mathematics

Article

Exploring an Intelligent Classification Model for the Recognition of Automobile Sounds Based on EEG Physiological Signals

Jingjing Guo [1], Tao Xu [2], Liping Xie [2,*] and Zhien Liu [2,*]

[1] Wuhan Vocational College of Software and Engineering (Wuhan Open University), Wuhan 430205, China; guojingjng@163.com

[2] Hubei Collaborative Location Center for Automotive: Components Technology, Wuhan University of Technology, Wuhan 430070, China; 293016@whut.edu.cn

[*] Correspondence: lpxie@whut.edu.cn (L.X.); lzen@whut.edu.cn (Z.L.)

Abstract: The advancement of an intelligent automobile sound switching system has the potential to elevate the market standing of automotive products, with the pivotal prerequisite being the selection of automobile sounds based on the driver's subjective perception. The subjective responses of diverse individuals to sounds can be objectively manifested through EEG signals. Therefore, EEG signals are employed herein to attain the recognition of automobile sounds. A subjective evaluation and EEG signal acquisition experiment are designed involving the stimulation of three distinct types of automobile sounds, namely comfort, power, and technology sounds, and a comprehensive database of EEG signals corresponding to these three sound qualities is established. Then, a specific transfer learning model based on a convolutional neural network (STL-CNN) is formulated, where the method of training the upper layer parameters with the fixed bottom weights is proposed to adaptively extract the EEG features related to automobile sounds. These improvements contribute to improving the generalization ability of the model and realizing the recognition of automobile sounds fused with EEG signals. The results of the comparison with traditional support vector machine (SVM) and convolutional neural network (CNN) models demonstrate that the accuracy of the test set of the STL-CNN model reaches 91.5%. Moreover, its comprehensive performance, coupled with the ability to adapt to individual differences, surpasses that of both SVM and CNN models. The demonstrated method in the recognition of automobile sounds based on EEG signals is of significance for the future implementation of switching driving sound modes fused with EEG signals.

Keywords: automobile sound; recognition; specific learning model; convolutional neural network; EEG signal

MSC: 37M99

Citation: Guo, J.; Xu, T.; Xie, L.; Liu, Z. Exploring an Intelligent Classification Model for the Recognition of Automobile Sounds Based on EEG Physiological Signals. *Mathematics* **2024**, *12*, 1297. https://doi.org/10.3390/math12091297

Academic Editor: Georgios Tsekouras

Received: 8 March 2024
Revised: 8 April 2024
Accepted: 19 April 2024
Published: 25 April 2024

1. Introduction

The intelligent sound mode control system for automobiles is designed to enhance driving satisfaction by intelligently controlling the sound modes in accordance with the driving requisites of both the driver and passengers. In order to achieve the above automotive functions, it becomes imperative to construct a classification recognition algorithm to achieve the switching of pre-designed sounds. Furthermore, a foundational prerequisite for achieving precise sound mode switching involves an investigation into the subjective evaluations provided by evaluators. Nevertheless, it has been reported that the evaluator's subjective evaluation of the sound quality is still perceived based on the physical and psychological acoustic indicators [1–4]. Traditional subjective evaluations face challenges in accurately capturing the genuine sentiments of evaluators when dealing with sounds characterized by intricate semantics, exemplified by terms like "comfort" [5], "powerful" [6], and "luxury" [7] emanating from sound design. Therefore, it is imperative to introduce a novel mechanism for the recognition of sound quality characterized by complex semantics.

Recent studies have suggested that EEG signals are a physiological indicator that contributes to reflecting individual variability and objectively mapping the intuitive perception of different evaluators for sound [8,9]. There are existing studies that have proved that EEG signals can be utilized to evaluate automobile sound quality [10,11]. In addition, different drivers have different demands for the sound quality of different automobiles, and the design of switching modes with multiple types of driving sounds helps to enhance the brand image and competitiveness of automobiles [12]. Therefore, EEG signals as a physio-acoustic indicator to realize the classification of diverse types of vehicle sound qualities are of great significance for the system development of switching between different driving sound modes. EEG signals are of importance in applying ergonomics/human factors to the development of automobiles.

In this paper, three types of automobile sound qualities, namely the comfort, power, and technology sound quality, are focused on, and an intelligent classification model (STL-CNN) that combines the convolutional neural network (CNN) and the individual-specific transfer learning model is constructed to realize the above three types of automobile sounds based on EEG signals. This effort will provide a research basis for the future realization of switching between driving sound modes fused with EEG signals.

1.1. Related Work

The brain is the most complex component of the central nervous system in the human body and is the basis for the control of human activities [13]. The EEG signal is a relatively objective physiological indicator, and it has gradually evolved from clinical research to the field of brain–computer interfaces [14]. In terms of research on auditory evoked EEG signals, in 2007, Lenz D et al. [15] analyzed changes in γ waves under familiar and unfamiliar audio stimuli and found that audio with long-term memory was subjectively recognized more quickly due to the subject's prior cognition. In 2013, Cong F et al. [16] investigated the correlation between EEG features and audio features, and an independent component clustering analysis was used to reveal that α waves in the occipital region and θ waves in the parietal region were significantly correlated with audio features. They also demonstrated that the fluctuating center of audio was the most effective feature for evoking EEG signals. In 2014, Li Z G et al. [17] investigated the relationship between EEG signals and subjective annoyance. Transient and steady-state stimulation comparison experiments were set up using pure tones of different frequencies as the stimulus, and they found that the sum of the average power of θ and α waves could be used to assess noise-induced subjective annoyance. In 2021, Zhang R et al. [18] proposed a "brain-ID" framework based on a hybrid deep neural network with transfer learning (HDNN-TL) to deal with individual differences in a four-class task. In 2022, Xie L et al. [19] focused on the evaluation of sound quality, where the evaluation method was equated to a 10-classification problem to achieve the score prediction of acceleration sound quality fused with EEG signals.

A summary of the above relevant studies demonstrates that the EEG signals have some certain regularity under different sound stimuli, and the qualitative relationship between different audio stimuli and EEG regularity can be obtained by quantitatively calculating the statistical characteristics of the EEG signal. Therefore, it is plausible to explore research on the identification of automobile sounds based on EEG signals through a rational design of brain-evoked tests.

1.2. Critical Issues

In recent years, the research on the sound-induced EEG laws has been deepened. EEG signals have the characteristics of a high temporal resolution; thus, the processing of EEG signals is a challenging research task [20]. The analysis methods of EEG signals mainly include EEG signal pre-processing, feature extraction and selection of EEG signals, and construction of classification models [21]. The pre-processing of EEG signals mainly involves the removal of irrelevant noise and the optimization of the data format after acquisition of the raw EEG signal [22]. The EEG signal features are mainly extracted from time domain, frequency domain, and entropy features. Most EEG devices acquire EEG time–domain signals, and the time–domain features are most intuitive and easy to obtain, including skewness [23], kurtosis [24], first- and second-order differences [25], Hjorth parameters [26,27], etc. In addition, the brain can be divided into the frontal, temporal, parietal, and occipital lobes according to its fissures, which all serve different perceptual functions and govern different physiological states [19,28]. EEG signals in the frequency domain can be divided into five frequency bands, namely δ, θ, α, β, and γ. The frequency domain features are obtained mainly by converting the acquired time domain signals into the frequency domain, decomposing the frequency bands into five sub-frequency bands, and then calculating the frequency features from them [29,30]. Differential entropy is a common entropy feature [31]. Unfortunately, there is no uniform standard for the extraction and selection of EEG features.

On the other hand, the construction of a classification algorithm that contributes to extracting effective features from a large amount of EEG signal data is the key to realizing the recognition of vehicle sounds based on EEG features. EEG signals are used in a variety of fields, such as emotion recognition, emotion computing, and audio recognition. In terms of classification algorithms, mathematical models such as support vector machines (SVMs) [32], linear discriminant analysis [33], and neural networks [34,35] have relatively good classification effects on the validation set. Nevertheless, the generalization ability is poor on the whole, and the individual differences between multiple subjects are often overlooked. Simultaneously, in the construction of data sets based on EEG signals, the "dimensional disaster" is often encountered for the high-dimensional EEG feature vectors in pattern classification research, which easy results in the data redundancy in the feature matrix, increases the computational load of the computer, and decreases the accuracy of the model's identification [21]. To achieve this, a transfer learning model based on a CNN model is built to extract potential features related to automobile sound from massive EEG features. We aim to construct an intelligent classification model for automobile sound to enhance the current level of accuracy and the generalization by combining deep learning and EEG signals.

1.3. Our Contribution

In this investigation, the main aim is to identify the types of automobile sounds based on EEG signals using a constructed intelligent classification model. The feasibility of mapping human subjective perception of vehicle sound based on EEG signals will be further explored in this paper. This exploration will lay the groundwork for future research on switching between different driving sound modes using EEG signals and provide guidance for data analysis methods that fuse EEG signals to evaluate sound.

The remaining structure of this paper is organized as follows: the experimental procedure for EEG signal acquisition is described in Section 2. The next section presents the principles of the construction of the STL-CNN model proposed in this paper, and the performance of the constructed STL-CNN model is evaluated to validate the recognition effect of automobile sounds fused with EEG signals in Section 4. In the next section, the results, methodological innovations, shortcomings, and future research directions are discussed, and a summary of the main findings are described in Section 6.

2. Data Collection and Experimental Setup

2.1. Sound Stimulus

The sound test is organized to acquire the acceleration sounds in the internal combustion engine vehicle that meet the requirements of sound quality. Equally, the automobile acceleration sounds with a good quality are downloaded from internet links and game software. The above acquired sounds are gathered to form a sound sample library with a total of 50 sounds. In this paper, three sound qualities of comfort, power, and technology are chosen as the subjective evaluation indicators of abstract semantics, and 22 evaluators with acoustics experience who are automobile engineers, teachers, or PhD students are recruited to evaluate the above three sound qualities of 50 selected sounds using the method of a rating scale.

The scores of the above three sound qualities are calculated for each sound after the evaluators have completed their evaluation, and the top three sounds of each sound quality type are screened. A total of 9 sounds are used as the sound stimulus for the next EEG test. The semantic descriptions of the three sound qualities of comfort, power, and technology and information regarding the 9 sound stimuli are shown in Table 1.

Table 1. Information regarding sound stimuli.

Types	Semantic Description	Sound Source
Comfort	Smooth acceleration, noiseless, soft, and comfortable sounds	Driver's right ear, sound of Peugeot 4008 4th gear Driver's right ear, sound of Golf 5th gear Driver's right ear, sound of Peugeot 4008 5th gear
Power	Thick sound, strong acceleration, no metallic clatter sounds	Driver's right ear, sound of Audio R8 3th gear Driver's right ear, sound of Audio R8 4th gear Engine sound of Peugeot 4008 3rd gear
Technology	High acceleration frequency, rapid sounds, science fiction feels	Web resource 1 Web resource 2 Web resource 3

2.2. EEG Experimental Setup

2.2.1. EEG Data Acquisition

15 subjects are recruited to participate in the EEG experiment, including 12 males and 3 females. The information regarding the subjects is shown in Table 2. All the subjects are right-handed, have normal vision and hearing, and are free of brain disorders such as epilepsy. Adequate sleep and a clean scalp are required to ensure a good signal-to-noise ratio during the procedure of EEG data acquisition.

Table 2. Information regarding the subjects.

Subject Characteristics		Quantity	Age	
			Mean	Standard Deviation
Gender	Male	12	24.81	5.32
	Female	3	23.2	4.3
Occupation	postgraduate	10	20.0	0
	PhD student	3	24.0	2.42
	Professor	2	43.5	1.52

The EEG signal is a physiological signal that is susceptible to external interference; thus, an environment with a low background noise must be ensured during the process of the EEG signal acquisition. The ActiCHamp EEG signal acquisition amplifier developed by Brain Products in Gilching, Germany is used as the EEG signal acquisition equipment, as shown in Figure 1a, and the data acquisition system is connected directly to the computer via a USB for real-time data transfer. To ensure sufficient spatial resolution of the EEG signal, an Ag/Agcl EEG cap (as shown in Figure 1b) that has 64 electrodes is utilized to synchronously acquire the EEG signal. The electrode arrangement meets the international lead 10–20 standard [36].

The EEG evoked experiment is designed using the E-Prime 2.0 software, and the 9 sound stimuli are stored in the audio playback module of the E-Prime software. The software is run by the data communication host during the experiment. The display text prompts the corresponding operation process information, and the subject completes the experimental process based on the presented information while their EEG signals are acquired. The transmission and storage of EEG data are completed in real time by the EEG acquisition equipment. The EEG acquisition process is shown in Figure 2.

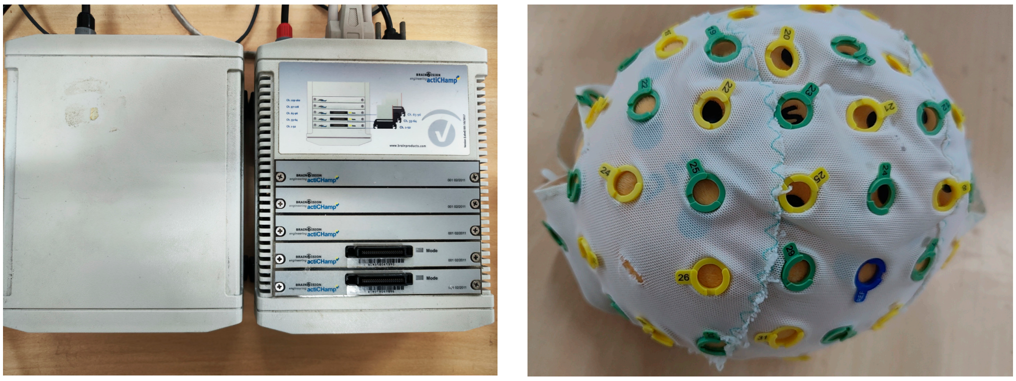

(**a**) EEG signal acquisition amplifier (**b**) 64-channel electrode cap

Figure 1. EEG acquisition equipment.

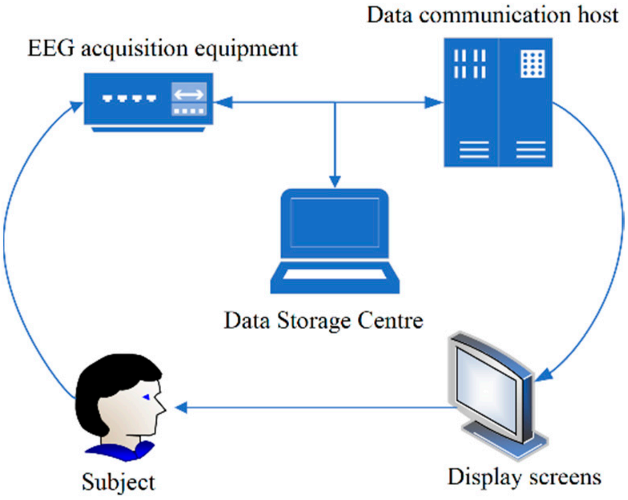


Figure 2. EEG acquisition process.

The entire EEG experiment is divided into three groups of experiments. The first group of experiments consists of three sounds with a comforting sound quality, and the subjects control the playback of the sounds using a designated button. The subjects take a short break after one sound is played, then the next sounds are presented using the designated button based on the text prompted in the display. Three comfort sounds are played randomly, and each sound is repeated 27 times. There are a total of 81 sound stimuli in each group of trials for each subject. Then, the second group of experiments employing three powerful sounds is initiated after a short break based on the state of the subjects. The procedure is the same as for the previous group. The EEG experiment is complete after the third group of experiments when the three technology sounds have been cyclically played.

A total of 243 (81 × 3) sets of EEG signals are obtained throughout the entire duration of the EEG experiment with the three types of sound stimuli. The duration of each sound is 5s, and there are 27 repetitions for each sound stimulus. The entire experimental procedure theoretically lasts at least 20 min for each subject, excluding the stage of wearing the EEG equipment and preparation before the formal trial. The EEG signal test scenario is shown in Figure 3.

2.2.2. EEG Data Preprocessing

In this paper, EEG data preprocessing is conducted using the EEGLAB V2019.1 toolkit on the MATLAB platform, which is primarily employed for EEG signal analysis. The toolkit supports multiple data formats, including .edf, .bdf, and .vhdr. A finite impulse response filter (FIR) is utilized in this study for EEG signal processing. Considering the high-frequency sensitivity of EEG signals to auditory stimulation, a bandpass filter of 0.1~100 Hz is applied. Simultaneously, to eliminate power frequency interference at 50 Hz and its harmonics, frequencies of 49~51 Hz and 99~101 Hz are attenuated. Additionally, the EEG signal is decomposed into independent components using independent principal component analysis in this experiment. The artifact signals are then removed based on feature screening of the ocular artifacts.

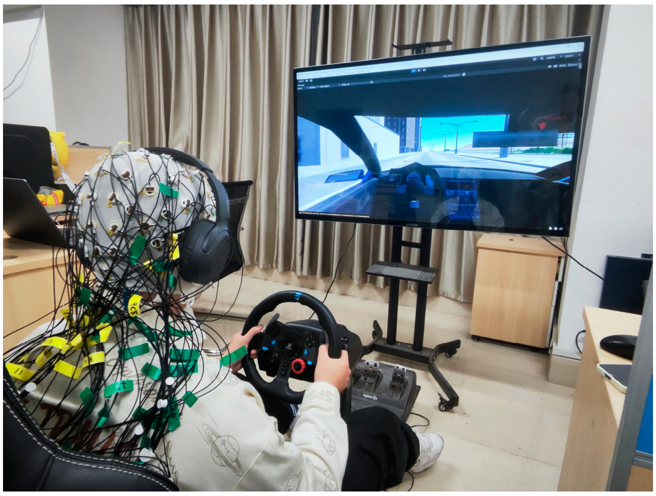


Figure 3. EEG signal test scenario.

3. A Method for the Recognition of Vehicle Sounds Fused with EEG Signals

The aim of the present research is to implement EEG-based recognition of automobile sounds when the individual variability of the subjects is synchronously considered. In order to solve the problem of a poor generalization ability on a new subject data that traditional classification models encounter, a transfer learning model based on a CNN model with deep learning theory is built.

3.1. Architecture of the CNN Model

The potential features associated with the target object can be automatically extracted from the raw EEG dataset by a CNN, which usually consists of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. The design and parameter selection of the CNN model constructed in our study will be expounded from two perspectives: the design of the feature extraction module and the network structure.

3.1.1. Design of the Feature Extraction Module

In the evoked EEG experiments described in Section 2, the sound stimulus is time-sensitive, while the EEG signal, as a physiological signal that reflects the subject's state, also shows some regularity in the temporal dimension. Thus, the EEG data are processed along the temporal dimension, which is beneficial for the extraction of potential EEG features associated with automobile sounds. The time-varying characteristics of the time series are considered, and EEG feature extraction is designed sequentially from the bottom layer and the upper layer.

Firstly, the bottom layer of the feature extraction module is designed as shown in Figure 4, where the dimension of the input EEG data is $n \times 5000$, n is the number of data samples, and 5000 is the number of data points for a 5s sound stimulus with a sampling rate of 1000 Hz. To expand the data sample, a 1s Hamming time window is used to segment the raw EEG data without overlap, and the dimension of each convolutional layer input data becomes $n \times 1000 \times 5$ after segmentation. The reconstructed EEG data is batch normalized after the non-linear transformation of the activation function and the pooling operation, and the first round of feature extraction is completed. The bottom layer of the feature extraction module is consisted with 5 identical modules.

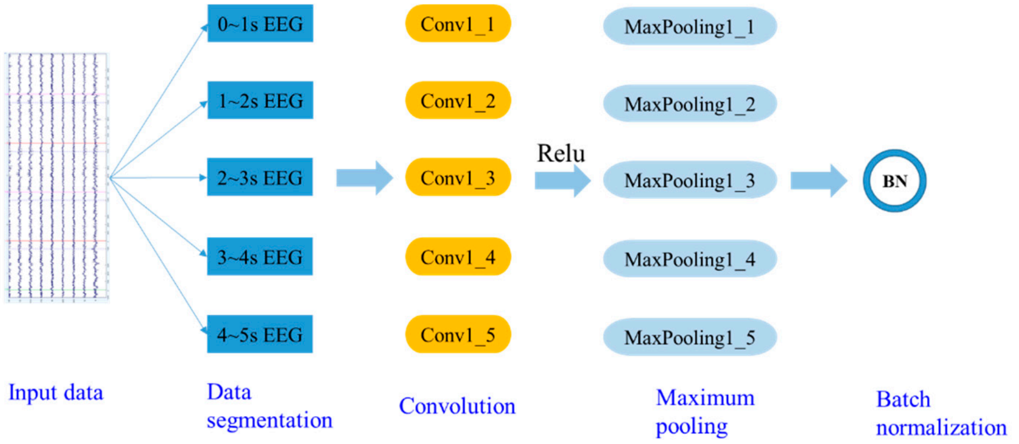


Figure 4. Schematic diagram of the bottom layer of feature extraction module.

The underlying low-dimensional features of the EEG data are extracted from the perspective of the time domain through the bottom layer of the feature extraction module, and the feature dimension is further increased by continuously expanding the convolutional kernel dimension. The upper layer of the feature extraction module is built to merge with the feature matrix extracted from the bottom layer after several iterations, as shown in Figure 5.

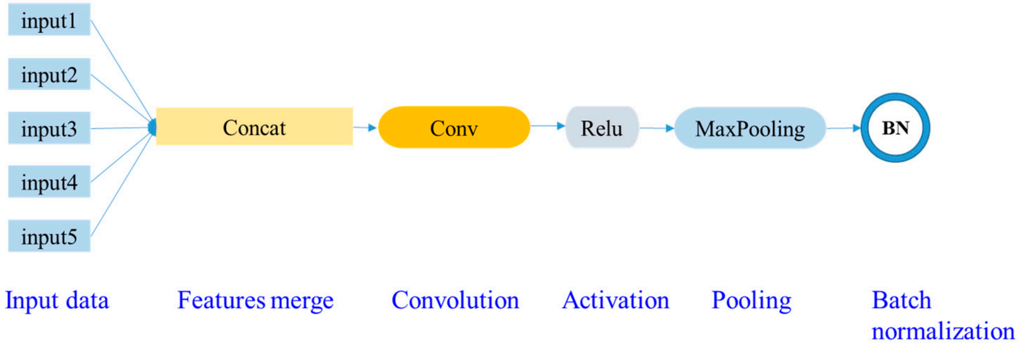


Figure 5. Schematic diagram of the upper layer of the feature extraction module.

The input data of the upper layer are the output feature matrix of the bottom layer of the feature extraction module, and 5 input feature matrices are merged in the upper layer to further increase the dimensionality of the feature matrix. The time-varying features of the EEG signals are extracted based on the merged feature matrix. In addition, due to the dimensional expansion from the time–domain superposition, there may be some redundancy in the output feature matrix. Thus, a suitable number and size of convolution kernels are selected to further convolve the merged feature matrix, and maximum pooling is used for down sampling after the action of the Relu activation function. The batch normalization is finally completed to achieve the feature extraction of the upper layer.

3.1.2. Design of Network Architecture

The feature extraction of the EEG data through the bottom layer and the upper layer of the feature extraction module is completed. Subsequently, the final feature matrix is integrated through the fully connected layer to achieve the output of the type of vehicle sound based on the EEG signals. The architecture of the CNN model designed in this paper is shown in Figure 6.

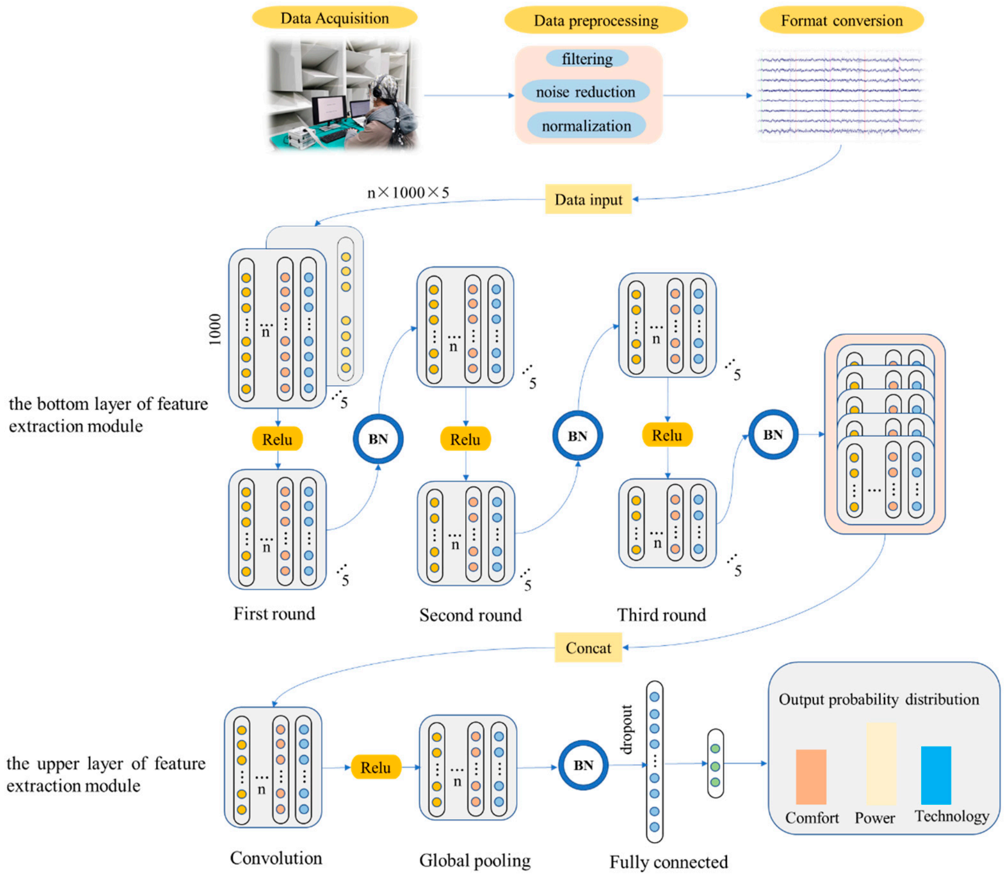


Figure 6. Architecture of the CNN model.

The construction of the CNN model is mainly divided into three stages, namely data preprocessing, feature extraction, and classification. The EEG data acquired in Section 2.2.1 are pre-processed using the method presented in Section 2.2.2. An EEG signal matrix with a dimension of 61 × 5000 × 243 × 15 is first constructed, where 61 represents the number of EEG channels, 5000 is the number of sample points for 5s data with a sampling rate of 1000 Hz, 243 is the total number of sounds played with each of the 9 sounds repeated 27 times, and 15 is the number of subjects. Then, to avoid the effects of the redundancy of the EEG channels, only the EEG data of the 6 channels related to auditory perception are selected according to related studies [37], including the T7, T8, F7, F8, P7, and P8 channels. The final EEG feature matrix for a single subject is 1458 (243 × 6) × 5000 dimensions, where 1458 represents the number of samples. Then, a 1s time window is used to segment the data into the dimension of 1458 × 1000 × 5.

Furthermore, 5 entrances of model input are created after completing the stage of data pre-processing, and the segmented EEG data are simultaneously input into the bottom layer of the feature extraction module for feature extraction by the internal convolution kernel. A total of 5 feature matrices with the time dimensions are merged after three rounds of iterations of the bottom layer of feature extraction module. Subsequently, the feature matrix extracted by the bottom layer enters the upper layer of the feature extraction module, and the strategy of global average pooling is used for down sampling by the convolution kernel designed for feature extraction. Then, the data are integrated in the fully connected layer, where some neurons are randomly deactivated based on the dropout method to reduce the risk of model overfitting. Finally, the final extracted features are fed into the Softmax layer to complete the label output of vehicle sounds, and the recognition of comfort, power, and technology sound qualities based on the EEG signals is achieved. The model parameters of the CNN architecture are shown in Table 3.

Table 3. Model parameters of the CNN architecture.

Model Component Name		Component Size	Number of Components	Output Dimension	Component Parameters
Input layer	Input data	\	5	1458 × 1000 × 1	0
Bottom layer	Convolutional layer 1_1	3 × 1	16	1458 × 1000 × 16	48 × 5
	Pooling layer 1_1	2 × 1	\	1458 × 500 × 32	0
	Convolutional layer 1_2	5 × 1	32	1458 × 500 × 32	2560 × 5
	Pooling layer 1_2	2 × 1	\	1458 × 250 × 32	0
	Convolutional layer 1_3	7 × 1	64	1458 × 250 × 32	14,336 × 5
	Pooling layer 1_3	2 × 1	\	1458 × 125 × 64	0
Feature merge		\		1458 × 125 × 320	0
Upper layer	Convolutional layer 2_1	3 × 1	128	1458 × 125 × 128	122,880
	Pooling layer 2_1	125 × 1	\	1458 × 128	0
Classification	Fully connected layer	\	10	1458 × 10	1290
	Softmax	\	3	1458 × 3	33

3.2. Developing the CNN with Specific Transfer Learning

In order to further improve the generalization ability of the CNN model, the idea of transfer learning is proposed to build an individual-specific transfer learning model based on parameter sharing, which is named specific transfer learning based on CNN (STL-CNN). The EEG data of 14 subjects are used as training samples to complete the construction of the CNN model, and the idea of transfer learning using the strategy of bottom weight sharing is used. The model is fine-tuned on the basis of the constructed CNN model to achieve the classification of vehicle sound based on the 15th subject, namely the target task.

As shown in the structure of the CNN model described in Section 3.1, it can be seen that the basic communal features of the subject's EEG signal are respectively extracted layer by layer from the time–domain dimension in the bottom layer of the feature extraction module. Then, the basic communal features are merged, and the feature dimension is increased in the upper layer of the feature extraction module to further complete the extraction of high-dimensional features.

The data of the new subjects are used as a new test sample. On the one hand, these data have the basic communal features of EEG signals, and on the other hand, the feature components of individual differences is also considered. Therefore, in this paper, the transfer learning model is designed, where the method of fixing the weights of the bottom layer of the feature extraction module and training the weights of the upper layer of the feature extraction module and the fully connected layer is proposed to build the STL-CNN model. The architecture of STL-CNN is shown in Figure 7.

3.3. Learning Rule of STL-CNN

The overall EEG data of the 15 subjects are divided into the training set and test set, and the 15th subject's data are used as the test set to test the stability of the constructed STL-CNN model on the new subject data. The EEG data of the remaining 14 subjects are selected as the training set to complete the parameter training of the STL-CNN model. The training is conducted in PyCharm 2021.3 on a mobile workstation equipped with an Intel i7-10875 CPU, Nvidia RTX 2060 GPU, and 16 GB of memory. The learning rule of STL-CNN includes mainly the forward propagation process and the backpropagation process.

There are three steps in the forward propagation process, namely the feature extraction of the bottom layer, the feature extraction of upper layer, and the full connection. Three rounds feature extraction of the convolutional pooling layer are performed sequentially in the bottom layer of the feature extraction module. Suppose X_t^l is the input matrix of the lth layer in the tth input channel, and the shared weight parameter matrix extracted is defined as W_t^l. Then, the result after convolution and activation of the lth layer is created by the following:

$$Y_t^l = \mathrm{Relu}(X_t^l W_t^l) \tag{1}$$

It can be given after passing the maximum pooling layer:

$$Y_t^l = \text{Max}(\alpha^l_{(t,i,j)}, \alpha^l_{(t,i,j+1)}) \tag{2}$$

where $\alpha^l_{(t,i,j)}$ is the jth element of the ith feature channel of the convolution output result in the tth input channel. The combined feature matrix is obtained after multiple rounds of feature extraction as follows:

$$Y_{contact} = \sum_{t=1}^{5} Y_t^3 \tag{3}$$

where ${Y_t}^3$ is the output of the third layer of the network in the tth input channel, and the final feature matrix is $Y_{contact}$.

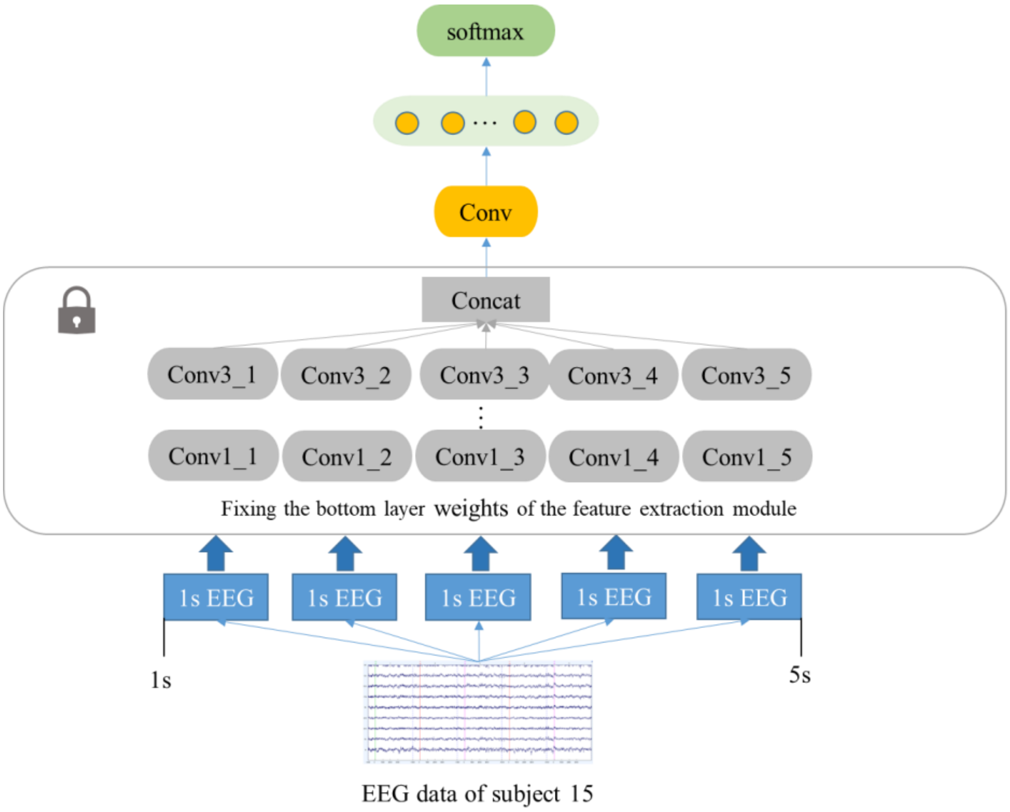


Figure 7. Architecture of STL-CNN.

Then, the internal parameter matrix of the upper layer is W_c. The feature matrix is expanded and integrated in the fully connected layer after finishing the feature extraction, and the final classification output is completed by Softmax function as follows:

$$Y_{final} = \text{Softmax}(\text{Relu}(Y_{contact} W_c)) \tag{4}$$

where Y_{final} is the final output vector of the model.

The backpropagation process is an update of the parameters using a loss function. The parameter matrix W_p of the bottom layer is shared in the CNN model, whose backpropagation process fails to be involved in the parameter update. However, the upper layer does involve. The parameter matrix of the upper layer of the feature extraction module is updated after the error is calculated, where the learning rate parameters are also updated according to the above-defined strategy until the iteration is completed.

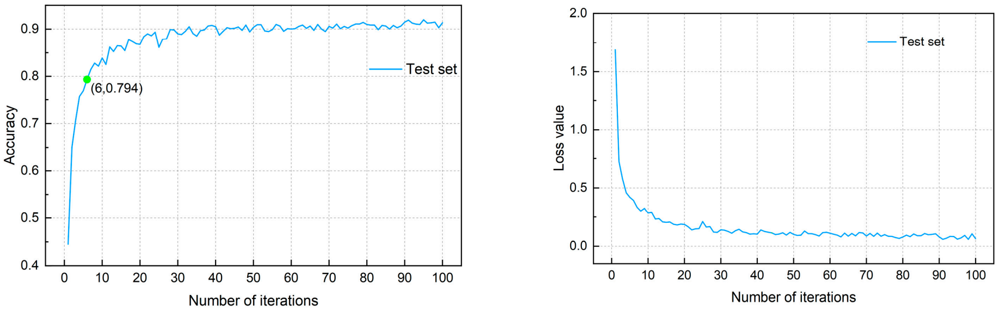
4. Performance Evaluation of STL-CNN

The hyperparameters of the convolutional and pooling layers in the upper feature extraction module of the STL-CNN model are the same as those of the CNN model, as shown in Table 3. The Adam optimizer is still used to improve the training efficiency. The initial learning rate is set to 0.05, the batch size is set to 64, and the Epoch is set to 100. The model training is activated after completing the format conversion of the input test data, in which the output labels "0", "1", and "2" are used as the representation of the comfort, power, and technology vehicle sound quality. The final parameters of STL-CNN model are shown in Table 4. The 84,720 weights in the convolutional kernel are same as the CNN model parameters after fixing the bottom feature extraction module, and the upper feature extraction module and classification module can be trained with a total of 124,203 parameters.

Table 4. STL-CNN model parameters.

Model Component Name		Component Parameters	Component Trainable Parameters
Bottom layer	\	84,720	0
Upper layer	Convolutional layer	122,880	122,880
	Pooling layer	0	0
Classification modules	Fully connected layer	1290	1290
	Softmax	33	33

The 15th subject's data are used as the test set, and the accuracy and loss values of the test set are shown in Figure 8. As can be seen from the analysis of the accuracy and loss values in Figure 8a,b, the accuracy is 44.5% in the first iteration of training. The reason for the above phenomenon is that the upper layer weights are randomly initialized at the beginning of training based on the bottom weights of the CNN model. Subsequently, an accuracy of 79.4% is reached at the sixth iteration, and the weights in the upper layer of the feature extraction module gradually fit the new test data as the number of iterations is increased. The test accuracy fluctuates upwards and finally stabilizes at 91.5%. The above results validate the ability of the STL-CNN model to adapt the new subject data and demonstrate that the STL-CNN model can take into account the individual variability of the new subject's EEG signals.



(**a**) Accuracy

(**b**) Loss values

Figure 8. Accuracy and loss values of the test set.

4.1. Comparison of Classification Models

To highlight the advantages of the constructed STL-CNN model, the traditional SVM and CNN are constructed as comparative experiments, which are effective intelligent models for solving pattern recognition problems. STL-CNN and the CNN are trained using the same datasets, and in each trial, the same parameters are used for them. The inputs of the SVM model are the extracted EEG features from three perspectives of the time domain, frequency domain, and entropy characteristics. Among them, there are seven time features, five frequency features, and one differential entropy feature. Thus, the input dimension of the SVM model is 1458 × 65 × 15, and 1458 = 243 × 6 represents the number of samples; 65 = (7 + 6) × 5 represents the dimension of the features, and 15 represents the number of subjects. The accuracies of the training set and test set using the SVM, CNN, and STL-CNN are shown in Figure 9.

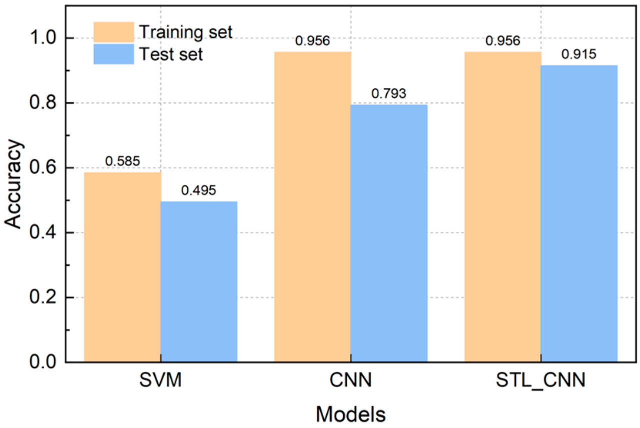


Figure 9. Accuracies of the training sets and test set using the SVM, CNN, and STL-CNN.

It can be clearly illuminated from the results in Figure 9 that the training set of the CNN model has a significant advantage over the typical SVM model, with an accuracy improvement of 37.1%. The accuracy of the test set of the CNN model is reduced by 16.3% compared to the accuracy of the training set; however, the accuracy is improved by 29.8% compared to the test set of the SVM model. Furthermore, it can be seen that the accuracy of the test set of the STL-CNN model is further improved by 12.2% compared to the CNN model, and there is a significant improvement of 42% that is out of the reach of the traditional SVM model. There is a reduction of only 4.1% compared to the training set of STL-CNN.

To further compare the performance of the above three models, the confusion matrix of the three models on the test set is analyzed, as showed in Figure 10. Each row of the confusion matrix represents the target category, and each column represents the output of the predicted category by the classification model. The classification performance of the three models is evaluated using precision, recall, and accuracy.

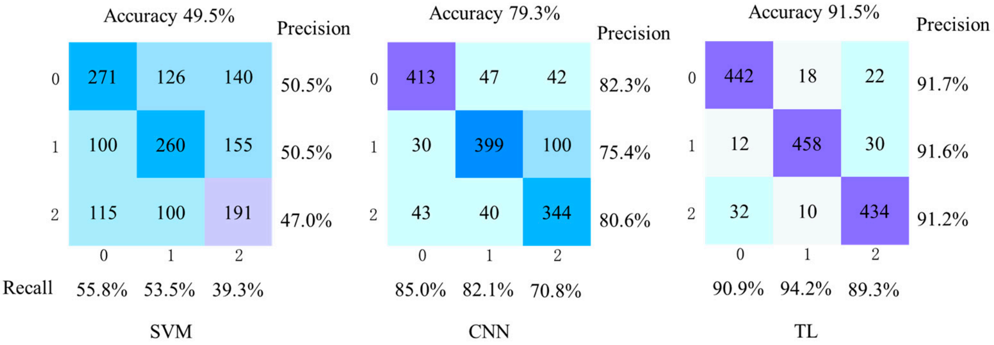


Figure 10. Confusion matrix for the test set of the three models.

The accuracy and recall of each category of the vehicle sounds are shown in Figure 10, where “0”, “1”, and “2” represent the vehicle sounds with comfort, power, and technology sound qualities, respectively. The SVM model is relatively effective in the recognition of the comfort and power sound qualities, while the lowest precision and recall rates are found for the recognition of the technology sounds: the recall rate of 39.3% is much lower than the classification results of the other two categories, which indicate that the SVM model is weak in identifying the vehicle sounds with the technology quality. The performance of the

CNN model is substantially improved compared to the SVM model; however, the recall rate for the technology sounds is still low, at 70.8%. The precision for the power sounds is also low, at 75.4%, which indicates that there are some instabilities in the CNN model. It can be clear seen from Figure 10 that the STL-CNN model has the best overall performance, where the recall rate of the technological sounds reaching 89.3%. There is a significant improvement in the overall stability of the model and a good recognition ability for the selected three categories of sounds, where the accuracy of the test set is 91.5%.

The above results validate the effectiveness of the constructed STL-CNN model and also demonstrate the advantages of the STL-CNN model in dealing with the problem of individual variability of EEG signals. Specifically, the changes in the new dataset can be adapted by the STL-CNN model to achieve the classification of EEG signals and the recognition of the vehicle sounds.

4.2. Comprehensive Evaluation of Classification Models

Five metrics are utilized to reflect the comprehensive performance of the model, including the minimum recall (min_recall), minimum precision (min_presicion), training set accuracy (val_acc), test set accuracy (test_acc), and test set kappa index. Among them, the kappa index is calculated as shown in Formula (5), and the radar plot of the five metrics of the three models for the test set is shown in Figure 11.

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{5}$$

Here, p_o is the accuracy of the model, and P_e is calculated as follows:

$$P_e = \frac{\sum\limits_{i=1}^{N} a_{i+}a_{+i}}{N^2} \tag{6}$$

where i is the category number, N is the total number of categories, a_{i+} represents the number of true samples in category i, and a_{+i} is the number of predicted samples in category i. The range of the kappa metric is [0, 1], and the larger the value, the better the model classification effect.

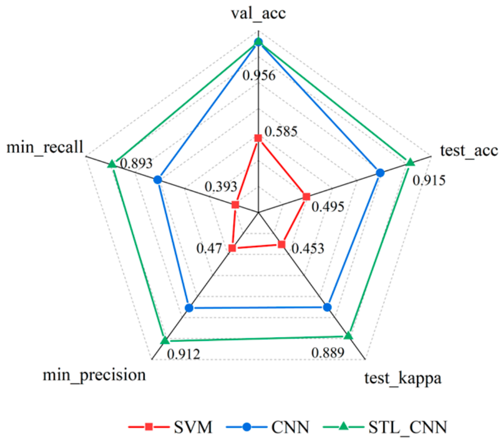


Figure 11. Comprehensive performance of the model.

It can be concluded from Figure 11 that the CNN model significantly outperforms the SVM model across all metrics, while the STL-CNN model has a better generalization ability and the best performance on the test set. This is due to the proposed weight sharing strategy, where the value of the kappa metric is 88.9%, which is out of the reach of the CNN and SVM models. In general, the STL-CNN model has the best overall performance compared the other two models and contributes to realizing the decoding of EEG-based vehicle sounds while simultaneously considering individual differences.

5. Discussion

In this paper, a deep learning hybrid model (STL-CNN) combining a CNN and specific transfer learning model is constructed to adaptively extract EEG features for the recognition of automobile sounds fused with EEG signals. Several important issues are explored. EEG signals can objectively characterize evaluators' subjective perceptions of sound [8,9], and it possible to utilize EEG signals as an evaluation index of automotive sound quality due to the popularity of miniaturization and portability of wearable devices equipped with brain electrophysiological signal sensors [38]. It has been demonstrated that there are some fluctuating changes in human brain signals that occur to map the differences between signals of different sound features. And human subjective emotions can be improved by driving in-vehicle audio systems to play pre-designed music based on EEG signals [17,39]. Therefore, the exploration of recognition of automobile sounds based on the EEG signals investigated in this paper is of great research significance in achieving the switching between driving sound patterns based on EEG signals.

In the present investigation, concerning target classification based on EEG signals, two primary challenges are evident: it is difficult to extract the EEG feature that is strongly related to the target object from the massive number of EEG signals and the suboptimal generalization performance exhibited by traditional machine classification models. In the STL-CNN model constructed in this paper, the bottom layer of the feature extraction module is designed to perform three rounds of feature extraction, where the basic communal features of the subject's EEG signal are extracted from the perspective of the time domain, layer by layer, then the extracted basic communal features are merged. The feature dimension is further increased in the designed upper feature extraction module to complete the adaptive extraction of potential EEG features related to vehicle sound. The constructed STL-CNN model uses a method of sharing the bottom weights to train the upper layer network, which significantly improves the model's generalization ability and increases the test set accuracy from 79.3% to 91.5%. It can be observed from the data analysis results presented in Section 4 that both the fundamental communal features of the EEG signals and the distinctive components of the individual differences have been taken into account when using subject data as a new test set. This enables the recognition of automobile sounds to consider individual differences. The advantages of using intelligent models in the extraction of latent features related to vehicle sounds have been elaborately demonstrated in the literature [40,41].

In addition, the superiority of the proposed STL-CNN model is also demonstrated by comparing the SVM and CNN models. The results presented in Sections 3 and 4.2 also manifest the advantages of the constructed STL-CNN model compared to the SVM and CNN models concerning the recognition accuracy of vehicle sound quality based on the EEG signals. As can be seen from Figures 10 and 11, the CNN model has obvious advantages compared with the typical SVM model. The accuracy of the CNN model on the training set is 37.1% higher than that of the SVM model, indicating that the EEG signal, as a physiological signal related to vehicle sound stimuli, possesses potential EEG features. The convolutional layer of the CNN model can complete the extraction of the potential EEG feature from the raw EEG data, and the constructed STL-CNN model further optimizes the structure of the CNN model and improves the ability of CNN deep learning to learn complex and non-linear EEG features. This is achieved through a fine-tuning process of the supervised parameter, which contributes to optimizing the model's performance. The low recognition efficiency of the technology sounds for the three models presented in Figure 10 is mainly due to the fact that the technology sounds come from online science fiction game sounds that may not be familiar to the subjects; thus, there is a lower recognition rate. Generally, the constructed STL-CNN model can adaptively extract EEG features to achieve the recognition of vehicle sounds combining EEG data.

Currently, there are different feature selections for different EEG matrices, and there are also various options for the model accuracy optimization methods in the research field of EEG data analysis. Neural networks designed with ad hoc interpretable elements can automatically identify the most relevant spatial and frequency neural signatures. These elements have been extensively investigated in Borra et al.'s studies [42–44]. The results presented in this paper are of importance for the future research of personalized sound design and the switching between preferred driving sound patterns based on EEG signals.

Research of the correlation between EEG signals and automobile sounds is still in its infancy. In future investigations, the number of sound quality types can be increased, and the range of subjects can be expanded to improve the comprehensiveness of the EEG experiment. In addition, a simulated driving pedestal can be built in future EEG signal acquisition experiments, where the changes in the subjects' perceptions for different scenarios can be considered and the real driving environment can be recreated as naturally as possible.

6. Conclusions

In this paper, EEG physiological signals are applied to recognize automobile sound quality with the complex semantics, and the construction of intelligent classification models is mainly explored. Then, in order to improve the generalization ability of the classification model, the bottom layer and upper layer of the EEG feature extraction modules are designed to adaptively extract the EEG features from the time–domain perspective. The method of sharing the bottom weights is proposed to train the upper layer network, where the STL-CNN model is established to achieve the recognition of vehicle sounds fused with EEG signals. The results of the performance comparison with traditional SVM and CNN models show that the deep learning model has obvious advantages. The test set accuracy of the CNN model is improved by 29.8% compared to the SVM model, and the accuracy of the constructed STL-CNN model is further improved by 12.2% compared to the CNN model. This demonstrates the effectiveness of the STL-CNN model constructed to deal with the problem of individual variability in EEG signals. The results presented in this paper provide a basis for future research on switching between different driving sound modes based on EEG signals, as well as a reference for data analysis methods fused with EEG signals to evaluate vehicle sound quality.

Author Contributions: J.G. and L.X. prepared the manuscript. This study was designed by L.X. and supervised by J.G. The psychological EEG test was designed by J.G. The data were collected by T.X. and L.X. Data analysis was performed by T.X., Z.L. and L.X. The language was improved by L.X. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by National Natural Science Foundation of China (No. 52175111).

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank all the subjects for their support.

Conflicts of Interest: Our study was supported by the National Natural Science Foundation of China (ID. 52175111, titled "Research on Active Sound Design of Automobile Integrating Brain Cognition Laws of Sound Quality"). In the current university research on brain–computer interface technology, most of the non-invasive head-mounted devices are used to collect EEG signals from the human brain, which is safe and non-invasive. In addition, all the subjects in this study understood the aims and procedures of the trial. The selected subjects have regular check-ups every year, and they are healthy and have normal hearing.

References

1. Lee, H.H.; Lee, S.K. Objective evaluation of interior noise booming in a passenger car based on sound metrics and artificial neural networks. *Appl. Ergon.* **2009**, *40*, 860–869. [CrossRef]
2. Schiffbänker, H.; Brandl, F.; Thien, G. *Development and Application of an Evaluation Technique to Assess the Subjective Character of Engine Noise*; SAE Technical Paper: Warrendale, PA, USA, 1991.
3. He, Y.; Tu, L.; Xu, Z. Review of Vehicle Sound Quality. *Automot. Eng.* **2014**, *4*, 391–401.
4. Liu, Z.; Xie, L.; Huang, T.; Lu, C.; Chen, W.; Zhu, Y. The objective quantification of door closing sound quality based on multidimensional subjective perception attributes. *Appl. Acoust.* **2022**, *192*, 108748. [CrossRef]

5. Murata, H.; Tanaka, H.; Takada, H.; Ohsasa, Y. *Sound Quality Evaluation of Passenger Vehicle Interior Noise*; SAE Technical Paper: Warrendale, PA, USA, 1993.
6. Ohsasa, Y.; Kadomatsu, K. *Sound Quality Evaluation of Exhaust Note During Acceleration*; SAE Technical Paper: Warrendale, PA, USA, 1995.
7. Chang, K.; Jeong, K.; Park, D. *A Study on the Strategy and Implementing Technology for the Development of Luxurious Driving Sound*; SAE Technical Paper: Warrendale, PA, USA, 2014. [CrossRef]
8. Engelke, U.; Darcy, D.P.; Mulliken, G.H.; Bosse, S.; Martini, M.G.; Arndt, S.; Antons, J.N.; Chan, K.Y.; Ramzan, N.; Brunnstrom, K. *Psychophysiology-Based QoE Assessment: A Survey*; IEEE Journal of Selected Topics in Signal Processing; IEEE: Piscataway, NJ, USA, 2017.
9. Geng, B.; Liu, K.; Duan, Y.; Song, Q.; Tao, X.; Lu, J.; Shi, J. A Novel EEG Based Directed Transfer Function for Investigating Human Perception to Audio Noise. In Proceedings of the 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 15–19 June 2020; pp. 923–928.
10. Xie, L.; Lu, C.; Liu, Z.; Yan, L.; Xu, T. Studying critical frequency bands and channels for EEG-based automobile sound recognition with machine learning. *Appl. Acoust.* **2022**, *185*, 108389. [CrossRef]
11. Xie, L.; Lu, C.; Liu, Z.; Yan, L.; Xu, T. Study of Auditory Brain Cognition Laws-Based Recognition Method of Automobile Sound Quality. *Front. Hum. Neurosci.* **2021**, *15*, 663049. [CrossRef] [PubMed]
12. Kalaganis, F.; Adamos, D.A.; Laskaris, N. A Consumer BCI for Automated Music Evaluation Within a Popular On-Demand Music Streaming Service "Taking Listener's Brainwaves to Extremes". In Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, Thessaloniki, Greece, 16–18 September 2016; Springer: Cham, Switzerland, 2016; pp. 429–444.
13. Friston, K.J.; Frith, C.D.; Dolan, R.j.; Price, C.j.; Zeki, S.; Ashburner, J.T.; Penny, W.D. *Human Brain Function*; Elsevier Science: Amsterdam, The Netherlands, 2004.
14. Cirett Galán, F.; Beal, C.R. EEG estimates of engagement and cognitive workload predict math problem solving outcomes. In Proceedings of the International Conference on User Modeling, Adaptation, and Personalization, Montreal, QC, Canada, 16–20 July 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 51–62.
15. Lenz, D.; Schadow, J.; Thaerig, S.; Busch, N.A.; Herrmann, C.S. What's that sound? Matches with auditory long-term memory induce gamma activity in human, EEG. *Int. J. Psychophysiol.* **2007**, *64*, 31–38. [CrossRef] [PubMed]
16. Cong, F.; Alluri, V.; Nandi, A.K.; Toiviainen, P.; Fa, R.; Abu-Jamous, B.; Gong, L.; Craenen, B.G.W.; Poikonen, H.; Huotilainen, M.; et al. Linking Brain Responses to Naturalistic Music Through Analysis of Ongoing EEG and Stimulus Features. *IEEE Trans. Multimed.* **2013**, *15*, 1060–1069. [CrossRef]
17. Li, Z.G.; Di, G.Q.; Jia, L. Relationship between Electroencephalogram variation and subjective annoyance under noise exposure. *Appl. Acoust.* **2014**, *75*, 37–42. [CrossRef]
18. Zhang, R.; Zong, Q.; Dou, L.; Zhao, X.; Tang, Y.; Li, Z. Hybrid deep neural network using transfer learning for EEG motor imagery decoding. *Biomed. Signal Process. Control* **2021**, *63*, 102144. [CrossRef]
19. Xie, L.; Lu, C.; Liu, Z.; Chen, W.; Zhu, Y.; Xu, T. The evaluation of automobile interior acceleration sound fused with physiological signal using a hybrid deep neural network. *Mech. Syst. Signal Process.* **2023**, *184*, 109675. [CrossRef]
20. Huang, G.; Hu, Z.; Zhang, L.; Li, L.; Liang, Z.; Zhang, Z. Removal of eye-blinking artifacts by ICA in cross-modal long-term EEG recording. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 217–220.
21. Zheng, W.L.; Zhu, J.Y.; Lu, B.L. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* **2017**, *10*, 417–429. [CrossRef]
22. Jenke, R.; Peer, A.; Buss, M. Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* **2014**, *5*, 327–339. [CrossRef]
23. Frederick, J.A.; Lubar, J.F. Skewness in the Time Series of EEG Magnitude and Spectral Correlation. *Soc. Neuronal Regul.* **2002**, *10*, 2377–4400.
24. Xiang, J.; Maue, E.; Fan, Y.; Qi, L.; Mangano, F.T.; Greiner, H.; Tenney, J. Kurtosis and skewness of high-frequency brain signals are altered in paediatric epilepsy. *Brain Commun.* **2020**, *2*, fcaa036. [CrossRef] [PubMed]
25. Hernández, D.E.; Trujillo, L.; Z-Flores, E.; Villanueva, O.M.; Romo-Fewell, O. Detecting Epilepsy in EEG Signals Using Time, Frequency and Time-Frequency Domain Features. In *Computer Science and Engineering Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 167–182.
26. Hjorth, B. EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* **1970**, *29*, 306–310. [CrossRef] [PubMed]
27. Carmen, V.; Nicole, K.; Benjamin, B.; Schlögl, A. Time Domain Parameters as a feature for EEG-based Brain-Computer Interfaces. *Neural Netw.* **2009**, *22*, 1313–1319.
28. Bokde, A.L.; Teipel, S.J.; Schwarz, R.; Leinsinger, G.; Buerger, K.; Moeller, T.; Möller, H.-J.; Hampel, H. Reliable manual segmentation of the frontal, parietal, temporal, and occipital lobes on magnetic resonance images of healthy subjects. *Brain Res. Protoc.* **2005**, *14*, 135–145. [CrossRef]
29. Hadjidimitriou, S.K.; Hadjileontiadis, L.J. Toward an EEG-based recognition of music liking using time-frequency analysis. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 3498–3510. [CrossRef]

30. Yoon, J.H.; Yang, I.H.; Jeong, J.E.; Park, S.-G.; Oh, J.-E. Reliability improvement of a sound quality index for a vehicle HVAC system using a regression and neural network model. *Appl. Acoust.* **2012**, *73*, 1099–1103. [CrossRef]
31. García-Martínez, B.; Martínez-Rodrigo, A.; Cantabrana, R.Z.; García, J.M.P.; Alcaraz, R. Application of entropy-based metrics to identify emotional distress from electroencephalographic recordings. *Entropy* **2016**, *18*, 221. [CrossRef]
32. Pan, Y.; Guan, C.; Yu, J.; Ang, K.K.; Chan, T.E. Common frequency pattern for music preference identification using frontal EEG. In Proceedings of the International IEEE/EMBS Conference on Neural Engineering, San Diego, CA, USA, 6–8 November 2013.
33. Nakanishi, M.; Mitsukura, Y.; Hara, A. EEG analysis for acoustic quality evaluation using PCA and FDA. In Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication, Atlanta, GA, USA, 31 July –3 August 2011.
34. WeiLong, Z.; BaoLiang, L. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [CrossRef]
35. Sobhan, S.; Zohteh, M.; Tohid, Y.R.; Farzamnia, A. Recognizing Emotions Evoked by Music using CNN-LSTM Networks on EEG signals. *IEEE Access* **2020**, *8*, 139332–139345.
36. Myslobodsky, M.S.; Coppola, R.; Bar-Ziv, J.; Weinberger, D.R. Adequacy of the International 10–20 electrode system for computed neurophysiologic topography. *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.* **1990**, *7*, 507–518. [CrossRef] [PubMed]
37. Zhang, K. *Emotion Recognition Based on EEG Signals under Music Induction*; Zhejiang University: Hangzhou, China, 2019.
38. Mikkelsen, K.B.; Kappel, S.L.; Mandic, D.P.; Kidmose, P. EEG recorded from the ear: Characterizing the ear-EEG method. *Front. Neurosci.* **2015**, *9*, 438. [CrossRef]
39. Adamos, D.A.; Dimitriadis, S.I.; Laskaris, N.A. Towards the bio-personalization of music recommendation systems: A single-sensor EEG biomarker of subjective music preference. *Inf. Sci.* **2016**, *343*, 94–108. [CrossRef]
40. Huang, H.B.; Wu, J.H.; Huang, X.R.; Yang, M.L.; Ding, W.P. The development of a deep neural network and its application to evaluating the interior sound quality of pure electric vehicles. *Mech. Syst. Signal Process.* **2019**, *120*, 98–116. [CrossRef]
41. Ma, C.; Chen, C.; Liu, Q.; Gao, H.; Li, Q.; Gao, H.; Shen, Y. Sound quality evaluation of the interior noise of pure electric vehicle based on neural network model. *IEEE Trans. Ind. Electron.* **2017**, *64*, 9442–9450. [CrossRef]
42. Borra, D.; Fantozzi, S.; Magosso, E. EEG motor execution decoding via interpretable sinc-convolutional neural networks. In Proceedings of the Mediterranean Conference on Medical and Biological Engineering and Computing, Coimbra, Portugal, 26–28 September 2019; Springer International Publishing: Cham, Switzerland, 2019; pp. 1113–1122.
43. Borra, D.; Magosso, E.; Castelo-Branco, M.; Simões, M. A Bayesian-optimized design for an interpretable convolutional neural network to decode and analyze the P300 response in autism. *J. Neural Eng.* **2022**, *19*, 046010. [CrossRef]
44. Borra, D.; Mondini, V.; Magosso, E.; Müller-Putz, G.R. Decoding movement kinematics from EEG using an interpretable convolutional neural network. *Comput. Biol. Med.* **2023**, *165*, 107323. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.