Developing Audio Zoom in Virtual Environments:
Real-World Soundscapes and Targeted Noise Detection

InfoTech -2024

12/09/2024

Stephen Stroud, Dr Karl Jones, Dr Gerard Edwards, Colin Robinson, Dr Sebastian Chandler-Crnigoj and Dr David Ellis

Applied Forensic Technology Research Group (AFTeR), School of Engineering,
Faculty of Engineering and Technology Liverpool John Moores University

Introduction

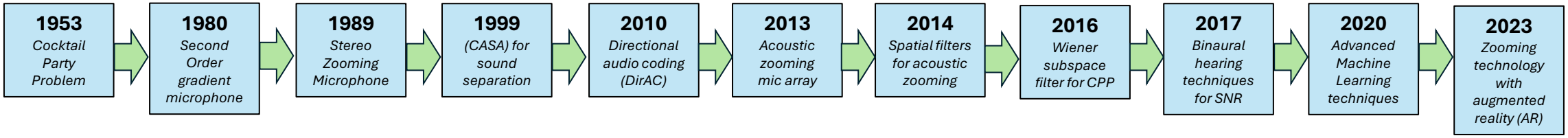
Overview: This research focuses on enhancing environmental sound recognition using advanced beamforming techniques.

Importance: Accurate sound identification is crucial for forensic evidence gathering and urban surveillance.

Objective: Develop a robust simulation environment that mimics real-world acoustic scenarios for better sound recognition and isolation.

Background

- Audio zooming allows users to focus on specific sounds within an auditory scene, similar to how a camera zooms in on a visual frame.
- The concept has been around since the 1950s, but replicating the human brain's ability to filter out unwanted noise remains challenging.
- Previous studies have shown various advancements in microphone technology and algorithms to improve audio zooming capabilities.



Key studies on The Cocktail Party Problem and audio zooming include works by Cherry [1], Van Waterschoot et al. [2], Thiergart, Kowalczyk, and Habets [3] and Fahim et al. [4]

[1] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," The Journal of the Acoustical Society of America, vol. 25, no. 5, pp. 975-979, 1953, doi: 10.1121/1.1907229.
[2] T. Van Waterschoot, W. Joos Tirry, and M. Moonen, "Acoustic Zooming by Multimicrophone Sound Scene Manipulation," *Audio Engineering Society,* vol. 61, 7/8, 2013.
[3] O. Thiergart, K. Kowalczyk, and E. A. P. Habets, "An acoustical zoom based on informed spatial filtering," 2014 2014: IEEE, doi: 10.1109/iwaenc.2014.6953348.
[Online]. Available: https://dx.doi.org/10.1109/iwaenc.2014.6953348
[4] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "Sound field separation in a mixed acoustic environment using a sparse array of higher order spherical microphones," in *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*, 1-3 March 2017 2017,
pp. 151-155, doi: 10.1109/HSCMA.2017.7895580.

The Problem

- Challenges in isolating specific sounds in cluttered environments.
- In real-world audio environments, interfering sounds can make it difficult to extract desired signals.
- There is a need for a system that effectively removes unwanted audio and retains only the desired sounds.
- This technology has potential applications in video surveillance and broadcast media.

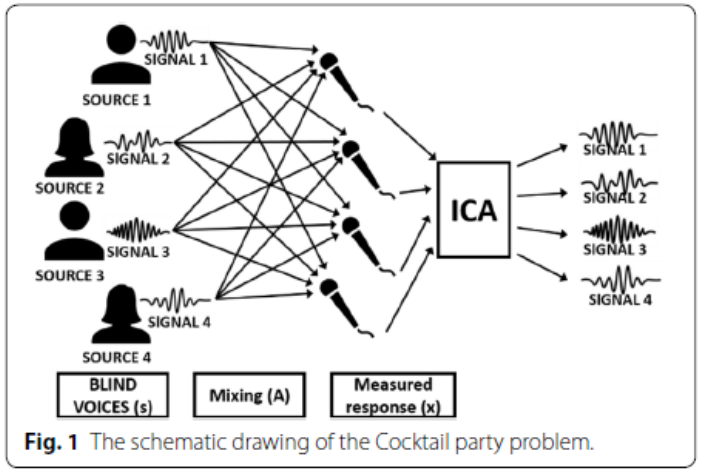
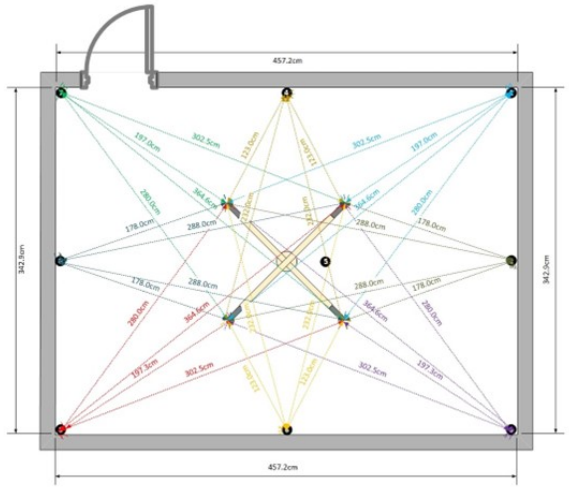
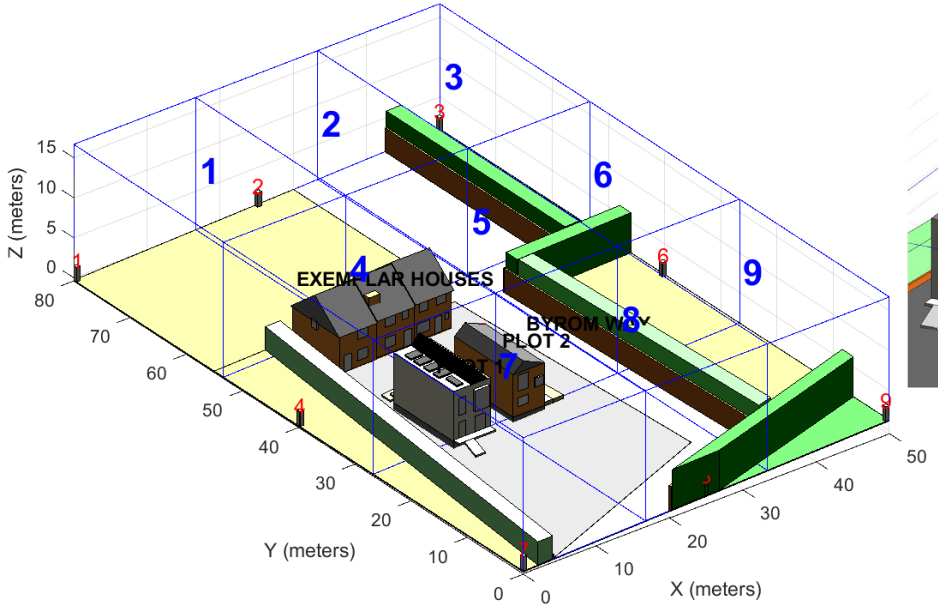


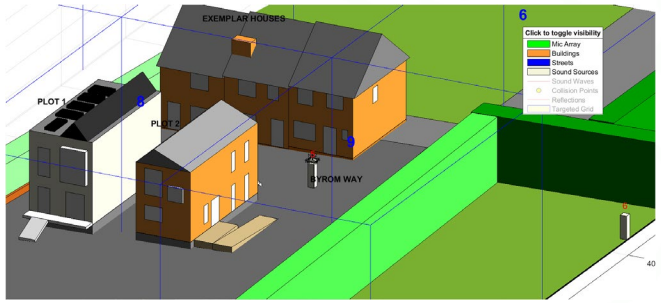
Fig. 1 The schematic drawing of the Cocktail party problem.



Methodology: Simulation Environment

- Developed a modular MATLAB-based simulator to enhance environmental sound recognition.
- Input quantities consist of actual experimental dimensions of the grounds surrounding 'Exemplar Houses' at LJMU.
- Generated a 3D scene using basic shapes and real-life Sabine acoustic coefficients to mimic real-world environments.

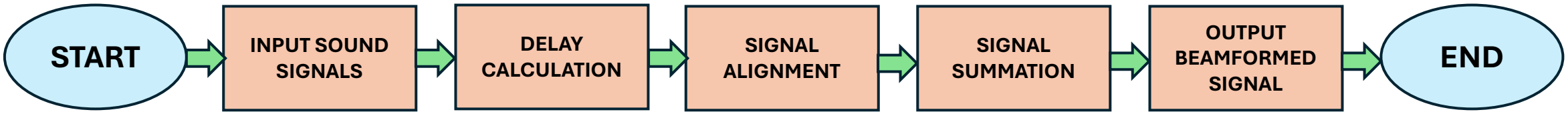




Methodology: Beamforming Algorithm

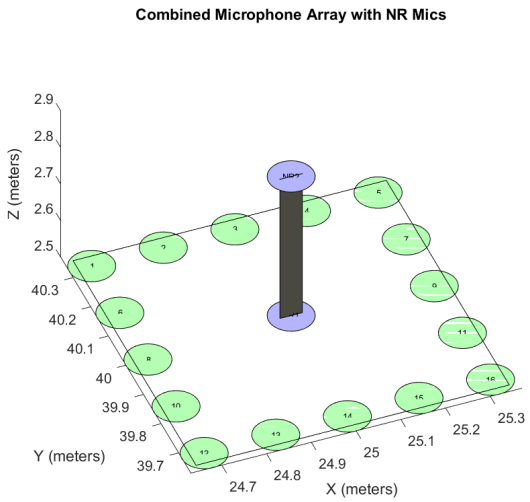
- The simulator allows users to specify the microphone array's location and shape.
- The algorithm focuses the array on precise sound sources within the chosen grid segment using time-delay beamforming.

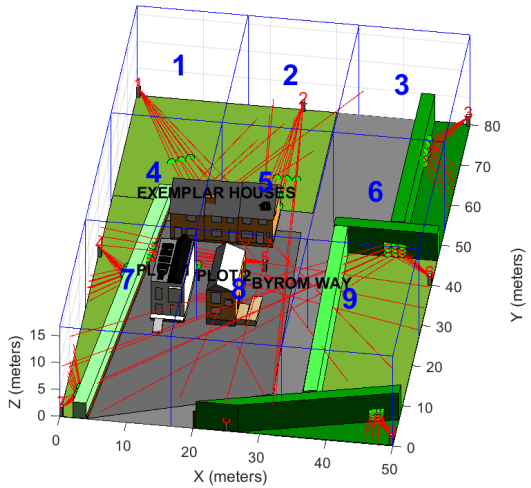
$$S_{out}(t) = \sum w_i S_{in}(t - \tau i)$$

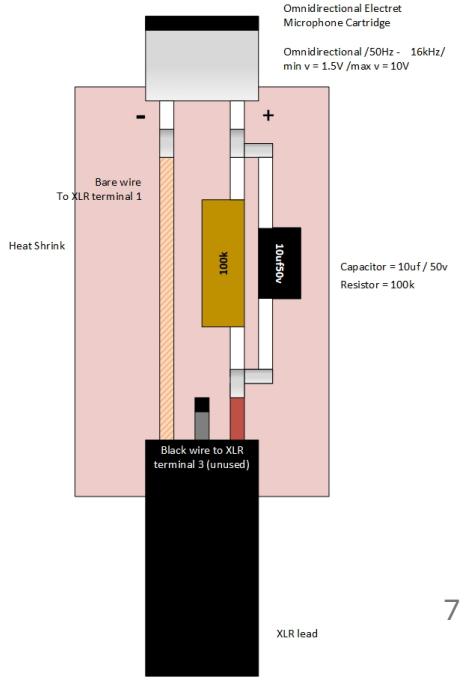


Methodology: Sound Capture and Noise Reduction

- A specialized microphone array designed for noise reduction is added on top of the main array.
- The subtractive noise reduction technique refines the beamformed signal using the feedforward microphone input.
- This process effectively reduces background noise, enhancing the overall sound quality.

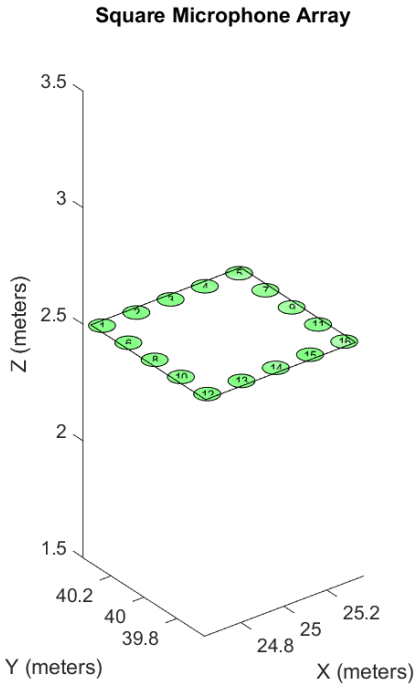


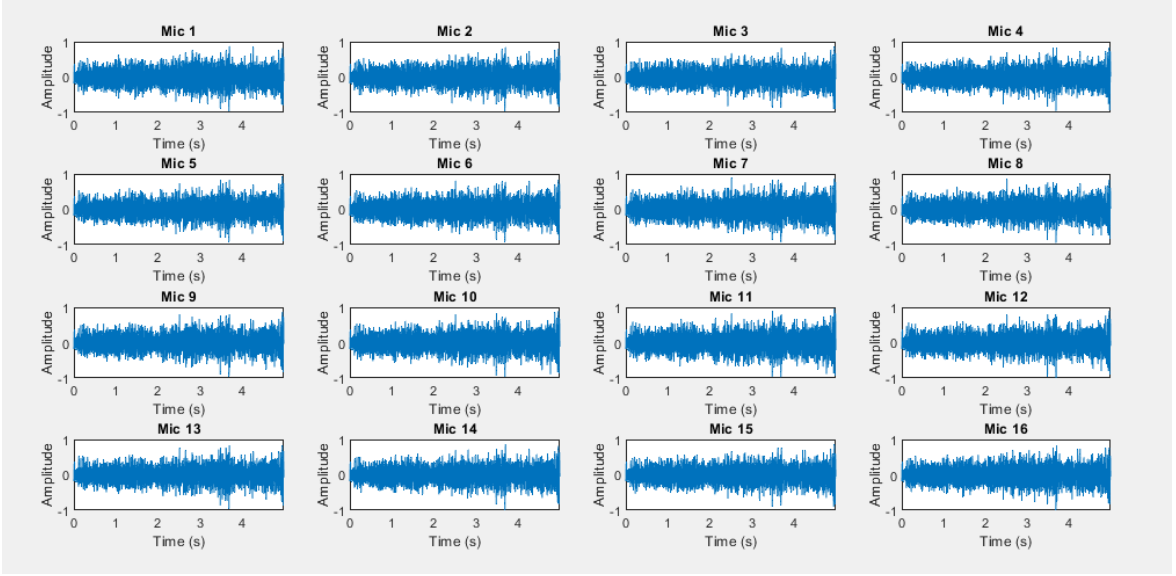


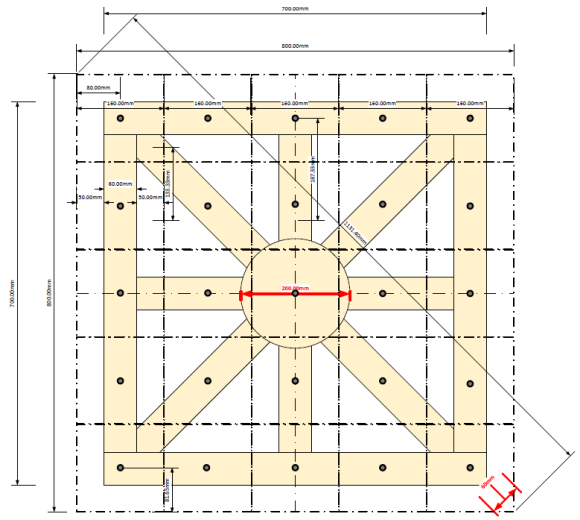


Results: Sound Capture

- The virtual omnidirectional microphones captured audio mixtures that closely emulate real-world scenarios.
- The algorithm considered critical parameters such as SPL, azimuth angle, elevation angle, and beam width.
- Recorded a 5-second audio file into a MATLAB array for each microphone, representing the intricate interplay of various audio components.

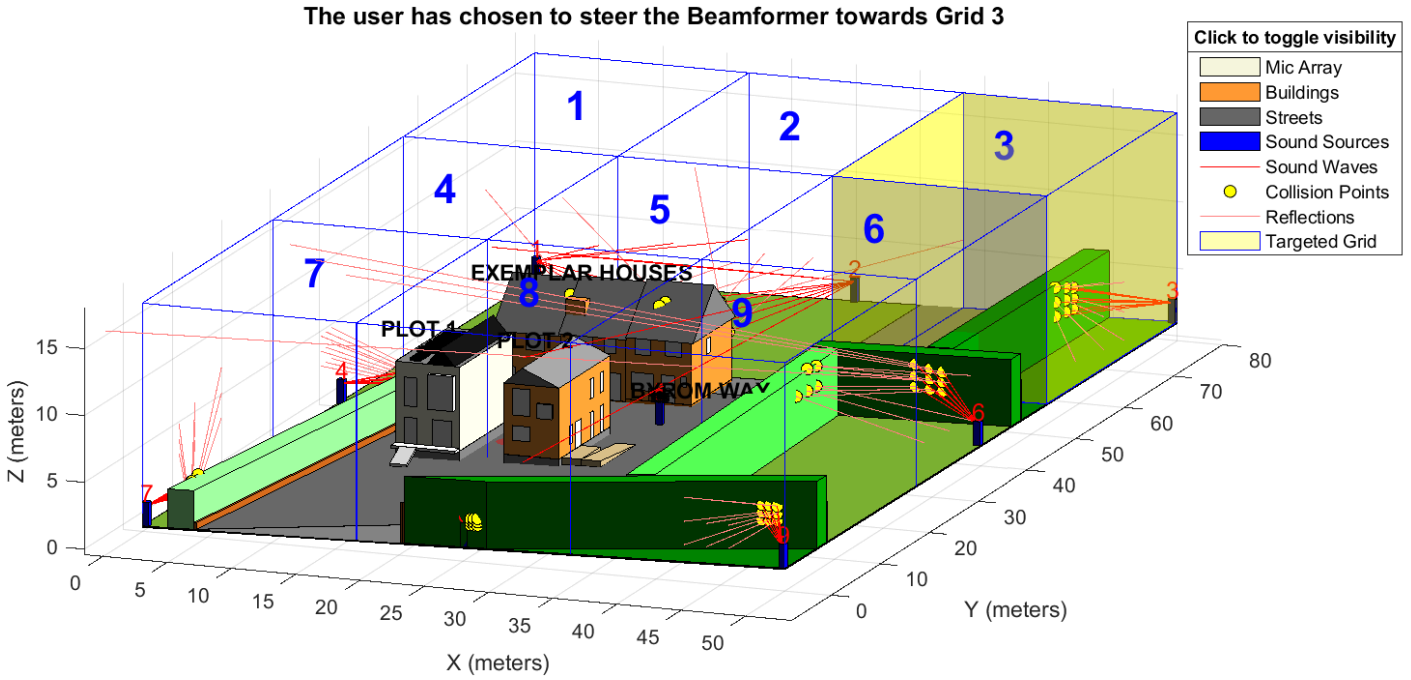


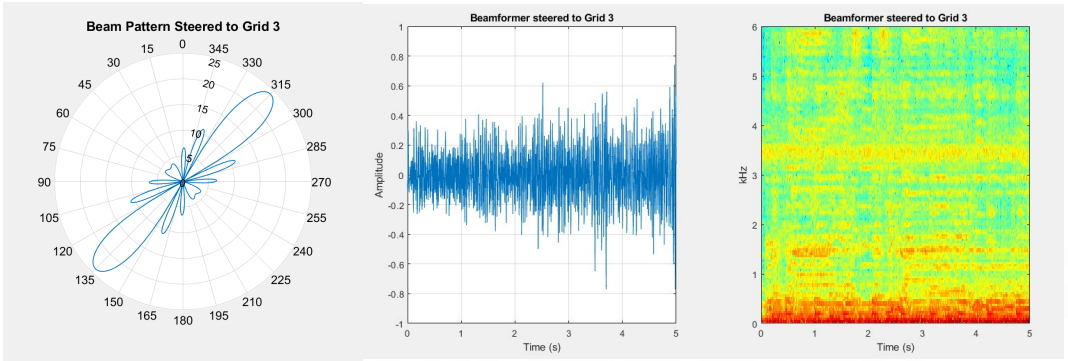




Results: Beamforming

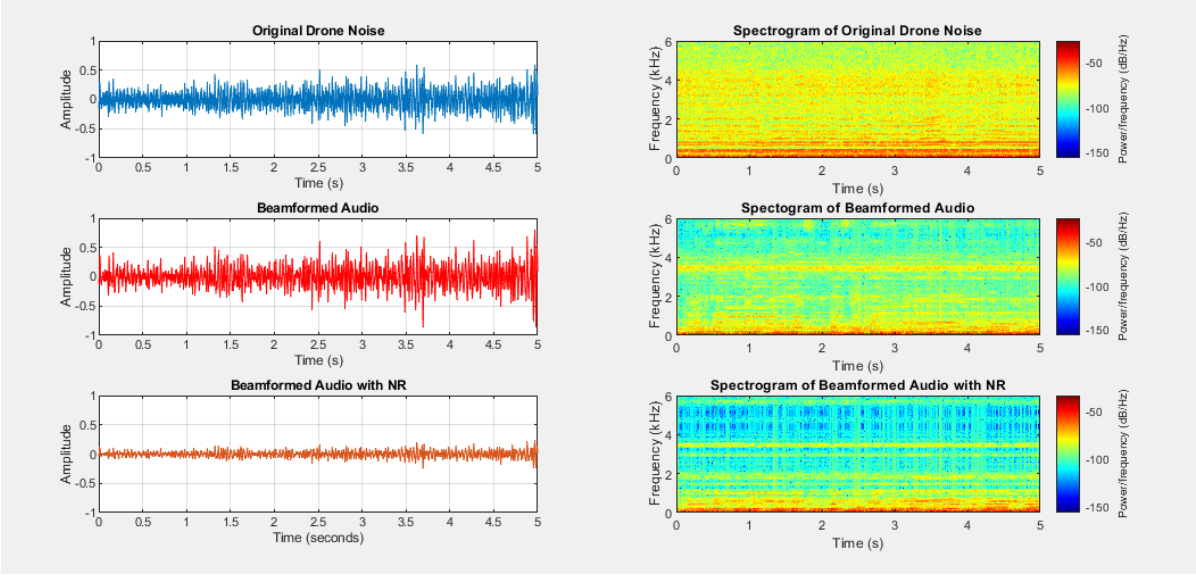
- The beamforming results demonstrated the algorithm's ability to selectively focus on desired sound sources.
- The algorithm enhanced the clarity and intensity of target sounds while reducing background noise.
- Processed beamformed audio was recorded and analysed in both time and frequency domains.

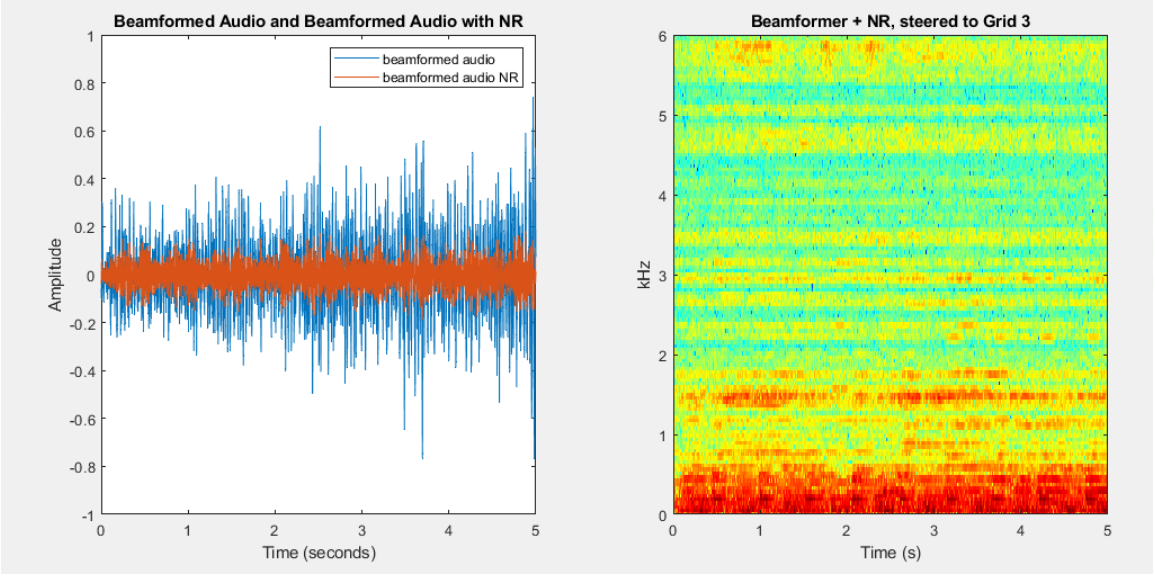




Results: Noise Reduction

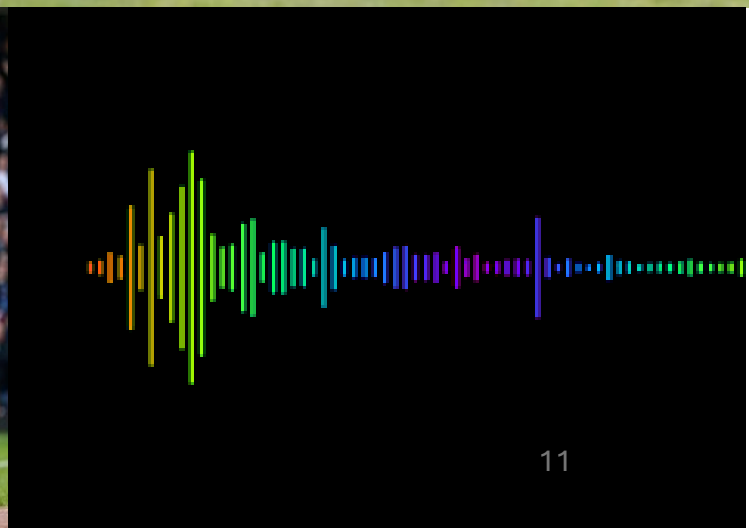
- Applied a subtractive noise reduction algorithm to the beamformed audio, significantly refining the clarity of the resultant sound.
- The algorithm isolated and subtracted ambient noise elements, enhancing human speech intelligibility within a chosen grid.
- This stage underscored the effectiveness of subtractive noise reduction techniques in improving audio recordings for surveillance purposes.





Applications

- The developed system has potential applications in forensic evidence gathering, environmental surveillance, and sound engineering.
- The technology could be used for synchronized audio and video surveillance systems.
- Future work aims to refine the beamforming algorithm further and explore additional applications.



Conclusion and Future Work

- The study successfully demonstrated a system capable of accurately steering captured audio and reducing unwanted noise.
- Future work will focus on enhancing the beamforming algorithm and reducing noise and interference levels.
- The goal is to achieve a level of audio quality and distinction that renders the technology invaluable for forensic surveillance and broadcasting applications.

Developing Audio Zoom in Virtual Environments:
Real-World Soundscapes and Targeted Noise Detection

Happy to answer questions.

s.stroud@2022.ljmu.ac.uk

Stephen Stroud, Dr Karl Jones, Dr Gerard Edwards, Colin Robinson, Dr Sebastian Chandler-Crnigoj and Dr David Ellis

Applied Forensic Technology Research Group (AFTeR), School of Engineering,
Faculty of Engineering and Technology Liverpool John Moores University