**Target Speech Extraction from Noisy Acoustic Mixtures: A Three-Channel Signal Processing Approach**

Dongliang Sun, University of Ottawa

***Abstract*—This report presents a signal processing framework for extracting a target speech signal from a noisy three-channel acoustic mixture. The mixture consists of three far-field directional sources: a target English-speaking female and two interfering speakers (a German-speaking female and an English-speaking male). Additionally, uncorrelated white Gaussian noise is present in all three microphone channels. The processing pipeline involves three main stages: direction of arrival (DOA) estimation, spatial filtering via beamforming, and post-filtering for noise reduction.**

**The DOA estimation stage leverages the Time Difference of Arrival (TDOA) between microphone pairs to estimate the angular positions of the sources. This information is used to design a Minimum Variance Distortionless Response (MVDR) beamformer, which isolates the target source by suppressing directional interferences. The resulting single-channel output is further processed using post-filtering techniques, such as** **Ephraim's MMSE log-STSA method, to attenuate residual background noise.**

**The system operates under simplified conditions, assuming plane wave propagation, stationary sources, anechoic environments, and offline processing capabilities. Results demonstrate the effectiveness of the proposed approach in enhancing the target speech signal while minimizing distortion and interference, making it suitable for controlled scenarios requiring high-quality speech extraction.**

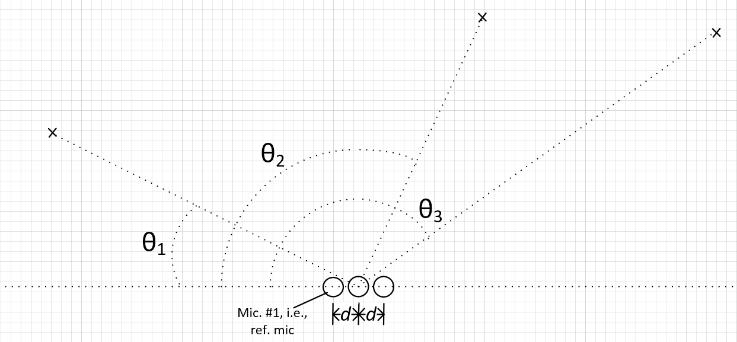
***Index Terms*—Beamforming, Minimum Variance Distortionless Response (MVDR), noise reduction, speech enhancement, Time Difference of Arrival (TDOA)**

# I. INTRODUCTION

T

he extraction of a target speech signal from a noisy acoustic mixture is a critical problem in many applications, including telecommunication, hearing aids, automatic speech recognition, and human-machine interaction. Real-world environments often involve multiple overlapping sound sources and background noise, making it challenging to isolate a specific speaker. This report addresses the problem of extracting a target speech signal from a three-channel acoustic mixture, where the mixture includes contributions from two interfering speakers and uncorrelated white Gaussian noise at each microphone.

The setup for this project involves three spatially fixed directional sources and three microphones arranged linearly with a spacing of cm, as shown in Fig. 1. The target signal is the speech of an English-speaking female, while the interferences are a German-speaking female and an English-speaking male. The sound propagation is assumed to follow plane wave behavior under anechoic conditions, with no reverberation or physical obstructions. These simplified conditions enable a focused exploration of signal processing techniques without the additional complexity introduced by real-world environments.



**Fig. 1.** Acoustic mixture and microphone array setup

The proposed solution employs three main stages. First, the DOAs of the sources are estimated using Time Differences of Arrival (TDOA), which provide critical spatial information for subsequent filtering. Second, an MVDR beamformer is designed to suppress directional interferences while preserving the target signal. Finally, post-filtering is applied using Ephraim’s MMSE Log-STSA method to reduce residual noise and further improve speech quality. The performance of the system is evaluated using metrics such as Signal-to-Interference Ratio (SIR) and Signal-to-Distortion Ratio (SDR), demonstrating significant improvements in speech intelligibility and quality.

This report is organized as follows. Section II describes the signal model, including the formulation of the multi-microphone acoustic mixture. Section III outlines the methodology, covering TDOA estimation, the MVDR beamformer, and single-channel speech enhancement. Section IV presents the performance evaluation, showcasing the effectiveness of the proposed approach. Finally, Section V concludes the report and discusses potential future directions.

# II. Signal model

In a uniform linear microphone array with microphones, the signal received at the -th microphone is modeled as the sum of contributions from one target source plus additive noise, considering two interferences as noise. The signal model can be expressed as:

Where is the target signal captured by the -th microphone, and is the noise including two interferences and white noise received at the -th microphone. is the target signal received at the reference microphone, i.e. Mic#1.is the relative time delay between the -th microphone and Mic#1.

In the frequency domain, the signal model given in (1) is written as

Where ,, and are the continuous time Fourier Transform of ,,andrespectively.

# III. Methodology

## A. TDOA

 TDOA provides critical information about the spatial location of sound sources, such as their direction of arrival (DOA), and is widely employed in applications like source localization, beamforming, and acoustic scene analysis.

When a sound wave propagates in a medium, its travel time to each microphone depends on the distance between the source and the microphones. In a uniform linear array (ULA) with  microphones, the relative time delay between the signals received at two microphones is known as the TDOA. Assuming far-field plane wave propagation, the TDOA between two microphones can be directly related to the angle of arrival  of the sound wave and the spacing  between the microphones.

The time delay between two microphones separated by  is given by[1]:

Where  is the speed of sound in the medium (typically  in air),  is the direction of target source signal arrival (measured relative to the endfire axis).

As a result, for the -th microphone, the time delaybetween it and Mic#1 is .

According to equation (2), we can define the steering vector for the target source as

Where the superscript represents transpose.

The time delay or TDOA can be estimated by locating the peak in the cross-correlation function of the signals received by a pair of microphones[2].

## B. Beamformer

The Minimum Variance Distortionless Response (MVDR) beamformer, also known as the Capon beamformer, is an advanced signal processing technique used to enhance the reception of signals from a desired direction while suppressing noise and interference from other directions. It is widely used in applications like audio signal processing, radar, sonar, and wireless communications.

Beamformer coefficients in the frequency domain can be given by[3]:

Where is the pseudo-coherence matrix, which in this project can be written as:

Where and are the steering vector of the two interferences respectively, which are defined in a similar way to , the superscript denotes the conjugate-transpose operator.

## C. Single Channel Speech Enhancement

Unlike multi-channel approaches, single-channel speech enhancement does not rely on spatial information or directional filtering to separate the speech from noise. Instead, it focuses on the statistical or spectral properties of the signal and noise to distinguish and enhance the desired speech component[4].

Since the background white Gaussian noise is stationary and its statistics in the time-frequency domain differ from the statistics of the target speech signal, it is possible to use single channel time-frequency filtering methods to attenuate the noise.

In this project, Ephraim's MMSE log-STSA method 1985 was implemented. This method is based on minimizing the mean square error between the logarithm of the clean speech spectral amplitude and its estimate[5]. By working in the log-spectral domain, the method aligns with how the human auditory system perceives sound, providing enhanced speech quality and naturalness compared to other techniques like spectral subtraction or Wiener filtering.

# IV. Performance Evaluation

## A. Measures TODAs



**Fig. 2.** Cross-correlation of the signals from Mic#1 and Mic#3

Fig. 2 illustrates the cross-correlation function of the signals received by Mic#1 and Mic#3. By identifying the values corresponding to the three peaks, the Time Differences of Arrival (TDOAs) are estimated as  seconds, respectively. Using these TDOAs, the Directions of Arrival (DOAs) are calculated as 1.0× degrees, respectively.

## B. Beampattern

A beampattern represents the spatial sensitivity of a microphone array or beamformer to sound waves arriving from different directions[6]. It is a graphical depiction of how the array amplifies or attenuates signals based on their direction of arrival (DOA). The beampattern is defined as the response of the beamformer as a function of the angle, typically measured relative to the array's axis.



**Fig. 3.** Polar beampattern



**Fig. 4.** Rectangular beampattern

Fig. 3 and Fig. 4 show the beampatterns of the MVDR beamformer at a frequency of 1 kHz. As observed from these figures, there are very small magnitude responses at the DOAs of the interferences (at angles 170° and 10°), while the response at the DOA of the target talker (at 80°) is unity (0 dB), indicating distortionless response for the desired source.

Fig. 5 shows the beamformer filter impulse responses in the time domain.



**Fig. 5.** Beamformer filter (impulse response)

## C. Post-filtering metrics

Post-filtering is a critical step in speech enhancement, particularly for noise reduction after spatial filtering (e.g., beamforming). Various metrics are used to evaluate the effectiveness of post-filtering techniques in enhancing speech quality and suppressing residual noise. Two commonly used metrics are Signal-to-Interference Ratio (SIR) and Signal-to-Distortion Ratio (SDR).

SIR measures the ratio of the target signal's power to the power of interference signals remaining after filtering. A higher SIR indicates better suppression of interference, which is crucial for isolating the target speech from competing sources.

SDR evaluates the overall quality of the enhanced signal by comparing the power of the clean speech signal to the total distortion introduced by the enhancement process, including both interference and noise. Higher SDR values indicate better overall speech quality and less distortion.

Using the second dataset, the SDR after applying Ephraim's MMSE Log-STSA method (1985) is -0.0405, while the SIR is 10.0180. Without any speech enhancement method, the SDR is -6.2809, and the SIR is 4.2139. These results demonstrate that Ephraim's MMSE Log-STSA method significantly improves the quality of the speech signal.

V. Conclusion

This report presents a comprehensive framework for extracting a target speech signal from a noisy three-channel acoustic mixture. The system employs a pipeline comprising direction of arrival (DOA) estimation, MVDR beamforming, and single-channel post-filtering using Ephraim's MMSE Log-STSA method (1985). While the beamformer effectively suppresses directional interferences, the post-filtering step complements it by attenuating non-directional and stationary noise. This two-stage approach ensures better overall enhancement compared to standalone methods.

However, the performance is subject to certain limitations, such as Simplified Acoustic Environment. The anechoic conditions and static sources may not reflect the complexities of real-world scenarios with reverberation and dynamic sources.

Future work can include extending the setup to include reverberant environments and dynamic sources to evaluate real-world performance and developing efficient implementations for real-time processing.

References

[1] “Acoustic location,” Wikipedia. Sep. 24, 2024. Accessed: Dec. 01, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Acoustic\_location&oldid=1247456233

[2] C. Blandin, A. Ozerov, and E. Vincent, “Multi-source TDOA estimation in reverberant audio using angular spectra and clustering,” Signal Processing, vol. 92, no. 8, pp. 1950–1960, Aug. 2012, doi: 10.1016/j.sigpro.2011.09.032.

[3] C. Pan, J. Chen, and J. Benesty, “Performance Study of the MVDR Beamformer as a Function of the Source Incidence Angle,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 1, pp. 67–79, Jan. 2014, doi: 10.1109/TASL.2013.2283104.

[4] R. C. Hendriks, T. Gerkmann, and J. Jensen, DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State-of-the-Art. in Synthesis Lectures on Speech and Audio Processing. Cham: Springer International Publishing, 2013. doi: 10.1007/978-3-031-02564-8.

[5] “MMSE log-STSA.” Accessed: Dec. 01, 2024. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/7655-mmse-log-stsa

[6] “Beam Pattern - an overview | ScienceDirect Topics.” Accessed: Dec. 01, 2024. [Online]. Available: https://www.sciencedirect.com/topics/engineering/beam-pattern