DOI: 10.61782/fa.2025.0728

FORUM ACUSTICUM EURONOISE 2025

REAL-TIME IMPLEMENTATION OF A USER-SELECTIVE ACTIVE SOUND CONTROL SEAT BASED ON AUDITORY SCENE ANALYSIS

Tahmida Islam* **Juan Garcia** **Maria Garzon** **Miguel Ferrer**
Maria de Diego **Laura Fuster** **Alberto Gonzalez**
Inst. of Telecommunications and Multimedia Applications (iTEAM)
Universitat Politècnica de València, València
Spain

ABSTRACT

Environmental noise can harm people's health, leading to issues such as hearing loss, increased stress levels, and reduced overall well-being. This noise pollution is a growing concern in urban environments, where constant exposure to sounds from traffic, construction, and other sources can create an uncomfortable living situation. We can implement active sound control techniques to address these challenges, which offer a promising solution to reduce unwanted noise. This method uses loudspeakers to change the sound in an area and create a specific sound field. However, real-time active sound control is challenging. There are limitations based on where we place microphones and loudspeakers and how we select sound sources to create a comfortable environment for users. This paper presents a user-selective active sound seat that addresses these design challenges. This seating solution combines the principles of active sound control with advanced techniques derived from the Auditory Scene Analysis framework. These techniques help separate individual sound sources from a mix by analyzing the position and characteristics of different sounds in a room. As a result, users of the active sound seat can selectively mask unwanted sounds in their vicinity.

Keywords: *active sound control, sound separation, spatial filtering*

**Corresponding author*: *tislam@upv.edu.es.*

1. INTRODUCTION

Active Sound Control (ASC) is an acoustic technology that manipulates sound waves using secondary sources (loudspeakers) to shape the acoustic environment. A key application of ASC is Active Noise Control (ANC), where the system generates a sound field that destructively interferes with unwanted noise, effectively reducing it. Beyond noise reduction, ASC can also enhance auditory experiences, such as through spatial audio reproduction and zonal sound control [1, 2]. These capabilities allow ASC to suppress noise and selectively amplify or modify specific sound fields, improving sound quality in complex environments. In automotive applications, ASC technologies are increasingly used to reduce in-cabin noise and enrich the driving experience through powertrain sound enhancement or infotainment features [3].

A key enabler of such user-focused ASC systems is Auditory Scene Analysis (ASA), a perceptual process by which the human auditory system separates and identifies individual sound sources in a complex acoustic environment [4]. ASA helps us understand and focus on specific sounds even in noisy surroundings [5] but selectively listening to different sounds in an acoustic mixture remains a challenge in machine hearing [6]. Spatial cues, such as interaural time and level differences, play a crucial role in localizing sound sources—analogous to how microphone arrays can be used in machines to separate overlapping sources based on spatial filtering [7].

Spatial filtering techniques like beamforming are widely used for sound source separation. Traditional beamformers focus on sounds from a specific direction, but modern implementations must address more complex environments with reverberation and multiple sources. In this broader context, beamformers are viewed as multiple-input single-output systems that combine spatially filtered microphone signals to enhance desired sources and suppress interference [8]. Recent advancements incorporate neural networks (NNs) to estimate beamforming parameters or directly predict filter coefficients [9–11]. Some architectures combine NNs with traditional spatial filters [12–14], followed by post-filtering stages for further enhancement [15, 16].

After separating sound sources using ASA-inspired techniques, ANC is applied to suppress unwanted noise and maintain an ambient acoustic environment around the listener. Among various ANC algorithms, the Filtered-X Least Mean Square (FxLMS) algorithm is favored for its simplicity and real-time adaptability [17–19]. Most real-time ANC systems use the FxLMS algorithm or its modifications [20–24] due to its proven effectiveness. In recent years, many researchers and engineers have chosen the Speedgoat platform to implement and test different acoustic models. Speedgoat supports machine-in-the-loop experiments, real-time simulation, and control validation under dynamic conditions [25, 26].

This work presents a real-time implementation of a user-selective Active Sound Control (ASC) seat that integrates Auditory Scene Analysis (ASA) principles. An NN-based spatial filtering system separates sound sources captured via a microphone array. After sound separation, a multi-channel FxLMS ANC system actively cancels unwanted noise around a selected user. The system is implemented and tested on the Speedgoat platform to validate performance under real-time conditions and ensure a targeted acoustic response around the seat.

2. METHODOLOGY

2.1 Multi-channel Active Noise Control System

The proposed system is based on a generic multi-channel Active Noise Control (ANC) framework consisting of I reference signals, J secondary sources, and K error sensors. The adaptive filtering approach used in this system follows the MultiChannel Filtered-X Least Mean Squares (MFXLMS) algorithm [27], which is detailed as follows.

The adaptive filter output at the j-th secondary source is expressed as:

$$\mathbf{y}_j(n) = \sum_{i=1}^{I} \mathbf{w}_{ji}^T(n)\mathbf{x}_i(n) \, . \tag{1}$$

The adaptive filter coefficient vector is given by:

$$\mathbf{w}_{ji}(n) = [w_{ji,0}(n), w_{ji,1}(n), \ldots, w_{ji,N-1}(n)]^T \, , \tag{2}$$

where N is the number of filter coefficients.

The input tap vector is defined as:

$$\mathbf{x}_i(n) = [\mathbf{x}_i(n), \mathbf{x}_i(n-1), \ldots, \mathbf{x}_i(n-N+1)]^T \, . \tag{3}$$

The update equation for the adaptive filter coefficients is given by:

$$\mathbf{w}_{ji}(n) = \mathbf{w}_{ji}(n-1) + \mu \sum_{k=1}^{N} e_k(n)(\hat{\mathbf{S}}_{kj}(n) * \mathbf{x}_i(n)) \, . \tag{4}$$

Here, $\hat{\mathbf{S}}_{kj}(n)$ represents the estimated secondary path between the kth error sensor and the jth secondary source.

For example, a 2×2 MFXLMS system block diagram is shown in Figure 1. The proposed control system consists of two main modules: an **ASA Module** and an **ANC Module**. Both modules work together to create a personalized acoustic environment by suppressing unwanted noise while preserving desirable sounds such as music, alarms, and voice alerts.

The ANC module implements real-time noise cancellation using the MFXLMS algorithm. It operates on a sample-by-sample basis to minimize the error signal at the error microphones, thereby ensuring an ambient acoustic zone around the user.

2.2 Auditory Scene Analysis and Source Separation

The system must first isolate desired and undesired sound sources to enable user-selective noise cancellation. This separation is accomplished using a *Filter-and-Sum Beamformer* (FASB) [28], which enhances signals originating from a particular direction while suppressing others.

The FASB beamformer processes the signals captured by an array of microphones as follows:

- The beamformer individually filters each microphone signal.
- The filtered signals are then summed to generate a reference signal representing the selected noise source.

This reference signal is fed into the ANC system to cancel user-specified noises selectively.

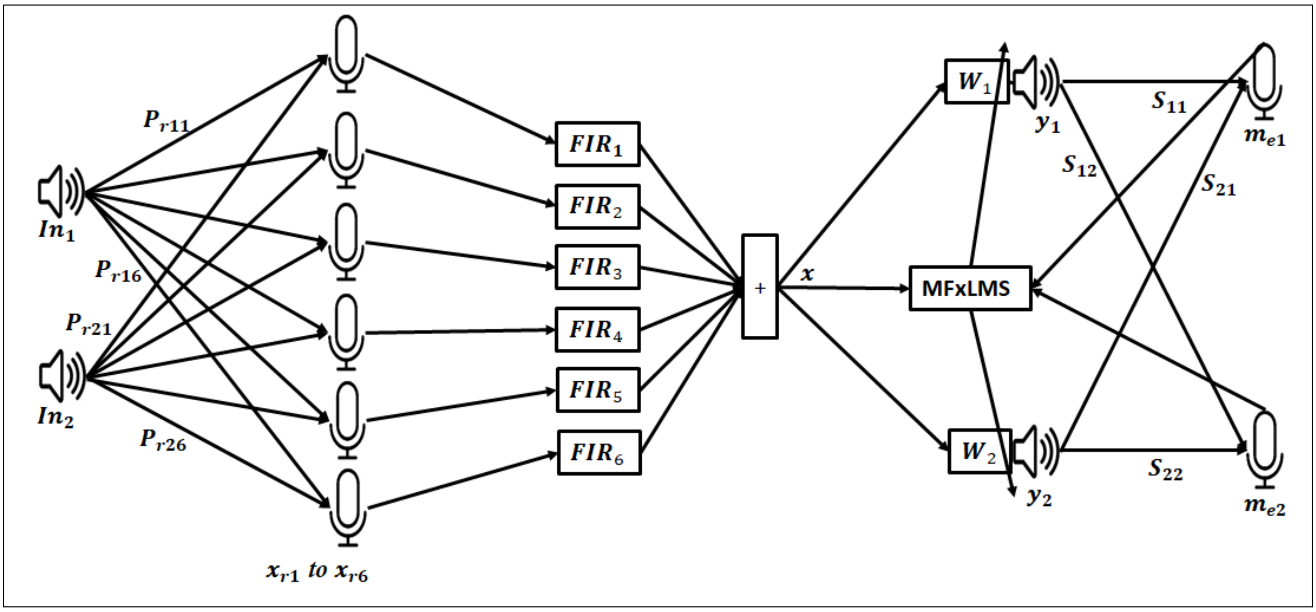

Figure 1. A block diagram illustrating an example of ASA with multichannel ANC system.

2.3 Neural Network-Based Beamformer Optimization

The beamformer filters are designed using an NN to optimize spatial filtering performance. The objective is to extract a single target sound source while suppressing others.

Consider an experimental setup with two independent sound sources located at different positions. The signals captured by the microphone array can be expressed as:

$$x_{rm}[n] = In_1[n] * \mathbf{P}_{r1m} + In_2[n] * \mathbf{P}_{r2m}\,, \quad (5)$$

where $\mathbf{P}_{rim}$ represents the acoustic path between the i-th source and the m-th microphone and $In_i[n]$ defines the primary signals at time instant n.

The signals from all microphones are linearly combined using a set of beamformer filters to extract the target source:

$$x[n] = \sum_{m=1}^{M} x_{rm}[n] * \mathbf{h}_m\,. \quad (6)$$

Here, $\mathbf{h}_m$ are the beamformer filter coefficients to be optimized.

The ideal objective is to recover only the target signal, i.e.,

$$x[n] = In_1[n]\,. \quad (7)$$

However, direct inversion of the acoustic propagation paths is highly complex. Instead, a relaxed objective is defined as:

$$x[n] = In_1[n] * \mathbf{P}_{r1m}\,. \quad (8)$$

This relaxed objective avoids the need for deconvolution while still achieving effective separation.

A neural network model is used to optimize the beamforming filters. The NN takes as input the raw signals from multiple microphones and processes them through a set of 1D convolutional layers. The outputs of these filters are summed to generate the beamformed signal.

The loss function used for training is the Mean Squared Error (MSE) between the network's output and the desired reference signal:

$$\mathcal{L} = \frac{1}{T}\sum_{n=1}^{T}\left(x[n] - In_1[n] * \mathbf{P}_{r1m}\right)^2\,. \quad (9)$$

The neural network learns optimal beamformer filter coefficients that maximize target source extraction by minimizing this loss function.

2.4 Real-Time Implementation on Speedgoat

The system, including the ASA-based source separation and MFXLMS-based ANC, is deployed in real time using the Speedgoat rapid control prototyping platform. Speedgoat allows for hardware-in-the-loop testing and real-time validation of the proposed algorithm under practical conditions.

The real-time implementation involves:

- Estimating multi-microphone acoustic paths $\mathbf{P}_{rim}$ to perform the offline design of beamformers using the trained neural network.
- Extracting the selected noise source using the designed beamformers and feeding it into the ANC system.
- Applying the MFXLMS algorithm to generate anti-noise signals in real time.
- Continuously adapting the ANC system to changing acoustic conditions to ensure effective noise suppression.

3. EXPERIMENTAL RESULTS

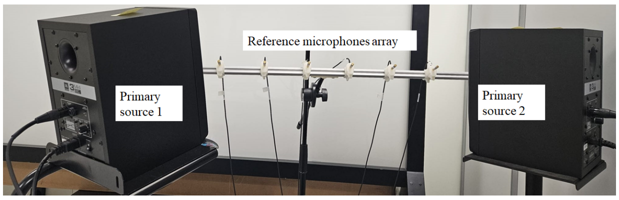

Figure 2. An array of six reference microphones for sound source separation.

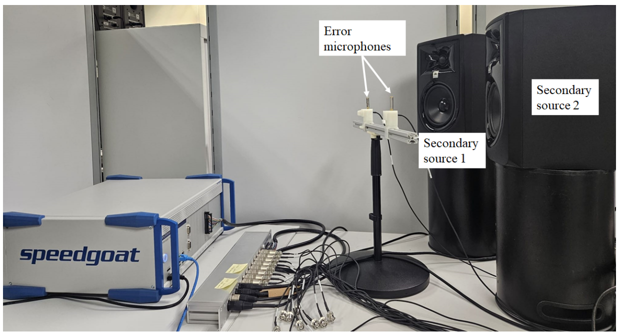

Figure 3. Control Area of the real time implementation.

This experiment utilized two primary sources to generate white noise and speech signals, where white noise was regarded as an undesired signal. An array of six microphones served as reference microphones (see Figure 2). The beamformers filtered the signals from these microphones to create a reference signal that represented the chosen noise source. In the control area, two secondary sources and two error microphones were employed to achieve user-selective Active Sound Control (ASC) at the car seat (see Figure 3).

The real-time experiments were conducted in three stages to assess the performance of the proposed ASA system, Only speech playing (Baseline Measurement): This scenario establishes a reference for clean speech signals without interference.

- Only speech playing (Baseline Measurement): This scenario establishes a reference for clean speech signals without interference.
- Speech and noise playing, ANC off: This represents a real-world noisy environment where the desired speech signal is masked by background noise.
- Speech and noise playing, ANC on: The ANC system is activated to suppress undesired noise while preserving the speech signal.

Figure 4. Power Spectral Density (PSD) of the signal at the error position with the ANC system Off, On, and with only a speech signal.

The measured signals from the error microphone in these scenarios are visualized in Figure 4. The dashed line represents when ANC is turned off, showing higher power spectral density due to noise. The dotted line corresponds to the clean speech signal without noise as a reference for ideal speech quality. The solid line represents the scenario with ANC enabled, illustrating the noise reduction. In the ANC off condition, the noise significantly raises the power spectral density across the frequency spectrum, confirming that speech intelligibility is negatively impacted. When ANC is activated, there is a noticeable reduction in noise levels, as seen by the decrease in power spectral density (solid line curve compared to dash line). This demonstrates that the ANC system effectively reduces unwanted noise while preserving the desired speech signal.

4. CONCLUSION

The experimental results confirm that the proposed Active Sound Amplification (ASA) system significantly improves the acoustic environment by reducing unwanted noise while preserving desired sounds. The offline neural network processing used to obtain beamformer information simplifies the ASA, making it less computationally demanding for real-time implementation. This design makes ASA a promising solution for applications that require personalized sound zones, such as in-vehicle communication or selective noise control in shared spaces. However, some residual noise is still present even when active noise cancellation (ANC) is activated. This suggests that while the ASA-based sound separation performs well, further refinements in spatial filtering or active noise cancellation could enhance its effectiveness.

5. ACKNOWLEDGEMENTS

This project is supported by Horizon Europe Marie Skłodowska-Curie Actions (MSCA) Doctoral Net-work under the European Union's funding scheme. The project title is 'Active reduction of noise transmitted into and from enclosures through encapsulated structures' and is identified by project number 101073037.

6. REFERENCES

[1] F. Rumsey, "Sound field control," *Journal of the Audio Engineering Society*, vol. 61, no. 12, pp. 1046–1050, 2013.

[2] F. Rumsey, *Spatial audio*. Routledge, 2012.

[3] J. Cheer, "Active sound control in the automotive interior," *Future Interior Concepts*, pp. 53–69, 2021.

[4] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[5] B. T. Szabó, S. L. Denham, and I. Winkler, "Computational models of auditory scene analysis: a review," *Frontiers in Neuroscience*, vol. 10, p. 524, 2016.

[6] I. Kavalerov, S. Wisdom, H. Erdogan, B. Patton, K. Wilson, J. Le Roux, and J. R. Hershey, "Universal sound separation," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 175–179, IEEE, 2019.

[7] D. Markovic, A. Defossez, and A. Richard, "Implicit neural spatial filtering for multichannel source separation in the waveform domain," *arXiv preprint arXiv:2206.15423*, 2022.

[8] E. Vincent, T. Virtanen, and S. Gannot, *Audio source separation and speech enhancement*. John Wiley & Sons, 2018.

[9] A. A. Nugraha, A. Liutkus, and E. Vincent, "Multichannel audio source separation with deep neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1652–1664, 2016.

[10] J. Heymann, L. Drude, C. Boeddeker, P. Hanebrink, and R. Haeb-Umbach, "Beamnet: End-to-end training of a beamformer-supported multi-channel asr system," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5325–5329, IEEE, 2017.

[11] W. Zhang, A. S. Subramanian, X. Chang, S. Watanabe, and Y. Qian, "End-to-end far-field speech recognition with unified dereverberation and beamforming," *arXiv preprint arXiv:2005.10479*, 2020.

[12] J. Wu, Z. Chen, J. Li, T. Yoshioka, Z. Tan, E. Lin, Y. Luo, and L. Xie, "An end-to-end architecture of online multi-channel speech separation," *arXiv preprint arXiv:2009.03141*, 2020.

[13] Y. Fu, J. Wu, Y. Hu, M. Xing, and L. Xie, "Desnet: A multi-channel network for simultaneous speech dereverberation, enhancement and separation," in *2021 IEEE Spoken Language Technology Workshop (SLT)*, pp. 857–864, IEEE, 2021.

[14] W. Liu, A. Li, C. Zheng, and X. Li, "A neural beam filter for real-time multi-channel speech enhancement," *arXiv preprint arXiv:2202.02500*, 2022.

[15] Z.-Q. Wang and D. Wang, "Multi-microphone complex spectral mapping for speech dereverberation,"

in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 486–490, IEEE, 2020.

[16] Z.-Q. Wang, P. Wang, and D. Wang, "Multi-microphone complex spectral mapping for utterance-wise and continuous speech separation," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 29, pp. 2001–2014, 2021.

[17] J. Lorente, M. Ferrer, M. de Diego, and A. Gonzalez, "The frequency partitioned block modified filtered-x nlms with orthogonal correction factors for multichannel active noise control," *Digital Signal Processing*, vol. 43, pp. 47–58, 2015.

[18] D. Shi, W.-S. Gan, B. Lam, and C. Shi, "Two-gradient direction fxlms: An adaptive active noise control algorithm with output constraint," *Mechanical Systems and Signal Processing*, vol. 116, pp. 651–667, 2019.

[19] M. Tufail, S. Ahmed, M. Rehan, and M. T. Akhtar, "A two adaptive filters-based method for reducing effects of acoustic feedback in single-channel feedforward anc systems," *Digital Signal Processing*, vol. 90, pp. 18–27, 2019.

[20] X. Mu, J. Ko, and J. Rheem, "Modified fxlms algorithm for active noise control and its real-time implementation," *Journal of the Institute of Electronics and Information Engineers*, vol. 50, no. 9, pp. 172–176, 2013.

[21] J. Lorente, M. Ferrer, M. De Diego, and A. González, "Gpu implementation of multichannel adaptive algorithms for local active noise control," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 11, pp. 1624–1635, 2014.

[22] M. Ferrer, A. Gonzalez, M. de Diego, and G. Pinero, "Convex combination filtered-x algorithm for multichannel active noise control,"

[23] D. Sachau and S. Jukkert, "Real-time implementation of the frequency-domain fxlms algorithm without block delay for an adaptive noise blocker," *Proceedings of the ISMA, Leuven, Belgium*, pp. 15–17, 2014.

[24] L. Liu, Q. Su, W. Li, and S. M. Kuo, "Real time implementation and experiments of multi-channel active noise control system for icu," in *2021 IEEE International Conference on Electro Information Technology (EIT)*, pp. 395–400, IEEE, 2021.

[25] A. Pelletier, P. Micheau, and A. Berry, "Implementation of a self-sensing piezoelectric actuator for vibro-acoustic active control," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2014*, vol. 9061, pp. 1093–1102, SPIE, 2014.

[26] X. Xu, Y. Lu, C. Lan, Z. Xing, and M. Shao, "Experimental research on global active rotor noise control using near-field acoustic holography and sound field reproduction," *Mechanical Systems and Signal Processing*, vol. 206, p. 110930, 2024.

[27] T.-H. Jung, J.-H. Kim, K.-J. Kim, and S.-W. Nam, "Active noise reduction system using multi-channel anc," in *2011 11th International Conference on Control, Automation and Systems*, pp. 36–39, IEEE, 2011.

[28] M. Kajala and M. Hamalainen, "Filter-and-sum beamformer with adjustable filter characteristics," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, vol. 5, pp. 2917–2920, IEEE, 2001.