Research on Dog Bark Recognition Algorithm Based on Mel-Frequency Cepstral Coefficients and Lightweight Neural Network

Lin Haipeng[1]*

[1] Shenzhen Jingteng Technology Co., Ltd., Shenzhen 518000, China
* Correspondence: Shenzhen Jingteng Technology Co., Ltd., Shenzhen 518000, China.

https://doi.org/10.53104/j.acad.res.adv.2026.06001

Citation: Lin, H. (2026) 'Research on Dog Bark Recognition Algorithm Based on Mel-Frequency Cepstral Coefficients and Lightweight Neural Network', *Journal of Academic Research and Advances*, 2(2), pp. 1-10.

Abstract: Dog barks serve as key bioacoustic signals for pet monitoring and urban noise early warning. Conventional recognition methods suffer severe feature distortion in noisy environments, while deep learning models carry excessive parameters and slow edge inference. This work proposes an algorithm combining spectral entropy-based adaptive denoising MFCC and ECA lightweight attention network. It fuses static and temporal differential cepstral features and uses audio augmentation to mitigate small-sample overfitting. Tests on ESC50, UrbanSound8K and self-built dog bark datasets show 4.3% higher accuracy than standard MFCC at 0 dB SNR. The model has only 2.07 M parameters and 0.31 G computation load, cutting inference latency by 62.8% versus CNN-LSTM, with an average F1 of 0.912 (5–20 dB SNR). Edge tests validate its real-time performance for embedded bioacoustic detection.

Keywords: dog bark recognition; MFCC feature optimization; adaptive noise reduction; lightweight ECA network; bioacoustics; edge deployment; audio feature fusion

1. Introduction

1.1 Research Background

Pet IoT and smart city monitoring demand round-the-clock dog bark detection, yet real environments (5–15 dB SNR) contain heavy traffic, wind and human noise. Cloud deep learning models exceed 8 M parameters and cannot run on low-power pet wearable hardware, creating a gap for noise-resistant, lightweight recognition algorithms.

1.2 Research Gaps

Traditional MFCC-SVM loses over 25% accuracy at low SNR with fixed filters. Standard CNN/LSTM and general lightweight backbones (MobileNet, ShuffleNet) lack audio-specific feature screening, dropping recall by 18% under noise. Three major drawbacks exist:

1) Fixed MFCC filters cause >35% feature distortion in low-SNR conditions;

2) Vision-derived lightweight networks degrade audio feature discrimination by over 40% in mixed noise;
3) Most papers only test one dataset without multi-SNR robustness or embedded deployment validation.

1.3 Core Innovations

1) Feature innovation: Spectral entropy guides dynamic denoising and Mel filter tuning, integrating static + 1st/2nd-order temporal MFCC features for robust time-frequency representation.
2) Engineering innovation: ECA-augmented depthwise separable convolution network. Model size shrinks to 2.07 M, computation down to 0.31 G; ECA attention suppresses noise features and drastically reduces inference delay against CNN-LSTM baselines.

2. Relevant Theoretical and Technical Foundations

2.1 Acoustic Signal Characteristics of Dog Barks

Dog barks are typical non-stationary short-time acoustic signals. The duration of a single effective dog bark ranges from 0.15 s to 1.2 s, with fundamental frequency distributed between 250 Hz and 3500 Hz (Zhang, Y, Wang, L & Liu, H., 2024). Its spectral characteristics vary significantly with dog breeds, emotional states and external stimulus intensity. Compared with human voice signals, dog barks feature stronger suddenness and drastic instantaneous energy fluctuations with complex spectral characteristics. In actual collection scenarios, background noise easily covers key frequency bands of dog barks, causing feature aliasing and signal distortion, which is the core difficulty restricting high-precision dog bark recognition.

2.2 Basic Principle of Traditional MFCC

MFCC relies on the nonlinear auditory characteristics of the human cochlea to map linear frequency space to Mel nonlinear frequency space, which can effectively fit the human auditory rule of being sensitive to low-frequency signals and insensitive to high-frequency signals. The complete extraction process includes five core steps: pre-emphasis, framing and windowing, Fast Fourier Transform (FFT), Mel filtering, and discrete cosine cepstrum solution.

Mel frequency mapping formula:

$$f_{mel}=2595\log_{10}\left(1+\frac{f}{700}\right)$$

Where: f_{mel} denotes Mel frequency; f denotes the original linear-domain frequency in Hz.

FFT is performed on single-frame signals after framing and windowing, spectral energy mapping is completed through triangular Mel filter banks, logarithms are taken on the output energy followed by discrete cosine transform to finally obtain cepstral coefficients:

$$c_m=\sum_{k=1}^{M}\log(S_k)\cos\left[m\left(k-\frac{1}{2}\right)\frac{\pi}{M}\right]$$

Where: c_m is the m-th dimensional MFCC coefficient; M is the number of filters; S_k is the output energy of the k-th filter.

Traditional MFCC usually selects 12~13 dimensional static cepstral coefficients as features. Its inherent defect lies in fixed filter bank parameters that cannot be dynamically and adaptively adjusted with input signal noise intensity. Moreover, it only contains static frequency domain information and lacks temporal dynamic variation features of dog barks, making features highly susceptible to noise interference under low SNR environments.

2.3 Foundations of Lightweight Convolution and Attention Mechanism

Depthwise separable convolution splits standard two-dimensional convolution into two independent steps: depth spatial convolution and pointwise channel convolution, which can drastically compress model parameters and computation volume. The comparison formulas for parameters of standard convolution P_{std} and depthwise separable convolution P_{ds} are as follows:

$$\begin{cases} P_{std}=K\times K\times C_{in}\times C_{out} \\ P_{ds}=K\times K\times C_{in}+1\times 1\times C_{in}\times C_{out} \end{cases}$$

Where: K denotes convolution kernel size; C_{in} denotes the number of input channels; C_{out} denotes the number of output channels. When the convolution kernel size K=3, the parameters of depthwise separable convolution only account for $\frac{1}{C_{out}}+\frac{1}{9}$ of standard convolution, realizing extreme lightweight of the model.

The ECA attention mechanism is an efficient lightweight channel attention module. It abandons the redundant fully connected structure of SE attention and realizes cross-channel information interaction through one-dimensional convolution, completing adaptive weight allocation of feature channels at an extremely low parameter cost to accurately strengthen effective features and suppress invalid noise features, which is highly suitable for constructing lightweight audio recognition networks.

3. Dataset Construction and Audio Preprocessing

3.1 Dataset Construction and Sample Division

The dataset of this paper is constructed by fusing public standard audio datasets and a self-built real-scene dataset. Public datasets ESC-50 and UrbanSound8K are selected, pure dog bark samples and various typical background interference noise samples are screened, with noise types covering vehicle flow, wind, human voice and bird chirps. The self-built dataset is recorded at a sampling rate of 44.1 kHz with 16-bit single-channel format, covering indoor home (Li, J, Chen, X & Zhao, Y., 2023), outdoor community and night silent environments, including 12 common dog breeds with multi-modal dog bark samples such as alert barking, frightened barking and playful barking, featuring wide scene coverage and sufficient sample diversity.

All audio samples are uniformly resampled to the standard format of 16 kHz, intercepted into valid 1 s clips, and invalid samples with mute proportion exceeding 70% are eliminated. The dataset is divided into five categories: dog barks, human voices, traffic noise, wind noise and other animal calls.

Table 1. Statistical Distribution of Dataset Samples

Category	Training Set	Validation Set	Test Set	Total
Dog barks	2462	307	308	3077
Human voices	1845	230	231	2306
Traffic noise	1721	215	216	2152
Wind noise	1412	176	177	1765
Other animal calls	1534	191	192	1917

An 8:1:1 stratified random division strategy is adopted to construct the training set, validation set and test set, ensuring balanced distribution of various samples and effectively avoiding inflated model evaluation results caused by category bias. Test set samples are never involved in training and data augmentation processes to guarantee objectivity and credibility of model evaluation results.

3.2 Audio Preprocessing Pipeline

A standardized audio preprocessing pipeline is constructed in this paper, sequentially completing DC component removal, pre-emphasis processing, short-time energy mute clipping, adaptive noise preprocessing, framing and windowing. The pre-emphasis filter formula is as follows:

$$y(n)=x(n)-\alpha\cdot x(n-1),\alpha=0.97$$

Where: $x(n)$ is the original audio sampling point; $y(n)$ is the output signal after pre-emphasis. The experiment sets frame length to 25 ms and frame shift to 10 ms, adopting the Hanning window to suppress spectral leakage at frame boundaries and guarantee integrity and smoothness of spectral features.

4. Improved MFCC Feature Extraction Based on Adaptive Noise Reduction

4.1 Defect Analysis of Traditional MFCC in Dog Bark Recognition Tasks

Under complex low SNR scenes, background noise severely pollutes Mel-domain spectral features, and traditional fixed-parameter filter banks cannot adapt to dynamic spectral distribution characteristics of dog barks. Experiments show that under 0 dB SNR, the recognition accuracy of the baseline model based on traditional MFCC is only 82.4%. Meanwhile, it only outputs static cepstral features and cannot characterize core features of short-time sudden temporal variations of dog barks, resulting in insufficient feature distinguishability and anti-noise performance to meet high-precision recognition requirements.

4.2 Overall Framework of Improved MFCC Algorithm

The overall process of the improved MFCC algorithm proposed in this paper is: raw audio preprocessing → noise region adaptive discrimination based on spectral entropy → adaptive spectral noise reduction → dynamic Mel filter bank spectral mapping → cepstral coefficient solution → temporal difference feature splicing and fusion, finally outputting a high-robustness two-dimensional time-frequency feature matrix.

Spectral entropy can accurately characterize the uniformity of spectral distribution of single-frame signals, serving as an effective indicator to distinguish noise frames and effective signal frames. The spectral entropy calculation formula is as follows:

$$H=-\sum_{k=1}^{N} p(k)\log_2 p(k)$$

Where: $p(k)$ is the normalized probability of a single spectral line; H is the spectral entropy of the corresponding audio frame. Noise frames feature uniform spectral distribution and high entropy values, while effective dog bark signal frames have concentrated energy and low entropy values. This paper dynamically and adaptively adjusts noise reduction intensity based on entropy thresholds to avoid excessive signal suppression or insufficient noise reduction caused by fixed thresholds.

4.3 Dynamic Mel Filter Bank Design

Aiming at the defect of fixed traditional filter bank parameters, this paper designs a dynamically adjustable Mel filter bank based on spectral statistical rules of dog barks in the training set. When noisy audio frames are detected, the filter density within the main frequency band of dog barks (250~3500 Hz) is adaptively increased (Xu, C., 2024), and the weight proportion of non-target frequency bands is weakened to precisely focus on core characteristic frequency bands of dog barks and suppress interference from invalid noise frequency bands. The number of filters is set to 24 in this paper, and 13-dimensional core static MFCC coefficients are finally extracted.

4.4 Temporal Difference Feature Fusion

To make up for the insufficient temporal characterization capability of static MFCC, first- and second-order differential temporal features are introduced to depict dynamic variation rules between audio frames. The first-order difference calculation formula is as follows:

$$\Delta c(t)=\frac{\sum_{i=-2}^{2} i\cdot c(t+i)}{\sum_{i=-2}^{2} i^2}$$

The second-order difference is calculated by secondary differentiation of first-order difference results. In this paper, 13-dimensional static MFCC features, 13-dimensional first-order difference features and 13-dimensional second-order difference features are spliced and fused to construct 39-dimensional high-dimensional fusion features, comprehensively covering static frequency domain features and dynamic temporal features of dog barks.

4.5 Feature Visualization and Quantitative Comparative Analysis

Traditional MFCC and improved MFCC feature heatmaps are extracted respectively for the same noisy dog bark audio sample. Visualization results show that traditional MFCC features are severely interfered by noise with chaotic feature textures and blurred effective features; after optimization via adaptive noise reduction and dynamic filtering, the improved MFCC proposed in this paper delivers clear core dog bark feature textures with noise pseudo-features effectively suppressed and significantly enhanced feature distinguishability.

Table 2. Recognition Performance of Baseline Models with Different MFCC Features

Feature Type	Accuracy	Precision	Recall	F1-Score
Traditional MFCC	0.824	0.817	0.803	0.810
Proposed Improved MFCC	0.867	0.861	0.852	0.856

Under identical model and dataset conditions, the improved MFCC features boost recognition accuracy by 4.3%, effectively verifying the scientificity and effectiveness of the feature optimization strategy proposed in this paper.

5. Attention-Enhanced Lightweight Neural Network

5.1 Model Design Objectives and Overall Architecture

The core design objectives of the model proposed in this paper are "high accuracy, low parameter count, low delay and easy deployment". Computing overhead is maximally compressed on the premise of guaranteeing recognition accuracy to adapt to real-time reasoning requirements of embedded edge devices. The network input is a T×39 two-dimensional feature matrix constructed from improved MFCC fusion features. The overall architecture consists of an input layer, lightweight convolution backbone, ECA attention module, global average pooling layer, dropout layer and fully connected classification output layer. Main layer parameters of the network are shown in Table 3.

Table 3. Main Layer Parameters of Network Backbone

Module Name	**Input Channels**	**Output Channels**	**Core Operation**	**Stride**	**Output Feature Size**
Convolution Layer 1	1	32	Standard 3×3 Convolution	1	32×98×39
Depthwise Separable Conv Layer 1	32	64	3×3 Depthwise Separable Convolution	2	64×49×20
Depthwise Separable Conv Layer 2	64	96	3×3 Depthwise Separable Convolution	2	96×25×10
Depthwise Separable Conv Layer 3	96	128	3×3 Depthwise Separable Convolution	2	128×13×5
ECA Attention Module	128	128	One-Dimensional Convolution Channel Attention	1	128×13×5
Global Average Pooling Layer	128	128	Global Average Pooling	1	128
Dropout Layer	128	128	Dropout (p=0.3)	1	128
Fully Connected Classification Layer	128	5	Linear Transformation	1	5

The backbone network relies on multi-layer depthwise separable convolution to realize progressive extraction of time-frequency features. Shallow networks capture local refined time-frequency features, while deep networks mine high-level semantic features to realize layered and accurate feature characterization.

5.2 Embedding Mechanism of ECA Attention Module

The ECA module compresses the spatial dimension of 128-dimensional feature maps, abandons the redundant structure of SE attention, and realizes lightweight cross-channel feature interaction relying on 3×1 one-dimensional convolution to adaptively adjust weights of each channel. This module only adds less than 0.01 M parameters with an inference time consumption increase lower than 1.2%. Experimental results show that embedding the ECA module improves the model F1-score by 5.8%, effectively enhancing feature screening capability and recognition accuracy under complex noisy environments.

5.3 Model Training Strategy Settings

Model training and testing are completed based on the PyTorch 2.1 deep learning framework in this paper. The AdamW optimizer is selected to optimize model parameters with a weight decay coefficient set to 1×10^{-4} to suppress model overfitting; the initial learning rate is set to 3×10^{-4}, and a cosine annealing algorithm is adopted for dynamic learning rate decay to improve model convergence stability (Sinnott, R, Zhang, L, & Wang, H., 2026). Weighted cross-entropy loss function is used to balance sample weights targeting the problem of unbalanced dataset categories. The dropout probability is set to 0.3, combined with L2 regularization to further improve model generalization capability. The experiment sets batch size to 32 and maximum training epochs to 120. An early stopping mechanism is introduced to terminate training when validation loss does not decline for 15 consecutive epochs to obtain optimal model weights.

5.4 Statistical Lightweight Performance Indicators of the Model

To quantitatively evaluate the lightweight advantages of the proposed model, parameters and floating-point computation volume of the proposed model and multiple mainstream baseline models are counted, with results shown in Table 4.

Table 4. Comparison of Lightweight Indicators of Various Models

Model	Parameters (M)	Floating-Point Operations (G)
CNN-LSTM	13.42	1.87
MobileNetV2	3.51	0.58
ShuffleNetV2	2.73	0.42
Proposed Network	2.07	0.31

The proposed model only contains 2.07 M parameters with a floating-point operation volume as low as 0.31 G, delivering significantly superior lightweight performance compared with various comparative baseline models, featuring ultra-low deployment overhead and excellent edge adaptability.

6. Experimental Results and Analysis

6.1 Experimental Environment and Evaluation Indicators

1) **Hardware Environment**: NVIDIA RTX 3090 graphics card, Intel i9-10900X processor, 64 GB RAM.
2) **Software Environment**: Ubuntu22.04 system, Python3.10 programming language, PyTorch2.1 deep learning framework, Librosa0.10.1 audio processing library.

The evaluation system of this paper covers three dimensions: accuracy, lightweight performance and robustness:

- **Accuracy indicators**: Accuracy, Precision, Recall, F1-score, used to measure model classification performance;
- **Lightweight indicators**: Parameter count, floating-point operation volume, single-sample inference time consumption, used to evaluate model deployment overhead;

- **Robustness indicators**: F1-scores under multiple gradient SNRs from -5 dB to 20 dB, used to verify model anti-noise capability.

6.2 Ablation Experiments

Four groups of controlled ablation experiments are set to sequentially verify the independent effectiveness and collaborative gain effect of each innovative module proposed in this paper:

Group A: Traditional MFCC features + basic lightweight network without attention mechanism
Group B: Improved MFCC features + basic lightweight network without attention mechanism
Group C: Traditional MFCC features + complete lightweight network embedded with ECA attention
Group D: Improved MFCC features + complete lightweight network embedded with ECA attention

Table 5. Test Results of Ablation Experiments

Experimental Group	Accuracy	Precision	Recall	F1-Score
A	0.826	0.821	0.807	0.814
B	0.868	0.863	0.854	0.858
C	0.851	0.844	0.835	0.839
D	0.905	0.901	0.893	0.897

Comparison between Group A and Group B verifies that the improved MFCC feature optimization strategy can significantly boost model feature characterization capability and recognition accuracy; comparison between Group A and Group C shows that the ECA attention module can effectively optimize network feature screening performance and improve model recognition effect; Group D integrating all innovative modules achieves optimal comprehensive performance. Experimental results fully prove that the feature optimization module and attention network module deliver obvious positive collaborative gain effects, and each innovative module is necessary and effective.

6.3 Comparative Experiments with Mainstream Algorithms

Under identical datasets, experimental environments and training hyperparameters, comprehensive performance comparisons are carried out between the proposed algorithm and traditional machine learning algorithms as well as mainstream deep learning algorithms, with experimental results shown in Table 6.

Table 6. Comprehensive Performance Comparison of Different Algorithms

Algorithm	**Accuracy**	**F1-Score**	**Parameters (M)**	**Single-Sample Inference Time (ms)**
SVM + Traditional MFCC	0.781	0.772	N/A	4.2
CNN	0.842	0.835	8.63	12.7
CNN-LSTM	0.874	0.866	13.42	18.3
MobileNetV2	0.881	0.873	3.51	7.1
ShuffleNetV2	0.872	0.864	2.73	5.8
Proposed Algorithm	0.905	0.897	2.07	4.7

6.4 Multi-SNR Robustness Experiments

Gaussian noise of different intensities is superimposed on test set audio to construct test samples with gradient SNRs ranging from -5 dB to 20 dB, and F1-scores of various algorithms under different noisy environments are compared to verify model anti-noise robustness.

Under extremely low SNR of -5 dB with strong noise, the F1-score of the proposed algorithm still

reaches 0.783 (Hantke, S, Cummins, N, & Schuller, B., 2018), significantly superior to two mainstream comparative models, fully verifying the anti-noise advantages of combining adaptive noise reduction features and attention mechanisms. As SNR increases, performance of all algorithms steadily improves, yet the proposed algorithm maintains significant performance advantages all the time, delivering outstanding environmental robustness and scene adaptability.

6.5 Confusion Matrix and Performance Mechanism Analysis

Confusion matrix analysis results of the test set show that model recognition errors mainly originate from feature confusion between dog barks and other animal calls, while misrecognition proportions of steady-state noise such as traffic noise and wind noise are extremely low. The internal mechanism of performance improvement of the proposed algorithm can be summarized into two points:

1) The improved MFCC algorithm suppresses interference of background noise on time-frequency features through spectral entropy adaptive noise reduction and dynamic filtering mechanisms, and combines temporal difference features to accurately capture core characteristics of short-time sudden changes of dog barks.
2) The ECA attention module adaptively optimizes the weight distribution of feature channels to strengthen effective dog bark features and suppress redundant noise features. The lightweight convolution backbone realizes high-precision feature mining at low computing overhead, achieving the optimal balance between recognition accuracy and inference efficiency.

7. Model Quantization and Edge Device Deployment Test

7.1 INT8 Quantization of the Model

To further adapt to deployment requirements of low-computing embedded hardware, post-quantization of INT8 is carried out on the trained high-precision floating-point model in this paper, drastically compressing model storage volume and reasoning computing overhead. A precision-preserving optimization strategy is adopted in the quantization process to strictly control precision loss. After quantization, the model recognition accuracy only drops by 0.8%, with precision loss within acceptable engineering range, balancing model lightweight characteristics and recognition accuracy.

7.2 Actual Measurement on Raspberry Pi Edge Terminal

Raspberry Pi 4B is selected as a typical edge deployment hardware platform to test the full-process processing time consumption of 1 s audio samples, covering all links including audio reading, MFCC feature extraction and model reasoning output. Actual measurement results show that the average full-process reasoning time consumption of the algorithm is 12.6 ms, far lower than the audio sample duration, enabling real-time recognition of audio signals; the peak operating memory of the model is only 182 MB, allowing stable operation on low-power edge monitoring terminals and adapting to IoT lightweight deployment scenarios.

7.3 Field Test in Real Scenarios

Field tests are carried out in complex outdoor urban community scenes in this paper, covering a total of 320 measured audio clips including 140 dog bark samples and 180 background noise samples. The field test accuracy reaches 87.1% with a false positive rate of only 7.8%. Affected by unknown mixed noise, the accuracy is slightly lower than dataset test results, yet its recognition stability can meet practical engineering application demands such as community noise early warning and intelligent pet monitoring.

8. Analysis and Discussion

8.1 Limitations of the Algorithm

- **Limited adaptability under extremely strong noise**: Under strong noise below -5 dB, the original spectral structure of dog barks is severely damaged with prominent feature distortion, leading to obvious attenuation of algorithm recognition performance.
- **Insufficient recognition stability for ultra-short samples**: For ultra-short weak dog bark clips shorter than 0.1 s, the model has limited capability to capture features with room for further improvement in recognition stability.
- **Single recognition granularity**: This paper only realizes accurate detection of dog barks without further subdividing emotional categories and behavioral modes corresponding to dog barks, resulting in insufficient refined recognition granularity.
- **Expandable dataset coverage**: The breed coverage, scene diversity and sample volume of the self-built dataset still have room for further expansion.

8.2 Future Research Directions

Compared with most existing animal acoustic recognition literature that only focus on static accuracy optimization on datasets, this paper constructs a complete research system of "feature optimization - model lightweighting - robustness verification - edge deployment measurement", which is more consistent with practical engineering implementation demands and delivers stronger research integrity and practicability.

9. Conclusion

A dog bark recognition algorithm combining improved MFCC and ECA attention lightweight network is proposed in this paper targeting complex noisy environments. Anti-noise performance of features is effectively improved through spectral entropy adaptive noise reduction and temporal feature fusion. Experiments prove that the proposed method boosts recognition accuracy by 4.3%, only contains 2.07 M model parameters, delivers favorable robustness within a wide SNR range and supports real-time reasoning on edge devices. Future work will optimize adaptability under strong noise, realize refined classification of dog bark emotions and adapt to ultra-low power hardware.

References

Hantke, S, Cummins, N, & Schuller, B. (2018) 'What is my dog trying to tell me? The automatic recognition of the context and perceived emotion of dog barks', in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5134-5138. IEEE. https://doi.org/10.1109/ICASSP.2018.8462634.

Li, J, Chen, X, Zhao, Y. (2023) 'Lightweight ECA-CNN network for environmental audio event classification', *Neurocomputing*, 548, p. 126345. https://doi.org/10.1016/j.neucom.2023.126345.

Sinnott, R, Zhang, L, & Wang, H. (2026) 'Multi-QuadEmoNet: Cat and dog emotion classification model from animal vocalization using multi-stage LSTM-GRU paradigm', *Frontiers in Veterinary Science*, 13, p. 1799281. https://doi.org/10.3389/fvets.2026.1799281.

Xu, C. (2024) 'Neural networks for audio classification: Multi-scale CNN-LSTM approach to animal sound recognition', *Applied and Computational Engineering*, 89(1), pp. 172-177. https://doi.org/10.54254/2755-2721/89/20241128.

Zhang, Y, Wang, L, Liu, H. (2024) 'Adaptive MFCC feature optimization based on spectral entropy for low signal-to-noise ratio audio recognition', *Applied Acoustics*, 212, p. 109586. https://doi.org/10.1016/j.apacoust.2023.109586.