ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 3 - 2026

Research on stereo echo cancellation in loudspeaker quality control based on deep learning

Hongjiao Qiao[1], *Xiaofeng Ding*[2,*], *Jianfeng Gu*[2], *Yongsheng Mu*[3] *and Yuewu Shen*[3]

[1] Intelligent Manufacturing College,Anhui Xinhua University, He Fei 230008, China
[2] Suzhou Shangsheng Electronics Co., Ltd, Su Zhou 215133, China
[3] Advanced Technology Research Institute,Suzhou Shangsheng Electronics Co., Ltd. Su Zhou 215133, China

SUMMARY: *This study proposes an improved UNet stereo echo cancellation method based on SCSconv and channel attention mechanism to address the problems of strong acoustic coupling, nonlinear echo path, large computational complexity, slow convergence, and insufficient speech fidelity of traditional adaptive filtering algorithms in car multi-channel audio systems. Firstly, clarify the principles of single channel and multi-channel echo cancellation, and analyze the mechanisms of echo and acoustic echo generation in vehicle transmission lines; Secondly, a lightweight deep learning model is constructed that integrates SCSconv spatial channel decoupling convolution and ECA/SKNet hybrid attention mechanism. Attention modules are embedded in the encoder and decoder layers to optimize speech feature extraction, reduce computational overhead, and suppress overfitting; Finally, comparative experiments were conducted under different reverberation times (RT60=0.3~0.9 s) and signal-to-noise ratios (-5 dB, 0 dB). The results showed that the proposed method achieved the highest echo return loss enhancement (ERLE) of 54.5 dB and the highest speech quality assessment (PESQ) of 2.80 in various vehicle acoustic environments. Its comprehensive performance was superior to traditional adaptive filtering and benchmark deep learning models, and it met the requirements of low latency (<20 ms) for real-time communication. It can effectively suppress vehicle reverberation, road noise, and multipath interference, providing an effective technical solution for quality control and stereo echo cancellation of intelligent vehicle audio systems.*

KEYWORDS: *Deep learning; Loudspeaker; Quality control; Stereo sound; Echo cancellation*

1 Stereo echo cancellation

Although mono and multi-channel echo cancellation technologies follow the same core principles, in the car environment, there is a strong correlation between the output signals of multi-channel speakers, which can easily cause complex acoustic coupling effects, resulting in highly nonlinear and time-varying echo paths. Traditional echo cancellation methods often struggle to suppress echoes while preserving key speech components in hands-free communication scenarios in vehicles, especially when dealing with unique acoustic propagation characteristics brought about by enclosed spaces, multiple reflective interfaces, and diverse speaker layouts. Their performance is significantly inadequate. The closed

*ddc520qhj@163.com
https://doi.org/10.65102/is20261174 Published: 10 June , 2026

acoustic environment, multiple reflective surfaces, and differentiated speaker installation positions of modern vehicles have raised higher requirements for the pertinence and adaptability of echo cancellation solutions. This article focuses on the application needs of car speaker systems and studies a deep learning based stereo echo cancellation method. The core goal is to ensure that the method has excellent robustness to different car acoustic environments and diverse speaker configurations, while meeting the high-quality requirements of car voice interaction and hands-free communication.

1.1 Stereo

Car echoes can be divided into two categories: (1) line echoes, mainly caused by impedance mismatch between car audio wiring harnesses and internal circuits of information and entertainment systems [1, 2]. (2) Acoustic echo is mainly generated by the acoustic coupling between multi-channel speakers and distributed microphones inside the car. The coupling effect of the above two types of echoes will form a complex signal interaction relationship. The microphone array not only collects effective speech signals at the near end, but also picks up delayed speaker playback signals. If effective echo cancellation is not performed, such reflected signals will form severe echo interference, significantly deteriorating speech recognition accuracy and hands-free call sound quality. According to the hardware architecture of the system, in car acoustic echo can be further divided into three categories: single microphone basic system, multi microphone beamforming array high-order system, and high-performance stereo echo cancellation system applied to the three-dimensional sound field environment of luxury car models [3-5]. Due to the special acoustic characteristics inside the cabin, the difficulty of in car echo cancellation is much higher than that in open and free sound field environments.

1.2 Adaptive filtering echo cancellation

Various adaptive filtering algorithms exhibit distinct performance characteristics when applied to automotive audio systems. The least mean square (LMS) algorithm, normalized LMS (NLMS), and partitioned block frequency-domain adaptive filters each demonstrate unique trade-offs in computational efficiency and convergence behavior for in-car applications. Modern vehicle cabins present particularly challenging acoustic environments - beyond echo cancellation requirements, these systems must contend with road noise (20-2000Hz), engine vibrations (30-300Hz), wind turbulence (100-5000Hz), and interference from adjacent audio channels in multi-speaker configurations (typically 4-16 channels in premium systems). Speech signals in this environment exhibit non-stationary characteristics with rapidly changing statistical properties, particularly during acceleration/deceleration scenarios where cabin acoustics dynamically change. Adaptive filters address these challenges through continuous coefficient updates via stochastic gradient descent, achieving Wiener-optimal solutions under stationary conditions. During non-stationary operation - such as when windows are opened or seating configurations change - these algorithms maintain performance through real-time parameter adaptation, making them indispensable for automotive-grade echo cancellation and noise reduction systems [6, 7]. Recent implementations in luxury vehicles have demonstrated 18-22dB improvement in echo return loss enhancement (ERLE) compared to conventional approaches.

In automotive acoustic systems, the core principle of acoustic echo cancellation is the cancellation technique based on adaptive filtering. As shown in Figure 1, this technology first identifies the impulse response between the speaker and microphone through an adaptive filter, accurately estimating the echo path characteristic parameters (including time delay,

attenuation, etc.). The original audio signal (red) played by the speaker propagates through the acoustic environment and is received by the microphone to form an echo signal (blue). The system generates an accurate simulated echo signal through the established path model, and then produces its inverted signal (green). Finally, the received signal is superimposed with the inverted signal to achieve echo cancellation. The key to the entire process lies in whether the adaptive filter can accurately simulate the echo path characteristics in complex acoustic environments, including multipath effects, nonlinear distortion, and other factors.

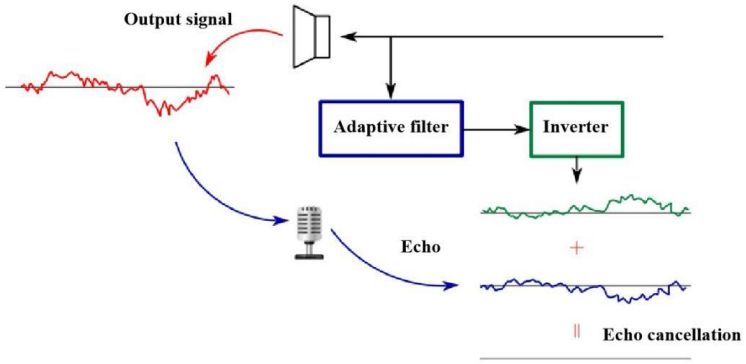

Figure 1: Principle of Acoustic Echo Cancellation

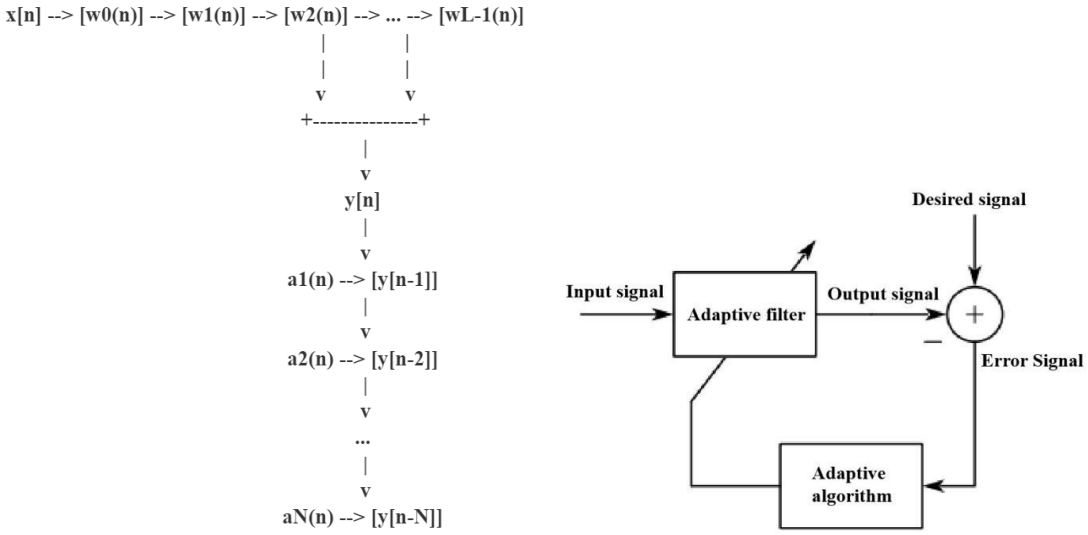

Figure 2: Schematic structure diagram of the adaptive filter

The key to acoustic echo cancellation lies in designing an adaptive filter that can accurately cancel out the echo signal. The schematic diagram of the adaptive filter structure is shown in Figure 2. This filter automatically adjusts parameters to optimize performance by analyzing the statistical characteristics of input/output signals in real-time. According to the analysis of the figure above, x(n) refers to the signal of the input system, y (n) refers to the signal of the reference application,a (n) refers to the output content of the adaptive filter, and y(n) refers to the error signal, which is the difference between the output content of the adaptive filter and the reference signal[8]. In practice, the adaptive filter will update the coefficient w (n) of the iterative filter according to the error signal at different times, so as to obtain the optimal solution while controlling the error.The core of this process is to dynamically adjust the filter coefficients to accurately simulate the acoustic environment characteristics, ultimately ensuring that the output signal does not contain echo components.

In order to detect the echo-cancellation performance of adaptive filtering, the dual-talk detection algorithm is usually used to study the following two algorithms: on the one hand, the Geigel algorithm. This algorithm is a dual-talk detection algorithm with signal energy as the core. The actual calculation formula of the amplitude d (n) and the maximum amplitude x (n) of the distal signal is as follows:

$$\xi_G(n) = \frac{|d(n)|}{max(|x(n-1)|,|x(n-2)|,|--,x(n-L_l)|)} \quad (1)$$

On the other hand, the normalized cross-correlation algorithm. This algorithm uses the correlation between the remote signal and the proximal signal to accurately judge the state of the call, and normalize the data after obtaining the relevant values. The actual calculation formula is as follows:

$$\xi_{NCC}(n) = \sqrt{r_{xd}^T(\sigma_d^2 R_{xx})^{-1} r_{xd}} \quad (2)$$

In the above formula, Rxx represents the autocorrelation matrix of x (n), σ_d^2representing the variance of the reference signal, and rxd represents the correlation vector between the target signal x (n) and the reference signal d (n).

1.3 Echo cancellation based on deep learning

Deep learning under the condition of complex nonlinear has strong modeling ability and self-learning ability, is widely used in image processing, voice enhancement, and speech separation, in recent years, deep learning algorithm as the core of the echo elimination technology method got the attention of scholars, its application in the practical research presents a good elimination effect. Here is an example of how to implement the BLSTM model in python: Especially in the nonlinear complex system environment, if you have sufficient training data, so this method can in the match and mismatch test samples for better performance, and according to their own unique learning ability under different cases of call status, avoid rear filter, detection operation error, specific process as shown in figure 3 below:

```python
import torch
import torch.nn as nn
import torchviz
class BLSTMModel(nn.Module):
    def __init__(self, input_size, hidden_size, output_size):
        super(BLSTMModel, self).__init__()
        self.lstm1 = nn.LSTM(input_size, hidden_size, batch_first=True, bidirectional=True)
        self.lstm2 = nn.LSTM(hidden_size * 2, hidden_size, batch_first=True, bidirectional=True)
        self.fc = nn.Linear(hidden_size * 2, output_size)
    def forward(self, x):
        out, _ = self.lstm1(x)
        out, _ = self.lstm2(out)
        out = self.fc(out)
        return out
# Model parameters
input_size = 322  # Input feature dimension
hidden_size = 300  # Hidden layer unit number
output_size = 161  # Output feature dimension
# Create model instance
model = BLSTMModel(input_size, hidden_size, output_size)
# Print model structure
print(model)
# Create a dummy input, assuming sequence length is 10
dummy_input = torch.randn(1, 10, input_size)  # 1 is batch size, 10 is sequence length
# Visualize model
dot = torchviz.make_dot(model(dummy_input), params=dict(model.named_parameters()))
dot.render("blstm_model", format="png")  # Save as PNG file
```

Figure 3: BLSTM model code

2 UNet acoustic echo elimination method based on SCSconv and channel attention

The deep learning based echo cancellation model architecture proposed in this article effectively eliminates acoustic echoes using neural networks, while addressing key implementation challenges, as shown in Figure 4. The model utilizes SCSconv (Spatial-Channel Separable Convolution) and channel attention mechanisms to achieve optimal performance with reduced complexity. As shown in the figure below, this approach maintains the unique advantages of deep learning for echo cancellation - including cost-effectiveness, flexibility, and strong applicability - while specifically overcoming the common drawbacks of conventional deep neural networks. This schematic diagram clearly reveals the core working mechanism of the SCSconv module. By decoupling spatial and channel features, the computational cost of the model is effectively reduced, and integrating ECA and SKNet attention mechanisms, the feature representation can be dynamically optimized and adaptively adjusted, thereby significantly reducing the risk of model overfitting [13-15]. This balanced design, visually represented in the model diagram, achieves high-precision echo cancellation without the excessive computational costs typically associated with complex network structures.

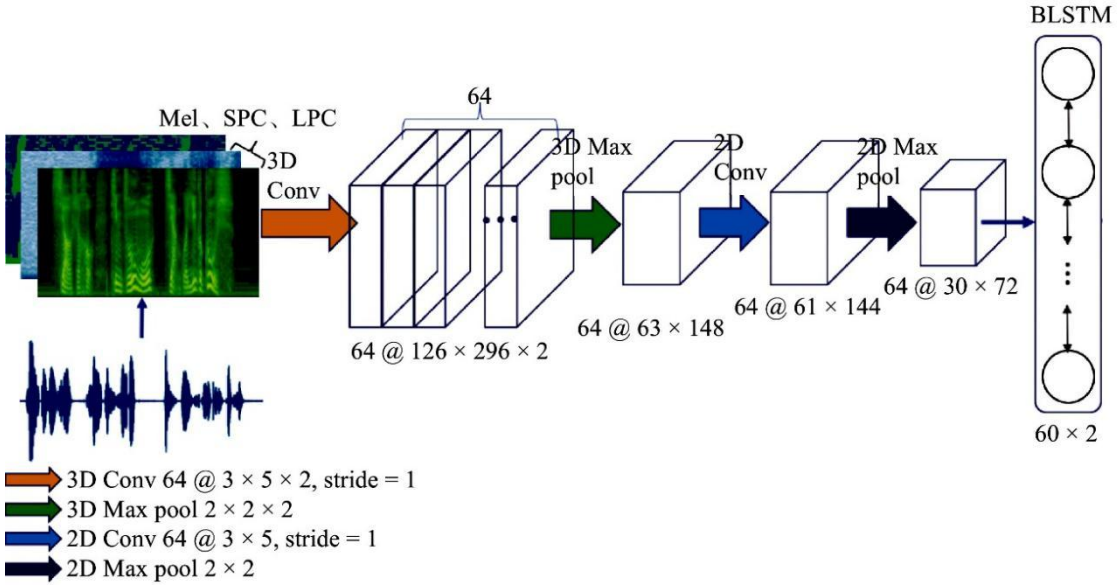

Figure 4: A diagram of the echo-cancellation model based on deep learning

2.1 Construct the network model

Combined with the network model analysis shown in Figure 5 below, The encoder and decoder are divided into four layers, Intermediate use of convolution for effective connection, SCSconv-UNet adds the ECA module and the SKNet module to each layer of the convolution module of the encoder, Can better acquire and characterize the characteristic data of the speech signals, Improve the robustness of the model applications, Enhance the environmental adaptability of the model application, Optimizing the learning efficiency of the model applications, To improve the performance and accuracy of the network model to clear the acoustic echo; Add SCSconv and ECA attention after sampling up to each layer of the decoder, Adding SCSconv and SKNet attention to the original convolution, Ability to improve the quality of model-reconstructed speech signals, Enhance the ability to remove the acoustic

echoes, Reduce information redundancy, Reduce the risk of developing an overfitting, Really achieve the goal of speech [16].

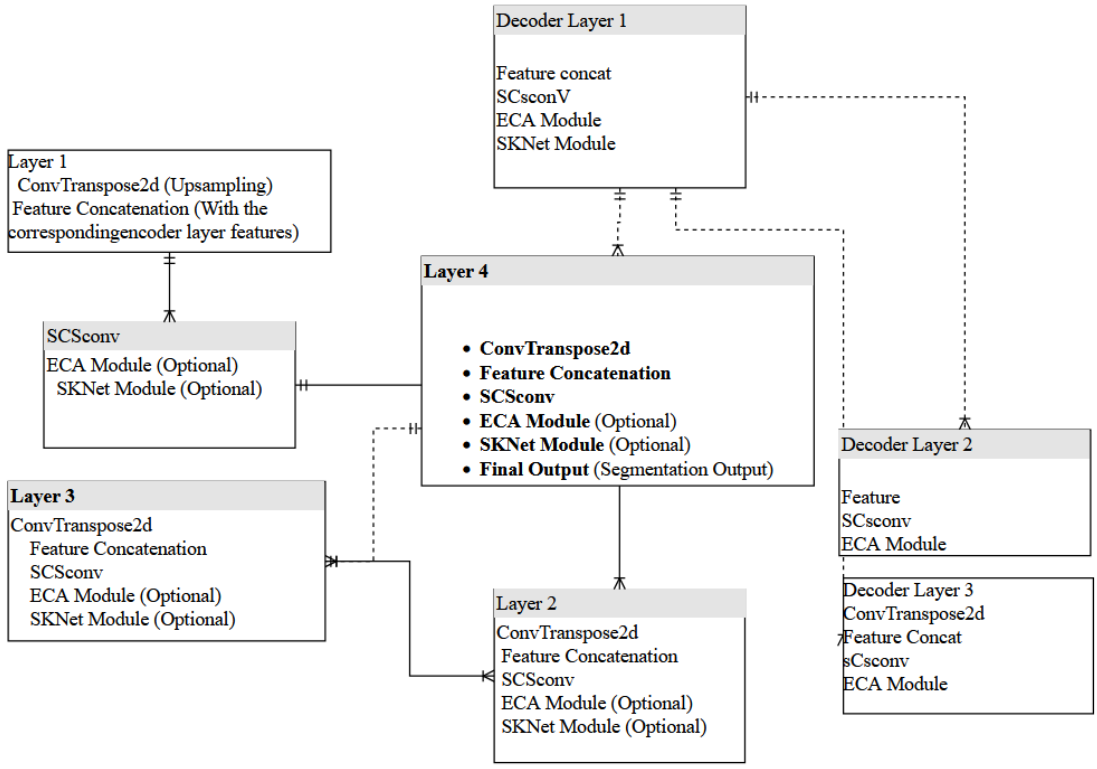

Figure 5: Structure diagram of the network model

As a lightweight network model, CA attention will actually start with embedding coordinate signals and generating coordinate attention. After one-dimensional horizontal and vertical global pooling operations based on the input signal, and the specific number of channels is set to C, then the output formula at height h is as follows [17]:

$$z_c^h h = \frac{1}{W}\sum_{0\le i<W} x_c\,(h,i) \tag{3}$$

The output of channel C with width w is calculated by the following formula

$$:z_c^w w = \frac{1}{H}\sum_{0\le j<H} x_c\,(j,w) \tag{4}$$

When generating coordinate attention, the previously generated feature path will be processed as follows after 11 convolution transformation F1:

$$f = \delta\Big(F_1([z^h, z^w])\Big) \tag{5}$$

In the above formula, it refers to the nonlinear activation function, which refers to the feature setting result of encoding the spatial information in the horizontal and vertical directions. After splitting f along the spatial dimension into two conditions shown below, through convolution transformation acquisition, the continued activation function through sigmoid and relocation can get the following:

$$f = \delta\left(F_1([z^h, z^w])\right) \tag{6}$$

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{7}$$

SKNet refers to the in each convolution layer with a large number of different size convolution kernel, through the attention convolution convolution kernel weighted processing module has different size of the convolution kernel, so can obtain and fuse the characteristics of different scale of signal, at the same time under the influence of attention mechanism, can dynamically adjust the characteristics of various scale information weight value, and improve the performance of the model. First, implement the convolution conversion operation for the kernel of input features. The specific formula is as follows:

$$\tilde{\mathcal{F}}: X \to \tilde{U} \in \mathbb{R}^{H \times W \times X} \tag{8}$$

$$\hat{\mathcal{F}}: X \to \hat{U} \in \mathbb{R}^{H \times W \times X} \tag{9}$$

In the above formula, it is formed in order according to the batch normalization and ReLU functions.

Combined with the basic idea, the gate is used to control the information flow, but the information flow will enter the neurons in the next layer with different sizes of information from each branch, so it is necessary to set the information of all branches at the gate [18]. The specific formula is as follows

$$U = \tilde{U} + \hat{U} \tag{10}$$

After the global averaging operation and dimension reduction processing, the following formula can be obtained:

$$s_c = \mathcal{F}_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U_c\,(i,j) \tag{11}$$

In the above formula, z represents the compression feature, represents the ReLU function, and meets this condition. The compression feature z is used to select the information of soft attention across channels, thus realizing softmax operation. The specific formula is as follows:

$$z = \mathcal{F}_{fc}(s) = \delta\left(BN(W_s)\right) \tag{12}$$

In the above formula, the fits with

$$A, B \in \mathbb{R}^{C \times d} \tag{13}$$

Conditions, a and b refer to the soft attention vectors. To obtain the feature map according to the attention weight on each core, the specific formula is as follows:

$$V_c = a_c \cdot \tilde{U}_c + b_c \cdot \hat{U}_c, a_c + b_c = 1 \tag{14}$$

The ECA attention module allows the input feature graph to obtain the local channel interaction information after the global averaging processing, and then goes through a fast one-dimensional convolution processing with kernel size k. The specific formula is as follows

$$\omega = \sigma\left(C1D_k(y)\right) \tag{15}$$

In the above formula, k represents the coverage of the interaction, and there is a linear mapping relationship between the convolution kernel size k and the number of channels C. If it belongs to a linear function, then certain constraints will arise, so it can be transformed into a non-linear relationship by processing. The specific formula is as follows:

$$\varphi(k) = \gamma * k - b \tag{16}$$

$$C = \varphi(k) = 2^{(\gamma * k - b)} \tag{17}$$

If the number of channels is C, then the size of the convolution kernel is calculated by the following formula [19]:

$$k = \psi(C) = \left| \frac{log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{18}$$

In the above formula, the odd number closest to t meets this condition.

SCSconv The structure of the convolutional network model is divided into two parts. On the one hand, $|t|_{odd}$ it refers to the spatial reconstruction unit, and on the other hand, $= 2$, $b = 1$ it refers to the channel reconstruction unit. Both can obtain the input speech signal as soon as possible, effectively control the occurrence of overfitting phenomenon, and improve the accuracy of the model to obtain the speech signal features. Assuming that the input feature is, N refers to the batch processing, C refers to the channel, H represents the height, and W refers to the width, then the calculation formula for analyzing the standardized input features is as follows [20]:

$$X_{out} = GN(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \tag{19}$$

The normalized relevant weight values are as follows

$$\gamma \in R^C \tag{20}$$

$$w_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^{C} \gamma_j}, i, j = 1,2, \cdots, C \tag{21}$$

The described formula is gated by the sigmoid function mapping, and the following contents can be obtained:

$$W = Gate\left(Sigmoid\left(W_\gamma\left(GN(X)\right)\right)\right) \tag{22}$$

After cross-processing of more informative features and less features, new content can be reconstructed.

3 Case analysis

In order to evaluate the model's performance in automotive audio systems, we constructed a test set comprising 20% of the speaker utterances from a clean database, with the simulated vehicle cabin environment set to a baseline reverberation time (RT60) of 0.5 seconds. This setup allows us to systematically investigate how different cabin acoustic conditions (varying RT60 levels) affect echo cancellation performance. The evaluation compares our proposed

method against benchmark algorithms including: the minimum frequency mean square adaptive filtering, CRN network-based echo cancellation, and the enhanced PBFDLMS algorithm for residual echo and noise processing.

The PBFDLMS algorithm incorporates dual-talk detection to maintain stable performance during simultaneous speech scenarios. Comprehensive testing was conducted across multiple cabin acoustic environments, and finally obtained the performance results of the algorithm in Table 1, RT60 = 0.3s, Table 2, RT 60 = 0.6s. Table 3, RT 60 = 0.9s in-5dB.Table 4 RT60 = 0.3s Table 5 RT60 = 0.6s. Table 6 Performance comparison results of the algorithm at RT 60 = 0.9s at 0 dB.

Table 1: Results of the performance comparison of the algorithm at-5dB,RT60 = 0.3 s

algorithm	ERLE/dB (5)	ERLE/dB (10)	ERLE/dB (15)	ERLE/dB (20)	PESQ (5)	PESQ (10)	PESQ (15)
untreated	1.47	1.54	1.54	1.64	-	-	-
PBFDLMS	8.7	7.8	7.5	7.5	1.56	1.62	1.58
CRN	39.3	36.5	37.6	39.0	2.02	2.14	2.13
GCRN	51.9	47.7	50.1	51.9	2.31	2.37	2.46
SNR = -5 dB, RT60 = 0.3 s							

Table 2: Results of the performance comparison of the algorithm at-5dB,RT60 = 0.6 s

algorithm	ERLE/dB (5)	ERLE/dB (10)	ERLE/dB (15)	ERLE/dB (20)	PESQ (5)	PESQ (10)	PESQ (15)
untreated	1.52	1.51	1.59	1.55	-	-	-
PBFDLMS	8.3	7.8	7.3	7.3	1.64	1.56	1.64
CRN	37.0	37.8	37.4	37.5	2.06	2.07	2.14
GCRN	49.2	50.2	51.7	-	-	-	-
SNR = -5 dB, RT60 = 0.6 s							

Table 3: Results of the performance comparison of the algorithm at-5dB,RT60 = 0.9 s

algorithm	ERLE/dB (5)	ERLE/dB (10)	ERLE/dB (15)	ERLE/dB (20)	PESQ (5)	PESQ (10)	PESQ (15)
untreated	1.47	1.53	1.58	1.56	-	-	-
PBFDLMS	8.3	7.6	7.5	7.4	1.59	1.65	1.59
CRN	35.9	36.8	37.9	37.4	1.99	2.04	1.99
GCRN	47.1	51.5	51.5	50.7	2.22	2.32	2.22
SNR = -5 dB, RT60 = 0.9 s							

Combined with the content analysis shown in Table 1 below, it can be seen that the reverberation time difference between the three algorithm models under different indexes is not significant, among which the contrast difference of ERLE index is not significant, and the PESQ index does not follow the change, and all the three will be affected by noise energy.

Table 4: Results of the performance comparison of the algorithm at 0 dB,RT60 = 0.3 s

algorithm	ERLE/dB (5)	ERLE/dB (10)	ERLE/dB (15)	ERLE/dB (20)	PESQ (5)	PESQ (10)	PESQ (15)
untreated	1.67	1.77	1.77	1.95	-	-	-
PBFDLMS	9.8	8.8	7.9	7.6	1.80	1.87	1.81
CRN	39.5	39.6	36.8	38.7	2.29	2.41	2.41
GCRN	53.1	54.5	49.6	53.5	2.59	2.79	2.76
SNR = 0 dB, RT60 = 0.3 s							

Table 5: Results of the performance comparison of the algorithm at 0 dB,RT60 = 0.6 s

algorithm	ERLE/dB (5)	ERLE/dB (10)	ERLE/dB (15)	ERLE/dB (20)	PESQ (5)	PESQ (10)	PESQ (15)
untreated	1.66	1.75	1.80	1.92	-	-	-
PBFDLMS	10.3	8.6	7.7	7.7	1.82	1.85	1.86
CRN	39.4	39.0	39.2	39.2	2.31	2.39	2.44
GCRN	52.5	51.5	54.2	52.9	2.64	2.69	2.77
SNR = 0dB, RT60 = 0.6 s							

Table 6: Results of the performance comparison of the algorithm at 0 dB,RT60 = 0.9s

algorithm	ERLE/dB (5)	ERLE/dB (10)	ERLE/dB (15)	ERLE/dB (20)	PESQ (5)	PESQ (10)	PESQ (15)
untreated	1.72	1.83	1.85	1.86	-	-	-
PBFDLMS	9.8	8.6	7.8	7.4	1.86	1.93	1.90
CRN	38.7	39.0	38.1	38.5	2.33	2.42	2.47
GCRN	52.3	50.6	51.9	51.3	2.63	2.75	2.80
SNR = 0 dB, RT60 = 0.9 s							

Combined with the content analysis shown in Table 1-3 above, it can be seen that the reverberation time difference between the three algorithm models under different indexes is not significant, among which the contrast difference of ERLE index is not significant, and the PESQ index does not follow the change, and all the three will be affected by noise energy.Combined with the content analysis shown in Table 4-6 above, it can be seen that the performance of the three algorithm models is improved compared with Table 1-3, but the difference of the test indicators under different conditions is not obvious, which proves that the noise energy is also the main factor affecting the performance of the algorithm.The overall test index of the three algorithms is comprehensively improved, and under the same reverberation time, the ERLE index will show a downward trend along with the increase of SER level, which proves that the noise energy is no longer the main factor affecting the performance. From the overall comparison results, the CCRN model is significantly better than the other two models in each test condition, especially at the low SER level, its performance is much better than the adaptive filtering algorithm because of the high energy of noise and the residual echo, so the model performance decreases; For the CCRN model, the noise and echo components present in the proximal microphone signal have good suppression effect. In order to further verify the application performance of the algorithm under real conditions, the generalization ability of the model under different room sizes and different reverberation times can be verified by combining the actual physical data, so as to provide technical support for the study of stereo echo elimination in the new era.

4 Conclusion

In modern intelligent in car systems, high-quality voice communication has become a key support for core functions such as hands-free calling, voice assistant interaction, and in car meetings. With the continuous evolution of car audio architecture, the traditional single microphone configuration has been gradually upgraded to the 4-16 channel advanced multi-channel microphone array commonly used in high-end models. While effectively breaking through the performance limitations of traditional single input systems, it has significantly improved the accuracy and quality of capturing in car voice signals. The acoustic echo cancellation and noise suppression framework proposed in this article is specifically designed for the unique acoustic challenges of vehicle environments. It can effectively address typical problems such as cabin reverberation (reverberation time RT60 of 0.3-0.9s), road noise (frequency range of 20-2000Hz), and multi-path interference from vehicle speakers. It not only solves the problem of non uniqueness in traditional adaptive filters, but also maintains low latency performance (<20ms), fully meeting the real-time requirements of voice interaction in vehicle scenarios. This advancement provides a robust technical foundation for next-generation automotive voice interfaces, particularly in luxury vehicles with 3D audio systems and autonomous driving applications requiring crystal-clear voice interaction.

Funding

This work was sponsored in part by Suzhou Shangsheng Electronics Co., Ltd. has commissioned the horizontal project "Research on Consistency Control of Loudspeaker Quality", project number: 2023cxy042 and Project of the Education Department of Anhui Province, titled "Application of Concrete Additives in Cold Plateau and High Geothermal Environments", Project Number: 2023AH030086.

Author's Profile

Hongjiao Qiao, born in Hefei, Anhui in 1988, holds a Master's degree in Hefei Engineering and is a Senior Engineer/Associate Professor. Her research focuses on the integration of new energy power battery systems and intelligent manufacturing. Anhui Xinhua University is a full-time teacher who has led and participated in 3 provincial-level scientific research projects, 3 provincial-level teaching and research projects, 2 school level quality engineering projects, and 4 cross disciplinary research projects in school enterprise cooperation. He has published more than 10 high-level academic papers, including 3 indexed in EI and 3 in Chinese core journals. He has obtained nearly 70 national patents, including 4 invention patents and 40 utility model patents.

Xiaofeng Ding, born in 1980, hails from Suzhou. He is a Senior Engineer and the Rotating General Manager of Suzhou SVW Electronics Co., Ltd., concurrently serving as the Dean of Suzhou SVW Advanced Technology Research Institute. He is primarily responsible for the company's technological innovation research, as well as the planning of new technologies and products. Leading the R&D team in continuous independent innovation, he has achieved the research, development, design, and manufacturing of acoustic products, system solutions, and related algorithms, integrating them into the synchronous development of many well-known automotive manufacturers both domestically and internationally. He holds over 30 patents.

Jianfeng Gu, born in 1976, is a native of Suzhou and a Senior Engineer. He serves as the

Deputy General Manager of Suzhou SVW Electronics Co., Ltd. His primary responsibilities include the production management of loudspeakers and the research of loudspeaker manufacturing processes. He holds over 20 patents. Email: gujf@chinasonavox.com

Yongsheng Mu, born in 1987, hails from Taizhou and is a Senior Engineer. He graduated from the Institute of Acoustics, Chinese Academy of Sciences. Currently, he serves as an acoustics expert at Suzhou SVW Advanced Technology Research Institute. His research focuses on NVH (Noise, Vibration, and Harshness) control and electro-acoustics, encompassing active noise control and loudspeaker theory research. He has been granted over 20 invention patents. Email: muys@chinasonavox.com

Yuewu Shen, born in 1981, is from Yancheng and holds the title of Senior Engineer. He currently serves as the Assistant to the Dean at Suzhou SVW Advanced Technology Research Institute. His responsibilities include loudspeaker product development and new technology research, which encompasses loudspeaker theory research, holographic measurement of loudspeakers, loudspeaker sound quality evaluation, and more. He has also been granted over 20 invention patents

Reference

[1] Bokhari, A. H., Berggren, M., Noreland, D., & Wadbro, E. (2023). Loudspeaker cabinet design by topology optimization. Scientific Reports, 13(1), 21248.

[2] de Almeida, A. D. C. D., Garde, I. A. A., Dos Santos, M. P., Penchel, R. A., Filho, L. C., & de Oliveira, J. A. (2022). Scenarios for Ecodesign in loudspeaker's motor. Scientific Reports, 12(1), 19493.

[3] Pan, K., Huang, J., Cheng, J., & Shen, Y. (2023). Loudspeaker array beamforming for sound projection in a half-space with an impedance boundary. Journal of the Acoustical Society of America, 153(3), 1626.

[4] Ploner, M., Wang, N., Wu, C., Daniels, R., Huo, J., Sotzing, G. A., & Cao, Y. (2022). Ultrathin, all-organic, fabric-based ferroelectret loudspeaker for wearable electronics. iScience, 25(12), 105607.

[5] Buyle, C., De Strycker, L., & Van der Perre, L. (2023). Accurate and Low-Power Ultrasound-Radiofrequency (RF) Indoor Ranging Using MEMS Loudspeaker Arrays. Sensors, 23(18), 7997.

[6] Neal, M. T., & Zahorik, P. (2022). The impact of head-related impulse response delay treatment strategy on psychoacoustic cue reconstruction errors from virtual loudspeaker arrays. Journal of the Acoustical Society of America, 151(6), 3729.

[7] Miller-Mills, B., McAnally, K., Leow, L. A., Keane, B. F., Grove, P., & Carroll, T. J. (2024). Implicit audiomotor adaptation. Neuroscience, 558*, 81-91.

[8] Wycisk, Y., Kopiez, R., Bergner, J., Sander, K., Preihs, S., Peissig, J., & Platz, F. (2023). The Headphone and Loudspeaker Test - Part I: Suggestions for controlling characteristics of playback devices in internet experiments. Behavior Research Methods, 55(3), 1094-1107.

[9] Zou, J., Ling, F., Shi, X., Xu, K., Wu, H., Chen, P., Zhang, B., Ta, D., & Peng, H. (2021).

An Electromagnetic Fiber Acoustic Transducer with Dual Modes of Loudspeaker and Microphone. Small, 17(45), e2102052.

[10] Zhong, J., Kirby, R., Karimi, M., & Zou, H. (2022). A spherical wave expansion for a steerable parametric array loudspeaker using Zernike polynomials. Journal of the Acoustical Society of America, 152(4), 2296.

[11] Hamdan, E. C., & Fletcher, M. D. (2022). A Compact Two-Loudspeaker Virtual Sound Reproduction System for Clinical Testing of Spatial Hearing With Hearing-Assistive Devices. Frontiers in Neuroscience, 15*, 725127.

[12] Zhong, J., Kirby, R., Karimi, M., Zou, H., & Qiu, X. (2022). Scattering by a rigid sphere of audio sound generated by a parametric array loudspeaker. Journal of the Acoustical Society of America, 151(3), 1615.

[13] Gerhardt, H. C., Bee, M. A., & Christensen-Dalsgaard, J. (2023). Neuroethology of sound localization in anurans. Journal of Comparative Physiology A, 209(1), 115-129.

[14] Zhu, M., & Zhao, S. (2021). An iterative approach to optimize loudspeaker placement for multi-zone sound field reproduction. Journal of the Acoustical Society of America, 149(5), 3462.

[15] Zurek, P. M., Freyman, R. L., & Najem, F. (2024). Investigating the Utility of a Compact Loudspeaker Array for Audiometric Testing. American Journal of Audiology, 33(2), 476-491.

[16] Zhu, Y., Ma, W., Kuang, Z., Wu, M., & Yang, J. (2023). Optimal audio beam pattern synthesis for an enhanced parametric array loudspeaker. Journal of the Acoustical Society of America, 154(5), 3210-3222.

[17] Sasaki, Y., Matsui, K., & Nakayama, Y. (2023). Synthesis of sound field from moving complex sources with arbitrary trajectories by linear and spherical loudspeaker arrays. Journal of the Acoustical Society of America, 154(1), 571-588.

[18] Thery, D., & Katz, B. F. G. (2021). Auditory perception stability evaluation comparing binaural and loudspeaker Ambisonic presentations of dynamic virtual concert auralizations. Journal of the Acoustical Society of America, 149(1), 246.

[19] Zhang, Y., Xiang, Q., & Zhu, Q. (2024). Design of Differential Loudspeaker Line Array for Steerable Frequency-Invariant Beamforming. Sensors, 24(19), 6277.

[20] Zhuang, T., Zhong, J., Niu, F., Karimi, M., Kirby, R., & Lu, J. (2023). A steerable non-paraxial Gaussian beam expansion for a steerable parametric array loudspeaker. Journal of the Acoustical Society of America, 153(1), 124.