Journal of Engineering, Project, and Production Management
2026, 16(3), 2025-228

A Perceived Acoustic Landscape Model for Zhangjiajie National Forest Park

Qing Sun[1], Qianni Cheng[2], Jianlin Tian[3], and Yanan Zhou[4]
[1] Teacher, School of Intelligent Construction, Jishou University, Jishou, 416000, China
[2] Teacher, School of Liberal Arts, Zhangjiajie College, Zhangjiajie, 427000, China, E-mail: chenqni@outlook.com (corresponding author).
[3] Teacher, School of Fine Arts, Jishou University, Jishou, 416000, China
[4] Teacher, School of Intelligent Construction, Jishou University, Jishou, 416000, China

Project Management
Received October 4, 2025; revised November 26, 2025; accepted December 18, 2025
Available online April 8, 2026

Abstract: To explore tourist's perception of soundscapes, this study selects the core scenic area of Zhangjiajie National Forest Park and takes a microphone array system to collect environmental sound data throughout the four seasons during both day and nighttime hours, while simultaneously collecting tourist's subjective perception, evaluation, and behavioral data. By extracting and preprocessing acoustic features and combining rectangular and dilated convolution kernels, an environmental sound classification model is proposed. Subsequently, a soundscape perception model is constructed. The proposed environmental sound classification model achieved an average accuracy of 93.5% in natural sound source classification, demonstrating the best robustness across different signal-to-noise ratio levels. The parameter size was only 6.8M, and the average inference speed was 12.3ms. The sound landscape perception model demonstrated better predictive performance in dimensions such as pleasure and tranquility than the baseline model. The natural sound source area had high pleasure, naturalness, and tranquility, and low interference. The highest interference level in the traffic impact area was 4.39, with engine noise accounting for 58% of the total noise, and significant spatial and temporal differences between different regions. This model can effectively predict tourist's perception and preferences of soundscapes, with significant differences in perception in different acoustic environments. The research provides a scientific acoustic basis for optimizing soundscapes, scenic area planning, and noise management, which helps to enhance tourist experience and promote the sustainable development of ecotourism.

Keywords: Acoustic environment modeling, ecotourism and sustainability, rectangular convolution kernel, self-distillation architecture, soundscape perception, spatiotemporal differences, visitor experience analysis.

DOI 10.32738/JEPPM-2025-228

1. Introduction

With the rapid urbanization and growth of the tourism industry, the natural scenic spot's acoustic environment faces challenges such as increased noise interference from human activities and the weakening of ecological acoustic features. How to scientifically measure acoustic landscape perception and uncover its underlying mechanisms has become a key research focus in environmental psychology and ecological protection (Tokaç et al., 2025). Currently, studies on soundscapes have gradually shifted from the simplicity of describing basic acoustic parameters to developing multidimensional models that combine acoustic features, semantic understanding, and environmental factors (Dua et al., 2023). However, current research shows limitations. First, traditional acoustic models are unable to analyze the frequency domain features of complex environmental sounds. This is especially true in situations where natural sound sources and artificial noise are mixed, with no efficient methods available for separating and classifying these features (Gao et al., 2024). Second, soundscape perception models often rely on single-modal data, and the mechanisms for integrating acoustic physical characteristics, semantic sound tags, and environmental variables have not been thoroughly explored. Third, there is a lack of research on the spatiotemporal dynamics of acoustic landscapes in typical natural scenic spots, making it difficult to understand perceptual differences caused by interactions between ecological environments and tourism activities (Jadoul et al., 2024). Zhangjiajie National Forest Park, as China's first national forest park, has cultivated a unique natural soundscape with its quartz sandstone peaks and forest landscape. Yet, it also faces noise pollution from tourism development (Haghighi et al., 2024). Nevertheless, existing research on the region's acoustic landscape mainly remains

descriptive, lacking comprehensive data collection across multiple seasons and day-night cycles, as well as perceptual models that incorporate deep learning and attention mechanisms.

Acoustic signal processing is a technology that collects, preprocesses, extracts features, and analyzes sound wave signals to obtain useful information and achieve goals such as noise suppression and pattern recognition. Alghamdi et al. (2024) proposed extracting Mel-frequency cepstral coefficients and analyzing them with deep convolutional neural networks for respiratory sound recognition of lung diseases. They achieved an accuracy of 97.4% on the validation set and 95.1% on the independent test set. Raevskii and Burdukovskaya (2024) studied the comprehensive effects of random internal waves and formed wind waves on the spatial processing coherence and efficiency of narrowband acoustic signals in shallow water from both theoretical and numerical perspectives. A theoretical model for the correlation matrix of multi-mode signals at the horizontal array aperture is proposed based on the spatiotemporal scale differences of sound field fluctuations caused by wind waves and internal waves. The research focuses on analyzing the relationship between array gain and wind and wave intensity, as well as the distance between the sound source and the array. The results indicate that, despite summer hydrological conditions, wind and waves still have a significant impact on the gain of horizontal arrays over a wide distance range of 10 to 100 kilometers. Liu et al. (2024) used low-frequency acoustic sensors to collect sound and designed sparse Mayer filters to generate time-frequency maps for non-cooperative drone detection within 500 meters, followed by recognition with a deep residual network. The system achieved an accuracy of 99.7% in the test set. Zhang et al. (2024) established a coupled dynamic model of vehicle-track-bridge sound barriers to address the secondary structural noise problem of upright sound barriers in rail transit and integrated it with the acoustic boundary element method for low-noise optimization design. The results showed that barrier vibration and noise reduction were significant under the optimized scheme with a reinforced plate. Zhang et al. (2024) tackled the problem of scarce large-scale annotated data for respiratory audio by pre-training a respiratory acoustic baseline model using 136,000 unlabeled samples and creating 19 downstream task benchmarks. This model outperformed general audio pre-training models on 16 tasks.

The characteristics of soundscapes influence human perception, spatial behavior, and specie's risk response and behavioral patterns in ecology, through factors like sound source type and sound level. Li et al. (2025) analyzed the relationship between sound sources and tourist satisfaction, comparing soundscapes in urban and forest parks. They collected sound level measurements and questionnaire data from 1903 and Xishan Forest Parks. Results showed that the average sound level of 1903 Park slightly exceeded 55dB, while Xishan Forest Park met the standards. Wrege et al. (2024) proposed a passive acoustic monitoring grid to measure diurnal activity changes of forest elephants in response to early risk perception of human interference, recording acoustic data of elephant groups in a 1,250 square kilometer area in Congo. As the poaching risk increased, the proportion of nocturnal activity in elephant herds significantly rose. Shao et al. (2024) developed a visual prediction model based on the psychological perception mechanism of urban green spaces near highways, combining sound walking and geographic information systems. They used environmental sound measurements and subjective questionnaires from Chengdu Bailuwan Wetland Park as a case study. Results indicated that landscape feature regulation had a significant influence on psychological perception. Razani et al. (2025) applied qualitative content analysis to clarify the interaction between sound and hearing in landscape perception, addressing auditory neglect in landscape reading. Through literature review and deductive analysis, they integrated visual and acoustic elements. The synergy of natural sound sources and visual features enhanced spatial comfort and strengthened the user's sense of presence and landscape vitality. Li and Liu (2024) used a partial least squares structural equation model to analyze the overall impact of park audio-visual environments on restorative experiences, based on data from 861 tourists across five cities. Results showed that natural elements like trees, water sounds, and bird songs directly benefited emotional well-being, preferences, and perceptual resilience. The positive effect of bird songs was significantly correlated with acoustic quality indicators.

Although significant progress has been made in sound classification and perceptual prediction in existing soundscape research, key research gaps remain, primarily the lack of robust classification methods for complex natural sound environments. The existing environmental sound classification models often have insufficient classification accuracy and noise resistance when dealing with scenes where natural sound sources (such as flowing water and bird songs) are highly mixed with human noise (such as tourist conversations and traffic engines). Second, there is no exploration of multimodal feature fusion mechanisms. The current soundscape perception models mostly rely on a single acoustic physical feature and fail to fully integrate the semantic labels of sound (such as "bird chirping," "engine sounds") and environmental variables (such as season, day, and night). Third, there is a lack of research on the spatiotemporal dynamic patterns of soundscapes in typical natural scenic areas. There is a lack of systematic research to reveal the differences and causes of perceived soundscapes in natural scenic areas across different seasons and day and nighttime scales (Zhang et al., 2024). Based on this, the study aims to construct a perception model for the acoustic landscape of Zhangjiajie National Forest Park using acoustic models and environmental sound analysis, with the hope of providing theoretical support for optimizing the natural sound environment. The new contribution of the research lies in proposing a lightweight and robust environmental sound classification model that combines rectangular convolution kernels, dilated convolutions, and self-distillation architectures. At the same time, a sound landscape perception model was developed that integrates multimodal features (acoustic and physical features, semantic labels, environmental variables) and utilizes attention mechanisms for feature fusion, while completing regression (perception score prediction) and classification (preference level recognition) tasks. Finally, at the practical level, a systematic study was conducted on the soundscapes of Zhangjiajie National Forest Park in four seasons and day and night, revealing the spatiotemporal dynamic differences in soundscape perceptions and providing new quantitative tools and theoretical support for optimizing the sound environment of ecotourism scenic spots.

2. Research Methods

2.1. Research Area and Data Collection

The core scenic areas of Zhangjiajie National Forest Park, including Jinbian Creek, Yuanjiajie, Huangshizhai, and other regions, are selected as research subjects. Based on sound source types and human activity levels, these areas are further divided into three typical sound environment zones, as shown in Fig. 1.

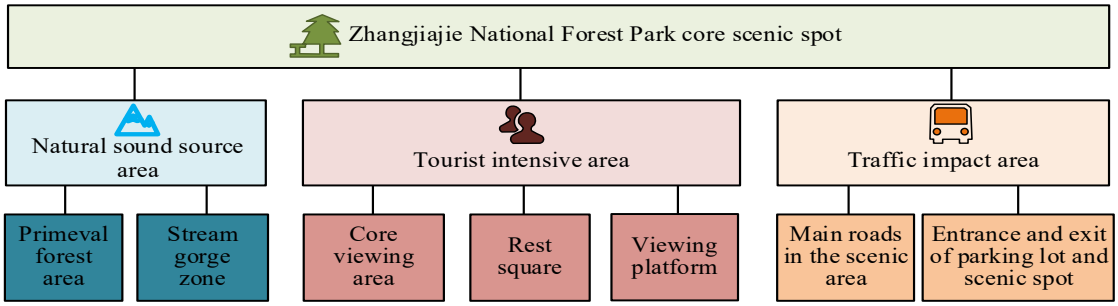


Fig. 1. Typical acoustic environment area

In Fig. 1, natural sound sources are mainly located in primitive forests and between streams and canyons far from tourist trails. The heavily visited tourist area includes the main viewing area, rest plaza, and observation deck of various scenic spots. The traffic impact zone covers the main roads within the scenic area, routes for electric bicycles, parking lots, and the areas around the entrance and exit of the scenic area (Yoon et al., 2023). Environmental sound data are collected using a microphone array system, which consists of 8 omnidirectional microphones arranged at the vertices and center of a square with a side length of two meters. The sound sampling rate is set to 44.1kHz, and each sample is recorded for at least 30 minutes. To capture changes in the acoustic environment across different seasons and times of day, data collection occurs throughout the four seasons (spring, summer, autumn, and winter), samples are taken during the daytime (6:00-18:00) and nighttime hours(18:00-6:00). Weather conditions and light intensity are also recorded simultaneously during the sampling time.

This study gathers tourist's subjective perceptions and evaluations of soundscapes through an on-site survey questionnaire and the Likert 5-point scale. The evaluation encompasses four key aspects: pleasure, tranquility, naturalness, and disturbance. Each aspect offers five response options, ranging from "very dissatisfied" to "very satisfied," with scores from 1 to 5 (Kaidouchi et al., 2023). During the survey, combined with an analysis of tourist behavior trajectories, data such as tourist's stay times and photo frequencies at various sampling points are recorded using cameras installed at key locations in the scenic area and positioning devices carried by tourists.

2.2. Acoustic Feature Extraction and Preprocessing

After collecting environmental sound data, acoustic feature extraction and preprocessing are necessary. First, basic acoustic parameters such as sound pressure level, equivalent continuous sound level. A sound level, spectral centroid, standard deviation, and others are calculated. The sound pressure level represents the intensity of sound, as shown in Eq. (1) (Moufid et al., 2024).

$$\mathrm{SPL}=20\log_{10}\left(\frac{p_{\mathrm{rms}}}{p_{\mathrm{ref}}}\right) \tag{1}$$

In Eq. (1), p_{rms} signifies the root mean square value of sound pressure. p_{ref} signifies the reference sound pressure. The equivalent continuous sound level (Leq) represents the average sound pressure level of a non-stable sound over a period of time, as shown in Eq. (2).

$$\mathrm{Leq}=10\log_{10}\left(\frac{1}{T}\int_0^{\tau}10^{\frac{\psi_I(\zeta)}{10}}dt\right) \tag{2}$$

In Eq. (2), T is the measurement time. SPL(t) is the sound pressure level that varies with time t. The A sound level takes into account the auditory characteristics of the human ear to sounds of different frequencies. The sound pressure level is corrected through an A-weighted network. The calculation is shown in Eq. (3).

$$L_A=\mathrm{SPL}+A(f) \tag{3}$$

In Eq. (3), $A(f)$ is the correction value of the A-weighted network at frequency f. A sound level is closely aligned with the subjective perception of sound volume by the human ear and is used for environmental noise assessment. The spectral centroid Centriod signifies the center frequency of the energy in the sound spectrum, as shown in Eq. (4) (Hossain et al., 2025).

$$\mathrm{Centroid}\ \frac{\sum_{i=1}^{N}f_iE_i}{\sum_{i=1}^{N}E_i} \tag{4}$$

In Eq. (4), N signifies the total frequency component. f_i signifies the frequency value of the i-th frequency component. E_i signifies the energy of the i-th frequency component. The spectral centroid can reflect the frequency distribution characteristics of sound, and sounds with more high-frequency components have higher spectral centroid values. The standard deviation SD is used to measure the fluctuation of the sound signal, as shown in Eq. (5).

$$SD=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{5}$$

In Eq. (5), n represents the quantity of sound signal samples. x_i signifies the i-th sample value. $\bar{x}$ is the average value of the sample. After calculating acoustic parameters, feature extraction must be carried out using deep learning. The study utilizes v

Mel-Frequency Cepstral Coefficients (MFCC) are used to process speech features and audio signals. The calculation process is shown in Fig. 2.

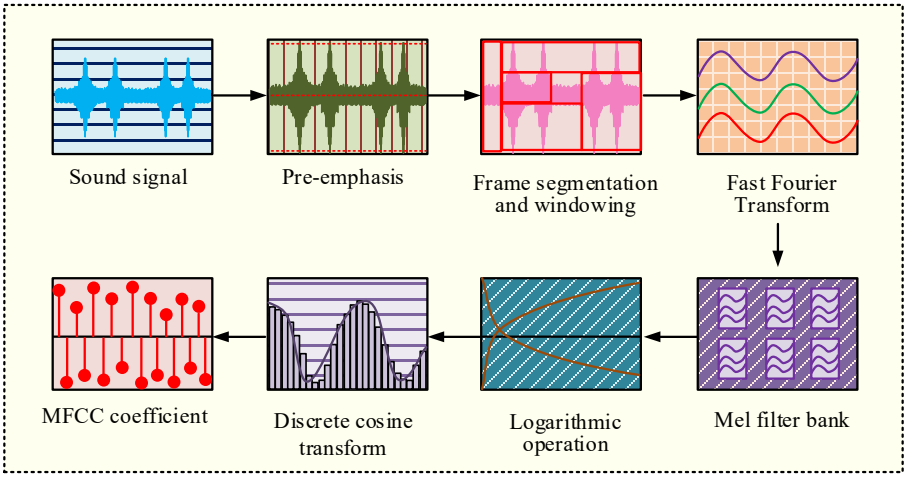


Fig. 2. MFCC calculation process

In Fig. 1, the audio signal is first converted to the Mel-frequency scale. Eq. (6) presents the conversion relationship between Mel-frequency and actual frequency.

$$\mathrm{Mel}(f)=2595\log_{10}\left(1+\frac{f}{700}\right) \tag{6}$$

In Eq. (6), $\mathrm{Mel}(f)$ is the Mel-frequency corresponding to frequency f. Then, the signal at the Mel-frequency scale is subjected to Discrete Cosine Transform (DCT) to obtain MFCC, as presented in Eq. (7).

$$\mathrm{MFCC}_n=\sum_{k=1}^{M}\log(\mathrm{En}_k)\cos\left(\frac{\pi n_1(k-0.5)}{M}\right) \tag{7}$$

In Eq. (7), n_1 is the index of the MFCC coefficient. M is the quantity of filter banks. En_k is the energy of the k-th filter group. The Mel spectrogram is a graphical representation of the energy distribution of an audio signal on the Mel-frequency scale, calculated based on the Short-Time Fourier Transform (STFT). Assuming the audio signal is $x(t)$, the STFT is shown in Eq. (8).

$$STFT(t,f)=\sum_{m=-\infty}^{\infty}x(m)w(m-t)e^{-j2\pi fm} \tag{8}$$

In Eq. (8), $w(m)$ is the window function. By converting the STFT results to the Mel-frequency scale through a Mel filter bank, the Mel spectrogram can be obtained. To optimize the generalization ability, the study takes noise injection and time stretching to expand the sample size.

2.3. Environmental Sound Classification Model

To efficiently extract high information density frequency domain features from audio spectrograms, rectangular convolution kernels and dilated convolution are combined to propose an environmental sound classification model, as shown in Fig. 3.

In Fig. 3, the environmental sound classification model uses a self-distillation architecture, which consists of a pre-trained label branch and a prediction model core. The model captures multi-frequency information with rectangular convolution kernels, broadens the receptive field through dilated convolution, and improves feature representation using 1D convolution, pooling, and batch normalization. The upper branch transfers knowledge to the prediction model via pre-trained labels, combined with a label smoothing strategy, to enhance classification accuracy and generalization. The rectangular convolution kernel has a large receptive field in the frequency domain and can simultaneously capture information from multiple frequency components. Its convolution operation is shown in Eq. (9) (Iyendo et al 2024).

$$y_{i,j}=\sum_{b=0}^{B-1}\sum_{c=0}^{C-1}x_{i+b,j+c}\cdot k_{b,c} \tag{9}$$

In Eq. (9), $y_{i,j}$ signifies the value of the output feature map of the convolutional layer at position (i, j). $x_{i+b,j+c}$ signifies the value of the input feature map at position $(i+b,j+c)$. $k_{b,c}$ signifies the value of the rectangular convolution kernel at position (b, c). B and C are the sizes of the convolution kernel in the time and frequency directions, respectively.

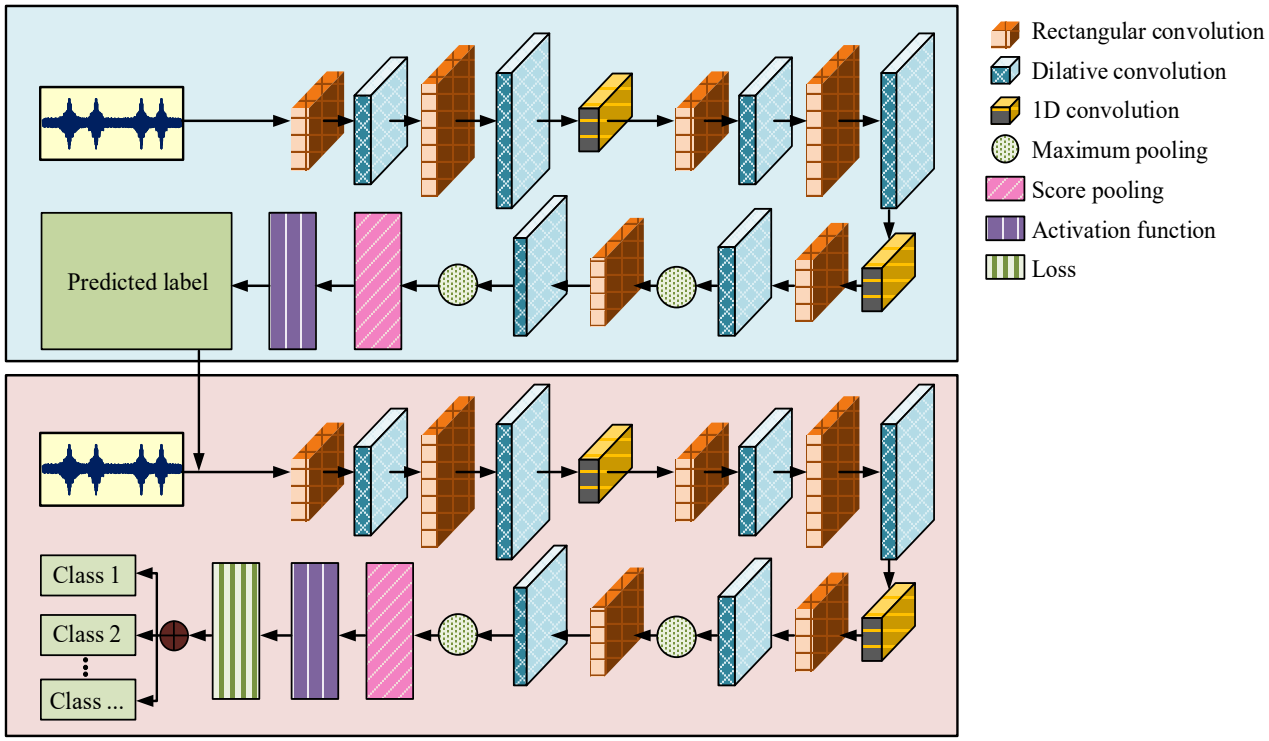


Fig. 3. Structure of the environmental sound classification model

Dilated convolution expands the receptive field of the convolution kernel by inserting intervals between elements, which can extract a wider range of frequency domain features without increasing the number of parameters. The calculation for dilated convolution is shown in Eq. (10).

$$y_{i,j}=\sum_{b=0}^{B-1}\sum_{c=0}^{C-1}x_{i+r\cdot b,j+r\cdot c}\cdot k_{b,c} \tag{10}$$

In Eq. (10), r is the expansion factor. To optimize the performance, the self-distillation soft labels and an improved Jensen-Shannon (JS) divergence loss function are adopted. Self-distillation soft labels use the predicted results of the model itself as soft labels to guide the model to learn more discriminative features (Peng et al., 2024). The improved JS divergence loss function is shown in Eq. (11).

$$L_{JS}=\frac{1}{2}[D_{KL}(p\|q)+D_{KL}(q\|p)] \tag{11}$$

In Eq. (11), $D_{KL}(q\|p)$ is the Kullback-Leibler divergence between p and q, used to measure the difference between two probability distributions. p is the probability distribution predicted by the model. q is the target probability distribution. By minimizing the loss function, the model can achieve efficient classification performance while reducing model complexity.

2.4. Construction of Acoustic Landscape Perception Model

After completing the sound data processing, the study integrates acoustic physical features, sound semantic labels, and environmental variables as input features for the soundscape perception model. The study uses an attention mechanism to weight multi-source fusion features, emphasizing the influence of key features on acoustic landscape perception. First, the similarity score between the feature vector and the query vector is calculated, and the attention weights for each feature vector are obtained through normalization. The weight magnitude indicates the importance of the features. Finally, the feature vectors are weighted and summed based on their weights to produce the final fused features. The similarity score between the feature vector and the query vector is calculated in Eq. (12).

$$s_i=\mathrm{Sim}(Q,X_i) \tag{12}$$

In Eq. (12), s_i is the similarity score between the i-th feature vector X_i and the query vector Q. Sim signifies the

similarity calculation function. Then, the similarity score is normalized to obtain attention weights, as shown in Eq. (13).

$$\alpha_d = \frac{\exp(s_i)}{\sum_{d=1}^{D} \exp(s_d)} \tag{13}$$

In Eq. (13), α_d signifies the attention weight of the d-th feature vector. D signifies the total number of feature vectors. According to the attention weights, the feature vectors are weighted and summed to obtain the final fused feature, as shown in Eq. (14).

$$Z = \sum_{d=1}^{D} \alpha_d X_d \tag{14}$$

The acoustic landscape perception model consists of two parts: a regression and a classification task. The regression task aims to predict tourist's perception ratings of soundscapes, and uses mean square error as the loss function, as displayed in Eq. (15).

$$\mathrm{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2 \tag{15}$$

In Eq. (15), y_i is the true perception score. $\hat{y}_i$ is the model's predicted perception score. The classification task divides the preference for soundscapes into three levels: high, medium, and low. The model is trained using the cross-entropy loss function, as displayed in Eq. (16).

$$\mathrm{L_{CE}} = -\sum_{i=1}^{N} \sum_{g=1}^{G} y_{ig} \log\left(\hat{y}_{ig}\right) \tag{16}$$

In Eq. (16), G represents the number of categories. y_{ig} signifies the true label of sample i, belonging to category g. $\hat{y}_{ig}$ is the probability that the model predicts sample i belongs to category g. The structure of the acoustic landscape perception model is shown in Fig. 4.

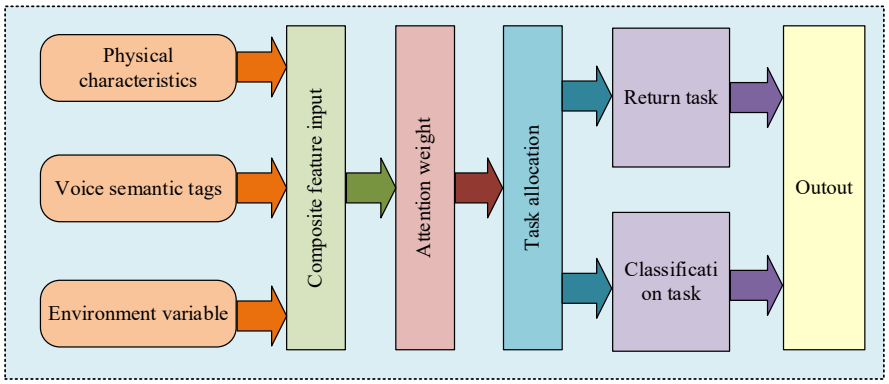


Fig. 4. Structure of acoustic landscape perception model

In Fig. 4, the acoustic landscape perception model combines physical acoustic features, sound semantic labels, and environmental variables to create a comprehensive feature input that includes sound physical properties, semantic meanings, and scene background information. Next, the similarity between the feature vector and the query vector is computed using an attention mechanism and then normalized to produce attention weights that indicate the importance of each feature. These multi-source features are weighted and summed accordingly. The model performs two tasks: regression and classification. The regression task predicts tourists' perceived ratings of soundscapes using the mean squared error loss function, while the classification task employs the cross-entropy loss function to categorize soundscape preferences into high, medium, and low levels. Finally, backpropagation is used to optimize parameters for accurately modeling and predicting tourists' perception of soundscapes.

3. Results

3.1. Performance Evaluation of the Environmental Sound Classification Model

The experiment is built on the PyTorch 2.0 framework and runs on an NVIDIA RTX 4090 GPU (24GB of video memory), an Intel i9-13900K CPU (32 cores), and 64GB of RAM. The data is divided into a training set (60%), a validation set (20%), and a test set (20%), with 5-fold cross-validation conducted. In the model training parameter settings, the batch size is 32, the initial learning rate is 0.001 (using the AdamW optimizer), the weight decay is 0.0001, the training lasts 300 epochs, and training stops if the validation loss does not decrease for 10 consecutive epochs. The data augmentation strategy includes ±15% time stretching and adding 10dB Gaussian white noise. The comparison models selected are ResNet-18, MobileNetV3-Large, and EfficientNet-B0, all using default parameter configurations. Fig. 5 shows the classification accuracy of different models on the test set. Fig. 5(a), 5(b), and 5(c) display the accuracy for natural sound sources, tourist voices, and traffic noise, respectively. The proposed model outperformed all three in the three sound categories, especially in natural sounds, with an average accuracy of 93.5%, surpassing EfficientNet-B0's 90.1%. The results suggest that the

ability of rectangular convolution kernels to capture multi-frequency features, along with the complex spectral structure improved by dilated convolution, enhances model performance.

Fig. 6 shows the number of model parameters and inference time. Fig. 6(a) displays the model's parameter count. The proposed model has 6.8 million parameters, which is fewer than those of ResNet-18. Fig. 6(b) illustrates the inference time. The average inference time of the proposed model is 10.5ms, faster than that of ResNet-18 and EfficientNet-B0. These results suggest that the proposed environmental sound classification model balances being lightweight and efficient. The rectangular convolution kernel reduces computational complexity in the time dimension by using wide convolution in the frequency domain, while the self-distillation architecture avoids the extra overhead of traditional distillation that requires a teacher model by transferring knowledge within a single model.

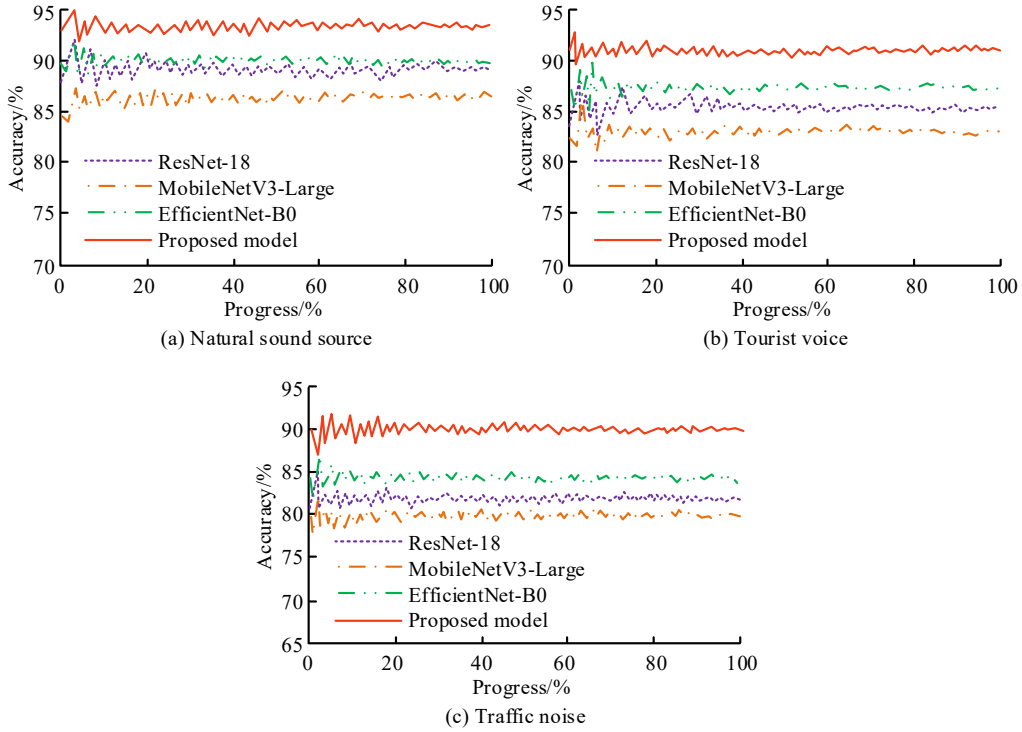


Fig. 5. Classification accuracy of different models on the test set

The robustness tests at different Signal-to-Noise Ratio (SNR) levels are shown in Table 1. The accuracy of each model increased as SNR (noise reduction) improved, and the proposed model maintained optimal performance across the entire SNR range. From 5dB to 40dB, ResNet-18's accuracy increased by 17.3%, MobileNetV3-Large by 18.4%, and EfficientNet-B0 by 17.5%, while the proposed model increased by 12.6%, showing the smallest gain and indicating lower sensitivity to noise. At 40dB, each model's accuracy approached that of clean data, confirming that sound feature recognizability improves in high SNR environments, and the advantages of the proposed model remain significant in these settings.

3.2. Prediction Results of the Acoustic Landscape Perception Model

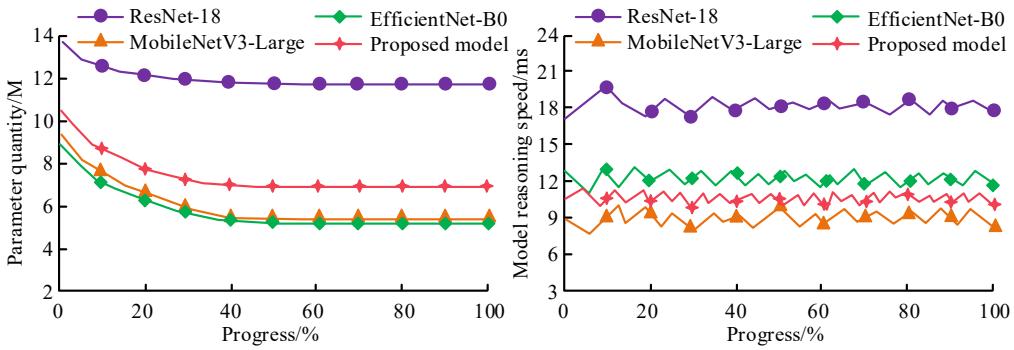
To validate the performance of the proposed acoustic landscape perception model, the experiment compares it with a baseline model based on acoustic features. The differences in perceptual rating prediction errors among various models are shown in Table 2. After integrating acoustic features, semantic labels, and environmental variables, the proposed model improved R2 by approximately 9% in predicting pleasure and tranquility. The attention mechanism emphasizes key features. The contribution weight of the sound of flowing water in summer afternoons to pleasure was 0.32, significantly higher than similar features. The naturalness and infection dimensions demonstrate the benefits of combining features and attention mechanisms. Compared to the baseline, the naturalness RMSE of the proposed model decreased from 0.80 to 0.63, with an R2 of 0.89, while the infection RMSE decreased from 0.85 to 0.68, with an R2 of 0.86.

Fig. 7 illustrates the differences in predicting day and night scenes. Fig. 7(a) shows the prediction results for daytime scenes, with a true mean interference degree of 3.21 and a predicted mean of 3.18, resulting in an error rate of approximately 0.93%. The actual mean naturalness degree was 4.15, with a predicted mean of 4.09 and an error rate of about 1.45%. The predicted values of both indicators closely matched the true values, with small errors. The model demonstrates high prediction accuracy for daytime soundscapes, and the features of the daytime sound environment are relatively stable and recognizable. Fig. 7(b) presents the prediction results for night scenes, with a true mean interference degree of 2.17 and a predicted mean of 2.22, with an error rate of around 2.30%, which is higher than that of daytime. The actual mean naturalness was 4.52, while the predicted mean was 4.46, with an error rate of approximately 1.33%, still relatively low. The predicted naturalness in night scenes was better due to the increased presence of natural sound sources and simpler

features, although the interference error was slightly higher.

Table 1. Model robustness testing at different signal-to-noise ratio levels

SNR	ResNet-18 accuracy (%)	MobileNetV3-Large accuracy (%)	EfficientNet-B0 accuracy (%)	Proposed model accuracy (%)
5	68.3	65.0	69.8	78.9
10	74.5	71.2	76.1	83.5
15	77.9	75.1	79.4	85.9
20	81.2	78.9	82.7	88.3
25	82.3	80.0	83.9	89.1
30	83.4	81.1	85.1	89.9
35	84.5	82.2	86.3	90.7
40	85.6	83.4	87.3	91.5



(a) Model parameter quantity (b) Model reasoning speed

Fig. 6. Comparison of model parameter quantity and reasoning tim

Table 2. Comparison of perception rating prediction errors of different models

Degree	Indicators	Baseline model	Proposed model
Pleasure	RMSE	0.82	0.65
	MAE	0.68	0.53
	R2	0.79	0.88
Tranquility	RMSE	0.79	0.61
	MAE	0.65	0.49
	R2	0.81	0.90
Naturalness	RMSE	0.80	0.63
	MAE	0.66	0.51
	R2	0.80	0.89
Infection	RMSE	0.85	0.68
	MAE	0.70	0.55
	R2	0.77	0.86

Fig. 8 presents the preference level classification results. Fig. 8(a) presents the classification accuracy. The accuracy of the proposed model was 86.2%, which was higher than that of the baseline model (78.5%). Fig. 8(b) shows the F1 value, with the proposed model and baseline model having F1 values of 87.2% and 76.3%, respectively. The results indicate that

the proposed acoustic landscape perception model has improved classification confidence of sensitivity to preference level features.

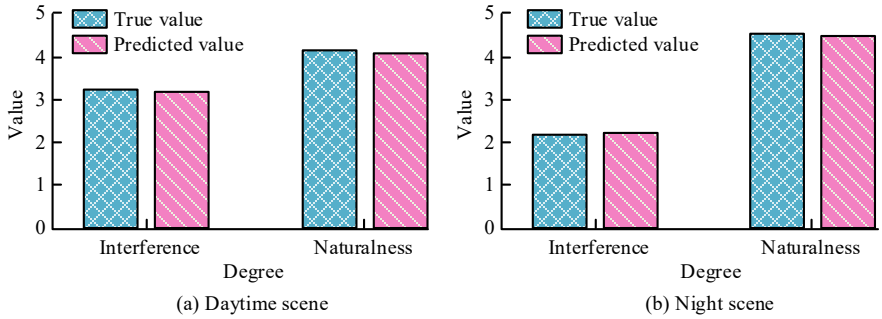


Fig. 7. Prediction difference between daytime and night scenes

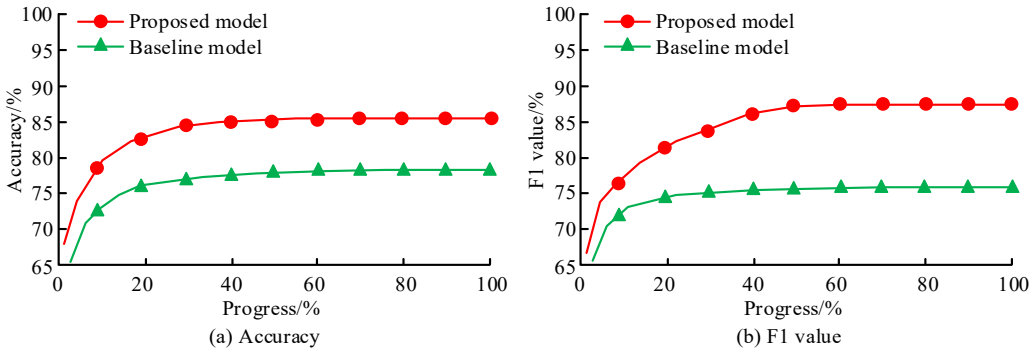


Fig. 8. Preference level classification results

3.3. Analysis of Perception Differences in Different Sound Environment Regions

Fig. 9 displays the perceptual scores and sound source contribution rates for each region. Part (a) of Fig. 9 shows the average perceived score of each region. The four indicators in the natural sound source area indicated high pleasure, naturalness, and tranquility, with low interference, aligning with the strengths of the original ecological soundscape. The highest interference level was observed in the traffic impact area, reaching 4.39, mainly due to high-frequency engine noise exceeding auditory comfort thresholds. The naturalness in tourist-heavy areas was the lowest, at 3.22, reflecting how human voices mask natural sounds. Part (b) presents the contribution rates of regional sound sources. In the natural sound source area, flowing water (42%) and bird songs (35%) were the main contributors. Human voices in tourist-heavy areas made up nearly 60%, mostly from conversations (45%) and camera shutter sounds (13%). Engine noise in the traffic impact area accounted for 58% of total noise and showed a strong positive correlation with perceived interference.

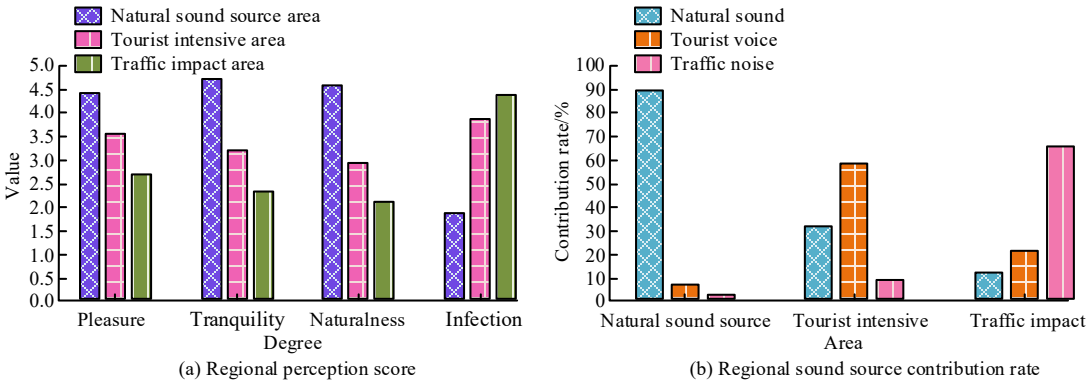


Fig. 9. Perceived scores and sound source contribution rates for each region

Fig. 10 shows the spatiotemporal differences among various regions. Fig. 10(a) displays the pleasure scores of different regions across seasons. The pleasure derived from natural sound sources was higher in spring and summer than in autumn and winter, aligning with seasonal fluctuations in bioacoustic activities such as bird and cicada calls. Tourist-heavy areas

had slightly higher scores in spring, while traffic-affected areas recorded the lowest scores in winter, due to extended engine idle times and increased noise from low temperatures. Fig. 10(b) illustrates the tranquility scores of different regions during day and night. The tranquility in natural sound source areas at night was 15.2% higher than during the day because low-frequency natural sounds like wind and streams become more prominent when human activities cease. The nighttime score in tourist-heavy areas increased by 22.4%, but remained lower than that of natural sound source areas, indicating residual human noises, such as camping conversations, continue to cause interference. Nighttime noise in the traffic impact area mainly originated from duty vehicles, and the improvement in tranquility was limited, approximately 23.8%.

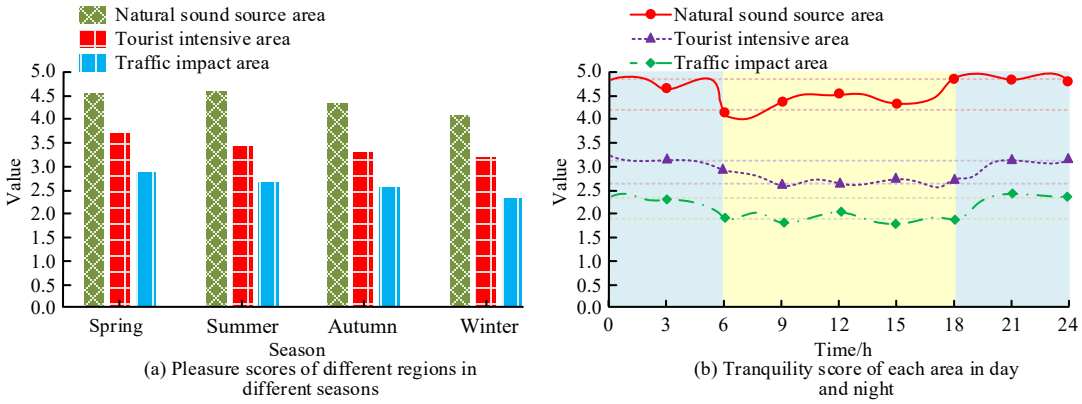


Fig. 10. Comparison of spatiotemporal differences among different regions

To comprehensively evaluate the perceptual characteristics of each acoustic environment region, Table 3 shows the statistical test results of different regions in four perceptual dimensions, with each experiment repeated 10 times. It can be seen that the natural sound source area performs the best in terms of pleasure, tranquility, and naturalness, with a small standard deviation, indicating a high consistency in tourist's perception of the area. The traffic impact zone has the highest score in terms of interference degree, with a small standard deviation, indicating that its noise interference is universal. The scores of various aspects in tourist intensive areas are in the middle, but the score in naturalness is the lowest, reflecting the significant masking effect of human voices on natural soundscapes.

Table 3. Statistical test results across four perceptual dimensions in different regions

Degree	Natural sound source zone	Areas with high tourist density	Traffic impact zone
Pleasure	4.52±0.31	3.45±0.38	2.98±0.47
Tranquility	4.48±0.29	3.21±0.41	2.75±0.44
Naturalness	4.61±0.25	3.22±0.39	2.81±0.46
Infection	1.89± 0.42	3.67± 0.51	4.39±0.38

4. Discussion

The findings of this study offer concrete and actionable insights for optimizing soundscapes, spatial planning, and visitor experience management in Zhangjiajie National Forest Park and similar natural scenic areas. The perceived soundscape model serves as a decision-support tool, enabling managers to move beyond simple noise level monitoring and towards a nuanced understanding of how sounds influence visitor perception and behavior. For instance, the results clearly advocate for a soundscape-based zoning strategy. Areas identified as natural sound sources, characterized by high pleasure, naturalness, and tranquility, should be designated as acoustic preservation zones where strict limits on human-generated noise can protect this critical ecological and experiential resource. In visitor-heavy zones, where human voices significantly reduce the naturalness of the environment, management efforts should focus on mitigation through landscape design that provides acoustic buffering or by implementing visitor flow control systems. The traffic impact zones, exhibiting the highest levels of interference, require targeted interventions at the noise source, such as transitioning to electric shuttle vehicles and routing traffic away from core scenic areas.

The identified spatiotemporal patterns of soundscapes provide a scientific basis for optimizing visitor experience across different times and seasons. The higher pleasure scores in natural sound source areas during spring and summer present an opportunity to promote seasonal ecotourism activities centered on "sound listening." The significant increase in tranquility at night in these areas underscores the value of developing low-impact night tourism, such as stargazing tours, while simultaneously requiring strict management of artificial light and camping noise to preserve the fragile nocturnal sound environment. Beyond spatial and temporal planning, this study connects soundscapes directly to visitor behavior and psychological experience. The correlation between high pleasure/tranquility scores, longer dwell times, and higher photo-

taking frequency in natural sound areas suggests that positive soundscapes enhance visitor immersion and satisfaction. Conversely, the high interference in traffic zones is linked to visitor stress and a potential reduction in visit duration, highlighting the tangible experiential cost of noise pollution. By leveraging these insights, park managers can make evidence-based decisions to enhance restorative experiences for visitors while safeguarding the natural acoustic environment, ultimately fostering the sustainable development of ecotourism.

The methods and results of this study have significant portability and provide a scalable framework for acoustic landscape management in global natural tourism destinations. From national parks in North America and Europe to forest reserves in Southeast Asia and heritage sites in Africa, the core challenge of balancing ecological conservation and visitor experience is widespread in protected areas around the world. The proposed two-stage modeling method, which adopts a robust environmental sound classification model followed by a perception model integrating acoustic, semantic, and contextual features, can adapt to any geographical location. The key is to recalibrate the model's input to reflect local sound sources (for example, replacing specific bird species or unique water features) and cultural nuances in sound perception, which can be achieved through localized data collection. For instance, the sound of cicadas may be regarded as natural and pleasant in one culture, but as noise in another. The established strategic principles, such as soundscape zoning, temporary visitor management, and targeted reduction of major noise sources, provide a multi-functional toolkit. Park managers around the world can apply this structured approach to diagnose their unique acoustic environment, identify conflict or opportunity areas, and implement evidence-based intervention measures to protect valuable natural acoustic landscapes and improve the quality of the visitor experience, thereby contributing to global sustainable tourism development.

5. Conclusion

A soundscape perception model based on acoustic modeling and environmental sound analysis was developed for Zhangjiajie National Forest Park. First, sound data was collected with a microphone array, combined with tourist Likert 5-point scale ratings, and acoustic features were extracted. A self-distillation convolutional neural network was designed for environmental sound classification, along with an attention mechanism perception model. The proposed classification model achieved an accuracy rate of 93. 5%, 88.3. 3%, and 91.5. 5% in identifying natural sound sources, tourist voices, and traffic noise, respectively, outperforming comparison models like ResNet-18. Additionally, it maintained 78. 9% accuracy at a low SNR of 5dB, demonstrating lightweight and high efficiency. By integrating multiple data sources, the sound landscape perception model achieved R^2 values of 0.88, 0.90, 0.89, and 0.86 for predicting pleasure, tranquility, naturalness, and infection, respectively, surpassing the baseline model by approximately 9%. In day and night settings, the prediction error rate for naturalness at night was 1. 33%, while the interference error rate was 2. 30%, higher than during the day. The pleasure level in natural sound areas during spring and summer was 12%- 15% higher than in autumn and winter, and nighttime tranquility was 15. 2%- 23. 8% higher than during the day. The study reveals that natural sound sources significantly enhance positive perceptions, whereas human noise primarily causes interference. Attention mechanisms effectively focus on key features. This model provides a quantitative tool for optimizing scenic acoustic landscapes, helping balance ecological protection with tourist experience. However, the study does not include special weather soundscape data, and there is limited semantic analysis of low-frequency natural sounds. Future research could expand multimodal data fusion to model the emotional aspects of soundscapes, offering a more comprehensive scientific basis for sound environment management in smart scenic areas.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Qing Sun, Qianni Cheng, Jianlin Tian, and Yanan Zhou. The first draft of the manuscript was written by Qing Sun. Manuscript review and revisions were performed by Qianni Cheng, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The authors confirm that no AI tools were used in the preparation of this manuscript.

References

Alghamdi, N. S., Zakariah, M., and Karamti, H. (2024). A deep CNN-based acoustic model for the identification of lung diseases utilizing extracted MFCC features from respiratory sounds. *Multimedia Tools and Applications*, 83(35), 82871–82903. doi: 10.1007/s11042-024-18703-0

Dua, M., Akanksha, and Dua, S. (2023). Noise robust automatic speech recognition: Review and analysis. *International Journal of Speech Technology*, 26(2), 475–519. doi: 10.1007/s10772-023-10033-0

Gao, C., Xu, J., Pang, F., Li, H., and Wang, K. (2024). Modeling and experiments on the vibro-acoustic analysis of ring stiffened cylindrical shells with internal bulkheads: A comparative study. *Engineering Analysis with Boundary Elements*, 162, 239–257. doi: 10.1016/j.enganabound.2024.02.007

Haghighi, M., Mirzaei, R., Putra, A., and Taban, E. (2024). A comprehensive review of advances and techniques in muffler acoustics and design. *International Journal of Environmental Science and Technology*, 21(13), 8695–8716. doi: 10.1007/s13762-024-05686-6

Hossain, M. R., Manohare, M., and King, E. A. (2025). Systematic review of indoor soundscape assessments: Activity-based psycho-acoustics analysis. *Building Acoustics*, 32(1), 123–141. doi: 10.1177/1351010X241294151

Iyendo, T. O., Welch, D., and Uwajeh, P. C. (2024). Soundscape and natural landscape as a design construct for improving psycho-physiological health in cities: A semi-systematic literature review. *Cities & Health*, 8(3), 447–485. doi: 10.1080/23748834.2023.2280288

Jadoul, Y., De Boer, B., and Ravignani, A. (2024). Parselmouth for bioacoustics: Automated acoustic analysis in Python. *Bioacoustics*, 33(1), 1–19. doi: 10.1080/09524622.2023.2259327

Kaidouchi, H., Kebdani, S., and Slimane, S. A. (2023). Vibro-acoustic analysis of the sound transmission through aerospace composite structures. *Mechanics of Advanced Materials and Structures*, 30(19), 3912–3922. doi: 10.1080/15376494.2022.2085348

Li, M., Gao, M., Zhang, M., Chen, J., Dai, P., Li, R., and Wang, Y. (2025a). Analysis of soundscape perception in different types of parks using structural equation modeling (SEM). *Journal of Resources and Ecology*, 16(3), 875–885. doi: 10.5814/j.issn.1674-764x.2025.03.023

Li, W. and Liu, Y. (2024b). The influence of visual and auditory environments in parks on visitors' landscape preference, emotional state, and perceived restorativeness. *Humanities and Social Sciences Communications*, 11(1), 1–18. doi: 10.1057/s41599-024-04064-4

Liu, L., Sun, B., Li, J., Ma, R., Li, G., and Zhang, L. (2024). Time-frequency analysis and recognition for UAVs based on acoustic signals collected by low-frequency acoustic-electric sensor. *IEEE Sensors Journal*, 24(12), 19601–19613. doi: 10.1109/JSEN.2024.3397163

Moufid, I., Roncen, R., Matignon, D., and Girier, J. (2024). Time-domain simulation of the acoustic nonlinear response of acoustic liners at high sound pressure level. *Nonlinear Dynamics*, 112(5), 3669 – 3698. doi:10.1007/s11071-023-09219-7

Peng, B., Gao, D., Wang, M., and Zhang, Y. (2024). 3D-STCNN: Spatiotemporal convolutional neural network based on EEG 3D features for detecting driving fatigue. *Journal of Data Science and Intelligent Systems*, 2(1), 1–13. doi: 10.47852/bonviewJDSIS3202983

Raevskii, M. A., and Burdukovskaya, V. G. (2024). The Combined Influence of Wind Waves and Internal Waves on the Coherence of Low-Frequency Acoustic Signals and the Efficiency of Their Spatial Processing in Shallow Water. Acoustical Physics, 70(4), 705 – 717. doi: 10.1134/S1063771024601547

Razani, F., and Sharghi, A. (2025). Landscape perception based on auditory processing (An investigation in the role of sound in landscape reading). *MANZAR, the Scientific Journal of Landscape*, 17(71), 34 – 43. doi: 10.22034/MANZAR.2025.481255.2315

Shao, Y., Lu, J., Zhang, Z., and Jin, T. (2024). Research on the influence mechanism of park landscape characteristics on psychological perception around expressway based on soundscape theory: A case study of Chengdu Bailuwan Wetland Park. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 270(8), 3244–3254. doi: 10.3397/IN_2024_3298

Tokaç, I., Heimes, A., Vorländer, M., and Brell-Çokcan, S. (2025). A rule-based framework for capturing geometric characteristics in design: A study of façade analysis for acoustic behaviour in urban space. International Journal of Architectural Computing, 23(2), 405–425. doi: 10.1177/14780771241260852

Wrege, P. H., Bambi, F. B. D., Malonga, P. J. F., Samba, O. J., and Brncic, T. (2024). Early detection of human impacts using acoustic monitoring: An example with forest elephants. *PLOS ONE*, 19(7), e0306932. doi: 10.1371/journal.pone.0306932

Yoon, B., Kim, J., Kang, C., Oh, M. K., Hong, U., and Suhr, J. (2023). Experimental and numerical investigation on the effect of material models of tire tread composites in rolling tire noise via coupled acoustic-structural finite element analysis. *Advanced Composite Materials*, 32(4), 501–518. doi: 10.1080/09243046.2022.2119832

Zhang, D., Su, X., Sun, Y., Chen, C., and Sun, X. (2024c). Mechanism analysis and experiment study for wire mesh-assisted ventilated acoustic metamaterials based on the acoustic analytical model and numerical acoustic-flow coupling model. *Journal of Vibration Engineering Technologies*, 12(4), 6649–6663. doi: 10.1007/s42417-024-01276-5

Zhang, X., Ren, X., Yang, L., Song, G., Wen, D., and Shi, G. (2024a). Acoustic–vibration characteristics and low–structural–noise optimization design of upright sound barrier for rail transit. *International Journal of Rail Transportation*, 12(6), 1020–1038. doi: 10.1080/23248378.2023.2301600

Zhang, Y., Xia, T., Han, J., Wu, Y., Rizos, G., Liu, Y., and Mascolo, C. (2024b). Towards open respiratory acoustic foundation models: Pretraining and benchmarking. *Advances in Neural Information Processing Systems*, 37(1), 27024–27055. doi: 10.17863/CAM.113972

Qing Sun was born in Zhangjiajie, Hunan. P.R. China, in 1986. He obtained a bachelor's degree from Central South University of Forestry and Technology in China. He is currently working at Jishou University. His research interests include Sustainable landscape architecture and landscape perception.

Qianni Cheng was born in Zhangjiajie, Hunan. P.R. China, in 1988. He obtained a bachelor's degree from Jishou University in China. He is currently working at Zhangjiajie College. His research interests include wisdom landscape and scenic area planning.

Jianlin Tian was born in Chaling, Hunan, P.R. China, in 1976. He obtained a Ph.D. in Science from Sun Yat Sen University in China. He currently works at Jishou University. His primary research focus is on environmentally remote sensing and urban-rural landscape planning.

Yanan Zhou was born in Fuzhou, Jiangxi, P.R. China, in 1995. She obtained a bachelor's degree from Yichun University in China. She is currently working at Jishou University. Her research interests are in landscape planning and design.