DOI: 10.61782/fa.2025.0343

FORUM ACUSTICUM EURONOISE 2025

ENHANCING IN-VEHICLE SPEECH INTELLIGIBILITY: A STUDY ON THE BENEFITS OF ACTIVE ROAD NOISE CANCELLATION SYSTEM

Tommaso Botti[1] **Marco Olivieri**[2] **Jacopo Gino**[1] **Alfonso Oliva**[2*]

[1] Teoresi S.p.A, Torino, Italy
[2] Ferrari S.p.A., Maranello, Italy

ABSTRACT

The acoustic quality inside vehicles has become a key factor in the overall evaluation of cars, particularly in luxury models. As a result, improving in-cabin communication and overall acoustic comfort has gained significant attention from the automotive industry. Recently technologies, such as Active Road Noise Cancellation (ARNC), have been extensively explored to reduce unwanted noise, effectively lowering acoustic pressure within the vehicle cabin. Beyond noise reduction, these algorithms have the potential to significantly enhance in-vehicle speech intelligibility. In this study, we investigate the impact of ARNC algorithms on person-to-person speech communication, simulating conversations between passengers. Experimental measurements have been conducted in a four-seat vehicle model equipped with a state-of-the-art ARNC system. Passenger communication has been simulated by convolving reference speech signals with measured acoustic impulse responses within the vehicle cabin. Additionally, we compare different metrics to assess speech intelligibility between front and rear seats, evaluating the effectiveness of the ARNC algorithm. The results demonstrate the potential of ARNC in enhancing in-vehicle speech intelligibility, offering valuable insights for future advancements in automotive audio technology.

Keywords: *speech intelligibility, in-car communications, active noise control*

**Corresponding author*: *Alfonso.Oliva@ferrari.com.*

1. INTRODUCTION

The ability to communicate effectively within a car is crucial for both comfort and safety, making acoustic optimization a key area of research and development [1, 2]. Traditionally, improving acoustic comfort in a vehicle has relied on passive noise control techniques, such as by improving insulation materials, acoustic glazing, and optimized cabin structures. Although these methods effectively reduce overall noise levels, they are often limited by weight, cost, and design limitations. Furthermore, passive solutions are typically more effective at high frequencies, while their insulating properties tend to degrade at lower frequencies, making them less effective against road-induced low-frequency noise.

As an alternative, active solutions have gained significant attention, encompassing a variety of approaches based on advanced signal processing, including personalized audio zones (commonly referred to as ”private bubbles”) and Active Road Noise Cancellation (ARNC). The latter, in particular, aims to reduce low-frequency road noise by generating anti-noise signals, typically through the vehicle's audio system. Although ARNC has been developed primarily to enhance passenger comfort by reducing background noise levels, its potential impact on speech intelligibility within the vehicle has not yet been fully explored.

In this paper, we assess the indirect benefits of ARNC on the in-car speech intelligibility by analyzing how reduced road noise contributes to clearer in-vehicle communication. We present a series of experimental measurements along with objective and subjective evaluations, aiming to provide insights into the role of ARNC not only as a comfort-enhancing or weight-reduction technology, but also as a potential enabler of improved speech communication among vehicle occupants.

The rest of the paper is structured as follows. Sec. 2 present the context and background of ARNC and speech intelligibility. In Sec. 3 we present the methodology and experimental scenario to evaluate the impact of ARNC in passanger-to-passanger communications. Results are provided in Sec. 4 along with objective metrics and subjective evaluations. In Sec. 5 we draw final conclusions and considerations.

2. BACKGROUND AND PROBLEM FORMULATION

2.1 Active Road Noise Cancellation

Active Road Noise Cancellation (ARNC) refers to the application of Active Noise Cancellation (ANC) techniques to mitigate low-frequency road noise inside vehicle cabins. ANC techniques actively generate anti-noise signals to counteract unwanted sound waves, leveraging the principle of destructive interference [3, 4]. Unlike passive solutions, which rely on insulating materials and structural modifications, active techniques require an external power source and the implementation of a control strategy.

Most control systems are based on the well-known Filtered-x Least Mean Squares (FxLMS) algorithm [5, 6], which is an extension of the LMS algorithm [7] belonging to the gradient descent [8] family of optimization techniques. These techniques minimize a cost function by iteratively moving in the direction of its negative gradient [9]. FxLMS is a hybrid-feedforward control strategy based on adaptive filtering [7] that iteratively adjusts the filter coefficients to minimize the residual error, i.e., the sound pressure level at the cancellation position. The algorithm employs reference sensors, error microphones, and actuators to find the optimal solution and apply the control action.

A typical ARNC architecture adopted accelerometers mounted on the chassis or suspension system as reference sensors to detect road-induced vibrations before they propagate into the cabin. To measure the residual error, microphones are strategically placed within the cabin. Finally, actuators are typically the loudspeakers of the car's sound system [10].

The ARNC system operates by modeling two main paths: the primary and the secondary [11]. The primary path refers to the sound propagation from the road to the cabin, while the secondary path represents the transfer function from the actuators to the microphones within the cabin. The FxLMS algorithm continuously adapts to these paths to optimize noise cancellation, accounting for plant variations.

2.2 Speech Intelligibility

Speech intelligibility within a vehicle cabin is a critical aspect of passenger communication, particularly in the presence of road noise [12, 13]. High noise levels can mask speech signals, reducing the ability of passengers to understand each other clearly. The assessment of speech intelligibility in noisy environments typically relies on both objective and subjective evaluation methods.

Several metrics are commonly used to quantify speech intelligibility. The Short-Time Objective Intelligibility (STOI) [14] metric provides a prediction of intelligibility by analyzing the correlation between clean and degraded speech signals. The Signal-to-Noise Ratio (SNR) [15] quantifies the relative level of speech against background noise, with higher values indicating improved clarity.

In this paper, we leverage these metrics to quantify the impact of ARNC on speech intelligibility, comparing objective measurements in different road conditions to assess its role in improving in-vehicle communication.

3. METHODOLOGY

In this study, a speaker located in the rear right seat of a car has been simulated, and the impact of an ARNC system on speech intelligibility at the front right passenger position has been assessed. To achieve this, a state-of-the-art active noise cancellation system based on the Filtered-x Least Mean Squares algorithm has been implemented.

3.1 Experimental scenario

In order to evaluate the potential benefits of an Active Road Noise Cancellation (ARNC) system on in-vehicle speech intelligibility, an FxLMS-based algorithm was implemented and tested on the vehicle. The car was instrumented with accelerometers placed at carefully selected locations, while microphones were positioned inside the cabin, near the right front passenger's ears.

We adopted the mouth simulator of a Head and Torso Simulator (HATS) [16] placed on the right rear passenger seat to simulate human speeches. Specifically, the impulse responses (IRs) from the artificial mouth to the left and right microphones at the co-driver's seat, with an occupant

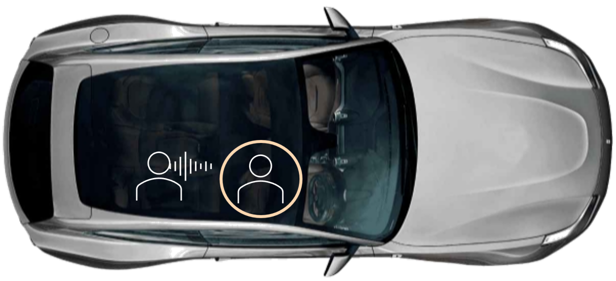

Figure 1. Measurement setup. Speaker at the rear-right seat and listener at the front-right seat along with the ARNC system.

in place, were measured in an anechoic chamber to characterize the acoustic transmission path through the Exponential Sine Sweep (ESS) method [17]. A schematic representation of the experimental setup is shown in Fig. 1. Notice that, in this study we assume the left-right microphones coincide with the co-driver ears, thus the measured IRs correspond to the transfer function from the rear-right talker to the ears of listener position under test.

Firstly, road noise measurements were performed on different road surfaces and at various speeds with the ARNC system deactivated. Then, the same measurements were repeated with the ARNC system enabled. In the rest of the paper, we will refer to those measurements as "*ARNC OFF*" and "*ARNC ON*", respectively. Therefore, this process provided road noise signals at the right front passenger's ears, both with and without active noise cancellation.

Moreover, a speech source at the rear-right passenger seat was simulated by convolving different clean speech signals with the previously measured IRs. The test signal was carefully designed to include both female and male speech segments, with progressively decreasing amplitude levels of $0\,\mathrm{dBFS}$, $-6\,\mathrm{dBFS}$, and $-12\,\mathrm{dBFS}$, respectively.

Finally, the resulting simulated speech signals at the right front passenger's ears were superimposed onto the road noise recordings, both with *ARNC OFF* and *ARNC ON*, allowing for a direct comparison of speech intelligibility under different noise conditions.

4. RESULTS

The impact of the ARNC system in terms of speech intelligibility has been evaluated with objective analysis and perceptual subjective analysis. The former, with the computation of different metrics. The latter, with the evaluation of two different perceptual listening tests. The following sections provide detailed explanations of the results along with insights and interpretations.

4.1 Objective analysis

To quantify the impact of the ARNC system on in-vehicle communication, we compute two different metrics presented in the literature of speech intelligibility, as explained in Sec. 2.2.

First of all, the Short-Time Objective Intelligibility (STOI) has been computed as reported in [14]:

$$STOI = \frac{1}{JM}\sum_{j,m} STOI_{j,m} \quad , \tag{1}$$

where:

$$STOI_{j,m} = \frac{(\boldsymbol{x}_{j,m} - \mu_{\boldsymbol{x}_{j,m}})^T \quad (\bar{\boldsymbol{y}}_{j,m} - \mu_{\bar{\boldsymbol{y}}_{j,m}})}{||\boldsymbol{x}_{j,m} - \mu_{\boldsymbol{x}_{j,m}}|| \quad ||\bar{\boldsymbol{y}}_{j,m} - \mu_{\bar{\boldsymbol{y}}_{j,m}}||}, \tag{2}$$

with

- J is the total number of one-third octave bands
- M denotes the total number of temporal frames
- $\boldsymbol{x}_{j,m}$ represents the clean speech short-time temporal envelope in the $j-th$ frequency band and $m-th$ time frame
- $\bar{\boldsymbol{y}}_{j,m}$ represents the normalized degraded speech envelope in the $j-th$ frequency band and $m-th$ time frame after applying a one-frame temporal clipping
- $\mu_{\boldsymbol{x}_{j,m}}$ and $\mu_{\bar{\boldsymbol{y}}_{j,m}}$ denote the sample average of $\boldsymbol{x}_{j,m}$ and $\bar{\boldsymbol{y}}_{j,m}$

The STOI index has been computed for each speech segment and evaluated at the three amplitude levels. Inspecting the results, we notice that for each segment the ARNC provides an improvement in terms of intelligibility with an increase up to $5\,\%$, and in general higher then $1.1\,\%$. However, to provide a single index for both ears position, the following binaural definition, denoted as BSTOI, has been formulated:

$$BSTOI = \sqrt{\frac{STOI_L^2 + STOI_R^2}{2}}, \tag{3}$$

where $STOI_L$ and $STOI_R$ represent the $STOI$ index computed at the left and right ear microphones position,

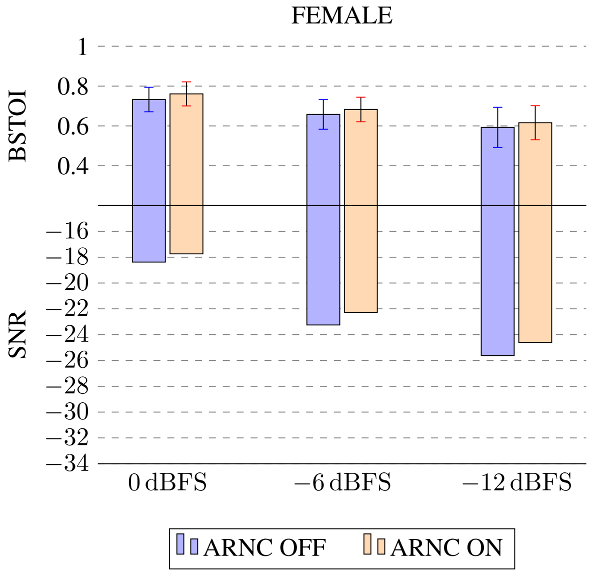

Figure 2. Average and standard deviation of BSTOI and average SNR for Female speakers at different dBFS values.

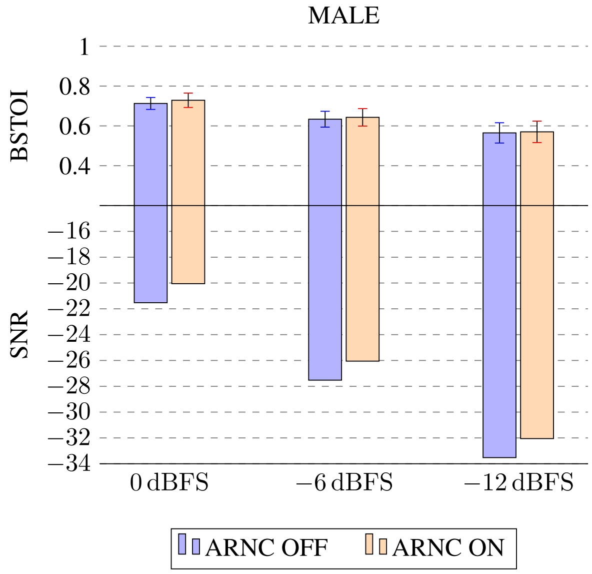

Figure 3. Average and standard deviation of BSTOI and average SNR for Male speakers at different dBFS values.

respectively. Therefore, as for STOI, also BSTOI values are in the range $[0, 1]$, with 1 referring to the highest intelligibility score.

Similarly, the Signal-to-Noise Ratio (SNR) has been computed in dB for all the speech segments and the different amplitude levels as follows:

$$SNR = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right), \tag{4}$$

where P_{signal} and P_{noise} denote the power of the signal and the noise, respectively. Moreover, we consider as SNR the mean values between the two microphones at the ear positions. It is worth noticing that in our scenario SNR values are negative due to the predominance of the noise compared to the signals.

Fig. 2 and 3 depict the average BSTOI (in the positive y-axis) and SNR data (in the negative y-axis) computed among female and male speakers, respectively, for each different dBFS values.

In general, female speech signals achieve higher values of BSTOI and SNR with respect to the male speech signals, both with ARNC system ON and OFF. This is due to the higher frequency content of female speeches, which is less affected by road noise. Indeed, male voices exhibit more spectral overlap with road noise, which is more dominant at low frequencies, resulting in lower scores, in general. Enabling the ARNC system for female speakers results in an increase of the BSTOI higher than 1.3% for all the three amplitude levels, with a maximum increase of 3.4%. For the case of male speakers, all amplitude levels show an increase of the index higher than 0.3%, with a maximum delta of 2%. In terms of SNR instead, we have a maximum increase of 1% for female speakers, while for the male case a maximum delta of 1.5% is registered.

Finally, the standard deviation of the BSTOI index between females decreases when the ARNC system is activated, indicating more uniform values, while an increment can be observed for the case of male speakers. Therefore, it can be noticed that the ARNC system provides an increase in terms of intelligibility for all the examined cases.

4.2 Subjective analysis

The benefits of the ARNC system on speech intelligibility has been evaluated also through a subjective evaluation campaign. Specifically, two blind tests were conducted:

- **Test 1:** A-B comparison of each speech signal with ARNC system enabled and disabled
- **Test 2:** Progressively increasing level test of speech signals.

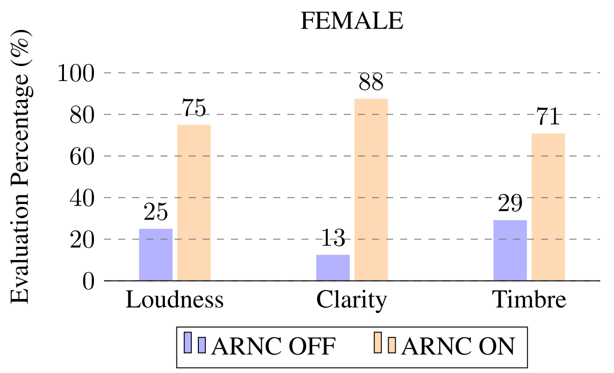

Figure 4. Average KPIs results of the A-B comparison for Female speech signals.

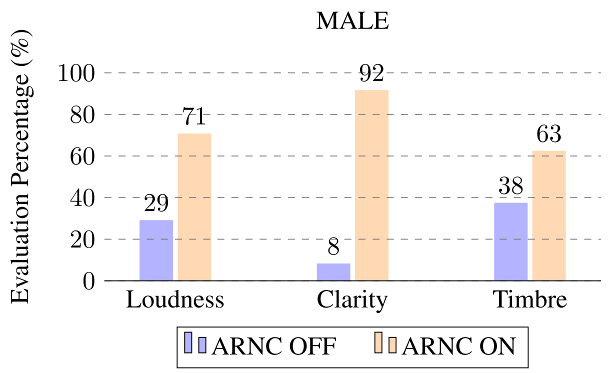

Figure 5. Average KPIs results of the A-B comparison for Male speech signals.

4.2.1 A-B comparison

For this experiment, we focus on four test signals from the simulated speech signals with background road noise described in Sec. 3.1. In particular, we considered the two female and the two male speakers. For each of the selected speech, signal A and signal B have been randomly generated considering *ARNC ON* and *ARNC OFF* conditions.

The blind listening test was conducted on 12 expert-listeners using Audeze high-quality professional headphones [18], and they were given the ability to switch between signal A and signal B for each of the four speech frames. Therefore, for each frame the listeners were asked to express their preference between A and B based on the following three key performance indicators (KPIs):

- Loudness
- Clarity
- Timbre

where *Loudness* refers to the highest perceived loudness of the speech, *Clarity* relies on the precise articulation and correct pronunciation perceived, and *Timbre* evaluated the perceived quality of the speech.

The average evaluation results expressed in percentage between female and male speakers are depicted in Fig. 4 and Fig. 5, respectively. In both cases, the same trend can be noticed with an overall preference to the *ARNC ON* condition for all the three KPIs. Therefore, the activation of the ARNC system implies a perceptual benefits on the rear-to-front speech intelligibility.

It is worth nothing that, considering the *ARNC ON* condition, the *Loudness* and *Timbre* indeces collect higher results for the female speakers with respect to male ones, respectively with 4 % and 5 % of difference. Conversely, the *Clarity* with *ARNC ON* has been selected by 92 % for the case of male speaker, lowering to 88 % for female speakers. Such results depend on the frequency range of the male and female speech. With lower frequency contents, the male spectra is more affected by the frequency range of the road noises, thus the presence of an ARNC system affects the speech *Clarity*.

The stacked bar plot depicted in Fig. 6 represents the detailed results for the two female speakers (refer to F1 and F2) and the two male speakers (refer to M1 and M2). Interestingly, the subjective results between *ARNC OFF* and *ARNC ON* present higher variance for the female case with respect to the male one. Such result confirms the impact of the ARNC system to the lower frequency range. Indeed, assessors considered easier the A-B test for male with respect to female speakers. Interestingly, although an overall preference for the case with the algorithm enabled can also be seen for the *Timbre* index, it has been described as the most difficult KPI to be evaluated by the majority of the listeners.

4.2.2 Progressively increasing level test

To further assess the impact of the ARNC system on in-vehicle speech intelligibility, an additional blind test was conducted. In this experiment, the same four speech frames used in the A-B comparison test of Sec. 4.2.1 have been evaluated. However, for each of the speech frames under consideration, assessors have been asked to listen to a progressively increasing level of the speech frames both with *ARNC OFF* and with *ARNC ON*, thus highlighting

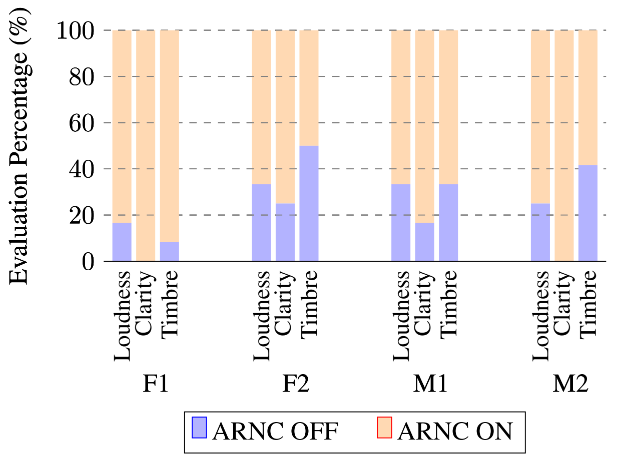

Figure 6. Overall KPIs results of the A-B comparison for Female (F1 and F2) and Male (M1 and M2) speech signals.

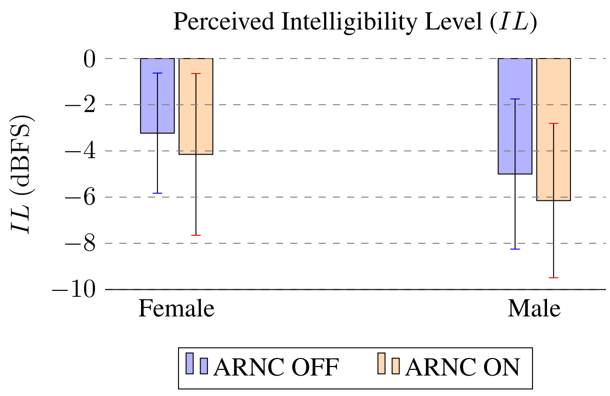

Figure 7. Results of perceived intelligibility level (IL). Avarage and standard deviation are shown with respect to Female and Male speeches with for the case of *ARNC OFF* and *ARNC ON*.

when the signals were intelligible. In particular, the simulated speech has been initially set to a level of -20 dBFS, thus progressively increasing the level at each repetition with a step of 2 dBFS of increments up to 0 dBFS.

A group of 13 expert-listener was selected for the blind test. Assessors have been instructed to stop the audio playback as soon as they were able to understand all the words in each frame, thus identifying a level of perceived intelligibility expressed in dBFS and for the sake of simplicity denoted as IL (intelligibility level). The average results of the perceived IL along with the standard deviation are depicted in Fig. 7 for the case of *ARNC OFF* and *ARNC ON* and differentiating between male and female frames.

In general, assessors were able to fully comprehend the speech content at a lower level in the case of *ARNC ON*, thus confirming the benefits of the road noise reduction for the intelligibility. In particular, the delta level of perceived IL between *ARNC OFF* and *ARNC ON* is equal to 2 dBFS for male speeches and decreases to 1.77 dBFS for female speakers.

It is worth noticing that independently of the ARNC system, male talkers achieved a lower IL then female speeches. The former with -4.15 dBFS and -6.15 dBFS of IL for the case of *ARNC OFF* and *ARNC ON*, respectively. For female speeches, the average results reduced to -3.23 dBFS and -5 dBFS, respectively. Therefore, the delta IL between the disabled and enabled ARNC system increases from 1.77 dBFS for female speakers to 2 dBFS for male speakers. As expected, such result confirms the benefits of the ARNC system in terms of speech intelligibility signals with lower frequency contents that overlap with the ARNC operating frequency range. Moreover, notice that the outcome of this listening test is aligned with Fig. 4 and Fig. 5 of the A-B listening test in Sec. 4.2.1, where male speeches with *ARNC ON* achieved a higher *Clarity* score compared to the female counterpart test.

Another interesting interpretation of the advantages of the ARNC system provided by the results of the devised progressively increasing level test lies in allowing the rear passengers to speak at a lower level but keeping the same speech intelligibility at the front passengers when the ARNC system is enabled.

5. CONCLUSIONS

In this study, the benefits of an Active Road Noise Cancellation (ARNC) algorithm on in-vehicle passenger-to-passenger speech intelligibility have been evaluated through both objective and subjective assessments. Road noise measurements were conducted at the front-right passenger's ear positions in a vehicle instrumented with accelerometers and microphones, using a state-of-the-art ARNC algorithm.

To simulate a speaker inside the vehicle, a Head and Torso Simulator was placed on the rear right-seat, and the impulse response from its artificial mouth to the microphones at the front passenger's ear position has been measured. A reference speech signal, consisting of both male and female voices at different amplitude levels, was then convolved with the measured impulse responses to replicate how speech from the rear passenger seat would be perceived by the front passenger. Finally, this processed speech signal was superimposed with the previously recorded road noise, both with and without the ARNC system enabled.

Speech intelligibility was assessed using two objective metrics, specifically STOI and SNR, at both ears. Additionally, two blind listening tests were conducted with more than 10 expert-listeners. Specifically, both objective and subjective assessments revealed an increase in speech intelligibility when the ARNC algorithm is enabled, with a stronger perceptual improve for the case of male speakers, where the frequency content of the voice has a major overlap with the one of the road noise. Therefore, results of devised experiments consistently demonstrated the positive impact of the ARNC system on in-vehicle speech intelligibility, confirming its potential in enhancing communication within the vehicle cabin.

6. REFERENCES

[1] C. Lüke, G. Schmidt, A. Theiß, and J. Withopf, "In-car communication," in *Smart Mobile In-Vehicle Systems: Next Generation Advancements*, pp. 97–118, Springer, 2013.

[2] B. Căşeriu and P. Blaga, "Automotive comfort: State of the art and challenges," in *International Conference Interdisciplinarity in Engineering*, pp. 375–387, Springer, 2022.

[3] D. Miljković, "Brief introduction to active noise control," pp. 977–982, 05 2023.

[4] K. Ng, *The Challenges of Active Noise Control*. 11 2023.

[5] S. Elliott and P. Nelson, "Active noise control," *IEEE Signal Processing Magazine*, vol. 10, no. 4, pp. 12–35, 1993.

[6] E. Bjarnason, "Analysis of the filtered-x lms algorithm," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 6, pp. 504–514, 1995.

[7] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *Neurocomputing: foundations of research*, pp. 123–134, MIT Press, 1988.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[9] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[10] Z. Jia, X. Zheng, Q. Zhou, Z. Hao, and Y. Qiu, "A hybrid active noise control system for the attenuation of road noise inside a vehicle cabin," *Sensors*, vol. 20, no. 24, 2020.

[11] S. Kuo and D. Morgan, "Active noise control: a tutorial review," *Proceedings of the IEEE*, vol. 87, no. 6, pp. 943–973, 1999.

[12] N. Samardzic and C. Novak, "In-vehicle speech intelligibility for different driving conditions using the speech transmission index," *Noise Control Engineering Journal*, vol. 59, no. 4, pp. 397–407, 2011.

[13] N. Samardzic and C. Novak, "In-vehicle application of common speech intelligibility metrics," *International Journal of Vehicle Noise and Vibration*, vol. 7, no. 4, pp. 328–346, 2011.

[14] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.

[15] P. C. Loizou, *Speech Enhancement: Theory and Practice*. CRC press, 2013.

[16] H. Acoustics, "Head acoustics head and torso simulator (hats) hms ii.3 with ln." `https://www.head-acoustics.com/products/artificial-head-binaural-recording/hms-ii3-ln` [Accessed: (03/2025)].

[17] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Audio engineering society convention 122*, Audio Engineering Society, 2007.

[18] Audeze, "Audeze lcd-x." `https://www.audeze.com/products/lcd-x` [Accessed: (03/2025)].