Mechanical Systems and Signal Processing 84 (2017) 245–267

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Evaluation of vehicle interior sound quality using a continuous restricted Boltzmann machine-based DBN

Hai B. Huang [a,b], Ren X. Li [a], Ming L. Yang [a], Teik C. Lim [b], Wei P. Ding [a,*]

[a] Institute of Automotive Engineering Research, Southwest Jiaotong University, 610031 Cheng Du, Si Chuan, China
[b] Vibro-Acoustics and Sound Quality Research Laboratory, College of Engineering and Applied Science, University of Cincinnati, 45221 Cincinnati, OH, USA

ARTICLE INFO

Article history:
Received 10 April 2016
Received in revised form
6 July 2016
Accepted 8 July 2016
Available online 25 July 2016

Keywords:
Sound quality evaluation
Vehicle interior noise
Feature extraction
Continuous restricted Boltzmann machines
Deep belief networks

ABSTRACT

The perception of vehicle interior sound quality is important for passengers. In this paper, a feature fusion process for extracting the characteristics of vehicle interior noise is studied, and an improved deep belief network (DBN) that uses continuous restricted Boltzmann machines (CRBMs) to model continuous data is proposed. Six types of vehicles are used for recording interior noise under different working conditions, and a corresponding subjective evaluation is implemented. Psychoacoustic metrics and energy-based criteria using the wavelet transform (WT), wavelet packet transform (WPT), empirical mode decomposition (EMD), critical-band-based pass filter, and Mel-scale-based triangular filer approaches have been applied to extract interior noise features and then develop a fusing feature set combining psychoacoustic metrics and critical band energy based on comparisons. Using the obtained fusion feature set, a CRBM–based DBN (CRBM-DBN) model is developed through experiments. The newly developed model is verified by comparing its performance relative to multiple linear regression (MLR), backpropagation neural network (BPNN), and support vector machine (SVM) models. The results show that the proposed CRBM-DBN model has a lower prediction error and higher correlation coefficient with human perception compared to the other considered methods. In addition, CRBM-DBN outperforms BPNN and SVM in terms of stability and reliability. The presented approach may be regarded as a promising method for evaluating vehicle noise.

1. Introduction

Currently, the demands for vehicle quality have increased. Specifically, vehicle sound quality has become one of the most important characteristics to vehicle consumers and manufacturers. Passengers can be considered as part of a vibro-acoustical system. Vehicle interior noise is the most intuitive parameter felt by passengers and could affect their decision to purchase a vehicle. Technical competition with competitors urges vehicle companies to improve the interior sound quality of their vehicles. Furthermore, increasingly strict government standards for vehicle noise have been implemented because vehicle noise accounts for more than 60% of urban noise [1]. Moreover, vehicle interior noise, which consists of airborne noise and structure-borne noise, negatively influences the physical and psychological perception of passengers. The acoustic

[*] Corresponding author.
E-mail addresses: huanghaibo214@my.swjtu.edu.cn (H.B. Huang), rxli@home.swjtu.edu.cn (R.X. Li), yml@home.swjtu.edu.cn (M.L. Yang), limt@ucmail.uc.edu (T.C. Lim), dwp@home.swjtu.edu.cn (W.P. Ding).

http://dx.doi.org/10.1016/j.ymssp.2016.07.014
0888-3270/© 2016 Elsevier Ltd. All rights reserved.

characteristics of a vehicle are an integral part of product identity and could also be regarded as the 'sound DNA' of the vehicle [2]. Therefore, the methods used for evaluating the interior sound quality of vehicles should be investigated further.

The sound quality of vehicle interior noise is influenced by three variables: sound field, auditory perception, and auditory evaluation [3], which cause the evaluation of sound quality to be a multidimensional task. The A-weighted sound pressure level is widely used to objectively evaluate the quality of vehicle interior noise because it is simple to apply and can easily be understood. Evaluations using the A-weighted technique have demonstrated that human hearing has highly complex level-dependent evaluation mechanisms [3]. In general, low-frequency sound waves can be transmitted over a longer distance than high-frequency sound waves on the basilar membrane. Therefore, high-frequency noise can be masked more easily by low-frequency noise [4]. The perception of a sound event is affected by the sound pressure level and by psychoacoustic characteristics, such as loudness, sharpness, roughness, fluctuation strength and other extended metrics (Some of the psychoacoustic metrics and their attributes are summarized in Table 1). Thus, the perceived properties of a sound are not identical to those of a sound that is being emitted. Perception involves the physics and psychology of the human hearing process. Previous research has found that loudness and sharpness constitute at least 50% of the subjective assessment of a sound [5–7]. Particularly, loudness and sharpness could constitute more than 70% of the subjective assessment for stationary vehicle noise [8]. However, psychoacoustic metrics are largely designed to represent the specific perceived characteristics of a noise, and it is difficult to find a psychoacoustic model to extract noise features that can precisely describe the perception responses for all individuals. Therefore, the individual psychoacoustic criterion is not entirely adequate for the overall objective quantification of vehicle sound quality [2].

People will experience different hearing perceptions when they receive different sound waves with different frequencies. Vehicle interior noise, including stationary and non-stationary vehicle noise, can be represented in the time and frequency domains. Thus, it is important to select a signal processing approach for extracting sound perception features based on human auditory properties. Recently, many acoustic scholars and engineers have extracted noise features by using advanced signal processing methods, including wavelet transform (WT), wavelet packet transform (WPT), empirical mode decomposition (EMD), Wigner–Ville distribution (WVD), and other methods frequently mentioned in the literature [8,16–18]. WPT has been applied to extract sound features for evaluating vehicle interior noise [18] and can be used as a specific filter bank that is similar to the critical bands in the human hearing system. The WVD-based method was developed and used to assess noise resulting from vehicle suspension shock absorbers [16]. The proposed sound quality criterion, the Sound Metric based on the Wigner–Ville distribution (SMWVD), was highly correlated with the corresponding subjective evaluation of rattling noise. The EMD in the Hilbert–Huang transform (HHT) method was adopted to estimate the noise resulting from slamming a car door [19], which means that EMD can extract the main impact characteristics of the noise resulting from door slamming. In addition, the critical-band-based method [18] and Mel-scale-based approach [20] have been applied to extract noise characteristics and have potential advantages for acoustic modeling. Human auditory perceptions are related to sound characteristics; hence, the extraction of noise features is an important sound quality evaluation issue, and extraction methods depend on the original sound signals and practical applications.

Various intelligent pattern recognition methods have been introduced because of the complexity and nonlinearity of human hearing perceptions and noise characteristics. Multiple linear regression (MLR) is the most commonly used method for predicting vehicle interior sound quality and has two major advantages: the small amount of required experiment data and the ease of interpreting the results. Lee et al. [17] and Kim et al. [21] used the MLR method to evaluate vehicle impact noise and modified the vehicle suspension components to improve the sound quality based on the developed MLR model. However, MLR has limitations when fitting the highly nonlinear characteristics of the human auditory perception process because of its linear property. The support vector machine (SVM) is a statistical learning method that can be used to map the input data from a low-dimensional feature space to a high-dimensional feature space by using nonlinear mapping and then performing linear pattern recognition in the high-dimensional space. Liu et al. [22] used the SVM model to predict engine-radiated sound quality and found that the SVM was successful for establishing a nonlinear relationship between the subjective and objective evaluations. However, because the characteristics of sound are distributed widely, the predefined kernel function in the SVM model might not be sufficiently powerful to express the characteristics of all features simultaneously. Another well-known pattern recognition method is neural networks (NNs), among which the backpropagation neural network (BPNN) is the most

Table 1
Sound quality metrics and their attributes.

Sound metrics	Attributes
Loudness	The auditory characteristic related to the intensity of sensations [9].
Sharpness	The auditory characteristic involving the high-frequency portion of a sound [10].
Roughness	The auditory perception property related to a frequency modulation of approximately 70 Hz and the amplitude modulation of a sound [11].
Fluctuation strength	The auditory perception characteristic involving a frequency modulation of approximately 4 Hz and the amplitude modulation of a sound [10].
Articulation index	The quantitative measure of the intelligibility of speech [12].
Tonality	The auditory perception property related to the pitch strength of sounds [13].
Tone-to-noise ratio	Used to record the prominent discrete tone in a noise [14].
Noise criterion	The specific measurement of indoor background noise [15].

commonly used. Two advantages make NNs attractive for evaluating vehicle noise. First, NNs have general nonlinear mapping capabilities, enabling any continuous data to be approximated with arbitrary desired accuracy. Second, NNs are nonparametric data-driven models that do not require restrictive assumptions of the data distribution. Hence, NNs are capable of solving many complex and nonlinear problems. Wang et al. [23] developed a sound quality model to objectively evaluate nonstationary vehicle interior noise based on a BPNN and found that the BPNN was accurate and effective as a sound pleasure (or annoyance) model. However, the intrinsic weakness of the NN method is that it easily falls into local minima, making it more difficult to reach the globally optimal solution. Recently, some intelligent algorithms have been introduced to improve the performance of pattern recognition methods and develop some combination models, such as genetic algorithm optimal support vector machines (GA-SVMs) [24] and particle swarm optimization neural networks (PSO-NNs) [25]. However, the main limitation of the abovementioned methods is that they are shallow models with one or no hidden layer [26]. Consequently, the models can only consider linear weighted combinations of multiple features and cannot be used to explore the regularity of features effectively [27]. Therefore, this paper introduces the deep belief network (DBN) to fuse the extracted noise features and obtain global minima for predicting vehicle sound quality.

The DBN is a probabilistic generative model composed of many layers of hidden units. The DBN functions based on restricted Boltzmann machines (RBMs) and a greedy layer-wise learning algorithm. The DBN possesses the advantages of conventional NNs, has a strong information fusing ability because of its deep architecture, and can determine the global minima through its pre-training and fine-tuning phases. Recently, DBNs have gained considerable attention in the signal processing and machine learning community, with successful applications in acoustic modeling [27], voice activity detection [28], and face recognition [29]. However, the binary stochastic variables of the RBM would restrict the DBN from modeling continuous data. Therefore, in this paper, we improved the DBN to predict continuous data by substituting continuous restricted Boltzmann machines (CRBMs) for restricted Boltzmann machines (RBM) and then applying the developed CRBM-based DBN (CRBM-DBN) to predict the sound quality of vehicle interior noise.

In this paper, a feature fusion process of vehicle interior sound quality has been studied. This investigation is crucial because a noise feature set represents the characteristics of sounds being perceived and influences the performance of the evaluated model. Moreover, the evaluation results of the interior sound quality depend on the accuracy and robustness of the prediction models, therefore, an outstanding prediction model is needed. In this study, an improved DBN model, CRBM-DBN, has been proposed for evaluating vehicle interior sound quality. The CRBM-DBN model is trained, verified, and compared with three conventional models: MLR, BPNN and SVM.

2. Continuous restricted Boltzmann machine-based deep belief networks

2.1. Deep belief networks

DBNs employ a multilayered architecture that stacks a number of RBMs [30,31]. Thus, the learning procedure of RBMs must be understood. The topological structure of RBMs is shown in Fig. 1(a) and consists of an input layer that contains the visible units $\mathbf{v}=\{v_1, v_2,\ldots,v_i\}$ and a hidden layer that contains the hidden units $\mathbf{h}=\{h_1, h_2,\ldots,h_j\}$. Assuming that $\mathbf{v}\in\{0, 1\}$ and $\mathbf{h}\in\{0, 1\}$ are both binary and discrete stochastic variables, the energy function is defined as

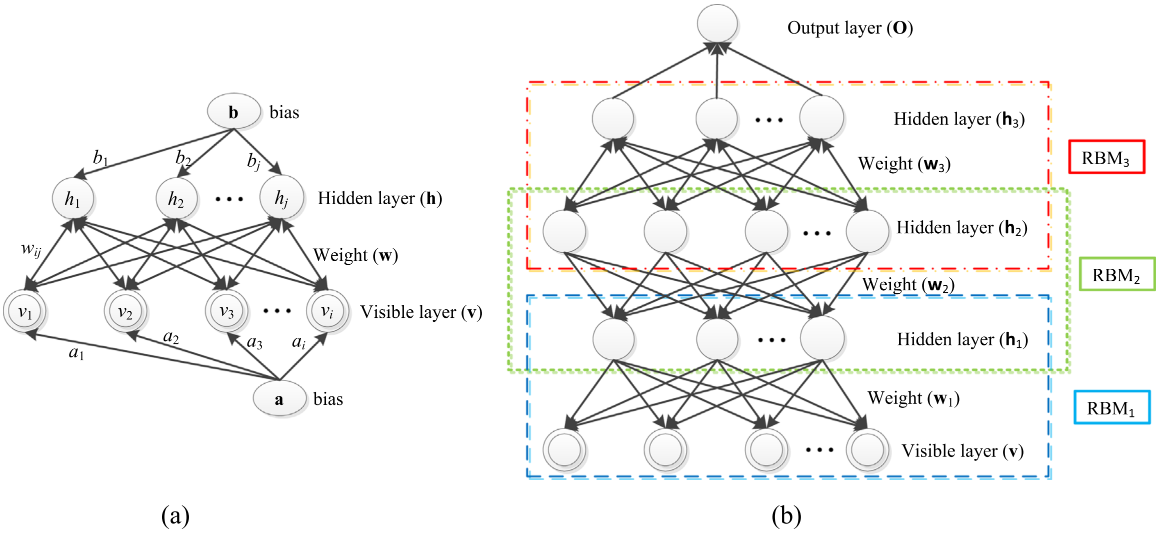

Fig. 1. Architectures for (a) an RBM and (b) a three hidden-layer DBN.

$$E(\mathbf{v},\mathbf{h};\theta) = -\sum_{i=1}^{V}\sum_{j=1}^{H} w_{ij}v_ih_j - \sum_{i=1}^{V} a_iv_i - \sum_{j=1}^{H} b_jh_j \tag{1}$$

where θ=(w, b, a) are the model parameters, w_{ij} is the symmetric weight between the visible unit v_i and hidden unit h_j, and a_i and b_j are their bias terms. V and H are the numbers of visible units and hidden units, respectively. The probabilities that an RBM assigns a visible vector **v** and a hidden vector **h** can be calculated via Eqs. (2) and (3), respectively.

$$p(\mathbf{v};\theta) = \frac{1}{Z(\theta)}\sum_{H} \exp(-E(\mathbf{v},\mathbf{h};\theta)) \tag{2}$$

$$p(\mathbf{h};\theta) = \frac{1}{Z(\theta)}\sum_{V} \exp(-E(\mathbf{v},\mathbf{h};\theta)) \tag{3}$$

where $Z(\theta)$ is a normalization factor and is defined as follows:

$$Z(\theta) = \sum_{V}\sum_{H} \exp(-E(\mathbf{v},\mathbf{h};\theta)) \tag{4}$$

Because no hidden-to-hidden connections occur in the RBM, the conditional probabilities are factorial and are given by

$$p(h_j = 1|\mathbf{v};\theta) = \sigma\left(\sum_{i=1}^{V} w_{ij}v_i + b_j\right) \tag{5}$$

$$p(v_i = 1|\mathbf{h};\theta) = \sigma\left(\sum_{j=1}^{H} w_{ij}h_j + a_i\right) \tag{6}$$

where σ=1/(1 + e^{-x}) is a sigmoid function. Eq. (5) can be regarded as a positive phase that transforms the input data from the visible layer to the hidden layer, whereas Eq. (6) is the negative phase that is implemented as a reconstruction of the neuron units of the previous visible layer. Following the gradient of the log likelihood, log $(p(\mathbf{v};\theta))$, the update rule for the synaptic weight can be obtained as follows:

$$\Delta w_{ij} = \eta(<v_ih_j>_{data} - <v_ih_j>_{\mathrm{mod}\ el}) \tag{7}$$

where η is the learning rate, which ranges from 0 to 1, $<v_ih_j>_{data}$ is the expectation with respect to the training set distribution, and $<v_ih_j>_{model}$ is the same expectation under the distribution defined by the model. The term $<v_ih_j>_{model}$ requires exponential time to be calculated precisely, therefore, the contrastive divergence (CD) approximation [21] of the gradient is used to replace $<v_ih_j>_{model}$ by running one full-step Gibbs sampler. The same update rule is applied for bias learning, but the individual visible unit v_i and hidden unit h_j are used instead of the pairwise products.

A DBN is formed by stacking a number of RBMs layer by layer, and a typical DBN model with three hidden layers is shown in Fig. 1(b). The DBN training process can be implemented in two steps, namely, pre-training and fine-tuning [31]. (1) Pre-training phase: Considering the first RBM, i.e., the RBM$_1$ in Fig. 1(b), the input data are fed into the visible layer first and then from the visible units to the hidden units using training parameters until the maximum number of training epochs is reached. The obtained weights and biases of RBM$_1$ are fixed and begin to train RBM$_2$ with the same method applied to RBM$_1$. The learning process is accomplished through the successive training of each individual RBM model. The pre-training procedure is unsupervised. (2) Fine-tuning phase: After the layer-wise learning process, the parameters of the DBN model will be updated using a backpropagation algorithm to further reduce the training error and improve the accuracy rate of pattern recognition. Unlike the pre-training procedure, which considers one RBM at a time, the fine-tuning procedure considers all DBN layers simultaneously. The training error is calculated using the expectation targets and model outputs. The supervised backpropagation process is continued until the maximum epoch has been achieved.

Conventionally, in the typical DBN approach, the output layer of the model is typically a logistic regression layer or softmax layer so that it can be applied to the classification problems [32–34]. However, few applications have used DBNs to solve regression problems because the original neurounits in the RBM are binary and discrete and thus cannot model continuous values in regression cases [35]. Therefore, an improved DBN model is proposed to evaluate continuous vehicle noise sound quality.

2.2. Development of deep belief networks to model continuous data

Because the subjective evaluation of vehicle interior noise is continuous, a DBN must be developed to model continuous data. According to the conventional DBN approach, the RBM can model continuous data if we treat the probabilities of visible units in Eq. (5) as approximations of the continuous values [36]. In this paper, we introduce a specific RBM called the CRBM, which tends to generate continuous data with high symmetry using a simple and reliable training algorithm proposed by Chen and Murray [37], to improve the conventional classification DBN model by substituting the CRBM for the RBM to develop a regression-based DBN.

The basic learning rule for the CRBM is similar to that of the RBM, except that the binary unit of the RBM is replaced with the continuous stochastic unit with zero-mean Gaussian noise added to the input of a sampled sigmoid unit [38]. If s_j is the output of hidden unit j, the inputs from the visible units with states $\{s_i\}$ are calculated as follows:

$$s_j = \varphi_j\left(\sum_i w_{ij}s_i + \mu N_j(0, 1)\right) \tag{8}$$

with

$$\varphi_j(x_j) = \theta_L + \frac{(\theta_H - \theta_L)}{1 + e^{(-\alpha_j x_j)}} \tag{9}$$

where μ is a constant and $N_j(0,1)$ represents a Gaussian random variable with a mean of zero and unit variance. $\varphi_j(x)$ is a sigmoid function with asymptotes at θ_L and θ_H. Parameter α controls the slope of $\varphi_j(x)$ and thus the nature and extent of the unit's stochastic behavior [35]. The developed CRBM allows for a smooth transition from noise-free, deterministic behavior to noise-controlled, stochastic behavior. The update rules for the weight w and noise-control parameter α are as follows:

$$\Delta w_{ij} = \eta_w(< s_i s_j >_{data} - < s_i s_j >_{mod\ el}) \tag{10}$$

$$\Delta \alpha_j = \frac{\eta_\alpha}{\alpha_j}(< s_j^2 >_{data} - < s_j^2 >_{mod\ el}) \tag{11}$$

where η_w and η_α are the learning rates and $<\cdot>_{data}$ and $<\cdot>_{model}$ refer to the expectations of the training data and the distribution by the model. Similar to the updated method of the RBM in Eq. (7), the one full-step Gibbs sampler is utilized in the term $<\cdot>_{model}$ to reduce the computation time in the training process of biases.

Because the newly developed CRBM-DBN is a regression model, the logistic output or softmax output layers are removed. Meanwhile, the training error of the fine-tuning step is calculated using the mean square error (MSE) of the estimators. Hereafter, the learning procedure of the CRBM-based DBN model is the same as that of the RBM-based DBN model. Fig. 2 shows the training procedure of the CRBM-DBN model.

3. Recording and evaluating the vehicle interior noise

3.1. Development of the vehicle interior noise database

The vehicle interior noise measurements were conducted on a straight test road, and no reflecting buildings or other objects were present within 20 m of the test site. Considering the direction and frequency-filtering properties of the human hearing system, the artificial head measurement system (HMS II.3) made by the HEAD acoustics company was employed in this experiment. Following the measurement method in the standard GB/T 18697 [39], the vehicle interior noise above the seats of the front and rear-left passengers was recorded through dummy heads, and the test conditions are given in Table 2. Because interior cavity of the vehicle is a closed and reverberant system, the diffused-field response equalization for noise signals in the acquisition system was selected. A signal length of 10 s for the stationary noise (idling state and constant speed state) and the speed tracking method for the non-stationary noise (acceleration state and deceleration state), with a sampling rate of 44.1 kHz, were adopted to measure the interior sound. Fifteen vehicles of six types were selected for this test, as summarized in Table 3. Each of the sample vehicles was tested under working conditions once, and all collected noise signals were saved on a mobile workstation for further analysis. The recorded sound signals have a signal-to-noise ratio greater than 20 dB because the measurement conditions were controlled well. Fig. 3 presents the noise signals recorded under different working conditions. The overall processes of acceleration and deceleration states vary with time, and the time required for the jurors to complete the evaluation is comparatively long. In this experiment, each of the stationary noise signals was cut to 5 s, and each of the non-stationary noise signals was divided into several sequential segments of 3–6 s using the *Cooledit* software application, with approximately 20 km/h speed intervals for acceleration and approximately 35 km/h speed intervals for deceleration. This procedure is used because the subjective and objective evaluations are more sensitive to vehicle speed than to time [40]. This process can make noise samples available with time that can be used to estimate and decrease the total evaluation time; meanwhile, for the acceleration and deceleration noise samples, the noise variation of approximately 3–6 s will not change dramatically, so the juries can evaluate their perceptions more accurately. We obtained 390 samples of noise from the preprocessing step.

3.2. Subjective evaluation of the vehicle interior noise

Human hearing is generated by signal transmission and frequency filtering through the physical system (the outer contour, inside portion, and auditory nerves of the ear) [4]. A sound first arrives at the outer contour of the ear (torso, shoulder, head and pinna), causing reflection and diffraction, and is then transmitted into the ear (outer ear, middle ear and inner ear). The outer ear functions as a direction filter that changes the sound pressure level at the ear drum by +15 to

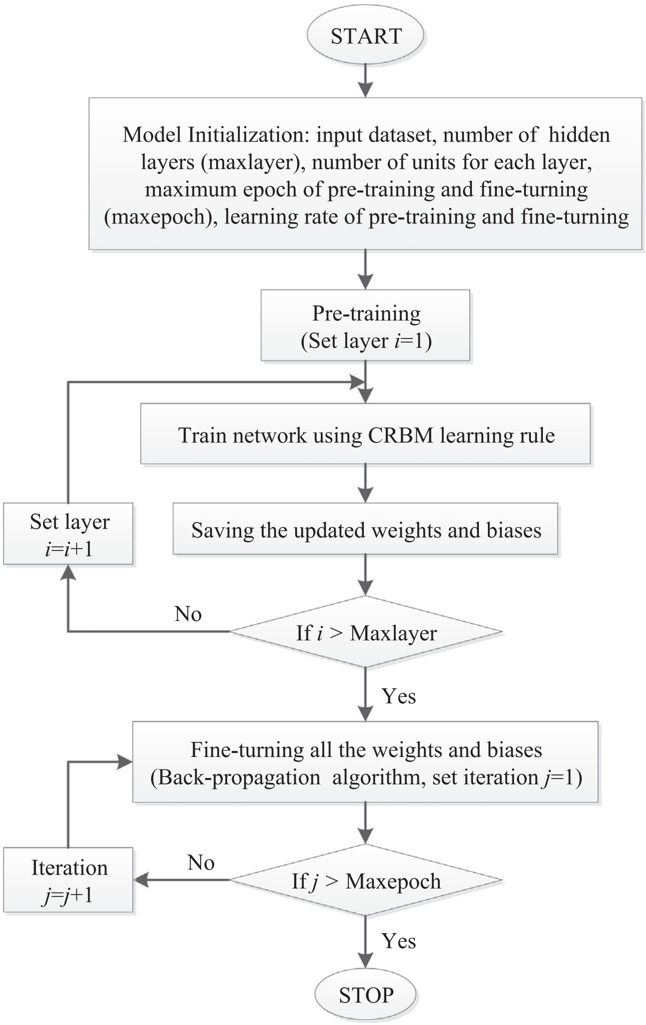
START
Model Initialization: input dataset, number of hidden layers (maxlayer), number of units for each layer, maximum epoch of pre-training and fine-turning (maxepoch), learning rate of pre-training and fine-turning
Pre-training (Set layer i=1)
Train network using CRBM learning rule
Saving the updated weights and biases
Set layer i=i+1
No
If i > Maxlayer
Yes
Fine-turning all the weights and biases (Back-propagation algorithm, set iteration j=1)
Iteration j=j+1
No
If j > Maxepoch
Yes
STOP

Fig. 2. Learning process of the CRBM-DBN model.

Table 2
Working conditions for the vehicle interior noise measurement.

# Index	State setting	Vehicle running conditions
1	Idling state	Neutral shift for MT (manual transmission) and “N” shift for AT (automatic transmission)
2	Constant speed state	Running speeds of 40 km/h, 60 km/h, 80 km/h, 100 km/h, 120 km/h. Time shift for MT and “D” shift for AT.
3	Acceleration state	Wide-open throttle from 40 to 120 km/h. Highest gear for MT and “D” shift for AT.
4	Deceleration state	Breaking from 100 km/h to 0

Table 3
Test vehicles and their attributes.

# Types	Vehicle level	Wheelbase (m)	Displacement (L)	Number
#1	Compact family car	2.30–2.45	1.0–1.6	3
#2	Medium family car	2.45–2.60	1.6–2.4	3
#3	Large family car	2.60–2.80	2.4–3.0	2
#4	Compact sport utility vehicle	2.45–2.60	1.0–1.6	3
#5	Medium sport utility vehicle	2.60–2.80	1.6–2.4	2
#6	Large sport utility vehicle	2.80–3.00	2.4–3.0	2

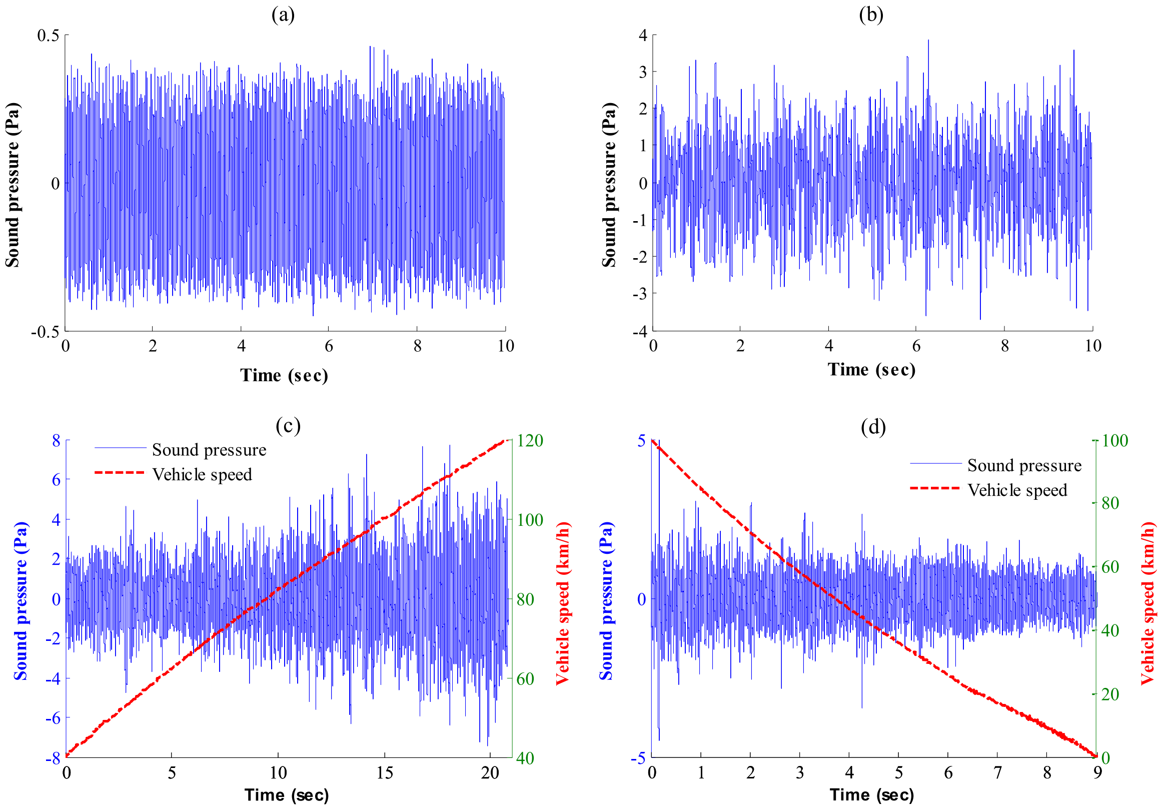

Fig. 3. Recorded interior noise signals in a sample vehicle while (a) idling; (b) traveling at a constant speed of 60 km/h; (c) accelerating from 40 km/h to 120 km/h; and (d) decelerating from 100 km/h to 0 km/h.

−30 dB depending on the frequency and direction of the sound incidence [3]. In the middle ear, sound waves are transferred to vibrations of the stapes, which improve the sound transmission efficiency. The vibration signals are transferred into electrical signals after nonlinear filtering in the inner ear. Finally, the human brain obtains the auditory perception via the nervous system.

Human hearing involves complex signal processing but has a short memory; hence, the sound perception of vehicle interior noise should be evaluated through a playback system. Sennheiser HD800 headphones were employed in this evaluation test, and the binaural noises were reproduced and displayed in real time. The subjective ratings of noise samples are best determined statistically due to the variability of human responses to a particular sound event. Twenty-six reviewers (14 males and 12 females with a mean age of 27.3 years and a standard deviation of 6.2) with normal hearing from universities or automotive companies participated in the experiment. The laboratory conditions for the jury test included a temperature of 25° and low background noise (< 35 dB). A rating method was used to provide semantic meaning to assess the interior sound quality; the subjective ratings and their relationships with human perceptions are shown in Table 4, which is recommended by the *China Automotive Technology & Research Center* in China. For flexibility, the reviewers can also rate the sounds between each continuous grade. The noise samples were randomly played to the participants, and the evaluation results are presented in Fig. 4. First, the subjective ratings were obtained by averaging the results of the reviewers. Then, the error bar corresponding to the standard deviation was added. To clearly show the results, the values on the *x*-axis are ordered by their mean values because the order of the sequence did not influence further analysis and the *y*-axis represents the subjective rating. Fig. 4 shows that the evaluation scores for each noise sample are relatively concentrated and can be used for following analysis.

Fig. 5 shows the subjective ratings of the interior noise with respect to different types and working conditions of vehicles. For a certain type of vehicle, its interior noise perceptions vary based on different working conditions, and the jury

Table 4
Grades for the subjective evaluation of vehicle interior sound quality.

Subjective perception	Excellent	Very good	Good	Acceptable	Not good	Bad
Rate scores	9	8	7	6	5	4

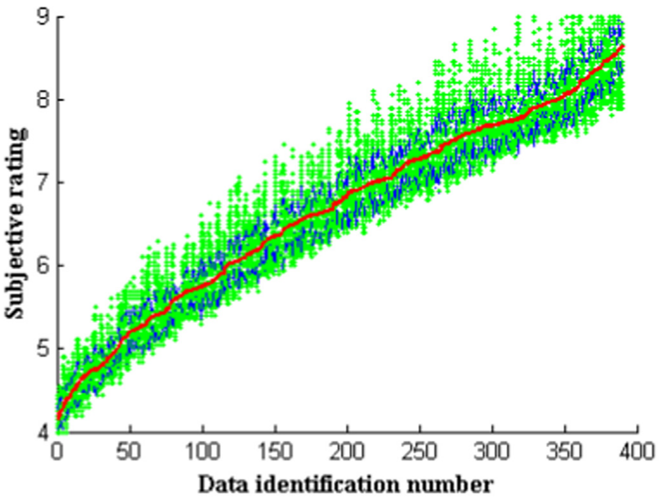

Fig. 4. Subjective evaluation of vehicle interior noise. ‘*’: subjective ratings of the jury; ‘—’: mean values of the subjective ratings; ‘– .’: error bar with ± standard deviation.

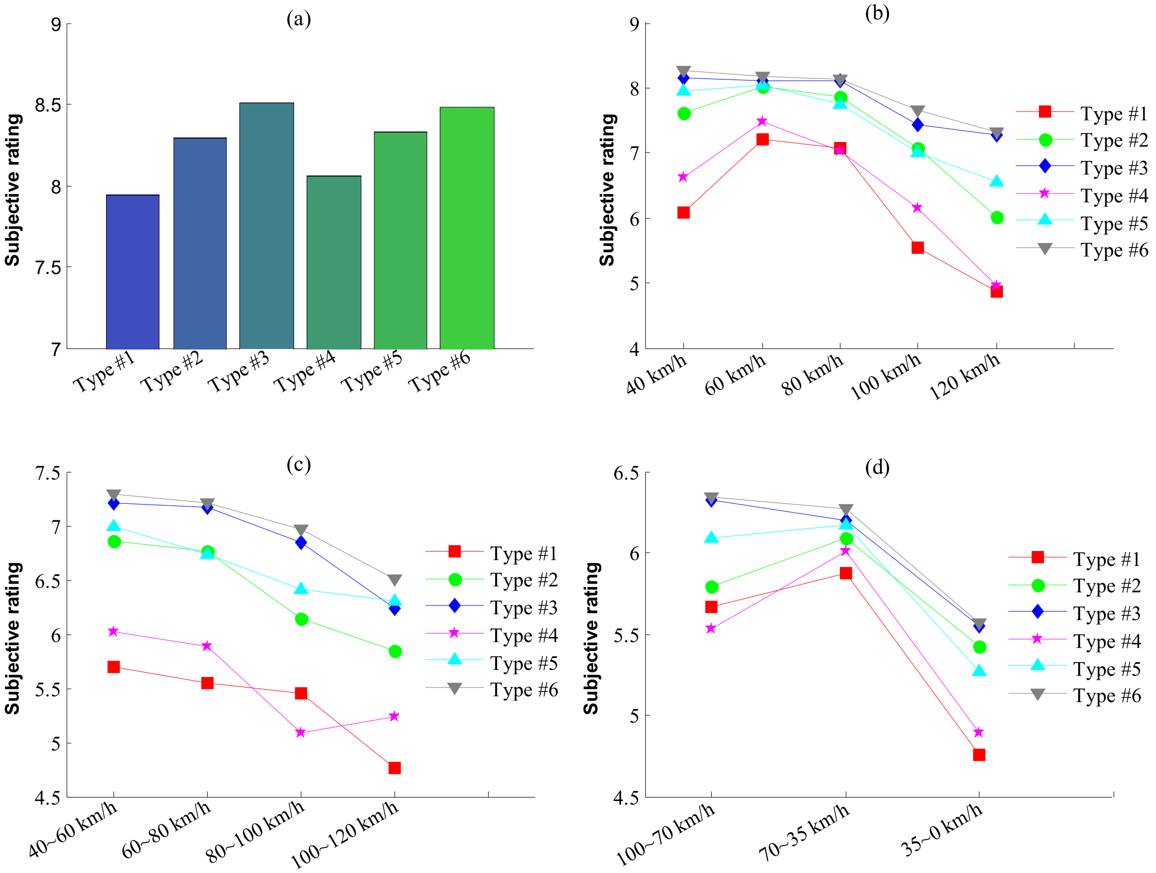

Fig. 5. Subjective ratings of the interior noise with respect to different working conditions and vehicle types. The bars and scatter points are mean values for the same type of vehicle at the same working condition. (a) idling state; (b) constant speed state; (c) acceleration state; (d) deceleration state.

evaluation of different vehicle interior noise differs greatly, even under the same working conditions. Fig. 5(a) illustrates that the subjective rating of the interior sound quality improves with increasing vehicle size in the idling state, and the evaluation scores are similar between the same levels of family cars and sport utility vehicles. In general, the sound package of higher-level cars is better than that of lower-level cars in design and material, which results in better performance in sound insulation and absorption. For the constant speed state shown in Fig. 5(b), the subjective ratings first increase or remain stable at middle-low speeds and then decrease sharply at high speeds. This trend occurs because of the relatively low transmission ratio and relatively high engine RPM (revolutions per minute) at low vehicle speeds, which might result in a lower interior sound quality compared with that at more moderate speeds. However, when the vehicle speed continues to increase, wind noise and tire noise become dominant, which results in a lower sound quality than at middle-low speeds. In addition, higher-priced cars present better interior noise performance than lower-priced cars due to the better effects of their sound packages. Fig. 5(c) illustrates that the subjective evaluation decreases during vehicle acceleration. The overall subjective estimation during acceleration is lower than that at a constant speed because the interior noise of the acceleration state contains the characteristic noise of the constant speed state and the noise orders emitted by the engine and transmission system. Fig. 5(d) shows that the sound quality of noise when decelerating from 35 km/h to 0 km/h is worse than the sound quality of noise in the other two speed ranges for each type of vehicle, which results from the squealing noise that is emitted when vehicles brake at low speeds. Compared with the other three working conditions, interior noise during deceleration shows relatively lower subjective ratings. The jury evaluation demonstrated that the subjective estimation of interior sound quality was highly related to vehicle type and working conditions.

4. Interior noise feature extraction

4.1. Psychoacoustic-based feature extraction

In terms of psychoacoustic theory, the characteristics of a sound perceived through the human auditory system could be described using a variety of psychoacoustic metrics, such as loudness, sharpness, fluctuation strength, and roughness, rather than its absolute level. In this paper, we calculated eight psychoacoustic sound quality metrics, including the four major metrics given above and four extended metrics, the articulation index (AI), tonality, tone-to-noise ratio and noise criterion, to extract sound features and objectively evaluate vehicle interior noise. The attributes of the psychoacoustic metrics are presented in Table 1.

The conventional psychoacoustic models are developed for stationary signals, and the time-varying analysis method should be applied for non-stationary signals. In this study, the interior noise samples of the idling state and constant speed state were directly processed using the stationary psychoacoustic models, and the interior noise samples of the acceleration and deceleration states were calculated by dividing the signals into several frames. Specifically, a non-stationary noise in a short time can be approximately regarded as a stationary noise [9]. Considering the forward and backward masking effects, the non-stationary signals were divided into frames of 100 ms with 50% overlap for each frame, so that the stationary psychoacoustic models could be used on these frames. The average values of the sound quality metrics in all frames were calculated for a non-stationary noise, and the average values of the left and right channels in the artificial head were used to obtain an overall objective rating of vehicle interior noise.

Scatterplots between the psychoacoustic metrics and subjective ratings are presented in Fig. 6, which illustrates that the correlation between loudness and the subjective rating is highest among the eight psychoacoustic metrics, with an absolute correlation coefficient (0.817). The correlations of sharpness, roughness, and fluctuation strength with the subjective evaluations are relatively strong, with absolute correlation coefficients of 0.777, 0.714 and 0.743, respectively. The articulation index, tonality, tone-to-noise ratio and noise criterion exhibit relatively weak correlations with the subjective ratings (i.e., correlation coefficients less than 0.7). Except for the articulation index, the other seven sound quality metrics exhibit a negative correlation with the subjective evaluation values. The above results illustrate that psychoacoustic metrics can extract the main features of vehicle interior noise. However, each sound quality metric has its own limitations to a certain characteristic of sound, and no dominant sound metric for evaluating vehicle interior noise has been identified. Therefore, a combination of psychoacoustic metrics and intelligent pattern recognition methods are helpful for evaluating vehicle interior sound quality.

4.2. Energy-based feature extraction

In this paper, the WT, WPT, EMD, critical-band-based bandpass filter [18], and Mel-scale-based triangular filter [20] have been applied and compared for sound feature extraction. Human hearing perception is more sensitive to low-frequency sound waves [3,18], and we selected 16 kHz as the upper frequency limit to extract the sound features in this study. To guarantee a dense resolution at a low frequency, the corresponding parameters of the five approaches were selected as shown in Table 5. Fig. 7 shows the frequency resolutions of the signal processing methods, which have been calculated based on their definitions [8,16–20] and the selected parameters listed in Table 5. All methods have a high resolution at low frequencies. The WT exhibits an exponential resolution in the middle and high frequencies, and the WPT presents a constant resolution in the entire bandwidth. The critical band and Mel scale both illustrate the nonlinear resolutions at middle and high frequencies. The EMD has a self-adaptive frequency resolution, so its resolution varies and depends on the analyzed signal.

Human auditory perception is related to the amplitude and frequency distribution of sound waves. An alternative means of expressing the feature of a sound is via its energy distribution. Therefore, the five aforementioned methods were used to

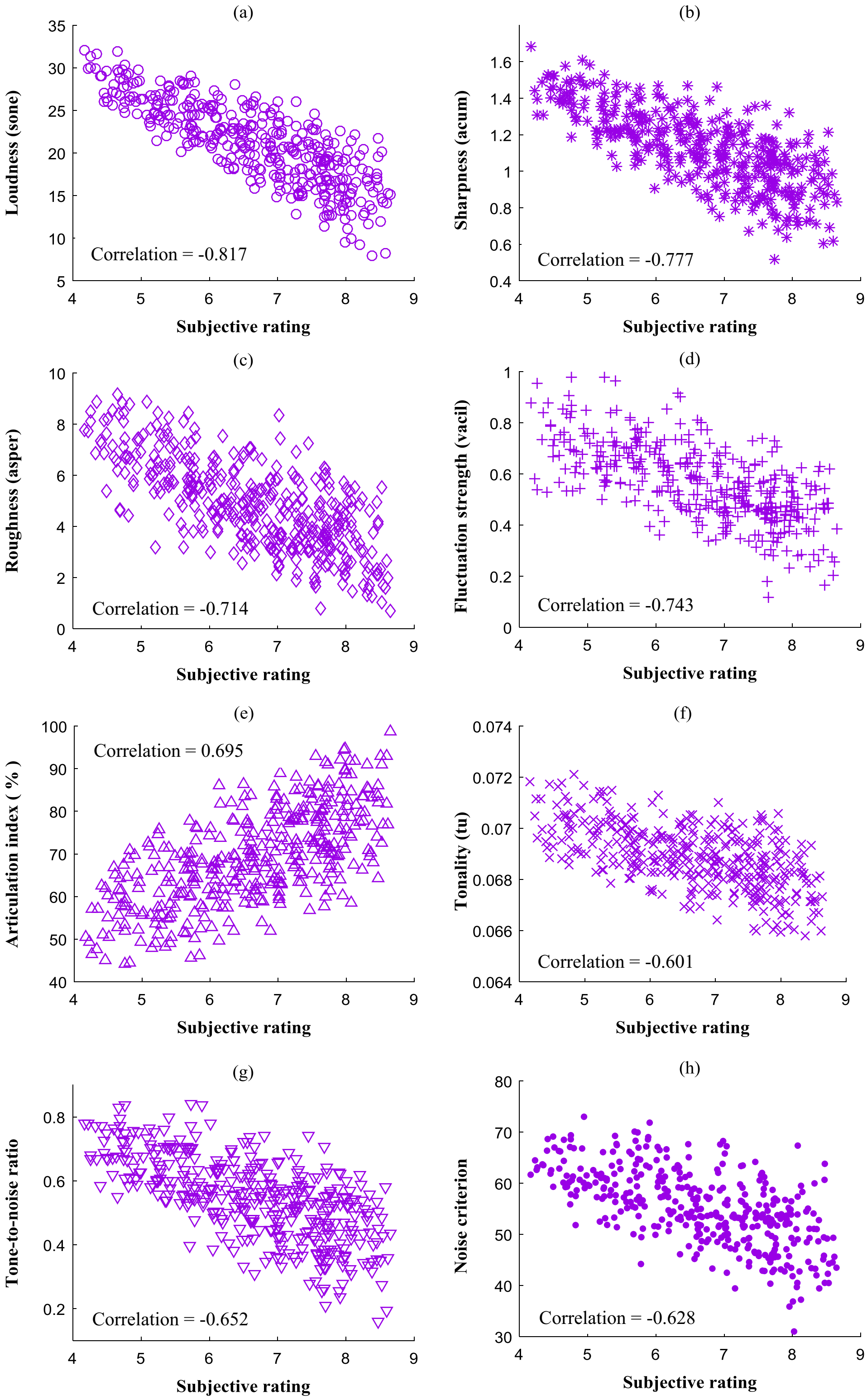

Fig. 6. Relationships between psychoacoustic metrics and subjective ratings for vehicle noise samples. (a) Loudness. (b) Sharpness. (c) Roughness. (d) Fluctuation strength. (e) Articulation index. (f) Tonality. (g) Tone-to-noise ratio. (h) Noise criterion.

Table 5
Parameter settings for the five methods used to analysis vehicle interior noise.

Method	Parameter settings
WT	Wavelet function: ‘db6’, analysis level: 7-level for 8 sub-bands
WPT	Wavelet function: ‘db6’, analysis level: 7-level for 128 sub-bands
EMD	Interpolation scheme: ‘spline’; max-modes: 15; maxiterations: 2000
Critical band-based bandpass filter	24 critical bands according to standard DIN 45631 [41]
Mel scale-based triangular filter	26 Mel scales, frequency interval of 150.9 Mel

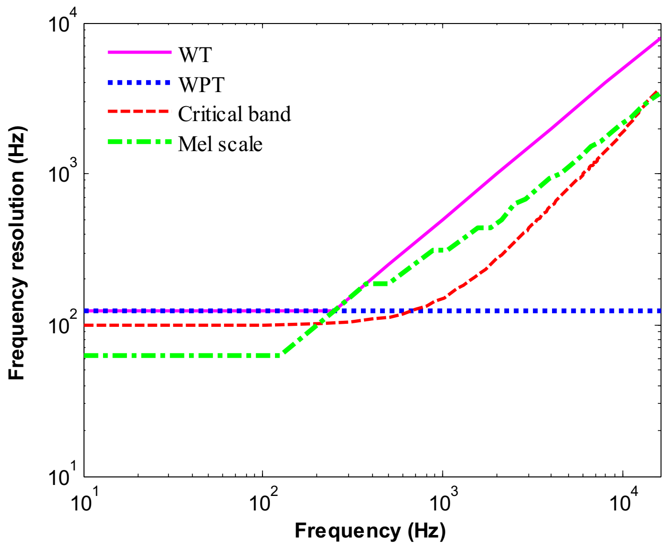

Fig. 7. Frequency resolutions of the signal processing methods according to the selected parameters.

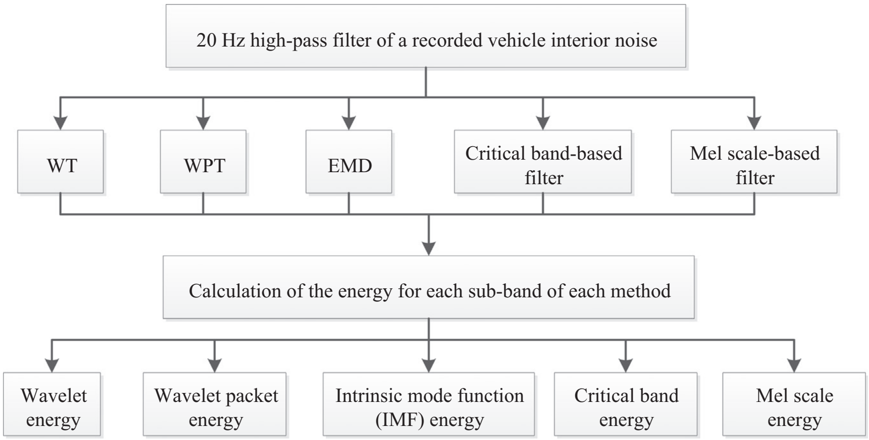
20 Hz high-pass filter of a recorded vehicle interior noise
WT
WPT
EMD
Critical band-based filter
Mel scale-based filter
Calculation of the energy for each sub-band of each method
Wavelet energy
Wavelet packet energy
Intrinsic mode function (IMF) energy
Critical band energy
Mel scale energy

Fig. 8. Feature extraction process of the vehicle interior noise.

extract the energy features of vehicle interior noise. Fig. 8 shows the procedure used to extract sound features from the measured interior noise samples. First, because humans cannot hear frequencies below 20 Hz, a three-order high-pass The Butterworth filter was designed to remove this infrasound. The five introduced signal processing methods were then applied to consider the preprocessed noise, and the corresponding sub-signals were obtained. For each series of sub-signals, the energy value can be calculated by summing the square of the signals with respect to the sub-bands, as defined in Eq. (12). According to the conservation of energy, the sum of the sub-energies should be equal to the total energy of the original signal.

$$E_i = \int_{-\infty}^{+\infty} |a_i(t)|^2 dt \overset{t \in finite}{=} \sum_t \left[a_i(t)\right]^2 \Delta\ t \tag{12}$$

where E_i is the energy of the ith sub-signal, $a_i(t)$ is the ith sub-signal, and Δt is the time interval of $a_i(t)$.

The energy feature extraction process for non-stationary noise signals is the same as that described in Section 4.1, and the average energy of each frame and channel was calculated. Figs. 9 and 10 provide the extracted energy features of vehicle interior noise for the constant speed and acceleration states, respectively. Fig. 9 illustrates that the noise generally has frequencies of less than 500 Hz. The WT shows the minimum number of energy features because its frequency resolution is increased exponentially, whereas the WPT has the maximum number of energy features because of its constant frequency

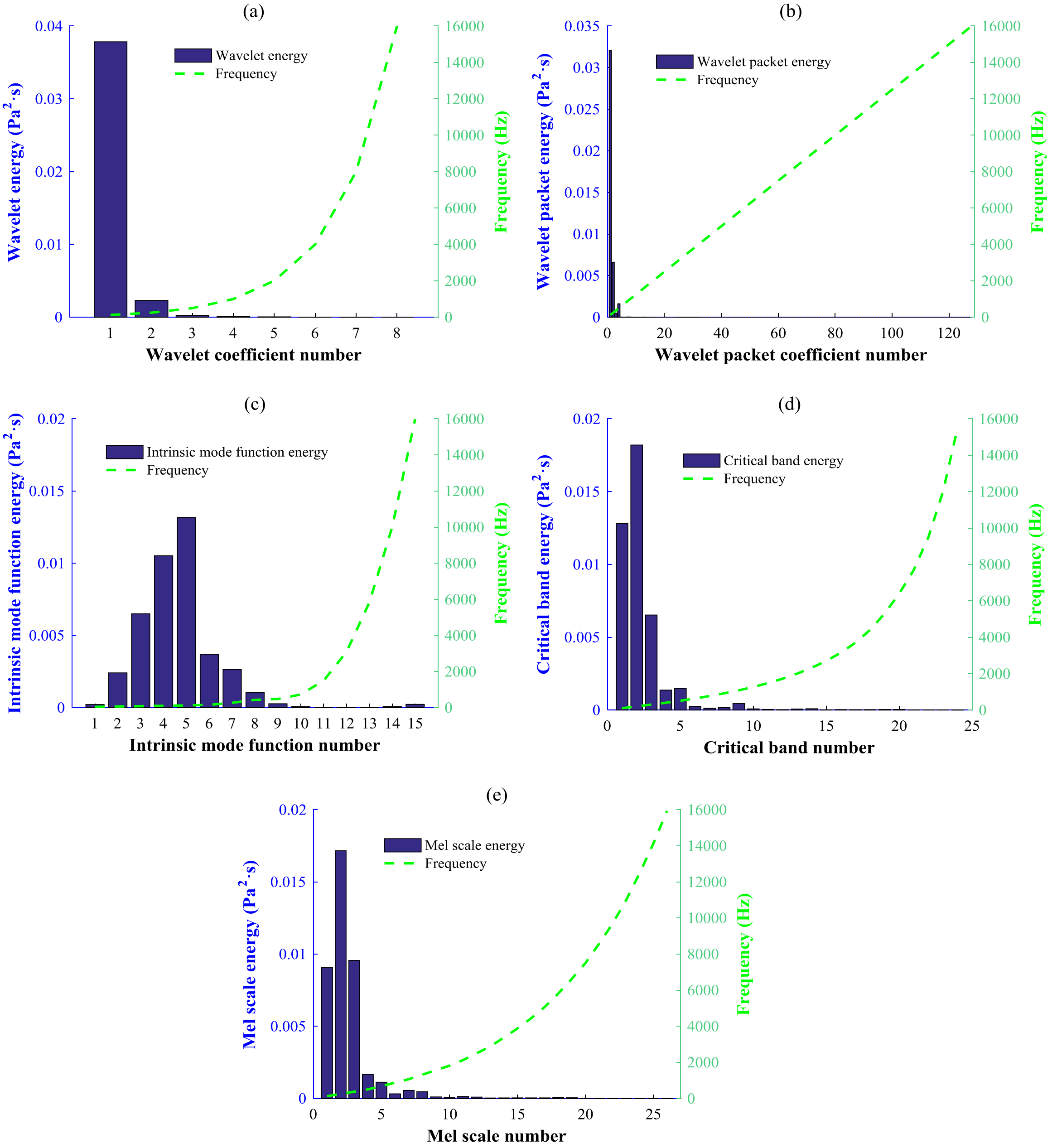

Fig. 9. Energy features of the vehicle interior noise when using a noise sample for a vehicle traveling at a constant speed of 80 km/h. (a) WT; (b) WPT; (c) IMF; (d) Critical-band-based bandpass filter; (e) Mel-scale-based triangle filter.

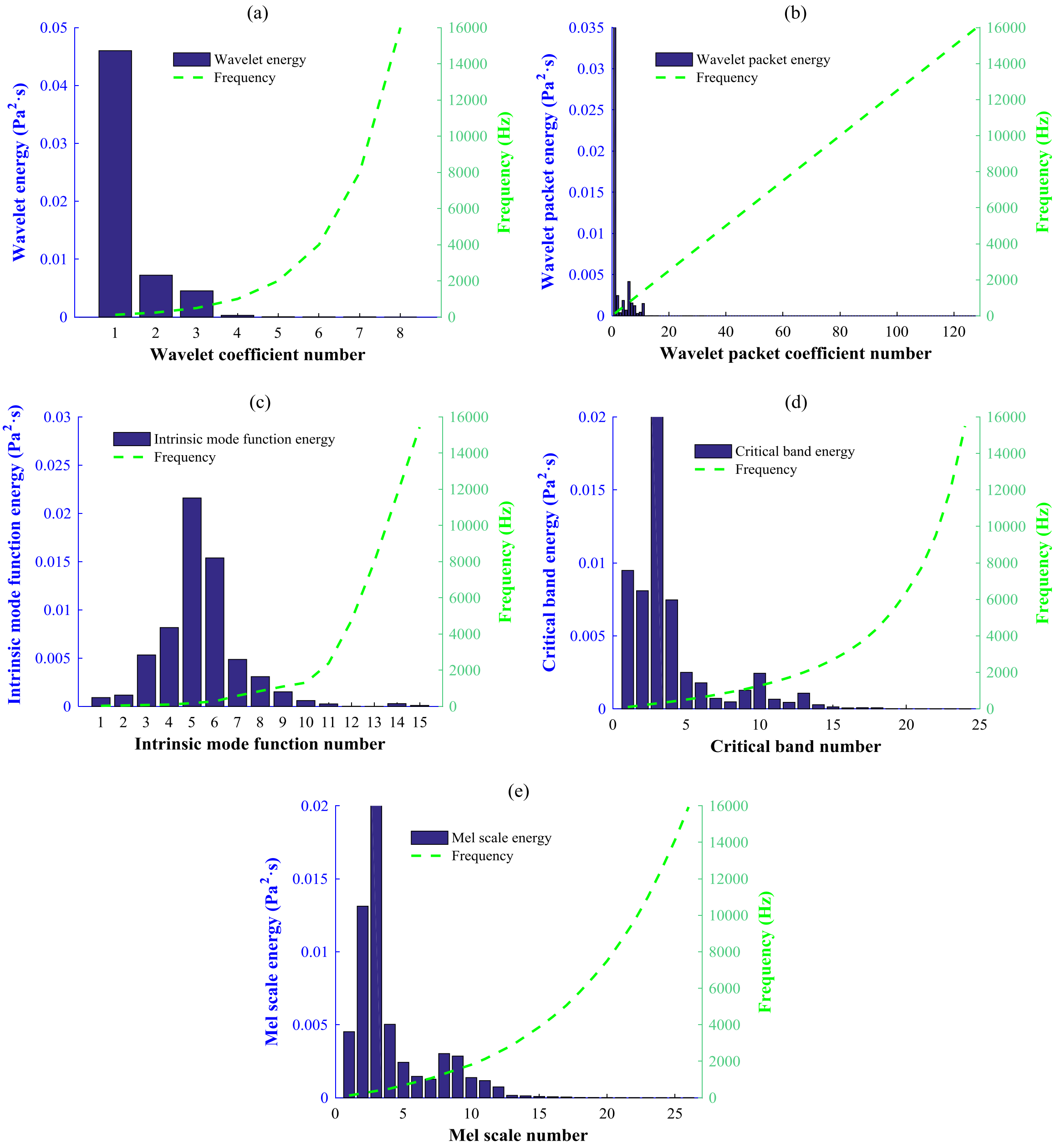

Fig. 10. Energy features of the vehicle interior noise using a noise sample from a vehicle accelerating to 80–100 km/h. (a) WT; (b) WPT; (c) IMF; (d) Critical-band-based bandpass filter; (e) Mel-scale-based triangle filter.

resolution. The EMD has an adaptive frequency resolution, and the original signal was decomposed into several intrinsic mode functions (IMFs), which consist of the natural frequencies of the decomposed signal with amplitude and frequency modulation. The EMD energy is mainly distributed in the range of 2–8 IMF modes. The critical band and Mel scale have frequency resolutions according to certain filters of the human hearing system. Additionally, compared with the WT and WPT, the frequency resolutions of the critical band and Mel scale are between constant and exponential, which could help them control the proportion of feature numbers in low and high frequencies. The critical band energy and Mel scale energy are both mainly distributed under the 10th sub-band; however, because of the different frequency resolutions, the energy features of the two methods differ in each sub-band. Although the energy features in the high frequencies appear lower than those in the low frequencies, they have specific influences on the auditory perception of people. Fig. 10 shows a pattern similar to that shown in Fig. 9, but the interior noise during the acceleration state contains more high components (e.g., up to 2000 Hz) than the interior noise during the constant speed state. This trend occurs because the interior noise components at the acceleration state are distributed in the relatively high frequency range, which is consistent with the experimental results presented in Section 3. However, the effectiveness of these energy features must be investigated further to apply a better noise feature and evaluate the sound quality of vehicle interior noise.

4.3. Feature comparison and selection

The psychoacoustic metrics and energy features can somewhat reflect the characteristics of the measured vehicle interior noise. One feature set typically requires a corresponding intelligent model to cooperate for pattern recognition. However, the selection of an appropriate model is complex and time consuming, especially for models with multiple hidden layers and many free parameters. To reduce the time costs of intelligent network modeling, the extracted sound features were compared with one another and then a fusing feature set was developed, as detailed in this section.

To compare the effectiveness of the extracted noise features and visualize the comparative result, the t-distributed stochastic neighbor embedding (t-SNE) [42] method was introduced to reduce redundancy and produce 2-D embeddings of the extracted features. t-SNE produces 2-D embeddings, in which points that are nearby in high-dimensional vector space are also nearby in the 2-D vector space [43]. First, t-SNE converts the pairwise distances d_{ij} in the original high-dimensional space to joint probabilities $p_{ij} \propto \exp\left(-d_{ij}^2\right)$. An iterative search for corresponding points in the 2-D space is then performed, which results in a similar set of joint probabilities. Because the volume near a high-dimensional point is higher than the volume near a low-dimensional point, t-SNE calculates the joint probability in the 2-D space by using a heavy tailed probability distribution, $q_{ij} \propto \exp\left(1+d_{ij}^2\right)^{-1}$. This method can lead to 2-D maps that exhibit structures at many different scales [42].

The sound quality metrics and energy features were processed via t-SNE to map the original features to the 2-D vector space as a guide [43]. Fig. 11 shows the t-SNE 2-D maps of each extracted vehicle interior noise feature. Figs. 11(c) and (d) illustrate that the energy features of WPT and EMD in 2-D maps are relatively discrete and that their noise samples of four working conditions are distributed without order. The energy features of the WT, critical band and Mel scale present better distributions in the corresponding 2-D maps. However, the noise samples of the idling state overlap with those of the constant speed state in Figs. 11(b), (e) and (f). Compared with the above three features, t-SNE mapping for psychoacoustic metrics can clearly differentiate the interior noise between the idling state and constant speed state presented in Fig. 11(a), however, in the t-SNE map causes confusion for other working conditions. Therefore, each type of extracted sound feature has its own defect and cannot classify different types of vehicle interior noise clearly. However, this knowledge is important for noise feature extraction and model development. Thus, a better feature set of vehicle interior noises should be studied further.

The psychoacoustic metrics and energy features are different properties of a sound, and they might complement each other according to the result of t-SNE 2-D maps. Therefore, in this study, we fuse noise characteristics by combining the psychoacoustic metrics and the energy features and then form new feature sets, which could integrate the advantages of both methods. Similarly, these newly formed fusing features are validated through the t-SNE method, and the visualization results are presented in Fig. 12. This figure illustrates that the effectiveness of the new features increased by varying degrees, especially for the feature combination of critical band energy and psychoacoustic metrics shown in Fig. 12(d), which separates the vehicle interior noise of four working conditions clearly except for one noise sample. Therefore, this feature set is better than other developed sets and can be selected as an input feature of the intelligent model for evaluating the sound quality of vehicle interior noise.

5. Development of the CRBM-DBN-based sound quality prediction model

5.1. Performance measurements

Four criteria are introduced to measure the performances of a developed CRBM-DBN model: mean absolute percentage error (MAPE), root mean square error (RMSE), variance (VAR) and Pearson's correlation coefficient (CORR). Each formula of the four measurements is summarized as follows [44]:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{x_i - x_i'}{x_i}\right| \times 100 \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_i')^2} \tag{14}$$

$$VAR = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2 \tag{15}$$

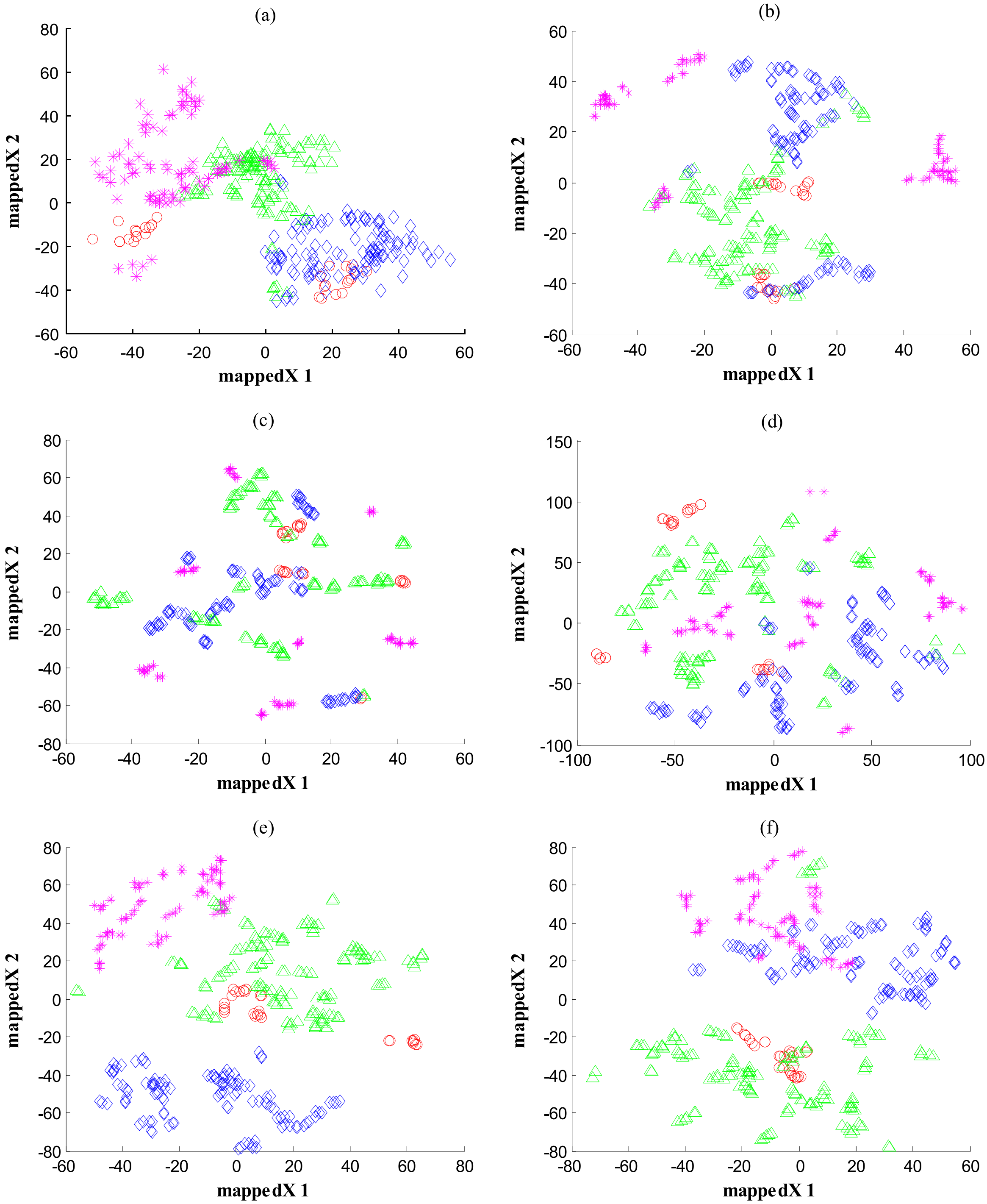

Fig. 11. t-SNE 2-D maps of the extracted interior noise features. (a) Psychoacoustic metrics. (b) WT energy. (c) WPT energy. (d) IMF energy. (e) Critical band energy. (f) Mel scale energy. '○': Idling state. '▵': Constant speed state. '◇': Acceleration state. '*': Deceleration state.

$$CORR = \frac{N\sum_{i=1}^{N} x_i x_i' - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i'}{\sqrt{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}\sqrt{N\sum_{i=1}^{N} x_i'^2 - \left(\sum_{i=1}^{N} x_i'\right)^2}} \tag{16}$$

where x_i is the actual value, x_i' is the predicted value, $\bar{x}$ is the mean value of x_i, and N is the number of samples.

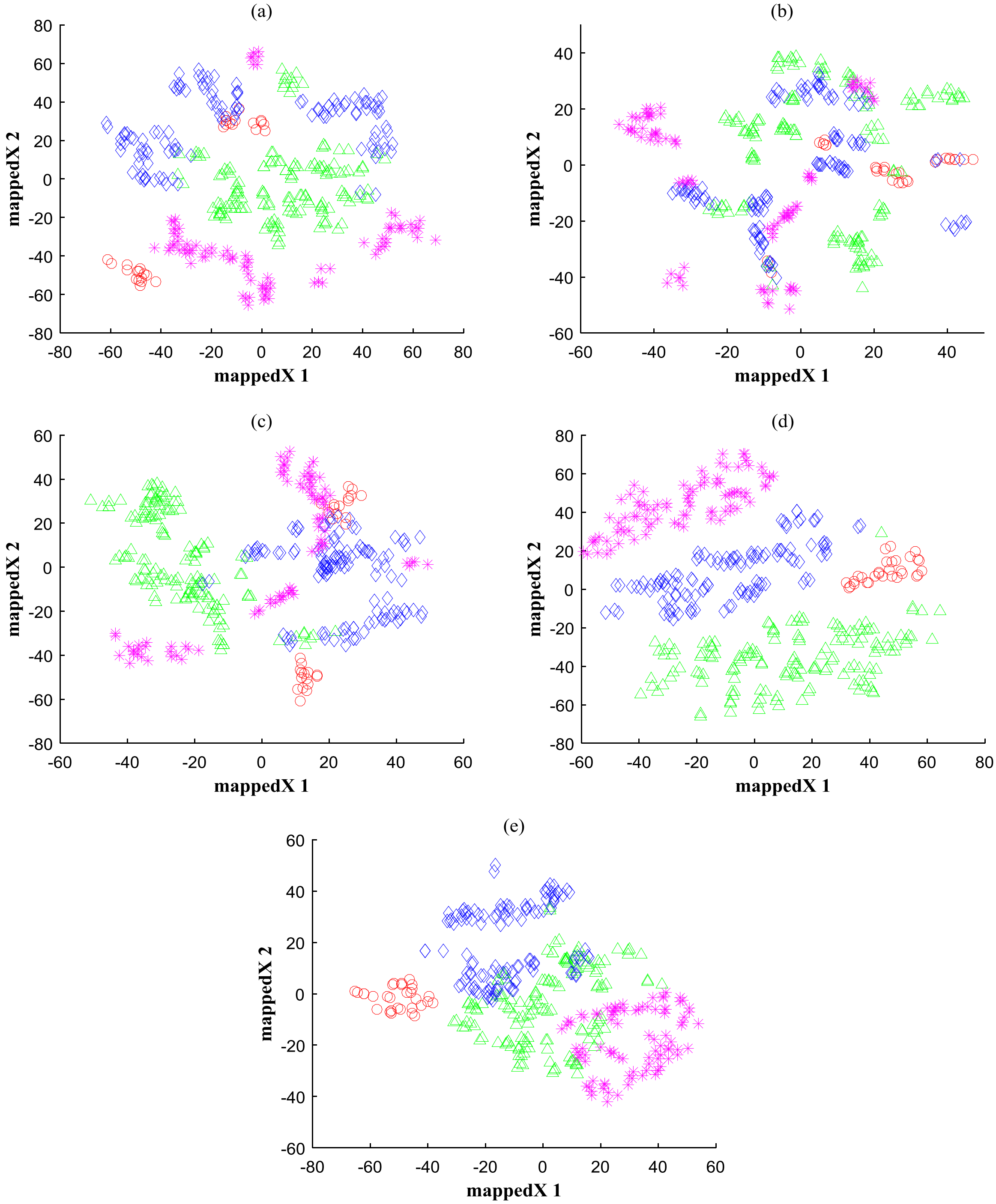

Fig. 12. t-SNE 2-D maps of the developed fusing feature sets. (a) WT energy + psychoacoustic metrics. (b) WPT energy + psychoacoustic metrics. (c) IMF energy + psychoacoustic metrics. (d) Critical band energy + psychoacoustic metrics. (e) Mel scale energy + psychoacoustic metrics. '○': Idling state. '▵': Constant speed state. '◇': Acceleration state. '*': Deceleration state.

These criteria are frequently applied in research to assess model performance. The MAPE and RMSE are real numbers greater than zero and are inversely related to the accuracy of the prediction. In addition, the MAPE and RMSE are mean-based measures and cannot illustrate the variations of the results of different runs. Because stability is also an important target for intelligent models, we calculate the variance of the mean-based measures of 50 runs to show stability. In other words, we run each model 50 times, obtain 50 MAPE and RMSE values and then compute their variances, VAR (MAPE) and

VAR (RMSE). CORR is an index used to measure the linear relationship between the predicted interior sound quality and subjective evaluation. In contrast to the mean-based measures, CORR is proportional to the model performance.

5.2. Designing the architecture of CRBM-DBN

There are two critical parameters in the design of a CRBM-DBN model: the number of hidden layers (maxlayer) and the number of units in each hidden layer (hidden units). In addition, many other free parameters should be decided. The hidden units allow the CRBN-DBN model to capture the nonlinear patterns from input data. An insufficient number of hidden units leads to under-fitting, which may not be suitable for modeling the data, whereas an excessive number of hidden units would lead to overfitting, which may result in poor prediction performance. CRBM-DBN is a probability generation model with many hidden layers, which allow the model to be more powerful for modeling the complex and abstract relationships of data. Therefore, selecting the maxlayer and hidden units provides researchers with a great deal of freedom. However, no general method exists to guide the determination of these parameters, and the architecture design of an appropriate multilayer CRBM-DBN model remains difficult.

Previous studies have shown that the performance of a deep belief network is more sensitive to the model architecture than to certain free parameters [36]. In this paper, we first determine the model structure (i.e., the maxlayer and hidden units) via a cut-and-try method with experimentally selected free parameters and then modify these free parameters by using the determined model structure. Ten levels of hidden units ranging from 4 to 40 with intervals equal to 4 were applied in this study. Because the number of input nodes corresponds to the dimensionality of the input feature set, the neural units of the model input layer can be fixed at 32 (Psychoacoustic metrics 8 + critical band energy 24). One node in the output layer is sufficient to meet our needs for predicting vehicle interior sound quality. Selection of the maxlayer is an important problem. Extensive experiments by Le Roux and Bengio [45] suggest that several hidden layers are better than one in most cases. However, there is not yet an optimal maxlayer in theory, and the best way to determine the optimal maxlayer is still based on the experimental method and task at hand. The data were divided into training, validation and testing sets, which consist of 60%, 20% and 20% of the database, respectively, and each dataset has a similar proportion of noise samples obtained from different vehicle types and under different working conditions. Before the experiment, each attribute of the data was normalized to the range of $[-1, 1]$.

To design an optimal CRBM-DBN architecture, we first initialize the CRBM-DBN model with 1 hidden layer, 32 input nodes and one output node. The model was trained by the training set, and the effects of hidden units on the prediction performance of the vehicle interior sound quality on the validating set are presented in Table 6. The MAPE, RMSE and CORR of each designed model listed in Table 6 are the average values of 50 runs. The best performance across the MAPE, RMSE and CORR occurs at 16 hidden units, with values of 3.6698 0.3060 and 0.9102, respectively. The lowest MAPE variance is 0.1307 at 28 hidden units, and the lowest RMSE variance is 0.0011 at 8 hidden units. Additionally, the five criteria change irregularly as the number of hidden units increased. Hence, we fix the number of units in the first hidden layer at 16 and continue to study the second layer.

Based on the analysis results above, we initialize the CRBM-DBN model with 2 hidden layers, with 16 hidden units in the first hidden layer and a combination of different numbers of units in the second hidden layer. The experimental results are summarized in Table 7, which illustrates that, except for the RMSE variance, the other 4 measurements reach the best performance at 20 hidden units. The average values of MAPE, RMSE, CORR and the variances of MAPE and RMSE in the model with 2 hidden layers are 3.3517, 0.2581, 0.9204, 0.1204 and 0.0016, respectively, which are better than the average performances of the CRBM-DBN model with only one hidden layer.

We then initialized CRBM-DBN with 3 hidden layers and with 16 and 20 units in the first 2 hidden layers. Table 8 shows the experimental results, and the best performance is obtained at 8 hidden units. A comparison of the performance among

Table 6
Effect of the number of hidden units in the first layer on the validating performance.

Hidden units	MAPE	RMSE	CORR	VAR (MAPE)	VAR (RMSE)
4	3.8410	0.3170	0.8884	0.1690	0.0022
8	3.9065	0.3199	0.9068	0.1681	**0.0011**
12	3.8784	0.3532	0.8960	0.1671	0.0048
16	**3.6698**	**0.3060**	**0.9102**	0.1573	0.0018
20	3.9357	0.3123	0.8998	0.3332	0.0083
24	4.1085	0.3229	0.8890	0.2625	0.0021
28	4.0880	0.3189	0.9050	**0.1307**	0.0020
32	3.7160	0.3281	0.8812	0.1461	0.0022
36	3.8505	0.3299	0.8922	0.2117	0.0029
40	3.9602	0.3149	0.8859	0.2931	0.0037
Average	3.8955	0.3223	0.8955	0.2039	0.0031

Table 7
Effect of the number of hidden units in the second layer on the validation performance.

Hidden units	MAPE	RMSE	CORR	VAR (MAPE)	VAR (RMSE)
4	3.3069	0.2637	0.9117	0.0985	0.0016
8	3.6473	0.2660	0.9283	0.1092	0.0016
12	3.1613	0.2410	0.9195	0.0971	**0.0008**
16	3.1856	0.2488	0.9244	0.1576	0.0019
20	**3.1527**	**0.2312**	**0.9297**	**0.0874**	0.0012
24	3.1951	0.2585	0.9150	0.1492	0.0014
28	3.3687	0.2585	0.9278	0.1176	0.0011
32	3.4881	0.2787	0.9120	0.1562	0.0031
36	3.4416	0.2778	0.9234	0.1139	0.0021
40	3.5695	0.2564	0.9128	0.1175	0.0012
Average	3.3517	0.2581	0.9204	0.1204	0.0016

Table 8
Effect of the number of hidden units in the third layer on the validation performance.

Hidden units	MAPE	RMSE	CORR	VAR (MAPE)	VAR (RMSE)
4	3.4080	0.2887	0.9079	0.1525	0.0014
8	**3.3260**	**0.2687**	**0.9228**	0.1769	**0.0013**
12	3.4936	0.2858	0.9054	0.2371	0.0072
16	3.6205	0.2786	0.9116	0.2190	0.0028
20	3.4430	0.2864	0.9109	**0.1172**	0.0018
24	3.4840	0.2771	0.9106	0.2286	0.0048
28	3.6527	0.2999	0.9060	0.1677	0.0017
32	3.6092	0.2766	0.9107	0.1714	0.0014
36	3.3528	0.2702	0.9131	0.1779	0.0024
40	3.5006	0.2741	0.9014	0.1816	0.0016
Average	3.4890	0.2806	0.9101	0.1830	0.0026

Tables 6–8 illustrates that the model with 2 hidden layers outperforms that with 3 hidden layers in terms of the best and average results, whereas the model with one hidden layer performs the worst.

To further analyze the effects of more hidden layers and hidden units, the foregoing experiment has been repeated for the CRBM-DBN model with 4, 5 and 6 hidden layers. The results of the experiment are shown in Fig. 13. The MAPE and RMSE values first decreased and reached their minimum values at the 2nd hidden layer before increasing with further increases in the number of hidden layers and finally decreasing at the 6th hidden layer. The correlation coefficients present the opposite trends regarding MAPE and RMSE, as shown in Fig. 13(c). The MAPE and RMSE variances changed irregularly with increasing maxlayer, but their average values were minimized at the 2nd hidden layer. Therefore, the CRBM-DBN model was selected with 2 hidden layers, and the optimal structure of the model is 32-16-20-1. Consequently, the best combination is 32 input units, 16 units in the first hidden layer, 20 units in the second hidden layer and one output unit, which we chose for the following analysis.

There are still other free parameters to determine in the CRBM-DBN model. The learning rates η_w and η_a in Eqs. 10 and 11 and the constant parameter μ in Eq. 8 should be predefined by the user. In this study, an experiential fourfold cross-validation method has been adopted on the training dataset to determine the values of these free parameters. The best combination of these parameters is selected as η_w=0.8, η_a=0.6, and $\mu = 0.3$. In addition, the boundary parameters θ_L and θ_H should be set to the minimum and maximum values of the training dataset [37]. Thus, the values of θ_L and θ_H are set to -1 and 1, respectively. The weight matrix $\{w_{ij}\}$ and noise control parameter $\{\alpha_j\}$ were randomly initialized and then updated during the pre-training and fine-tuning phases. The pre-training phase provides an optimal initial value in the training process but often cannot reach the minimum directly. Therefore, the fine-tuning process is introduced to achieve convergence of the minimum quickly via the backpropagation algorithm. Based on experiments, the numbers of epochs for the pre-training and fine-tuning phases are set to 200 and 300, respectively.

5.3. Model verification and comparison

To verify the effectiveness and generalization of the newly developed model, the trained CRBM-DBN model is applied to evaluate the vehicle interior sound quality on the testing set. Meanwhile, the MLR, BPNN and SVM have been introduced for

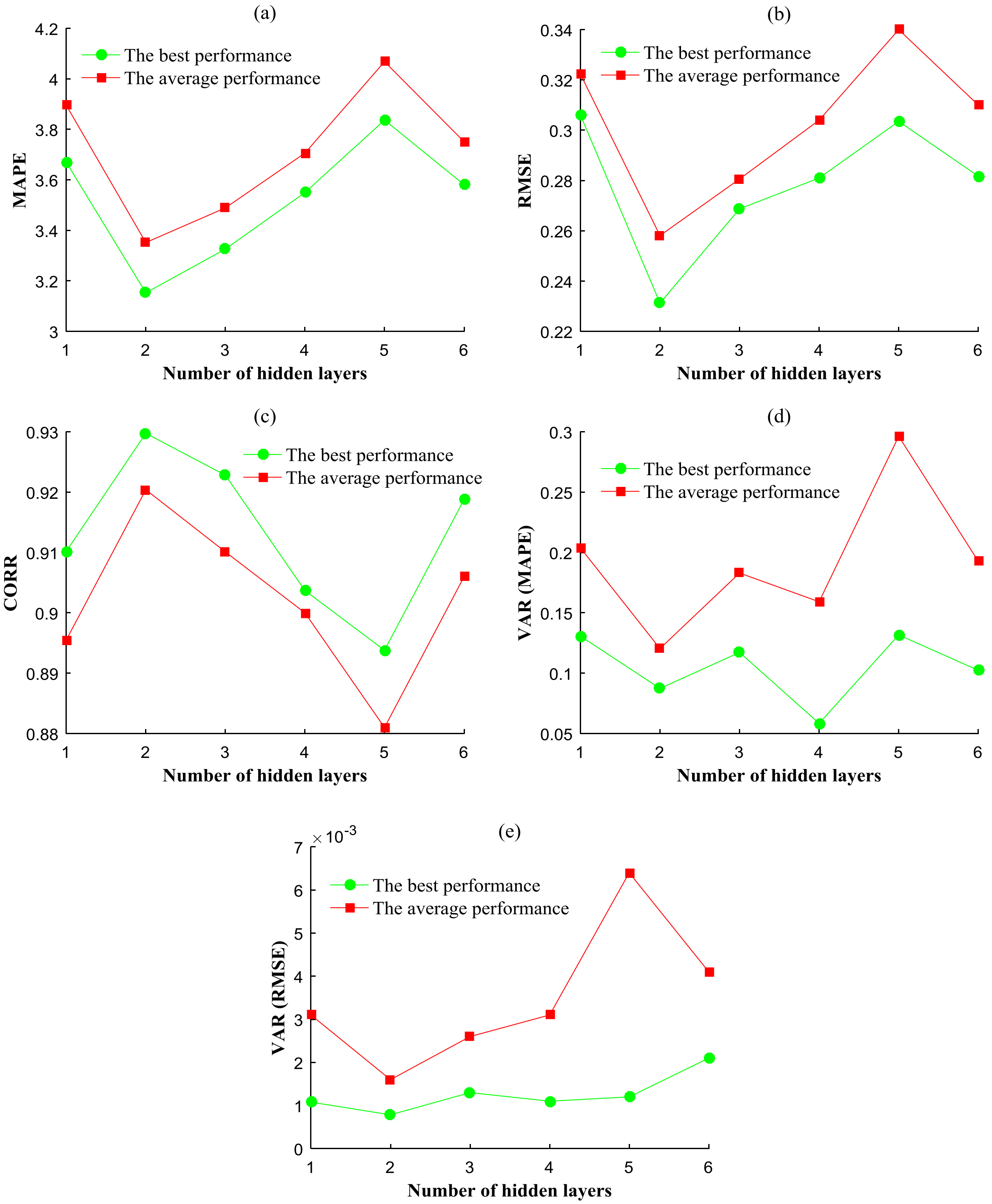

Fig. 13. Effects of the maxlayer for the five criteria. (a) MAPE error. (b) RMSE error. (c) CORR. (d) Variance of MAPE. (d) Variance of RMSE.

comparison to further verify the performance of the developed model. The optimal architectures of each comparison model used in this experiment are as follows. The MLR has 32 independent variables, and the ordinary least squares (OLS) estimator is adopted to learn the training data. The BPNN structure has three processing layers: one input layer, one hidden layer and one output layer, with 32, 40, and 1 neuron units in each layer, respectively. The transfer functions for the hidden layer and output layer are selected as ‘tansig’ and ‘purelin’, respectively, and the ‘LM’ algorithm is set as the training function. An SVM model with the Gaussian kernel function is applied, in which the soft margin parameter C and kernel

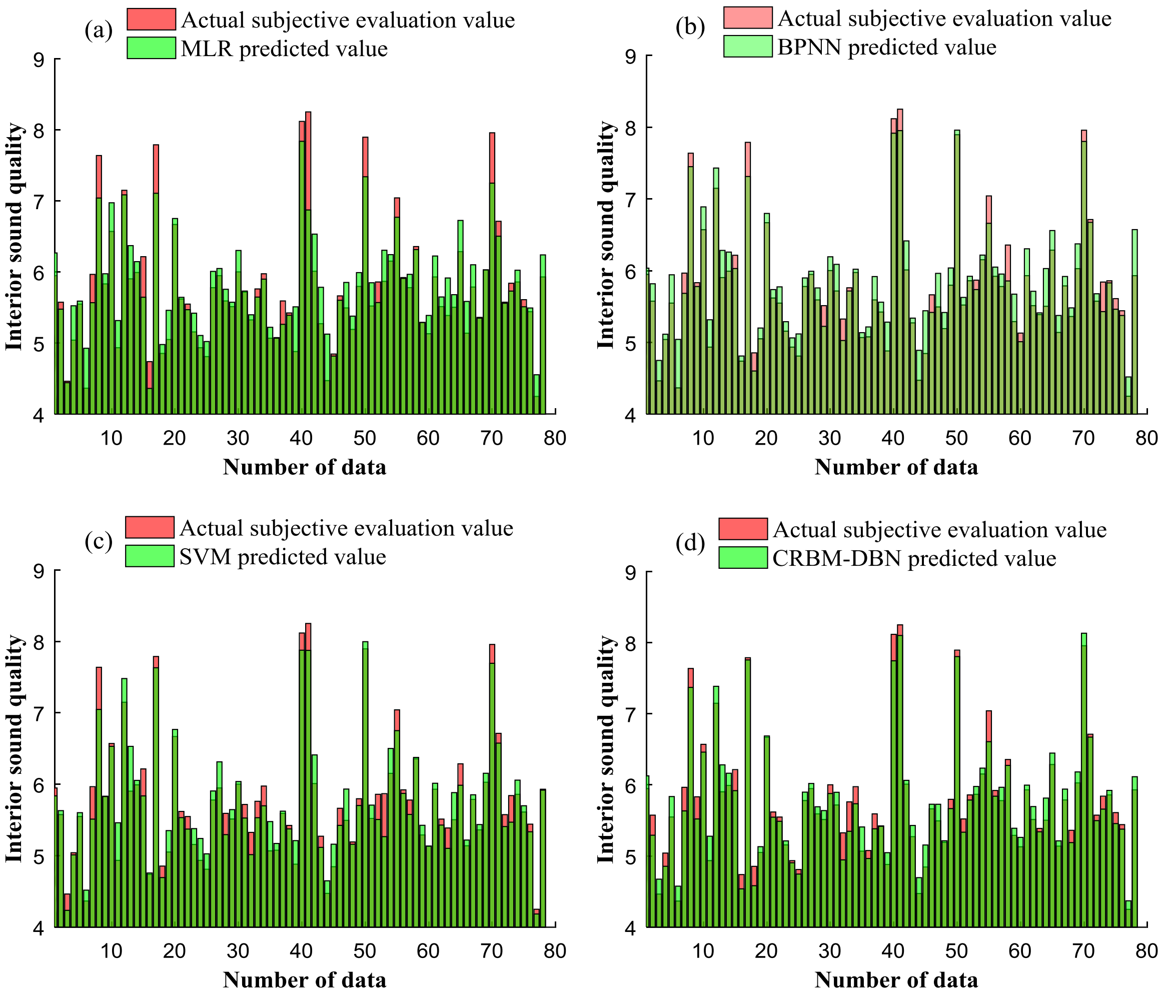

Fig. 14. Comparison results between the predicted and actual values in the testing dataset. (a) MLR. (b) BPNN. (c) SVM. (d) CRBM-DBN.

parameter γ are determined via a fourfold cross-validation and 10-level grid search method. To account for the stochastic characteristic of intelligent algorithms, the prediction processes for each method are repeated 50 times, and the average values are reported.

Fig. 14 shows the prediction results of the testing dataset compared with the actual subjective ratings, which indicates that all four methods have the capability to objectively evaluate the vehicle interior sound quality well. The relative error between the predicted and actual values is presented in Fig. 15. The scatter points of the MLR prediction error are more dispersed from the centerline (zero line) than the other three methods. In addition, the prediction errors of the MLR, BPNN and SVM are greater than ±0.5 for several noise samples, whereas the prediction errors of CRBM-DBN are all smaller than ±0.5. This illustrates that the CRBM-DBN model can evaluate the vehicle interior sound quality more accurately, possibly because of its capacity to learn the highly complex and nonlinear relationship between the noise features and the subjective ratings by encoding a richer and higher-order network architecture in the deep learning process with both unsupervised pre-training and supervised fine-turning phases.

The numeric calculation results are summarized in Table 9. The values of the error-based criteria for the MLR, BPNN, SVM and CRBM-DBN decreased gradually. The MAPE and RMSE decreased from 4.7128 and 0.3646 to 3.2722 and 0.2587, respectively. The correlation coefficients of the MLR, BPNN, SVM and CRBM-DBN are 0.8793, 0.8924, 0.9055 and 0.9187, respectively, which shows that the CRBM-DBN model outperforms the other three methods in the relationship with the subjective evaluation. The nonlinear models (BPNN, SVM and CRBN-DBN) outperform the linear model (MLR), which implies that the characteristics of a sound perceived by the human auditory system are different from those of the sound being produced. In other words, nonlinear models could fit these abstract characteristics more precisely than linear models. Regarding the stability of the four models, the variances of the MAPE and RMSE of the MLR model are both zero because MLR is a deterministic mathematical model when the experimental dataset (training dataset) is fixed. That is, the MLR result

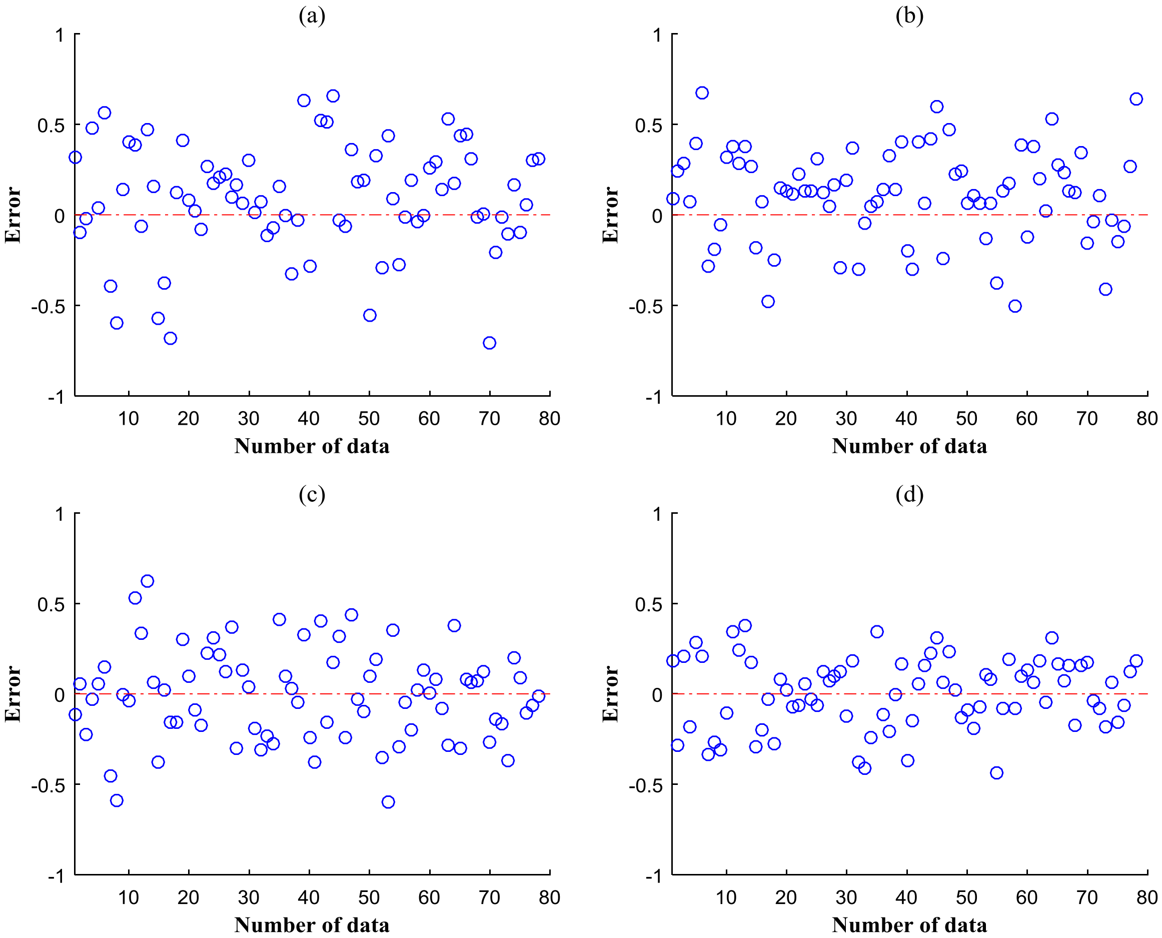

Fig. 15. Relative error between the predicted and actual values. Relative error = prediction value – actual value. (a) MLR. (b) BPNN. (c) SVM. (d) CRBM-DBN.

Table 9
Prediction results of the four methods in the testing dataset.

Method	MAPE	RMSE	CORR	VAR (MAPE)	VAR (RMSE)
MLR	4.7128	0.3646	0.8793	0	0
BPNN	4.1956	0.3267	0.8924	0.4353	0.0075
SVM	3.6434	0.2825	0.9055	0.2208	0.0051
CRBM-DBN	3.2722	0.2587	0.9187	0.1516	0.0033

will not change between different runs in this study, so a variance of zero is obtained. However, the data-driven models, such as the BPNN and CRBM-DBN, which have many hyper-parameters to be determined randomly or with probability, yield variable results in different runs. Similarly, the random cross-validation and grid search for parameter selection in the SVM may also result in different results for each run. The MAPE variances for BPNN, SVM and CRBN-DBN in 50 runs are 0.4353, 0.2208 and 0.1516, respectively, and there is a similar decreasing trend in the RMSE variance. Thus, CRBM-DBN is superior to BPNN and SVM in terms of stability and reliability and also outperforms the other three methods in terms of sound quality evaluation of vehicle interior noise.

6. Conclusion

In this paper, a vehicle road test was conducted and indicates that the subjective evaluation of recorded interior noises is highly related to the vehicle type and working conditions. The noise features of interior sounds were extracted using psychoacoustic and energy-based methods, and then a feature fusion process was applied to these extracted features. A fusing feature set combining psychoacoustic metrics and critical band energy was obtained, which represents the characteristics of interior noises more effectively than the other combinations considered. Furthermore, an intelligent approach, CRBM-DBN, which substitutes the CRBM for the RBM in the DBN to model continuous data, was developed to evaluate the sound quality of vehicle noise using the fusing feature set as input data. Experimental verification and comparisons demonstrated that the CRBM-DBN model is more accurate than the MLR, BPNN and SVM models, with MAPE, RMSE and CORR values of 3.2722, 0.2587 and 0.9187, respectively. In addition, CRBM-DBN outperforms the BPNN and SVM models in terms of reliability and stability, with MAPE and RMSE variances of 0.1516 and 0.0033, respectively. As an intelligent technique, the newly proposed approach can be expended to more general applications to evaluate sound quality and thus may be a promising technique for future use.

Acknowledgments

This work was supported by the Chinese National Science Foundation (Grant no. 51475387), the Key Science and Technology Research Program in Sichuan Province, China (Grant no. 2015GZ0126) and the Science Foundation of Sichuan Province, China (Grant no. 16ZA0010). Special thanks to *China Automotive Technology & Research Center* for their kind supply of the research conditions.

References

[1] Y.S. Wang, G.Q. Shen, H. Guo, X.L. Tang, T. Hamade, Roughness modelling based on human auditory perception for sound quality evaluation of vehicle interior noise, J. Sound Vib. 332 (2013) 3893–3904.

[2] G. Pietila, T.C. Lim, Intelligent systems approaches to product sound quality evaluations – a review, Appl. Acoust. 73 (2012) 987–1002.

[3] K. Genuit, The sound quality of vehicle interior noise: a challenge for the NVH-engineers, Int. J. Veh. Noise Vib. 1 (2004) 158–168.

[4] D. Västfjäll, M.A. Gulbol, M. Kleiner, T. Gärling, Affective evaluations of and reactions to exterior and interior vehicle auditory quality, J. Sound Vib. 255 (2002) 501–518.

[5] H.H. Lee, S.K. Lee, Objective evaluation of interior noise booming in a passenger car based on sound metrics and artificial neural networks, Appl. Ergon. 40 (2009) 860–869.

[6] H.B. Huang, X.R. Huang, R.X. Li, T.C. Lim, W.P. Ding, Sound quality prediction of vehicle interior noise using deep belief networks, Appl. Acoust. 113 (2016) 149–161.

[7] J.H. Yoon, I.H. Yang, J.E. Jeong, S.G. Park, J.E. Oh, Reliability Improvement of a sound quality index for a vehicle HVAC system using a regression and neural network model, Appl. Acoust. 73 (2012) 1099–1103.

[8] Y.S. Wang, C.M. Lee, D.G. Kim, Y. Xu, Sound-quality prediction for nonstationary vehicle interior noise based on wavelet pre-processing neural network model, J. Sound. Vib. 299 (2007) 933–947.

[9] E. Zwicker, H. Fastl, Psychoacoustics: Facts and Models, third ed. Springer-Verlag, Berlin, 2006.

[10] W.A. Aures, The sensory euphony as a function of auditory sensations, Acoustica 58 (1985) 282–290.

[11] W.A. Aures, Procedure for calculating auditory roughness, Acoustica 58 (1985) 268–281.

[12] K.D. Kryter, Methods for the calculation and use of the articulation index, J. Acoust. Soc. Am. 34 (1962) 1689–1697.

[13] W.A. Aures, Berechnungsverfahren fur den Wohlklang beliebiger Schallsignale, Ein Beitrag zur gehorbezogenen Schallanalyse, Munich University, Munich, 1984 (Insert of publication).

[14] Standard ECMA 74-2012, Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment, 2012.

[15] ISO 9568:1993: Cinematography – Background Acoustic Noise Levels in Theatres, Review Rooms and Dubbing Rooms, International Organization for Standardization, Geneva, 1993.

[16] H.B. Huang, R.X. Li, X.R. Huang, M.L. Yang, W.P. Ding, Sound quality evaluation of vehicle suspension shock absorber rattling noise based on the Wigner–Ville distribution, Appl. Acoust. 100 (2015) 18–25.

[17] S.K. Lee, H.W. Kim, E.W. Na, Improvement of impact noise in a passenger car utilizing sound metric based on wavelet transform, J. Sound Vib. 329 (2010) 3606–3619.

[18] Y.F. Xing, Y.S. Wang, L. Shi, H. Guo, H. Chen, Sound quality recognition using optimal wavelet-packet transform and artificial neural network methods, Mech. Syst. Signal Process. 66 (2016) 875–892.

[19] C. Yang, D.J. Yu, Research on the sound metric of door slamming sound based on pseudo Wigner-Ville distribution, J. Mech. Eng. 47 (2011) 91–96.

[20] A. Mohamed, G.E. Dahl, G. Hinton, Acoustic modeling using deep belief networks, IEEE Trans. Audio Speech 20 (2012) 14–22.

[21] H. Kim, S. Lee, E. Na, Sound quality evaluation of the impact noise induced by road courses having an impact bar and speed bumps in a passenger car, in: Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, vol. 224, 2010, pp. 735–747.

[22] H. Liu, J. Zhang, P. Guo, F. Bi, H. Yu, G. Ni, Sound quality prediction for engine-radiated noise, Mech. Syst. Signal Process. 56 (2015) 277–287.

[23] Y.S. Wang, G.Q. Shen, Y.F. Xing, A sound quality model for objective synthesis evaluation of vehicle interior noise based on artificial neural network, Mech. Syst. Signal Process. 45 (2014) 255–266.

[24] H.B. Huang, R.X. Li, X.R. Huang, T.C. Lim, W.P. Ding, Identification of vehicle suspension shock absorber squeak and rattle noise based on wavelet packet transforms and a genetic algorithm-support vector machine, Appl. Acoust. 113 (2016) 137–148.

[25] S. Ding, H. Li, C. Su, J. Yu, F. Jin, Evolutionary artificial neural networks: a review, Artif. Intell. Rev. 39 (2013) 251–260.

[26] Dong Yu, Li Deng, Deep learning and its applications to signal and information processing [exploratory dsp], IEEE Signal Process. Mag. 28 (1) (2011) 145–154.

[27] A. Mohamed, E.G. Dahl, G. Hinton, Acoustic modeling using deep belief networks, IEEE Audio Speech 20 (2012) 14–22.

[28] L.X. Zhang, J. Wu, Deep belief networks based voice activity detection, IEEE Audio Speech 21 (2013) 697–710.

[29] G.B. Huang, H. Lee, M.E. Learned, Learning hierarchical representations for face verification with convolutional deep belief networks, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, 2012, pp. 2518–2525.

[30] G.E. Hinton, S. Osindero, Y.W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (2006) 1527–1554.

[31] G.E. Hinton, A practical guide to training restricted Boltzmann machines, Momentum 9 (2012) 926.

[32] G.E. Hinton, L. Deng, D. Yu, G.E. Dahl, A.R. Mohamed, N. Jaitly, B. Kingsbury, Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups, IEEE Signal Process. Mag. 29 (2012) 82–97.

[33] V.T. Tran, F. AlThobiani, A. Ball, An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks, Expert Syst. Appl. 41 (2014) 4113–4122.

[34] P. Tamilselvan, P. Wang, Failure diagnosis using deep belief learning based health state classification, Reliab. Eng. Syst. Saf. 115 (2013) 124–135.

[35] E.M. Schmidt, Y.E. Kim. Learning emotion-based acoustic features with deep belief networks, in: Proceedings of the 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011, pp. 65–68. ⟨http://dx.doi.org/10.1109/ASPAA.2011.6082328⟩.

[36] F. Shen, J. Chao, J. Zhao, Forecasting exchange rate using deep belief networks and conjugate gradient method, Neurocomputing 167 (2015) 243–253.

[37] H. Chen, A.F. Murray, Continuous restricted Boltzmann machine with an implementable training algorithm, IEEE Proc. – Vis. Image Signal Process. 150 (2003) 153–158.

[38] H. Chen, A. Murray, A continuous restricted Boltzmann machine with a hardware-amenable learning algorithm, in: J.R. Dorronsoro (Ed.), Proceedings of 12th International Conference on Artificial Neural Networks, Springer Berlin Heidelberg, Berlin, 2002, pp. 358–363.

[39] GB/T 18697-2002: Acoustics – Method for Measuring the Vehicle Interior Noise, 2002.

[40] S. Lee, T.G. Kim, J.T. Lim, Characterization of an axle-gear whine sound in a sports utility vehicle and its objective evaluation based on synthetic sound technology and an artificial neural network, Proc. Inst. Mech. Eng. D – J. Automob. 222 (2008) 383–396.

[41] DIN, 45631-1991: Acoustics – Procedure for Calculaiting Loudness Level and Loudness, 1991.

[42] L. Maaten, G.E. Hinton, Visualizing high dimensional data using t-SNE, J. Mach. Learn. Res. (2008) 2579–2605.

[43] L. Maaten, Accelerating t-SNE using tree-based algorithms, J. Mach. Learn. Res. 9 (2014) 3221–3245.

[44] H.B. Huang, R.X. Li, X.R. Huang, M.L. Yang, W.P. Ding, Prediction of a suspension shock absorber sound metric based on sample entropy and ELM-adaboost, J. Shock Vib. 13 (2016) 125–133.

[45] N. Le Roux, Y. Bengio, Representational power of restricted Boltzmann machines and deep belief networks, Neural Comput. 20 (2008) 1631–1649.