical noise is a type of noise which comes occurs when noise reduction methods are applied. This is found when some methods are applied in this thesis so it is important to understand it.
Musical noise refers to an undesirable artifact that can occur during the process of noise suppression. Instead of suppressing noise uniformly, noise suppression algorithms may inadvertently create tonal or musical-like artifacts in the output signal. These artifacts are perceived as unnatural and disruptive to the listening experience. The name "musical noise" is derived from the fact that the resulting artifacts often resemble musical tones or whistling sounds. These tones may vary in frequency, intensity, and duration, leading to an unpleasant listening experience.
Musical noise is primarily caused by the excessive attenuation or over-adaptation of noise suppression algorithms. When noise is estimated and suppressed too aggressively, the adaptive filters used in these algorithms can start modeling the residual noise as part of the desired signal, leading to the creation of musical-like artifacts.
Several factors can contribute to the occurrence of musical noise, including:

- Over-Adaptation: When the noise suppression algorithm adapts too quickly or overestimates the noise, it may start to distort the desired signal, resulting in musical noise.
- Insufficient Regularization: In some cases, inadequate regularization of the adaptive filters can cause them to "overfit" the noise, leading to the generation of musical artifacts.
- Non-Stationary Noise: Musical noise can be more pronounced in the presence of non-stationary noise, as its characteristics change over time and challenge the adaptability of the algorithms.
- Insufficient Data: If the algorithm does not have sufficient data to accurately estimate the noise, it may produce inaccurate results and introduce musical noise.

## 2.6 Signal-to-Noise Ratio (SNR)

SNR is a metric that compares the power of the speech signal to the power of the noise signal. It is defined as:

$$SNR = 10 \log_{10} \frac{P_{signal}}{P_{noise}} \tag{2.14}$$

where $P_{signal}$ is the power of the speech signal and $P_{noise}$ is the power of the noise signal. A higher SNR indicates a higher-quality speech signal.
However, SNR is limited in its ability to accurately reflect speech quality, as it only measures the signal's power and does not consider the perceptual effects of noise on the speech signal.

# 3
# Methods

The non-stationary and low SNR speech enhancement always has a complex situation where traditional methods are not very useful. Therefore, several methods are implemented in this work. They include beamforming and combined noise cancellation with beamforming and combined noise suppression with beamforming.

## 3.1 Beamforming

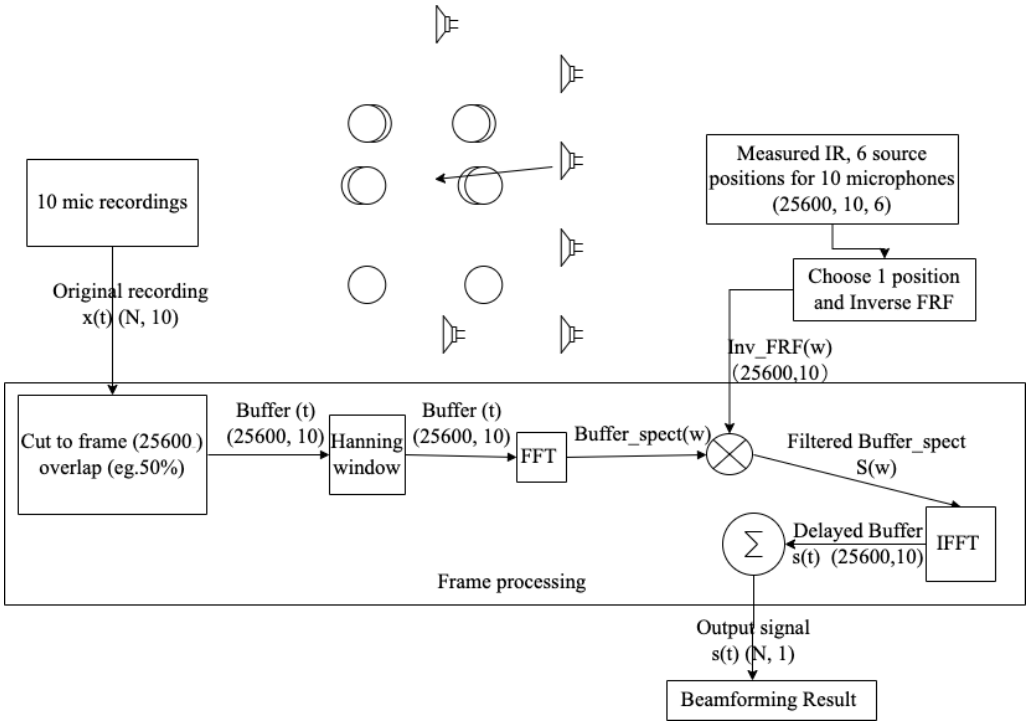


**Figure 3.1:** Flow chart of beamforming algorithm

The delay and sum beamforming is used in this method. Suppose the sound comes from location 2, the algorithm flow is as Figure. 3.1 shown. The 10 microphone positions follow the previous study which captures original signals when noise and speech are playing.
The process is a frame processing that cuts each of the original signals as a frame of 25600 samples with 50% overlap. This will then be processed as a buffer. Hanning window and FFT are performed and the buffer spectrum is then multiplied by the

inverse frequency response function(FRF) of the current transfer path. This step is performing the delay and sum. The inverse FRF is calculated by the measured impulse response of each source position to each mic position which contains the delaying and phase information of each transfer path. After this step, IFFT is performed on the filtered Buffer spectrum which gives the 10-channel delayed buffers. The final step is to sum up all ten channels and get the beamforming result.

## 3.2 Noise cancellation with Beamforming

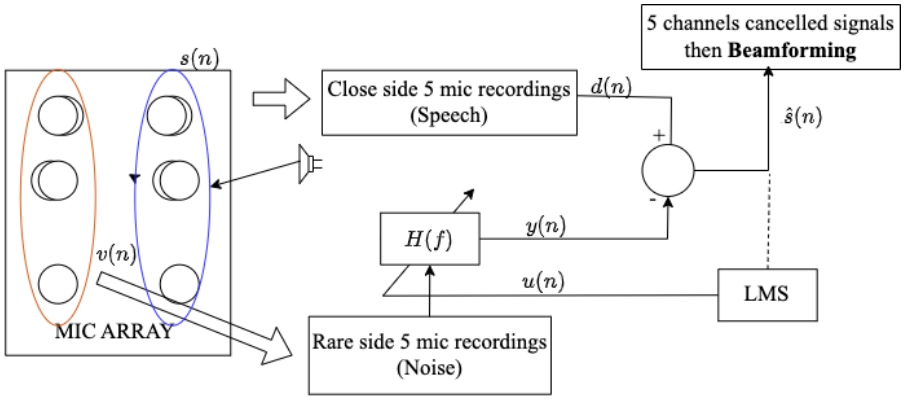


**Figure 3.2:** Flow chart of Noise cancellation with Beamforming algorithm

The noise cancellation method is performed with beamforming. This requests a pre-knowledge of the location property of the noise and the speech source. In this case, when the speech source comes from position 2, some assumptions are made to fulfill the noise cancellation method. As Figure. 3.2 shows the 5 microphones on the right-hand side with a blue mark as closer to the speech source. They are considered to receive both speech and noise signals. On the left-hand side, 5 microphones are considered only receiving noise signals. The noise cancellation algorithm is used among the left and right sides in pairs. The FXLMS method is used to determine the noise cancellation result. After processing, 10 channels of microphone signals become a 5-channel canceled signal. Then beamforming is performed using these cancelled signals. The beamforming process follows the procedure of the above section.

## 3.3 Noise suppression with beamforming

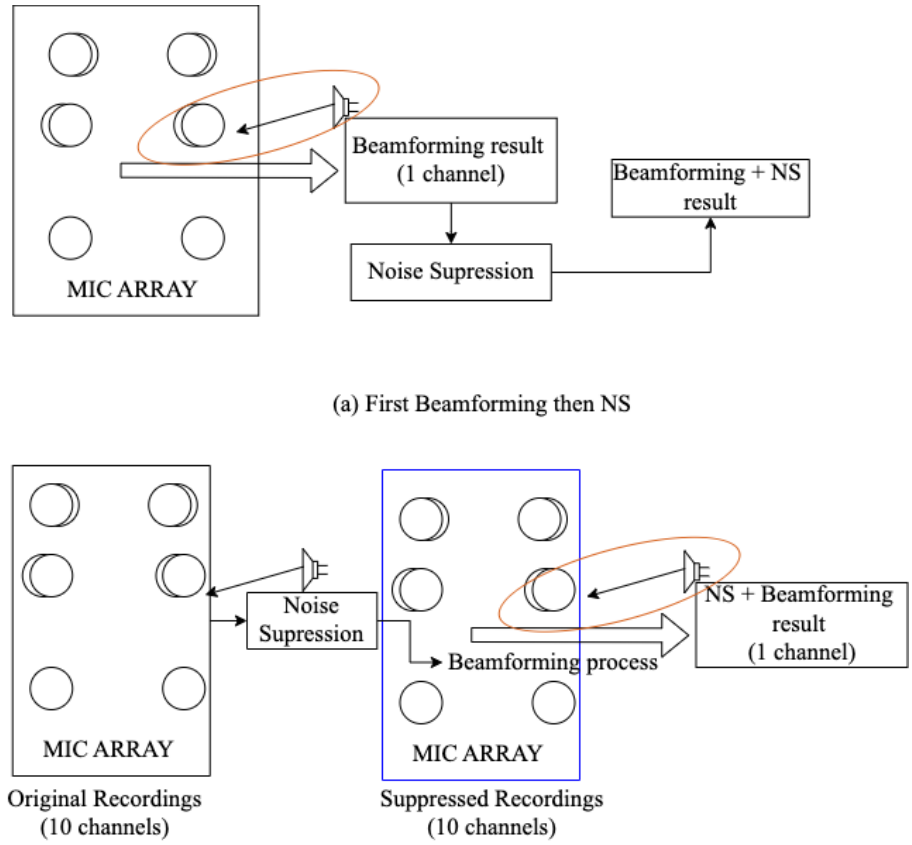


**Figure 3.3:** Flow chart of noise suppression with beamforming algorithm

Since the noise suppression method is used for a single channel, the combined noise suppression with the beamforming method is implemented in 2 different ways. The procedure can be seen in the Figure. 3.3. The above figure shows, the first use of 10-channel microphone recordings to do the beamforming. After getting the 1-channel beamforming result, the noise suppression algorithm is applied to get the result for the combined method.
The figure below shows another way in which the noise suppression method is first applied to 10 microphone recordings. After getting the 10-channel suppressed signals, a beamforming process is performed to get the combined result.
These 2 processes will introduce different effects and the result will be discussed in the later chapter.

# 4

# Results

With the aim of objectively evaluating the performance of different speech enhancement methods and their limits. The beamforming method is established based on Tomoya's thesis and the result includes the SNR of the beamforming output for 6 different positions. For the combined methods, position 2 was chosen to show the result to avoid repetition. After the SNR and spectrogram result, a robustness check is evaluated for each method to be able to check the limitations and the stability.

## 4.1 Beamforming

The beamforming method is performed in 6 different target positions, shown in Figure 4.1. The results of this section present the performance of each beamforming target point under the same level of mixed signals (Speech add noise signal) playing at each speaker position. A robustness check is performed using the result of target position 2.

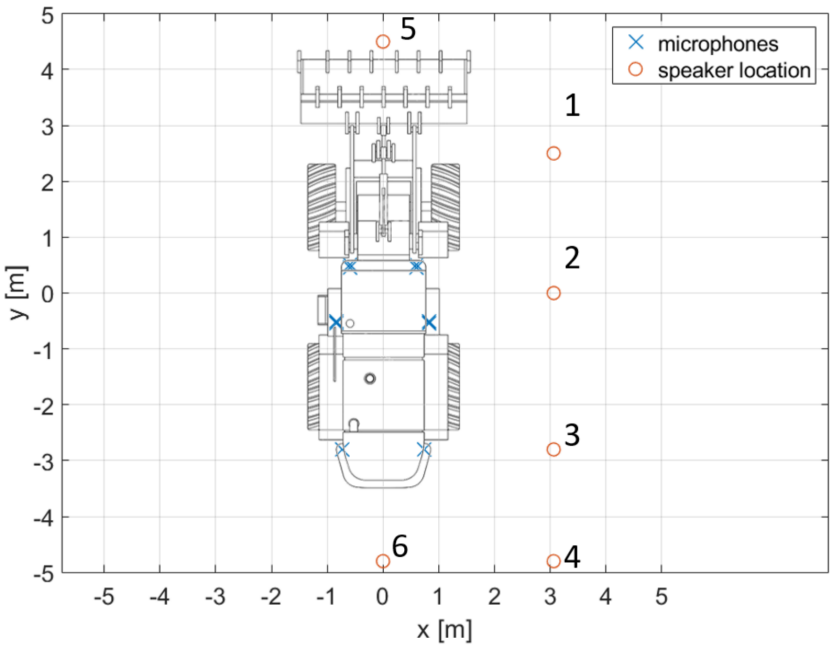


**Figure 4.1:** Beamforming target point and microphone positions

### 4.1.1 Beamforming SNR

This section compared the SNR before and after the beamforming. The mixed signals were played at each of the loudspeaker positions from position 1 to 6. 10 microphones were recorded for each set of measurements. The beamforming algorithm is applied to each set of recording conditions. The SNR before and after beamforming is calculated.
In Table 4.1., the first column shows the 6 positions. The second column is the best single mic signal with max SNR chosen from each of the positions. In this column, positions 1 and 2 have higher SNR, position 4 has the worst SNR in the unprocessed signals. By looking at the SNR values, we found they are all really low(all of them are lower than -9 dB), which means in these cases, the noise is strong and the speech is corrupted badly.
The obvious reason is that from position 1 and 2, the sound sources are closer to the side microphones. Another reason is from the recordings, it is found to be noisier in the middle to the rare side of the machine so a better SNR is obtained when the sound sources are in positions 1 and 2.
The third column is the SNR after beamforming. The output SNR in all positions increases around 2 dB to 3 dB. Among the results in 6 positions, position 2 gets the highest SNR(-6.5 dB) after beamforming.
When calculating the SNR difference before and after beamforming, we found the SNR in positions 2, 3, and 4 increased a lot. They have over 30% differences, and position 2 has the most significant difference which is 32.3 %. While position 5 has the least improvement(11.1 %).
From the SNR result, position 2 is shown to have the best beamforming performance, and position 5 has the worst beamforming performance.

**Table 4.1:** Comparison of SNR between single microphone and beamforming output.

| Position | Max. SNR(single Mic)[dB] | BF SNR [dB] | Difference % |
|---|---|---|---|
| 1 | -9.6 | -7.4 | 22.1 |
| 2 | -9.4 | -6.5 | 32.3 |
| 3 | -13.2 | -9.0 | 31.6 |
| 4 | -15.4 | -10.8 | 30.1 |
| 5 | -15.0 | -13.4 | 11.1 |
| 6 | -14.0 | -11.0 | 21.5 |

### 4.1.2 Robustness check by moving the source position

The robustness check of the beamforming method is very important. Since the fixed beamformer is used, slightly changing the source location will probably influence the performance of the beamforming. It is important to know what is the property of the beamformer and in which cases the result is not stable.
This section is to evaluate the beamformer by moving the source in different areas. From the SNR result in the last section, position 2 has the most difference among

all positions. We assume this beamformer has the best performance. According to this, the recording data from position 2 is chosen for the robustness check. In order to manipulate the source in different positions, the recordings of this measurement are adjusted. This is done by calculating the time of arrivals from each new position and applying different delays to each of the signals to virtually change the source positions according to Equation 2.3. The moving area is shown in the blue area of the plan drawing and the SNR is plotted using a surface plot with a resolution of 11 on both the x-axis and y-axis,i.e. from -1m to +1m, there will be 11 points being equally distributed.

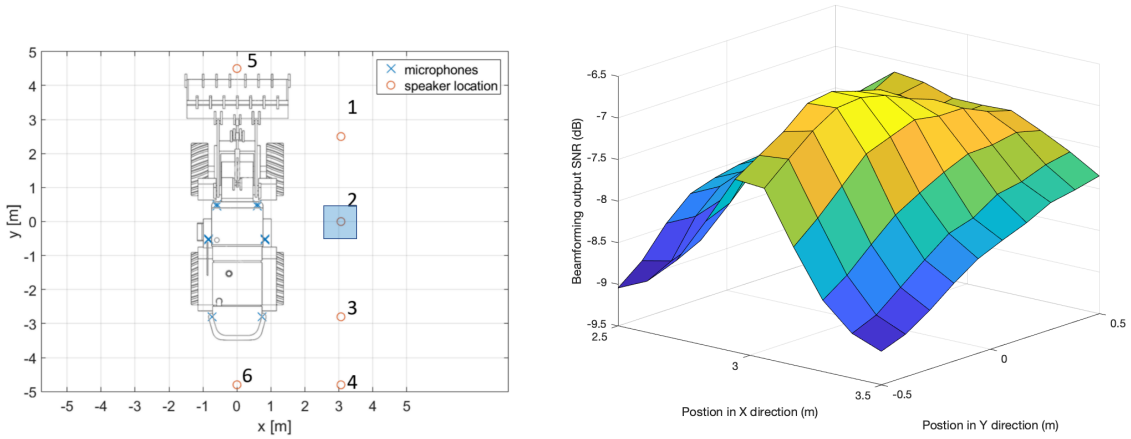


**Figure 4.2:** Output SNR when the source is moving in the blue square(0.5m shifting from center).

In the first case, the source is moving in a 1m* 1m square, the center point is location 2, see Figure 4.2. From the surface plot on the right side, it is obvious that the highest SNR(-7.4 dB) is in the (0,0) position which is the optimum position for beamforming. The interesting point is that, when the source is slightly moved in the x-axis direction (within 0.5m), the SNR decreases fast till around -9 dB. While the source slightly moving in the y-axis, the result is different. The SNR decreases slower when the source moves off the center and the minimum SNR is around -7.6dB.
This shows the beamformer has a better tolerance in the y direction than the x direction when moving the source slightly (within 0.5m from the center).

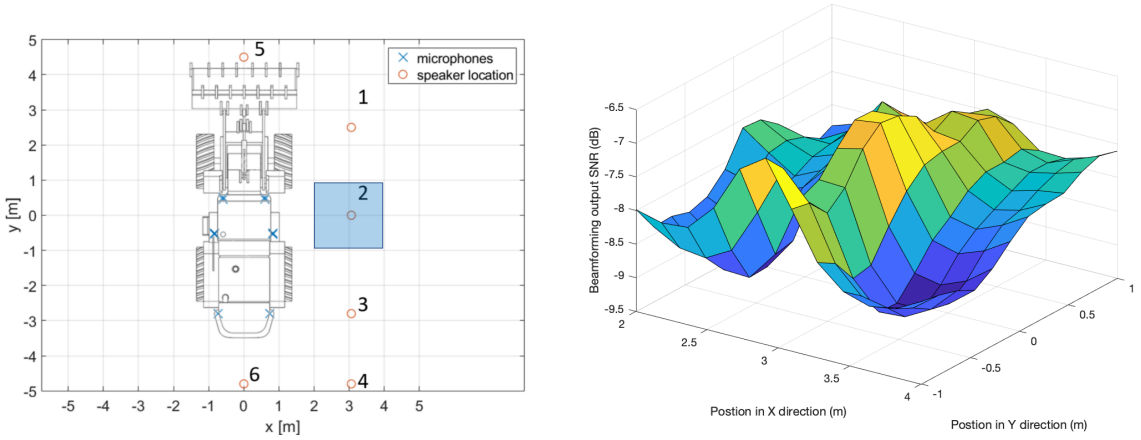


**Figure 4.3:** Output SNR when the source is moving in the blue square(1m shifting from center).

In the second case, the source is moving in a 2m* 2m square, the center point is location 2, see Figure 4.3.
When the source is moving in a larger area, the output SNR of beamforming changes differently. A periodic feature is observed in both the x-axis and y-axis. When the source is moving away from the center, the SNR will first decrease and then increase. The change in the x direction seems to have a stronger influence than the y direction.

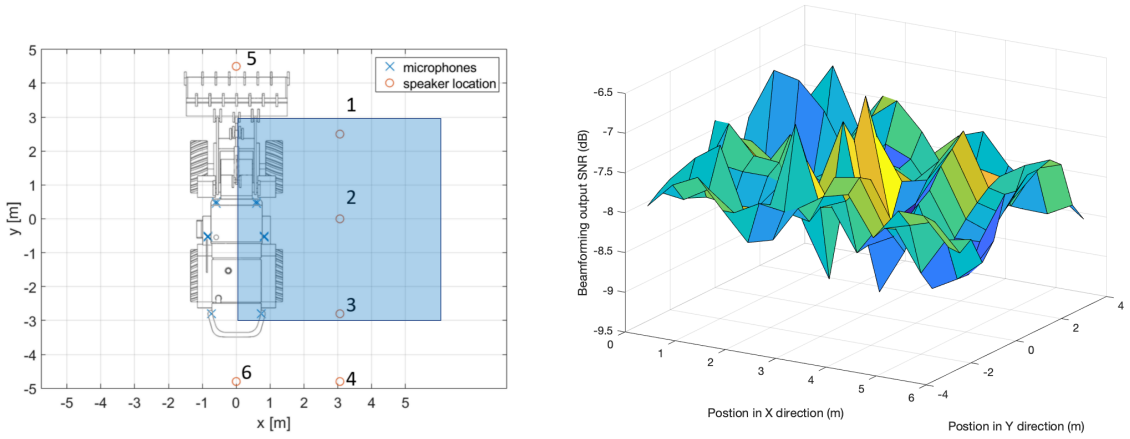


**Figure 4.4:** Output SNR when the source is moving in the blue square(3m shifting from center).

In the third case, the source is moving in a 6m* 6m square, the center point is location 2, see Figure 4.4.
In this situation the source moves across a large area, even reaching positions 1 and 3. The periodic feature is clear in the surface plot, and more peaks and dips appear in the graph. The output SNR fluctuates with a decreasing trend when the source is moving apart from the center.

## 4.2 Noise cancellation with beamforming

This section displays the result of the combined Noise cancellation and beamforming method. The first part plots the spectrogram result. The second part performs the robustness check of this method.

### 4.2.1 Spectrogram comparison

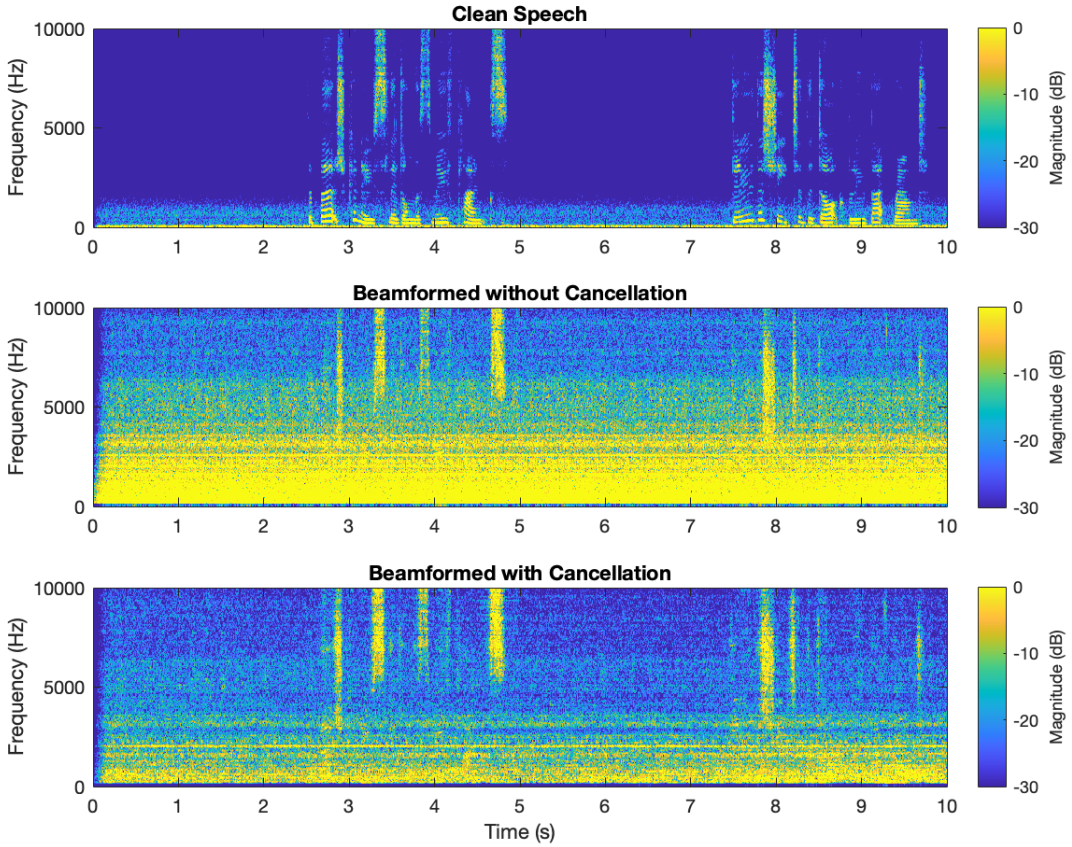


**Figure 4.5:** Spectrogram of noise cancellation with beamforming algorithm

The above Figure 4.5 compares the results of clean speech, beamforming without cancellation, and beamforming with noise cancellation. The top graph is the spectrogram of the clean speech recording, it also contains some low-level background noise. In this case, this signal can be noticed as the desired signal.
The middle graph shows the result after beamforming. It is clear that broadband noise remains in the result. Especially below 4k Hz, the noise level is high and most useful speech contents are corrupted.
The bottom graph is the result of combined noise cancellation and beamforming. The noise is reduced a lot in the middle frequency, which helps increase the sound quality. The low-frequency noise is still left in the result as well as some frequency tones and modulation noise. The speech is not clear after the process. This is shown by comparing the top and bottom graph in the middle frequency. The combined

noise cancellation and beamforming method can reduce both speech and noise levels. One reason for this is that the noise cancellation method requires a noisy signal channel and the noise channel is not correlated. But in this case, the noisy channels are still mainly receiving noise because of the low input SNR. This makes it hard for the LMS process to find the optimum solution.

### 4.2.2 Robustness check by changing the input noise percentage

The robustness check is done by increasing the percentage of the input noise and running this method under different input signal situations, i.e. 10 % of input noise means a mixed noisy signal with 10 % multiplied by the noise signal amplitude plus 90 % multiply by the clean speech signal. In this way, from 5 % noise (low noise level) to 95 % noise (high noise level) situations are simulated.

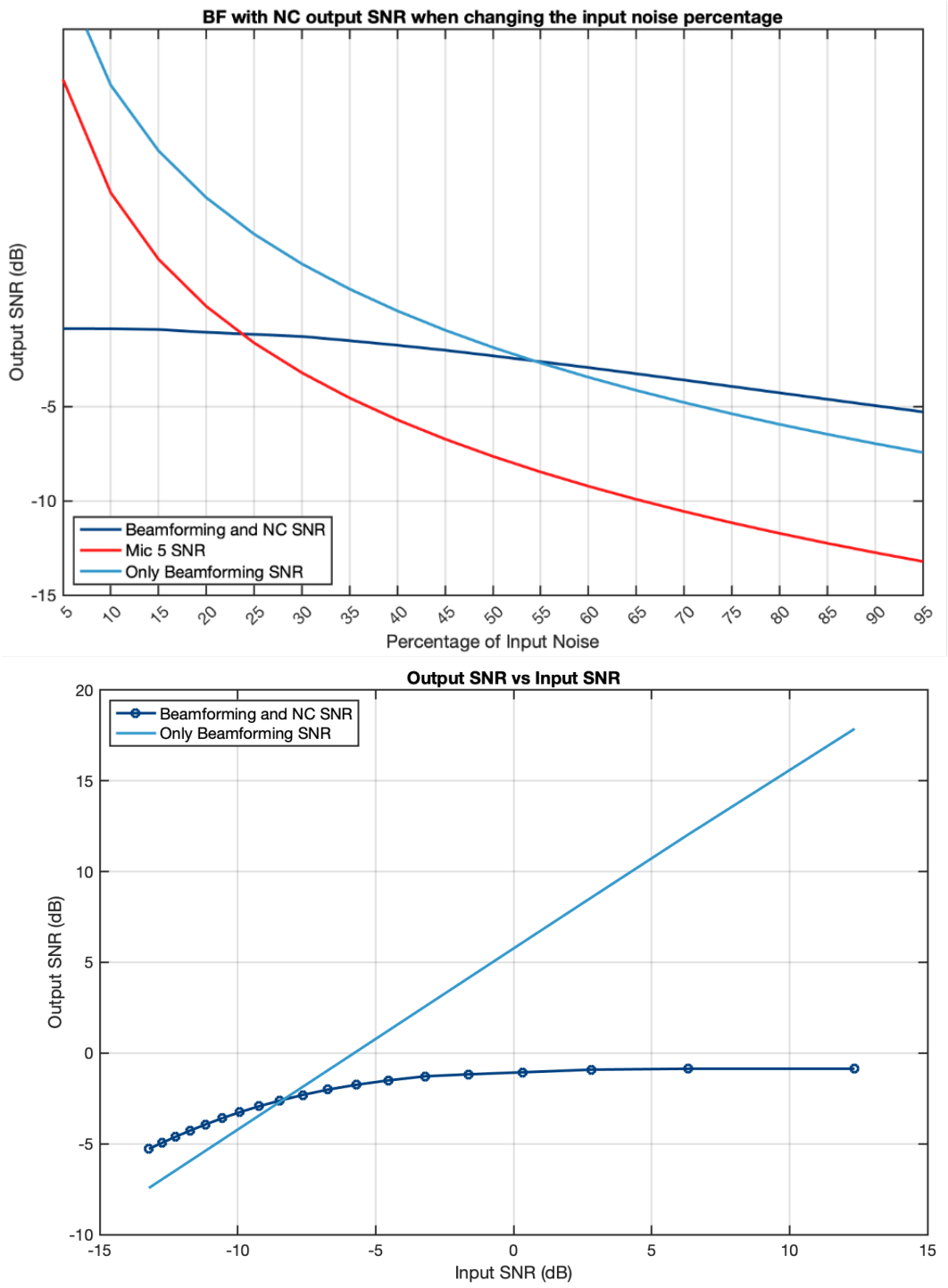


**Figure 4.6:** Robustness check for first Noise suppression then beamforming.

In Figure 4.6, the upper one is the output SNR of different noise percentages. The x-axis is the percentage of input noise increasing with a 5 % step size, from 5 % to 95 %. The y-axis represents the SNR level in dB. The red line is the SNR of the Mic 5 recording. This is the best single-channel recording with the highest SNR and this can be a reference value. If only using the microphone to record without any post-processing, the red line is the best result that can be achieved. When the

input noise percentage rises, the output SNR decreases, which is only due to the noise ratio being higher. The light blue line is the only beamforming result. We can see an overall increase in the SNR compared with the single-mic result. What is more important to mention is that the higher the noise percentage is in the input signal, the larger the SNR improvement can be achieved. The beamforming method is always bringing an increase to the SNR. The combined method result is shown with the dark blue line. The SNR is relatively stable and slowly decreasing from around -1 dB to around -5 dB. In this situation, the highest SNR result sounds not good, because most of the information is canceled. After around 23 % noise input, the SNR starts to be better than the single Mic 5 result. After 55 % noise has been added to the input signal, the combined BF and NC method starts to get better results than the beamforming result.

In the lower graph of Figure 4.6, the input SNR is set as the x-axis and the y-axis is still the output SNR. In this graph, the input SNR is calculated using the value of the Mic 5 recording. In other words, the Mic 5 SNR is considered to be the input SNR before using any algorithms. This is another way of showing the same result from the upper one. This graph shows the relationship between output SNR and input SNR. The light blue and the dark blue line represent the only beamforming SNR and the combined BF and NC method SNR is the same as the upper graph. From this graph, several facts can be observed. First, when the input SNR increases, the beamforming output SNR increases with a linear trend. Second, the combined method shows a slowly increasing trend when increasing the input SNR, and when the input SNR is larger than 0 dB, the combined method gets a stable SNR lower than 0 dB.

## 4.3 MCRA noise suppression with beamforming

In this section, the results of the combined MCRA Noise suppression and beamforming method are shown. One subsection shows the results of first beamforming then Noise suppression and first noise suppression then beamforming in spectrogram and coherence evaluation. The other subsection analyses the robustness of the first noise suppression and then the beamforming method.

### 4.3.1 Spectrogram comparison and Coherence

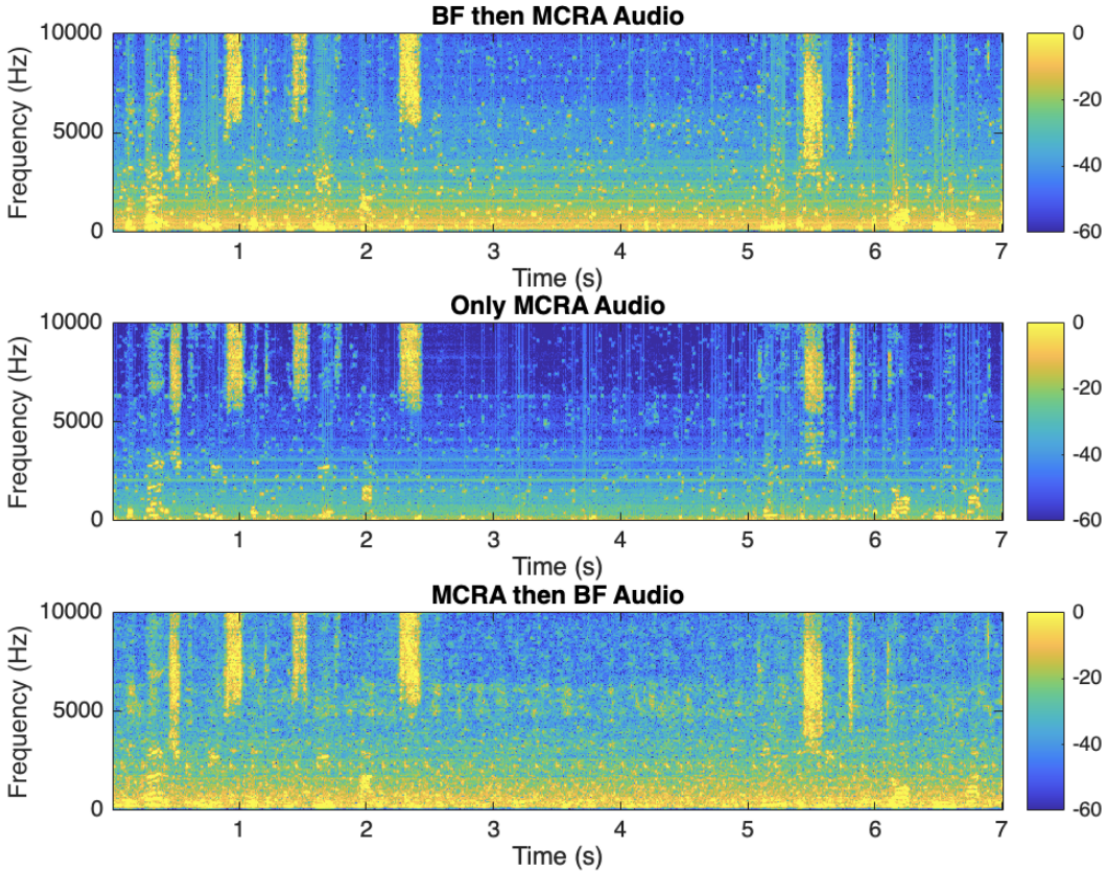


**Figure 4.7:** Spectrogram of noise suppression with beamforming algorithm

The above Figure 4.7 compares the results of first beamforming then MCRA Noise suppression, only MCRA for single channel Mic 5 result and first MCRA noise suppression then beamforming.

The top graph is the spectrogram of the first beamforming and then the MCRA Noise suppression result. The broad band noise in the middle frequency is reduced, also the speech frequency stays after the process. But the noise in the low-frequency range is still left. Another new problem shown in this graph is the musical noise. From the spectrogram, we are able to see many small dots in a wide frequency range. These tones vary in frequency, intensity, and duration, leading to an unpleasant listening experience. In this case, 'an underwater bubble sound is described' when many volunteers listen to this noise.
The middle graph shows the result of the single-channel MCRA noise suppression. The noise is suppression a lot and most of the noise in the middle frequency range is below -40 dB. However, the musical noise problem is even larger and becomes a tough problem when listening to the output signal. The frequency dots are more clear in the spectrogram.
The bottom graph is the result of first noise suppression for 10 recordings and then

beamforming for the 10 suppressed signals. The result is similar to the top one which reduces the broadband noise and maintains the speech.
The important improvement of this method is that first noise suppression and then beamforming will get rid of musical noise and improve the speech intangibility. This can be addressed both from the spectrogram and the listening experience. From the spectrogram, less clear dots in the frequency range are observed. A blurry amplitude noise is shown instead of those clear dots. When listening to this result, most people comment this is without much bubble noise and rate this result the best among all.
As for the musical noise problem, it only appears in this method. This means it is introduced by the noise suppression process. Several reasons for the musical noise have been mentioned in the theory part, like over-adaptation, insufficient regularization, non-stationary Noise, and insufficient data.
In this scenario, beamforming offers more channel recordings and aliens them together to get more useful data which may increase the performance of the combined method.

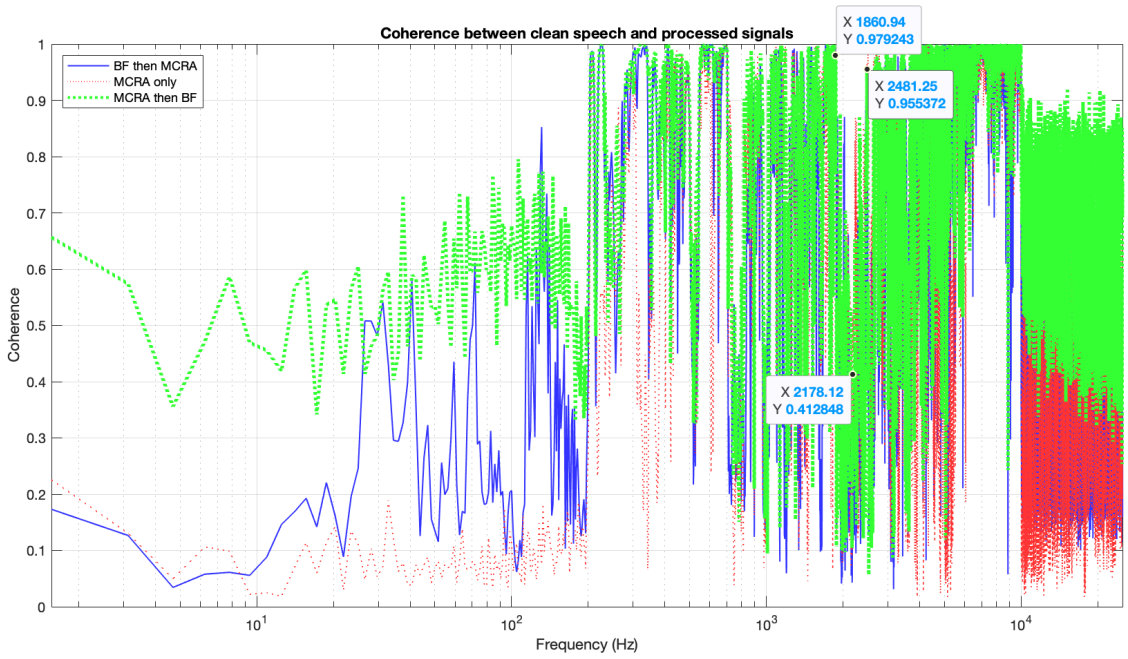


**Figure 4.8:** Coherence of clean speech and the result of different methods

To further evaluate the performances of different results, coherence between the processed signals and clean speech signal of different methods is plotted, see Figure 4.8. The aim of speech enhancement is to get clean speech, so the coherence between the clean speech and the processed signal can be seen as an indicator of the performance. From the figure, we are able to see that the MCRA then BF method(Green line) is the best among 3 different methods. We observe a large dip between around 1800Hz to 2500Hz in all 3 coherence results. This may be the noise frequency that is left in all 3 results.

### 4.3.2 Robustness check by changing the input noise percentage

The robustness check for this method is done by using the same way as the BF and NC methods, to increase the percentage of the input noise and run this method under different input signal situations. Also, input SNR is used for plotting the result.

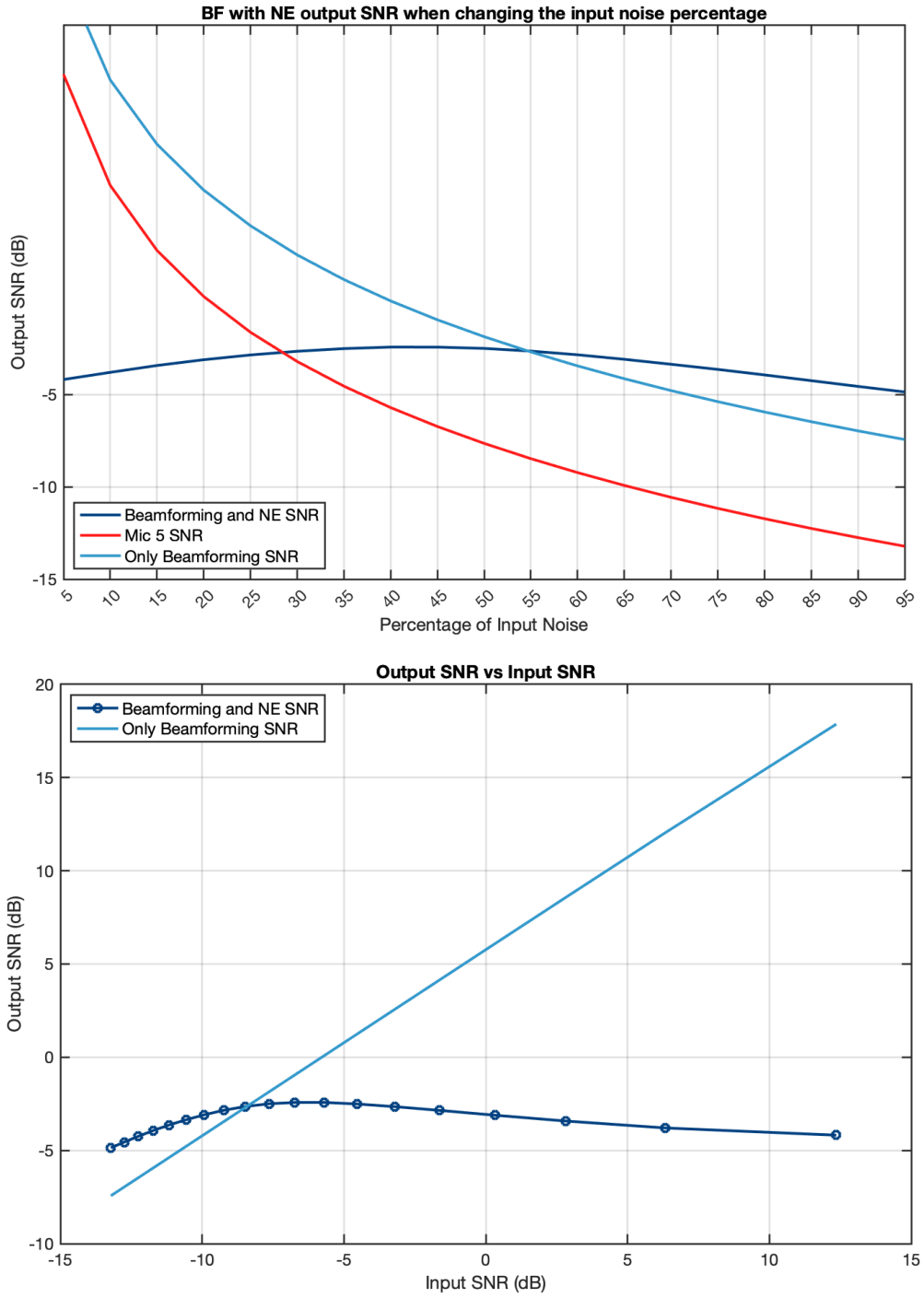


**Figure 4.9:** Robustness check for first noise suppression then beamforming.

In Figure 4.9, the upper one is the output SNR of different noise percentages. The x-axis is the percentage of input noise increasing with a 5 % step size, from 5 % to 95 %. The y-axis represents the SNR level in dB. The red line and the line blue line remain the same as Figure 4.6. The red line is the best single Mic SNR and the light blue line is the only beamforming result. The dark blue line in this case is the result of first Noise Suppression and then the beamforming method. The SNR is relatively stable and has an upside-down U shape. It first slowly increases from -4 dB to -2 dB and then decreases from around -2 dB to around -5 dB. In this situation, when the input SNR is low, the combined result is not good in the subjective listening. After around 27 % noise input, the SNR starts to be better than the single Mic 5 result. After 55 % noise has been added to the input signal, the combined BF and NS method starts to get better results than the beamforming result.
In the lower graph of Figure 4.9, the input SNR is set as the x-axis and the y-axis is still the output SNR. This graph shows the relationship of output SNR and input SNR. The light blue and the dark blue line represent the only beamforming SNR and the combined BF and NS method SNR as the same as the upper graph. From this graph, the combined method shows a slowly increasing trend when increasing the input SNR is lower than -5 dB input SNR. When the input SNR is larger than -5 dB, the combined method is getting a lower SNR.

## 4.4 Informal listening for different methods

Apart from objective data and spectrogram of different results, informal listening is also made to describe the perception of different results. I and my supervisor Nicklas addressed these evaluations after listening to different results, and these comments have been agreed upon by another 3 people in the Volvo CE team. The result is shown in the Table 4.2.

**Table 4.2:** Subjective evaluation for different methods.

| **Signal** | **Quality** | **Intelligibility** | **Comment** |
|---|---|---|---|
| Mic 4 | Bad | Bad | Low SNR with harsh noise hard |
| Only NS for Mic 4 | Good | Bad | Noise reduced, Output not clear |
| Only BF for all Mic | Bad | Bad | Modulation noise left |
| NC then BF | Medium | Medium | Good quality but speech is unclear |
| First BF then NS | Good | Good | Clear but with musical noise |
| First NS then BF | Good | Good | Clear without musical noise |

# 5

# Conclusion

## 5.1 Discussions

In this thesis, several methods are implemented and analyzed. Based on different properties and knowledge of the noise and Speech signal, different methods have their own performance. In short, beamforming is a way to generally improve SNR by using multiple microphones and it requires the knowledge of the locations of the target source. Noise cancellation also requires the location property, i.e. the noise source should be nearer to the noise microphone(s) than the noisy speech source. As for Noise suppression, it is different in that it does not require a pre-knowledge of the noise source. The method estimates noise by different strategies like statistics.

Regarding the results from the beamforming method, we can see an overall improvement of SNR in all position results. Among all positions, position 2 has the best SNR and the largest increase before and after beamforming. In the robustness check, when moving the source in a small area, the SNR decreases slower in the y direction than in the x direction. When moving the source to a larger area, the SNR fluctuates with a periodic feature.

This shows us the property of the beamforming process. The slight move in the y-axis is acceptable if the distance is within 0.5m, while the displacement on the x-axis will influence the result a lot.

As for the combined beamforming and noise cancellation method, it can remove most of the mid-frequency noise and make the sound quality better. It is shown in the robustness check that the SNR result is better than just using beamforming when the input SNR is low. The problem is that a lot of speech information is lost during the processing.

The combined beamforming and noise suppression method is the best among all. When we use BF first and then MCRA, the SNR is reduced a lot and the speech signal is much clearer than just using beamforming. But this method introduced musical noise which makes the listeners feel uncomfortable. While using MCRA first to process all 10 channel recordings and then perform beamforming gave us a better result. In this situation, although we added some low-level noise, a better intangibility is achieved. The result is less engine noise and less musical noise. Most of the listeners also rate this result as the best.

From the robustness check of this method, we address that this method is also better than only using beamforming when the input SNR is low. A meaningful conclusion is that using beamforming to multiple noise suppression results is able to reduce musical noise.

## 5.2 Limitations

There are limitations in this thesis that restrain the performance of different methods. For beamforming, a more concentrated microphone array is preferred rather than the setup now which has 10 microphones far away from each other. This will introduce a difficulty to perform beamforming and also reduce the accuracy of the result. The combined beamforming and noise cancellation method requires that the noise microphones are close to the noise source and the noisy speech microphones are close to the speaker. This is not achieved properly in the application. The 2 different types of microphones are getting mainly the same noise and speech signals so the output result is not good in this method. When it comes to the combined beamforming and noise suppression method, it has the best result among all. It is hard to estimate the rapidly changing noise in this method, that's why musical noise is left when using MCRA.

## 5.3 Real-world applications

This thesis is meaningful for some real-world applications. A combined system can be designed for the machine cabin. The system is with microphone arrays and uses several different methods according to the noise and speech situation. There can be 3 buttons on the control panel. The driver is able to turn on beamforming when seeing the speaker in a certain direction outside the cabin. For example, as Figure 5.1 shows, if the the speaker is in position 2, the operator is able to turn on beamforming at position 2 to capture the speech. If the input SNR is low the operator is able to open the NC or NS function to have a future speech enhancement.

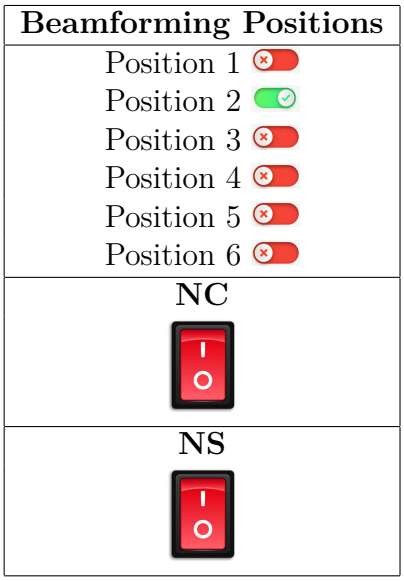


**Figure 5.1:** Control Panel for Application

# Bibliography


[1] Gustaver, M. (2020) A Chalmers University of Technology Master's thesis template for LaTeX. Unpublished.

[2] Barry D. Van Veen and Kevin M. Buckley (1988) Beamforming: A Versatile Approach to Spatial Filtering

[3] Benesty, J., Chen, J., and Huang, Y. (2009). Microphone array signal processing. Springer Science and Business Media.

[4] Ephraim, Y. and Malah, D. (1984). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Transactions on Acoustics, Speech, and Signal Processing, 32(6), 1109-1121.

[5] Alan V. Oppenheim, Alan S. Willsky, and S. Hamid Nawab. Signals and Systems

[6] Wiener, N. (1949). Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications. MIT Press

[7] Hu, Y., and Loizou, P. (2006) Evaluation of objective measures for speech enhancement.

[8] Antony W. Rix, John G. Beerends (2002) Perceptual evaluation of speech quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs.

[9] Cohen, , Baruch Berdugo. (2001). Speech enhancement for non-stationary noise environments. Signal Processing 81 (2001) 2403–2418.

[10] Loizou, P. C. (2013). Speech Enhancement: Theory and Practice, Second Edition. CRC Press.

[11] Greensted, Andrew. Delay Sum Beamforming(2012). URL:http://www.labbookpages.co.uk/audio/beamforming/delaySum.html.

[12] Jose Maria, Giron-Sierra (2017) Digital Signal Processing with Matlab Examples, Volume 2 Decomposition, Recovery, Data-Based Actions

[13] Tim Van den Bogaerta, Simon Doclo(2017) Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids

# A

# Appendix 1

FRF and Speech Signal Measurement Setup
This measurement will conduct four measurements: FRF between the white noise (from monopole source) and microphones, time signal between pink noise (GENELEC) and microphones, time signal speech signal vs time, and machine load recordings.
6 locations of points of measurements (these locations are on one side of the machine)
10 microphone inputs + 1 direct signal input

1. FRF Measurement Method (perfect case scenario)
Purpose: Get the best-case transfer function between microphones on the machine and the monopole source Excitation: white noise from MS
Outputs: FRF and Cross Correlation
How many averages for the FRF: 30
Machine: OFF
Post processing: Use TF to create a filter to use for the mixed signal use TF to correct the listen response since we want the listen response to be flat.

2. Pink Noise Measurement Method
Purpose: Used as working material to perform beamforming + and other filters as see fit
Excitation: pink noise from GENELEC speaker
Output: Coherence, times series and cross correlation Only need to measure one location at a time
Machine: OFF
Post processing: none conducted.

3. Speech Signal Measurement Method (point of comparison) Purpose: Retrieve speech signal so can be used as point of comparison between beamformed and unbeamformed.
Excitation: prerecorded speech from GENELEC speaker
Output: Times series
Machine: OFF
Post processing: Conduct beamforming on the speech signal and evaluate the performance.

4. Machine Load Measurement Method

Purpose: record noise
Excitation: Machine engine at various load states below. Low idle (lowest RPM)
High idle (highest RPM)
1400 rpm idle
Low idle + high hydraulic pressure (lowest RPM + highest pressure in the hydraulic pump)
Output: Times series
Machine: various load states
Post processing: mix with speech recording.

Equipment
- 10 microphones + 1 mic for source
- SCADAS Mobile (MOB 2)
- Power supply for MOB 2
- Extension cables for power to PC and MOB 2
- LMS monopole source speaker
- Monopole source power supply
- GENELEC speaker 1029A
- Sticking tape
- Spray can
- Measuring tape
- Monopole/Speaker stand
- Cables (see below for more detail)
- Computer
- Computer power
- L70H

Cables:
- Short BNC to 3.5 mm (BNC T to PC)
- 8 m XLR to BINC (BNC T GENELEC)
- BNC T connector (to split the source signal)
- BNC female to female (adapter for T connector)
- 1 m BNC to BNC (MOB 2 to BNC FF)
- 8 m BNC to BNC (MOB 2 to monopole source)
- 8x 3 or 4 m BNC to BNC (MOB 2 to cab mic)
- 2x 5 m BNC to BNC (MOB 2 to rear mic)
- 5 m BNC to BNC (MOB 2 to monopole Amp)
- LAN cable (connect MOB 2 to PC)

Excitation
Measurement 1: White noise 200 Hz 10000 Hz (enough for 30 avg.)
Measurement 2: Pink noise at 1.6 m height (enough for 30 avg.)
Measurement 3: 30 seconds of the same prerecorded speech
Measurement 3: 30 seconds of machine load states
Location/Date
Outside on test track, 2022-03-22

Spectral Testing Setup
Frequency range: 0 – 10024 Hz
Frequency resolution: 2 Hz
Sampling rate: 51200 Hz

**Department of Architecture and Civil Engineering**
**CHALMERS UNIVERSITY OF TECHNOLOGY**
Gothenburg, Sweden
www.chalmers.se


CHALMERS
UNIVERSITY OF TECHNOLOGY