Sound Quality Prediction of Vehicle Interior Noise using Deep Belief Networks

Hai B. Huang ^{a, b}, Xiao R. Huang ^a, Ren X. Li ^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, Ohio, USA

*Corresponding author, Tel: (+86)87600695, E-mail: dwp@home.swjtu.edu.cn

Abstract: The sound quality of vehicle interior noise strongly influences passengers' psychological and physiological perceptions. To predict the sound quality of interior noise, a vehicle road test with four compact cars has been conducted. All recorded interior noise signals have been denoised via a discrete wavelet transform (DWT) denoising procedure and subsequently evaluated subjectively through the anchor semantic differential (ASD) test by a jury. In addition, a novel prediction method, namely, regression-based deep belief networks (DBNs), which substitute the support vector regression (SVR) layer for the linear softmax classification layer at the top of the general DBN's structure, has been proposed to predict the interior sound quality. The parameter selection of the DBN model has been compared and studied using a grid search. In addition, four conventional machine-learning-based methods have been introduced to enable a comparison of the performance with the newly developed DBNs. Furthermore, the feature fusion ability of DBNs has been studied by varying the amount of information that the dataset offers. The results show the following: 1) The accuracy and robustness of the proposed DBN-based sound quality prediction approach are better than those of the 4 other referenced

methods. 2) The multiple-feature fusing process can strongly affect the prediction performance. 3) Finally, the unsupervised pre-training process of the DBNs can enhance the information fusing ability. Finally, the newly proposed regression-based DBN approach may be extended to address other vehicle noises in the future.

Key words: Sound quality, Vehicle interior noise, Deep belief networks, Machine learning, Feature fusion.

1. Introduction

Vehicles improve efficiency and have become one of the most important non-human entities in individuals' daily lives. Many people spend substantial amounts of time in vehicles; thus, the comfort and sound quality, i.e., NVH (Noise, Vibration and Harshness), of a vehicle have become increasingly important to passengers. In addition, vehicle noises, which constitute more than 70% of urban noise [1], have been studied for many years. Vehicle interior sound quality evaluation (SQE), or sound quality prediction (SQP), has accordingly become increasingly important in vehicle design [2]. A bad vehicle interior sound quality would not cause hearing damage; however, it would negatively affect passenger psychology and physiology [3]. In addition, the sound quality of a vehicle is a key factor in customers' purchasing decisions. Therefore, it is necessary and reasonable to apply sound quality prediction to passenger cars, as has been performed by many researchers and engineers in recent decades [2, 4-10].

Different types of cars have different types of interior sound quality, including sporty, conservative, powerful, and refined. To improve the sound quality in vehicle designs, researchers and engineers should first understand how to evaluate interior noise [11]. Consequently, a substantial number of research studies related to sound quality evaluation have been conducted

and have produced a relatively satisfactory subjective evaluation system [7]. The characteristics of interior noise perceived by passengers have been found to be slightly dissimilar to the characteristics of the sound being emitted [12]; to some extent, the characteristics are the result of a process related to the physical structural effects of the human hearing system. Therefore, many psychoacoustics-based metrics, such as loudness, sharpness, fluctuation strength, and roughness, have been developed and applied to predict interior sound quality in, for example, vehicles, trains [13] and airplanes [14], due to their good quantitative explanation of noise characteristics. However, further studies on the prediction of interior sound quality have shown that the relationship between human sensations and acoustical motivation presents strong nonlinearity, and as such, no dominant psychoacoustics-based metric representation for sound quality prediction has emerged.

The strong nonlinear relationship between vehicle interior noise and subjective perception makes it an especially active research problem for the application of signal processing and feature learning. Thus, various non-stationary, or transient, analysis methods and machine learning techniques have been introduced to address this issue. Correspondingly, a wide range of practical applications for vehicle interior sound quality prediction have been reported in the literature, where successful methods include the wavelet transform (WT) [6], empirical mode decomposition (EMD) [15-17], artificial neural networks (ANNs) [18] and statistical learning models (SLMs) [19, 20]. In addition to general interior noise, there are various other types of vehicle noise, such as heating, ventilation and air conditioning noise (HVAC) [9]; suspension shock absorber rattling noise [3]; and axle gear noise [21, 22], which have been objectively predicted based on the above advanced methods. Despite their success, the more effective prediction of vehicle interior sound

quality remains an intricate problem, and substantial work remains for the following three reasons.

First, the representational ability of a sound quality feature is limited. Although there are many sound quality metrics for estimating human hearing perception, features seldom perform overwhelmingly better than other features. Second, most generally used models, such as ANNs, support vector machines (SVMs) and logical regression (LR), are shallow models, namely, these models only have one or even no hidden layer (Multi-hidden-layer ANNs will not be discussed in this paper due to their defect of gradient diffusion) and thus cannot fully consider the space distributions of features or learn the deeper feature representations alone. Third, the information fusion performance obtained based on the advantages of using multiple sound quality features is not fully utilized. Machine-learning-based methods have been found to be superior to non-machine-learning-based methods in certain regards when applied to the feature fusion task; however, fusion methods continue to lack adequate study.

Despite the challenges in predicting vehicle interior sound quality, a new deep machine learning method, called deep belief networks (DBNs), brings new hope to addressing the above problems. DBNs were first proposed by G. E. Hinton in 2006 [23] and currently receive substantial attention in both the signal processing field and machine learning field while being successfully applied to pattern recognition problems, especially image classification [24], speech recognition [25], and natural language feature learning [26]. DBNs represent a probability generation model with a hierarchical structure and are stacked by many restricted Boltzmann machines (RBMs) [27]. In addition, DBNs provide various advantages such as the ability to encode deeper and higher order network topological structures and the ability to prevent over-fitting and falling into local minimum through a special unsupervised pre-training phase [28].

Analysis of the similarity between sound quality prediction and pattern recognition in the above successful applications has motivated collaboration with their analysis process. Therein lies the great potential in utilizing state-of-the-art deep machine learning techniques to address the challenges faced in the prediction of vehicle interior noise. However, generally, almost all DBN research is focused on classification problems; the application of regression issues has received very little attention. Thus, in this paper, we propose a novel regression-based DBN method that replaces the top classification layer of DBNs by a regression layer. Moreover, this method could be easily extended to any regression-based machine learning problem.

This paper is organized as follows. In section 2, we briefly review related work on sound quality prediction methods. In section 3, the basic techniques of DBNs are introduced, and a regression-based DBN model is proposed. In section 4, we describe the experiments and methods used in sound quality prediction. Section 5 reports the experimental results and further analyzes the superior performance of DBNs. Finally, section 6 gives the conclusion and presents some future research topics.

2. Related works

The prediction of vehicle interior sound quality has been studied for many years. Roughly speaking, the reported methods for sound quality prediction in the literature can be briefly categorized into two classes, namely, psychoacoustics-based methods and machine-learning-based methods, both of which will be introduced below.

2.1 Psychoacoustics-based interior sound quality prediction

To numerically quantify acoustical stimuli, the ISO (International Standardization Organization) has generalized objective sound quality evaluation methods into two types [29]. The

first type is the octave-band-based analysis for linear or weighted sound pressure levels. The most generally used analysis using this method is the one-third octave-band analysis with A-weighting due to its high calculation speed and simplicity. The second type is the critical-band-based analysis for the human sensation of loudness. The masking effect considered in the latter method makes it more effective and accurate than the former method. Typically, the standard ISO 532B [30] has adopted the second method to calculate loudness levels and has been proven to be a very powerful method for sound quality prediction in automotive engineering [7, 31]. In addition to loudness, there have been many other psychoacoustic metrics developed by scientists and engineers, such as sharpness, roughness, fluctuation strength, articulation index, and tonality, although they have not been completely recognized. These sound quality metrics have been widely utilized in research on interior noise.

Y.S. Wang etc. [8] used a roughness index to evaluate vehicle interior noise and proposed a modified roughness metric, HAP-RM, based on human auditory perception; their method provided good results in analyzing stationary and non-stationary sound signals. E. Parizet etc. [32] analyzed the car door closing sound quality based on some psychoacoustic metrics and found that two timbre parameters, namely, frequency balance of the sound and its cleanness, were the most effective indices for evaluating this special type of noise. H. H. Lee etc. [33] found that it is difficult to predict the gear whine sound quality when only using the A-weighted sound pressure level but that the loudness, articulation index and tonality can be used to objectively and efficiently evaluate the gear whine noise. H. B. Huang etc. [34] adopted four psychoacoustic metrics to evaluate the vehicle suspension shock absorber rattling noise and found that all generally used sound quality metrics cannot accurately reflect this specific noise. Other studies have been

conducted on psychoacoustic metrics and sound quality problems, as noted in references [8, 35].

In addition, many researchers have integrated the characteristics of practical noises with human hearing procedures as well as advanced signal processing methods, therein developing new sound quality metrics. S. K. Lee etc. [6] introduced a special sound quality index, namely, HFEC (High Frequency Energy Contribution), using WT to evaluate the impact and shock noise of a car. To predict the squeak and rattle (S&R) noise of vehicle suspension shock absorber, H. B. Huang etc. [3] proposed a quantifiable metric, namely, SMWVD (Sound Metric based on the Wigner-Ville distribution), that can extract the S&R parts of suspension shock absorber noise. L. Zhou etc. [36] presented an IEI (Impact Energy Index) metric that uses order analysis and spectral analysis to evaluate and improve the sound quality of vehicle powertrain noise during acceleration. Additionally, FINI (Friction-Induced Noise Intensity) and SQPWVD (Slamming Quality based on Pseudo WIGNER-VILLE distribution) have gradually emerged for the prediction of vehicle wiper system noise [37, 38] and car door slamming noise, respectively [39].

2.2 Machine-learning-based interior sound quality prediction

Due to the complex relationship between the human hearing process and acoustic performance, an individual psychoacoustics-based sound quality metric cannot fully reflect the characteristics of noises. Consequently, machine learning techniques have often been used to solve this problem. Machine-learning-based pattern recognition can be broadly divided into classification-based learning methods and regression-based learning methods; this paper focuses on regression-based techniques. A number of state-of-the-art vehicle interior sound quality prediction methods based on machine learning have been applied and seen significant advances. For instance, the relatively simple multiple linear regression (MLR) and the relatively complicated ANN and SVM

techniques have been adopted because of their outstanding prediction ability.

MLR is a typical statistic regression method. MLR provides two advantages: a small amount of data is required, and the method can provide reasonable outputs through simple calculations.

The mathematical expression of MLR is given by

$$y_i = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n + \varepsilon_i \quad i = 1, 2, \dots, m \quad (1)$$

where y_i is the output objective; x is the input feature; n and m are the number of input dimensions and the number of output data, respectively; α is the weighting coefficient related to x ; α_0 is the constant term; and ε_i is the error term. MLR can effectively solve problems whose inner relationship is not very complex; however, it performs poorly when addressing strongly nonlinear issues. On the other hand, ANN and SVM present much better advantages when addressing such issues. An ANN is a multi-layer perceptron (MLP) with three or more layers: one input layer, one or more hidden layers and one output layer. ANNs approach expected targets by optimizing the weights between the layers and the biases of the layer. It has been proven that an ANN with one hidden layer structure can approach any form of output. Similar to ANNs, SVMs also have three layers: the input layer, the interface layer and the output layer. Note that many optimization methods have been developed for SVMs; this paper applies the least squares algorithm due to its excellent performance. SVMs were originally developed for binary classification; when given training data and their corresponding targets $\{(x_n, y_n)\}$, $n=1, 2, \dots, N$, $x_n, y_n \in \mathbb{R}$, $t_n=\{-1, +1\}$, the SVM optimization function with constraints can be written as follows:

$$\begin{aligned} \min_{w, \xi_n} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{n=1}^N \xi_n^k \\ \text{s.t.} \quad & \forall n, \mathbf{w}^T x_n t_n \geq 1 - \xi_n \\ & \forall n, \xi_n \geq 0 \end{aligned} \quad (2)$$

where $\mathbf{w}$ is the normal vector of the hyper plane, C is the soft margin parameter, ξ_n is the

orthogonal distance away from the hyper plane, and k is a positive integer. In general, the input data will be normalized between $[0, 1]$, and the corresponding unconstrained optimization model can be transformed into

$$\min_w \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{n=1}^N \max(1 - \mathbf{w}^T x_n t_n, 0)^k \quad (3)$$

In **Eq. (3)**, if the parameter k is set to 1, the objective of this equation is known as L1-SVM, and if the parameter k is set to 2, the objective of **Eq. (3)** is known as L2-SVM, which is differentiable and with a squared hinge loss. The predicted value y of input data x can be calculated through **Eq. (4)**:

$$y_i = \sum_{j=1}^N \alpha_j K(x, x_k) + b \quad (4)$$

where α_j is the support vector, $K(x, x_k)$ is the kernel function that maps the input data from a low-dimensional feature space to a high-dimensional feature space, and b is the offset value.

The above-mentioned methods as well as other similar techniques have been widely applied to many aspects of vehicle noise. J. H. Yoon etc. [9] used MLR and back-propagation neural networks (BPNN) models to predict HVAC system sound quality, and their results demonstrated that both the MLR and BPNN can be used to evaluate the acoustic quality but that the BPNN obtains a better performance. Y. S. Wang etc. [11] combined the WT technique and a BPNN to predict the sound quality of non-stationary vehicle interior noise. The input noise signal being pre-processed by WT and then fed into the BPNN was found to produce more accurate results compared to only using the psychoacoustic metric as input. K. Manimala etc. [40] and G. R. Ibrahim etc. [41] applied advanced signal processing methods, namely, WT, EMD and WVD, to extract the discriminable features from noise signals to compare their discriminability. They

subsequently applied machine learning methods, namely, an SVM, probabilistic neural network (PNN) and fuzzy k-nearest neighbor classifier (FKNN), as feature fusing tools for regression. Finally, they found that the features extracted from the original signals and the fusing ability of the adopted learning methods were both very important for pattern recognition. In addition, other research reported in the literature [5, 33] presents the similar results.

Despite the achievements obtained by the general machine learning methods, there remains significant potential for improvement. MLP models can easily become trapped in local extremum if the initial weights and biases are not selected appropriately, and increasing the number of hidden layers and units can make the problem computationally intractable. Because different features distribute differently, the predefined kernel combination functions in SVMs and models derived from SVMs poorly express the characteristics of all features simultaneously. More importantly, because the general machine-learning-based methods are shallow models, they are not sufficiently powerful to learn new information on their own. Moreover, because the original features are merely concatenated linearly in the original feature space or kernel-induced feature spaces [42], they cannot fully fuse the diversity of features.

The limited capability of general machine learning methods prevents them from learning and fusing streams of information sourced in noise signals, and solving this issue may require deeper and richer architectures. Therefore, to further improve performance, a deep machine learning technique, namely, DBNs, the advantages of which could overcome most current challenges through the advanced unsupervised pre-training phase and supervised fine-tuning phase, will be introduced in this paper to predict interior noise sound quality. The following sections will detail the procedure for applying this advanced method to predict interior sound quality.

3. Deep belief networks

Simply speaking, DBNs are densely connected hierarchical structures that are stacked by multiple restricted Boltzmann machines (RBMs). In other words, they are composed of many nonlinear hidden layers. Therefore, in this section, we first introduce the architecture of an RBM and then detail the DBN architecture.

3.1 Restricted Boltzmann machine architecture

RBMs are a special type of Markov random field (MRF), which consists of one layer (typically Bernoulli or Gaussian) of binary stochastic visible units $\mathbf{v} \in \{0, 1\}^V$ and one layer (typically Bernoulli) of binary stochastic hidden units $\mathbf{h} \in \{0, 1\}^H$. As noted in the name, the connections between the units within different layers are restricted. In this model, all the visible nodes are concatenated to the hidden nodes; however, there are no connections within the visible-visible or hidden-hidden nodes. The topological structure of the RBM is shown in **Fig. 1**.

Fig. 1 Topological structure of the RBM

In the RBM model, the weights between each unit of different layers and the biases on the individual units define a probability distribution via an energy function. For a Bernoulli-Bernoulli RBM (BB-RBM), the energy function is given as

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\sum_{i=1}^V \sum_{j=1}^H w_{ij} v_i h_j - \sum_{i=1}^V a_i v_i - \sum_{j=1}^H b_j h_j \quad (5)$$

where $\theta = (w, a, b)$ are the model parameters; v_i and h_j are the visible unit i and hidden unit j , respectively; w_{ij} is the weight connection between v_i and h_j , with a_i and b_j their biases; and V and H are the number of visible and hidden units, respectively.

Given the energy function and the model parameters, the joint distribution over the visible units and the hidden units is defined as follows:

$$p(\mathbf{v}, \mathbf{h}; \theta) = \frac{\exp(-E(\mathbf{v}, \mathbf{h}; \theta))}{\sum_v \sum_h \exp(-E(\mathbf{v}, \mathbf{h}; \theta))} \quad (6)$$

Because the visible-visible or hidden-hidden unit connections do not exist, the probabilities of the visible unit v_i and the hidden unit h_j are independent. Hence, the condition distributions $p(\mathbf{h}|\mathbf{v})$ and $p(\mathbf{v}|\mathbf{h})$ can be computed via **Eq. (7)** and **Eq. (8)** as follows:

$$p(h_j = 1 | \mathbf{v}; \theta) = \sigma\left(\sum_{i=1}^V w_{ij} v_i + b_j\right) \quad (7)$$

$$p(v_i = 1 | \mathbf{h}; \theta) = \sigma\left(\sum_{j=1}^H w_{ij} h_j + a_i\right) \quad (8)$$

where $\sigma(x) = 1 + \exp(x)^{-1}$. If the input data are not binary values but rather real-valued stochastic variables, then the Gaussian-Bernoulli RBM (GB-RBM) should be introduced. Similar to the BB-RBM, the energy function of GB-RBM is defined as

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\sum_{i=1}^V \sum_{j=1}^H \frac{w_{ij} v_i h_j}{\eta_i} - \sum_{i=1}^V \frac{(v_i - a_i)^2}{2\eta_i^2} - \sum_{j=1}^H b_j h_j \quad (9)$$

where η_i is the standard deviation of v_i . The joint distribution over the visible units and the hidden units of GB-RBM is the same as **Eq. (6)**. Then, the condition distributions of GB-RBM can be expressed as

$$p(h_j = 1 | \mathbf{v}; \theta) = \sigma\left(\sum_{i=1}^V \frac{w_{ij} v_i}{\eta_i} w_{ij} v_i + b_j\right) \quad (10)$$

$$p(v_i = x | \mathbf{h}; \theta) = \frac{1}{\eta_i \sqrt{2\pi}} \exp\left(-\frac{(-\eta_i \sum_j w_{ij} h_j - a_i + x)^2}{2\eta_i^2}\right) \quad (11)$$

where x is the real-valued input data. Therefore, the GB-RBM can transform the real-valued data into binary variables. This enables the development of a deep architecture for feature learning.

The RBM model parameters θ can be estimated by steepest gradient descent or conjugate gradient descent to maximize the positive log-likelihood probability of training data. The update rule for the weights w_{ij} and biases a_i and b_j is

$$\begin{aligned} \Delta w_{ij} &= \varepsilon(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}) \\ \Delta a_i &= \varepsilon(\langle v_i \rangle_{data} - \langle v_i \rangle_{model}) \\ \Delta b_j &= \varepsilon(\langle h_j \rangle_{data} - \langle h_j \rangle_{model}) \end{aligned} \quad (12)$$

where ε is the learning rate, which ranges from 0 to 1. The term $\langle \cdot \rangle_{data}$ is the expectation over the training data, and the term $\langle \cdot \rangle_{model}$ is the same expectation over the distribution defined by the model [43]. In practice, the expectation $\langle \cdot \rangle_{model}$ is computationally intractable [28]; therefore, an approximation method called the contrastive divergence (CD) is introduced to substitute the $\langle \cdot \rangle_{model}$ term by performing Gibbs sampling to initialize the data. Using this method, the new update rule for Eq. (12) is given as

$$\begin{aligned} \Delta w_{ij} &= \varepsilon(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_k) \\ \Delta a_i &= \varepsilon(\langle v_i \rangle_{data} - \langle v_i \rangle_k) \\ \Delta b_j &= \varepsilon(\langle h_j \rangle_{data} - \langle h_j \rangle_k) \end{aligned} \quad (13)$$

where $\langle \cdot \rangle_k$ denotes the expectation in terms of the distribution of samples after Gibbs sampling for step k . Usually, the 1-step Gibbs sampler can obtain adequate results [23].

To increase the efficiency of the training procedure, the training set can be divided into several “batches” that contain fewer samples than the original set [44]. In addition, to stabilize the

learning process, a momentum term (m) [28] can also be introduced in the RBM update rule; Eq.

(13) can then be updated to

$$\begin{aligned}(\Delta w_{ij})_n &= m(\Delta w_{ij})_{n-1} + \varepsilon(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_k) \\(\Delta a_i)_n &= m(\Delta a_i)_{n-1} + \varepsilon(\langle v_i \rangle_{data} - \langle v_i \rangle_k) \\(\Delta b_j)_n &= m(\Delta b_j)_{n-1} + \varepsilon(\langle h_j \rangle_{data} - \langle h_j \rangle_k)\end{aligned}\tag{14}$$

With the momentum, the updating of the weights and the biases in the current epoch will be related to the corresponding terms in the previous epoch.

3.2 Deep belief network architecture

Based on the above description, it is easy to understand the architecture of DBNs. DBNs are a probabilistic generative model that are stacked by a number of RBMs, and a typical DBN model with three hidden layers is shown in **Fig. 2**. Note that the bottom layer of DBNs will receive the real-value dataset, the first RBM model should be the GB-RBM, and the remaining RBM models are BB-RBM. The learning procedure of the DBNs consists of two phases: pre-training and fine turning.

Fig. 2 Topological structure of the DBNs with three hidden layers (the bias items are omitted for simplicity)

In the pre-training phase, the DBNs will be greedily trained layer by layer from bottom to top. Each RBM is trained individually, and the weights between layers and biases of layers are fixed and saved for further analysis. This process can be expressed as **Eq. (15)**:

$$p(h_j^{(l-1)} = 1 | \mathbf{h}^{(l)}; \theta^{(l)}) = \sigma\left(\sum_{i=1}^V w_{ij}^{(l)} v_i^{(l)} + b_j^{(l-1)}\right) \quad (15)$$

where l is the number of layers, and the other variables are the same as in **Eq. (7)**. The states of the bottom units (visible layer) are written as $\mathbf{h}^{(0)}$; there is a similar form for GB-RBM in terms of **Eq. (10)**. This learning process will be continued until the last hidden layer is trained. The pre-training phase simplifies the optimization of the deep architecture and allows a number of parameters to be used to improve the feature fusing ability as well as prevents overfitting. In the fine-tuning phase, the weights and biases of the DBN model will be updated via the back-propagation algorithm using the labeled input data.

3.3 Regression-based deep belief networks

Classification-based DBNs have been successfully applied to failure diagnosis problems in mechanical engineering research [45, 46]; however, regression-based DBNs have yet to be introduced. Therefore, in this paper, we proposed a novel regression-based DBN model, which uses a linear support vector regression (SVR) layer to substitute the softmax layer on the top of the DBN structure. The SVR-DBNs continue to perform the pre-training phase and the fine-tuning phase in the same manner as the general DBNs. The difference is that the weights and biases of the lower layers are learned by back propagating the gradients from the top SVR layer. Therefore, the SVR objective should be differentiated with respect to the activation of the penultimate layer [47]. In practice, the L1-SVM and the L2-SVM are widely used to solve regression problems. In **Eq. (3)**, we let the objective be $f(w)$ and replace the input $\mathbf{x}$ with the penultimate activation $\mathbf{h}$; then,

the back-propagation error for L1-SVM and L2-SVM can be respectively obtained as

$$\left. \frac{\partial f(\mathbf{w})}{\partial \mathbf{h}_n} \right|_{k=1} = \mathbf{w} - C\mathbf{w}t_n (\prod\{1 > \mathbf{w}^T \mathbf{h}_n t_n\}) \quad (16)$$

$$\left. \frac{\partial f(\mathbf{w})}{\partial \mathbf{h}_n} \right|_{k=2} = \mathbf{w} - 2C\mathbf{w}t_n (\max(1 - \mathbf{w}^T \mathbf{h}_n t_n, 0)) \quad (17)$$

where the superscript T denotes the transposition of a matrix and $\prod$ is the indicator function.

Herein, the learning procedure of the back-propagation algorithm is the same as that of the general

DBNs. From the reports of [45, 47], the L2-SVM has been found to perform better than L1-SVM

in most cases; therefore, we selected the L2-SVM model as the top layer of the DBNs.

Furthermore, we adopted conjugate gradient descent to replace the steepest gradient descent

because the former typically performs slightly better than the latter [48]. Here, we have

summarized the training process for SVR-DBNs in **Fig. 3**. For simplicity, without notation, we use

DBNs to denote SVR-DBNs in the following text.

Fig. 3 Flowchart of the SVR-DBN training procedure

4. Experiments and methods

4.1 Recording and pre-processing of vehicle interior noise

In this measurement, four compact cars manufactured by Volkswagen automotive company, two of which being automatic transmission (AT) and two being manual transmission (MT), were selected as the sample vehicles. The interior noises above the seats of the driver, co-driver and rear passenger were recorded by G.R.A.S. 40HF microphones and an LMS 24-channel acquisition system, as shown in **Fig. 4**, with a signal length of 10 s and a sampling rate of 44.1 kHz. The vehicle tests have been conducted on a one-way and straight test track provided by *China*

Automotive Technology & Research Center, and there were no sound reflecting objects within 30 m. On the test day, the weather was sunny, and no wind was present during the measurements. Following the standard GB/T 18697-2002, which uses similar test settings as the standard ISO 5128 and was developed by *Chinese Environmental Protection Ministry* and *China Automotive Technology & Research Center*, the vehicle working conditions were set as shown in **Tab. 1**. Considering the collection positions for interior noise and the above-mentioned vehicle working conditions, the recording test was performed twice for each point for each vehicle, and a total of 216 noise signals were obtained.

As expected, the recorded interior noises were influenced by background noise and interference induced by the acquisition system. Therefore, the original signals should be denoised. The discrete wavelet transform (DWT) denoising method has been proven to be very effective in many applications; thus, we adopted this method. The DWT denoising method can generally be performed in three steps: 1) decomposing the original signal, 2) denoising the signal in each sub-band, and 3) reconstructing the processed signal. The parameters of the DWT denoising method were set as in **Tab. 2**. After denoising, all the noise signals were spliced into 5 s intervals using the CoolEdit software and saved for further analysis.

Fig. 4 The microphone positions for the vehicle road test

Tab. 1 Working conditions for the vehicle test

#	State setting	Vehicle running conditions
1	Idling state	Neutral shift for MT and “N” shift for AT
2	Constant speed	Running speed at 30 km/h to 100 km/h with the incremental interval equal to 10 km/h. Timely shift for MT and “D” shift for AT.

Tab. 2 Parameter settings of DWT denoising method

Parameter type	Settings	Descriptions
Wavelet function	Daubechies	“db5”
Analysis levels	5	Denoising in $2^5=32$ sub-band
Threshold determination	Soft threshold	Threshold = $2 \log(\text{length}(x))$, where x is the signal

4.2 Subjective evaluation of interior sound quality

To evaluate interior noise based on human perception, a jury test has been conducted with 24 university students (14 males and 10 females) between 20 and 30 years of age and with normal hearing ability. An anchor semantic differential (ASD) method was adopted in this evaluation work to eliminate the influence of individual bias. The ASD defined an anchor noise signal that was selected by experimental reviewers and ranked to the medial score. Then, the other noise signals were listened to by the jury and compared with the anchor signal. A noise signal was given a score based on the difference between the evaluated signal and the anchor signal. **Tab. 3** presents a subjective evaluation form using a five-grade scale ranging from 1.0 to 5.0. In addition, the evaluation participants could also rate the noise level between each continuous score if necessary. An example of a noise signal for subjective sound quality evaluation is provided in **Fig.**

5. Prior to the formal jury test, reviewers could select noise samples to allow them to adapt to the

overall situation.

Tab. 3 Definition of interior sound quality grades and the related scores

Grade	Definition of the scale	Scores
1	The noise is much more annoying than the anchor noise	1.0
2	The noise is slightly more annoying than the anchor noise	2.0
3	The noise is nearly the same as the anchor noise	3.0
4	The noise is slightly more pleasant than the anchor noise	4.0
5	The noise is much more pleasant than the anchor noise	5.0

Fig. 5 An ASD noise sample for jury evaluation

4.3 Methods for sound quality prediction

In this subsection, we will propose a novel regression-based DBN model to predict vehicle interior noise and compare the newly developed model with 4 other conventional machine learning models (MLR, BPNN, General Regression Neural Network (GRNN) and SVM) in terms of effectiveness and efficiency and then study the information fusing ability of each model with respect to different types and numbers of input features. All experiments have been implemented in MATLAB 2014a on a 2.40 GHz 4-core 8-thread Intel (R) Core (TM) i7-4700 MQ CPU mobile

workstation with 16 GB of memory.

4.3.1 Acoustic features of interior noise

A total of 10 commonly used acoustic features have been extracted from each noise sample to study the feature fusing ability of DBNs. They are linear sound pressure level (L-SPL), A-weighted sound pressure level (A-SPL), loudness, sharpness, roughness, fluctuation strength, articulation index (AI), tonality, wavelet transform energy [49], and wavelet transform entropy [50]. The attributes of these features are listed in **Tab. 4**. Note that many methods have been developed to calculate the psychoacoustic metrics; in this paper, the ISO 532B method, i.e., the Zwicker model, is adopted to compute the loudness, and the other referenced psychoacoustic metrics are calculated using the LMS Test.lab 13A. The parameters of the wavelet transform are selected as follows: Daubechies wavelet “db5”, three levels of analysis and soft universal threshold. Other parameters for the wavelet transform energy and wavelet transform entropy are set as in [49] and [50].

All the above features are normalized into a range of [0, 1] and concatenated into one sequence. Then, a new feature with a total of 24 dimensions is obtained for each noise sample.

Tab. 4 Features and their attributes

Index	Feature	Dimension
1	Linear sound pressure level	1
2	A-weighted sound pressure level	1
3	Loudness	1
4	Sharpness	1
5	Roughness	1

6	Fluctuation strength	1
7	Articulation index	1
8	Tonality	1
9	Wavelet transform energy _{3-level}	8
10	Wavelet transform entropy _{3-level}	8

4.3.2 Parameter settings and comparison schemes

The newly obtained dataset consists of 216 samples, each with 24 dimensions. We have randomly split the dataset into a training set and a testing set, which contain 75% and 25%, respectively, of the total samples. Concerning the sound quality prediction of vehicle interior noise, the better model parameters should be adopted in this experiment for further study. For the proposed DBN model, there are several parameters to be determined, as shown in **Fig. 3**, two of which, namely, the number of hidden layers (max-layer) and the number of units for each hidden layer (layer-unit), are critical to the prediction results of DBNs. Therefore, these two parameters should be studied in depth, and in this paper, we have selected different combinations of max-layer (with 1, 2, 3, 4, 5, and 6 hidden layers) and layer-unit (with 32, 64, 128, 256, 512, and 1024 units in each layer) to explore their influence on the prediction results. For simplicity, we denote the DBN model with i max-layer and j layer-unit as DBNs_i^j . According to the literature [28, 44, 51], the other model parameters are empirically selected as in **Tab. 5**.

Tab. 5 Parameter settings of DBNs

Learning rate of pre-training phase	0.001
Maximum epoch of pre-training phase	100
Learning rate of fine-tuning phase	0.01

Maximum epoch of fine-tuning phase	200
Momentum (start and stop)	[0.5, 0.9]
Batch-size	108

To demonstrate the prediction accuracy and time cost of the DBNs, 4 other widely used machine learning models, namely, MLR, BPNN, GRNN and SVM, have also been introduced for comparison with the DBNs. The parameter settings of the 4 models are presented below.

a) MLR

Because there are 24 dimensions in the extracted features, the MLR has 24 variables in terms of **Eq. (1)**, in which $n=24$ and m equals the total number of training samples. The ordinary least squares (OLS) estimator is adopted in this method for analyzing both the experimental (training set) and observational (testing set) data.

b) BPNN

A BPNN with three layers, namely, one input layer, one hidden layer and one output layer, is applied. The input layer is constructed with neurons denoting the dimension of the input dataset, and the output layer is constructed with one neuron to predict the network's output. A hidden layer with 15 neurons and a sigmoid transfer function are used to develop the BPNN model. The weights between the layers and biases of the layers are randomly selected initially. Learning is performed until either the number of iterations has reached 10,000 or the error has reached 0.001.

c) GRNN

In contrast to the BPNN, the learning procedure for the GRNN is a one-pass learning algorithm; thus, the iterative training process is unnecessary [52]. There are 4 layers in the GRNN model: the input layer, pattern layer, summation layer and output layer. Simpler than the BPNN,

the numbers of neurons of the input layer and pattern layer are both equal to the dimension of the input dataset, and the transfer function in the summation layer is the arithmetic sum with equal weights of 1. In the GRNN model, a key parameter “ σ ”, called the smoothing factor, strongly affects the network’s output. When the smoothing factor σ is large, the estimated variables would approximate the mean values of input-dependent variables, i.e., under-fitting. In contrast, a small smoothing factor σ would lead to over-fitting. Therefore, following the guide in [53, 54], we have selected a smoothing factor $\sigma=0.8$.

d) SVM

An SVM model with a Gaussian kernel function is applied in this paper. The soft margin parameter C and the kernel parameter γ have been selected via 4-fold grid search with the exponential grids $\{2^0, 2^1, \dots, 2^8\}$ and $\{2^{-3}, 2^{-1}, \dots, 2^3\}$, respectively. The LIBSVM toolbox [55] has been used to implement the regression-based SVM in this experiment.

Except for the comparison of the prediction accuracy and time cost, it is also important to study the information fusing ability of DBNs, i.e., fusing the advantages of new acoustic features. This will be further studied based on the comparison results of each model’s effectiveness and efficiency in the next section.

To account for the stochastic characteristics of the machine learning methods, we conduct all experiments 50 times, and the average performances for the training set and testing set are reported in the following section.

5. Results and discussion

5.1 Subjective evaluation

The jury evaluation has been conducted using the ASD method described in subsection 4.2,

and the rating scores were carefully collected. **Fig. 6** shows the subjective ratings and 95% confidence interval of all the interior noise samples, where the x-axis represents the interior noise evaluation scores ordered by their mean values. **Fig. 6** shows that the subjective ratings of the jury are relatively stable and can thus be used in further analysis. The subjective evaluation scores applied in the subsequent studies are the average values in the 95% confidence interval.

Fig. 6 Subjective ratings of the vehicle interior noises. ‘.’: subjective ratings of the jury. ‘--’: 95% confidence interval for the subjective ratings. ‘-’: average values of the subjective ratings with 95% confidence interval.

5.2 Prediction results and analysis

5.2.1 Parameter selection of DBNs

As described in subsection 4.3, 10 acoustic features, which have been concatenated into a long feature vector, are used as the input data for the DBNs, and the subjective evaluation scores have been used as the target of the DBNs. To choose better model parameters, two critical factors, namely, max-layer and layer-unit, are selected through a specific search, as described in subsection 4.3.2, and other parameters are fixed, as shown in **Tab. 5**. **Fig. 7 (a)** and **Fig. 7 (b)**

present the root mean square error (RMSE) of the prediction results obtained by varying the max-layer and layer-unit in the training set and testing set, respectively. **Fig. 8 (a)** and **Fig. 8 (b)** show the spearman correlation coefficient between the prediction output of the DBN models and the subjective evaluation scores. The significant difference in the experiments has been evaluated through the two-tailed t -test with a confidence interval of 95%.

Fig. 7 and **Fig. 8** show the main trend that, to some extent, an increased performance is obtained with an increasing number of hidden layers. Specifically, adding the second or third hidden layer to the models causes a larger proportional increment both in the training set and testing set compared to adding more layers. In addition, applying more hidden units for each layer also improves the prediction performance when max-layer is no greater than 3, and DBN models with 128 hidden units in each layer achieve the best performance. This may be because the DBNs are effective at exploring the potential manifold of a highly variant issue, such as natural language and image recognition; however, when the latent manifold of a problem is relatively apparent, a model with a deep structure cannot provide a better outcome than can a model that is well tuned with a shallow structure. However, when the structure of the DBNs is sufficiently deep and when the number of hidden units in each layer is sufficient, it is also possible for the model to acquire the manifold characteristics of the problem in a more abstract manner. From this point of view, the DBNs_6^{512} model and the DBNs_6^{1024} model have achieved a similar performance as the DBNs_3^{128} model on the testing set.

The computation times of each model applied to the training set and testing set are presented in **Fig. 9 (a)** and **Fig. 9 (b)**, respectively, from which we find that the run times increase with increasing max-layer and layer-unit, and the training times are an order of magnitude greater than

those during testing. Therefore, it is better to apply the DBN model with fewer hidden layers and hidden units for each layer to obtain similar prediction performance. Hence, in this paper, we selected the DBNs_3^{128} model for further analysis.

Fig. 7 Root mean square error (RMSE) of the DBN models' outputs with respect to the number of hidden layers

and the number of units for each hidden layer. (a) Training set. (b) Testing set.

Fig. 8 Correlation coefficients between the DBN output and the subjective evaluation scores with respect to the

number of hidden layers and the number of units for each hidden layer. (a) Training set. (b) Testing set

Fig. 9 Running times (CPU time) of each DBN model with respect to the number of hidden layers and the number of units for each hidden layer. (a) Training set. (b) Testing set.

5.2.2 Comparison of the effectiveness and efficiency

To demonstrate the advantage of the DBN-based sound quality prediction applied to the interior noise, a widely used experimental comparison between DBNs and 4 other machine-learning-based prediction methods has been introduced. The parameter settings of the referenced machine-learning-based models have been described in subsection 4.3.2, and the average RMSE and correlation coefficient of all the models are presented in **Fig. 10** and **Fig. 11**, respectively.

Fig. 10 and **Fig. 11** show that the MLR model obtained a higher RMSE and lower correlation coefficient on both the training set and testing set compared to all the methods due to the linear characteristic of this model. The performance of the BPNN, GRNN and SVM models improved gradually, therein resulting in RMSEs of 0.271, 0.187 and 0.109 and correlation coefficients of 89.54%, 91.32% and 92.22%, respectively, on the testing set. Therefore, for these three models, the SVM method provides the best performance. However, the performance of the proposed DBN-based sound quality prediction approach outperforms the 4 other machine-learning-based

prediction methods in terms of both the RMSE and the correlation coefficient, which are 0.064 and 94.37%, respectively, on the testing set. This is mainly the result of the deep learning capacity of DBNs to learn highly complex and non-linear characteristics of features through the pre-training and fine-tuning procedures with a layer-wise approach.

Fig. 12 summarizes the computation time of each model. The running times of the BPNN are several times those of the other models, namely, 13.32 s and 1.43 s during the training process and testing process, respectively. The MLR method has the lowest running time; unfortunately, this is inversely proportional to its performance. The running times of the GRNN, SVM and DBNs exhibit some difference on the training set but are of the same order of magnitude on the testing set. The SVM model has a smaller real-time working ratio (training time/testing time), and the DBN model presents a higher real-time working ratio. However, comparing with the performance achieved by the DBNs, the running time of the DBNs is acceptable because it is short and can be further reduced if the model is calculated on a better computer.

Fig. 10 Root mean square error (RMSE) comparison of the referenced machine learning models and the DBNs¹²⁸

model. (a) Training set. (b) Testing set.

Fig. 11 Comparison of correlation coefficients between the subjective evaluation scores and the referenced machine learning models as well as the $DBNs_3^{128}$ model. (a) Training set. (b) Testing set.

Fig. 12 Running times (CPU time in second) of the referenced machine learning models and the $DBNs_3^{128}$ model.

(a) Training set. (b) Testing set.

5.2.3 Analysis of the feature fusing ability

According to the above results, we can see that the DBN model outperforms the 4 other referenced machine-learning-based methods. However, it is also important to study the information fusing ability of DBNs for the following two reasons. First, one must determine if the superior performance of the DBN model depends on a certain feature group presented in **Tab. 4** or

if it is an intrinsic quality. Second, the superior ability of the unsupervised pre-training process in the DBNs, which other referenced machine learning models lack, must be studied further.

Addressing the first reason, to explore the regularity of the feature fusing performance with respect to the number of acoustic features and feature types, some feature assemblies have been developed by gradually adding new features to a feature subset from **Tab. 4**. Specifically, according to **Tab. 4**, we denote each feature as f_i , where the subscript i is the feature index; then, the feature assemblies can be selected as $\{ f_1 \}$, $\{ f_1, f_2 \}$, $\dots$ $\{ f_1, f_2, \dots, f_{10} \}$. Thus, 10 feature groups have been obtained, and the division task for the training set and testing set for each group is the same as that described in subsection 4.3.2.

Fig. 13 and **Fig. 14** show the RMSE and correlation coefficient of each model with respect to the feature groups, i.e., the number of acoustic features and feature types, on the training set and testing set. The two figures show that the RMSE is sharply decreasing and that the correlation coefficient is gradually increasing when the feature assembly increases. Therefore, clearly, the DBNs and the 4 other referenced prediction models can benefit from the new features. Among these methods, the GRNN and SVM models exhibit a better feature fusing ability than do the MLR and BPNN models; however, the DBN model exhibits the best performance in most of the feature groups. Moreover, these two figures show that even when there are fewer acoustic features in the feature assemblies, the DBN model continues to achieve a relatively high accuracy. This experimental result demonstrates the importance and advantage of the multiple-feature fusing process and also reflects that the feature fusing ability of DBNs is better than that of traditional machine-learning-based methods.

Fig. 13 Root mean square error (RMSE) of the $DBNs_3^{128}$ model and the 4 other referenced models with respect to the number of features and types. (a) Training set. (b) Testing set.

Fig. 14 Correlation coefficients between the subjective evaluation scores and the $DBNs_3^{128}$ model as well as the 4 other referenced models with respect to the number of features and types. (a) Training set. (b) Testing set.

Concerning the second reason, to elucidate upon the superior information fusing capacity of the unsupervised pre-training process in the DBNs, we varied the size of the training dataset to simulate the different amounts of information offered by features. Specifically, the proportion of the training set is randomly set to 35%, 45%, 55%, 65%, 75% and 85% of the entire dataset, and the remaining data serve as the testing set. In this experiment, all 10 acoustic features are used in the dataset, and the prediction performance of each model is shown in **Fig. 15** and **Fig. 16**.

Fig. 15 shows that the RMSE of the 5 models decreased with increasingly large training sets, and the performance rapidly improved especially when the proportion of the training set was below 55%. This is because the amount of information that is used to for learning is increased. The MLR model exhibits a less pronounced ability, and the BPNN and GRNN models obtain better performances. Notice that the SVM model obtains a better performance compared to the MLR, BPNN and GRNN methods on the training set and the testing set; however, the DBN model outperforms the 4 referenced machine-learning-based models even when less information is provided. Moreover, the outcome presented in **Fig. 16** is consistent with the results of **Fig. 15** in terms of performance. The above results indicate that the unsupervised pre-training process can learn more information from the features of datasets than can conventional machine-learning-based methods.

Summarizing, 1) the accuracy and robustness of the proposed DBN-based sound quality prediction approach are better than those of the 4 other referenced machine-learning-based methods. 2) The multiple-feature fusing work can greatly contribute to improving the prediction performance. 3) Finally, the unsupervised pre-training process of DBNs can enhance the information fusing ability.

Fig. 15 Root mean square error (RMSE) of the $DBNs_3^{128}$ model and the 4 other referenced models with respect to the proportion of the training dataset. (a) Training set. (b) Testing set

Fig. 16 Correlation coefficients between the subjective evaluation scores and the $DBNs_3^{128}$ model as well as the 4 other referenced models with respect to the proportion of the training dataset. (a) Training set. (b) Testing set.

6. Conclusion

In this paper, a vehicle road test has been conducted to record interior noise, and a total of 216 noise samples were obtained. A subjective evaluation test was conducted using the anchor semantic differential (ASD) method with 24 jurors. Then, to predict the sound quality of the collected interior noise, a DBN model that uses the L2-SVM layer to substitute the softmax layer at the top of the general DBN structure has been proposed. To improve the prediction performance of the DBN model, two crucial factors, namely, the max-layer and the layer-unit, have been carefully selected. Then, the developed DBN model was applied to predict the sound quality of interior noise and compared with 4 other traditional machine-learning-based methods, therein demonstrating the superiority of the DBNs. In addition, by changing the information that the training dataset offered, the DBN model has been found to exhibit a superior feature fusing ability compared to the 4 other referenced methods.

Despite the success of the DBNs, various problems remain to be addressed, including how to more effectively select the model parameters and how to improve the performance of DBNs with more advanced algorithms. In addition, it would be of great interest to apply the DBNs to the evaluation of non-stationary and non-linear vehicle noise in the future.

Acknowledgments

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

References

- [1] G. K, The sound quality of vehicle interior noise: a challenge for the NVH-engineers, *Int. J. Vehicle Noise Vibrat.* 1-2 (2004) 158-168. doi: 10.1504/IJNVN.2004.004079.
- [2] 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. doi: 10.1016/j.ymssp.2013.11.001.
- [3] 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. Acoustics* 100 (2015) 18-25. doi: 10.1016/j.apacoust.2015.06.018.
- [4] R. Johnsson, J. Odelius, M. Rantatalo, A new test track for automotive squeak and rattle (S&R) detection, *Appl. Acoustics* 80 (2014) 79-85. doi: 10.1016/j.apacoust.2014.01.010.

- [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. doi: 10.1016/j.apergo.2008.11.006.
- [6] S. Lee, H. Kim, E. Na, Improvement of impact noise in a passenger car utilizing sound metric based on wavelet transform, *J. Sound Vibration* 329 (2010) 3606-3619. doi: 10.1016/j.jsv.2010.03.001.
- [7] G. Pietila, T.C. Lim, Intelligent systems approaches to product sound quality evaluations – a review, *Appl. Acoustics* 73 (2012) 987-1002. doi: 10.1016/j.apacoust.2012.04.012.
- [8] 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 Vibration* 332 (2013) 3893-3904. doi: 10.1016/j.jsv.2013.02.030.
- [9] J. Yoon, I. Yang, J. Jeong, S. Park, J. Oh, Reliability improvement of a sound quality index for a vehicle HVAC system using a regression and neural network model, *Appl. Acoustics* 73 (2012) 1099-1103. doi: 10.1016/j.apacoust.2012.02.018.
- [10] 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, *P. I. Mech. Eng. D-J Aut.* 224 (2010) 735-747. doi: 10.1243/09544070JAUTO1296.
- [11] 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 Vibration* 299 (2007) 933-947. doi: 10.1016/j.jsv.2006.07.034.
- [12] D. Västfjäl, M.A. Gulbol, M. Kleiner, T. Gärling, Affective evaluations of and reactions to exterior and interior vehicle auditory quality, *J. Sound Vibration* 255 (2002) 501-518. doi:

10.1006/jsvi.2001.4166.

- [13] W. Zhang, G.X. Chen, B.X. Xiao, S.X., Jin, Objective evaluation of sound quality of noises inside high speed train, *Journal of the China Railway Society*, 006 (2011) 491-500.
- [14] C.A. Powell, B.M. Sullivan, Subjective response to propeller airplane interior sounds modified by hypothetical active noise control systems, *Noise Contr. Eng. J.* 49 (2001) 125-136. doi: 10.3397/1.2839648.
- [15] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, C. Nai, C.C. Tung, H.H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, *P. Roy. Soc. Lond. A Mat.* 454 (1998) 903-995.
- [16] N.E. Huang, M. Wu, W. Qu, S.R. Long, S.S.P. Shen, J.E. Zhang, Applications of Hilbert-Huang transform to non-stationary financial time series analysis, *Appl. Stochastic Models Bus. Ind.* 19 (2003) 245-268. doi: 10.1002/asmb.501.
- [17] R. Ricci, P. Pennacchi, Diagnostics of gear faults based on EMD and automatic selection of intrinsic mode functions, *Mech. Syst. Signal Process.* 25 (2011) 821-838. doi: 10.1016/j.ymssp.2010.10.002.
- [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] X.M. Shen, S.G. Zuo, L. Li, S.W. Zhang, Interior sound quality forecast for vehicles based on support vector machine, *J. Vibrat. Shock* 6 (2010) 66-68.
- [20] W. Caesarendra, A. Widodo, B. Yang, Application of relevance vector machine and logistic regression for machine degradation assessment, *Mech. Syst. Signal Process.* 24 (2010)

- 1161-1171. doi: 10.1016/j.ymssp.2009.10.011.
- [21] 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, *P. I. Mech. Eng. D-J. Aut.* 222 (2008) 383-396. doi: 10.1243/09544070JAUTO563.
- [22] W.J. Staszewski, K. Worden, G.R. Tomlinson, Time–frequency analysis in gearbox fault detection using the Wigner–Ville distribution and pattern recognition, *Mech. Syst. Signal Process.* 11 (1997) 673-692. doi: 10.1006/mssp.1997.0102.
- [23] G.E. Hinton, S. Osindero, Y.W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.* 18 (2006) 1527-1554. doi: 10.1162/neco.2006.18.7.1527.
- [24] D. Ciresan, U. Meier, J. Schmidhuber, Multi-Column Deep Neural Networks for Image Classification, in: *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3642-3649. doi: 10.1109/CVPR.2012.6248110.
- [25] A. Graves, A.R. Mohamed, G. Hinton. Speech recognition with deep recurrent neural networks, in: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6645-6649. doi: 10.1109/ICASSP.2013.6638947.
- [26] R. Collobert, J. Weston. A unified architecture for natural language processing: deep neural networks with multitask learning, in: *ICML '08 Proceedings of the 25th International Conference on Machine Learning*, 2008, pp. 160-167. doi: 10.1145/1390156.1390177.
- [27] G.E. Hinton, S. Osindero, Y.W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.* 18 (2006) 1527-1554. doi: 10.1162/neco.2006.18.7.1527.
- [28] G.E. Hinton, A practical guide to training restricted Boltzmann machines, in: G. Montavon,

- G.B. Orr, K.-R. Müller (Eds.), *Neural Networks: Tricks of the Trade*, Springer, Berlin, 2012, pp. 599-619. doi: 10.1007/978-3-642-35289-8_32.
- [29] Y.S. Wang, Sound quality estimation for nonstationary vehicle noises based on discrete wavelet transform, *J. Sound Vibration* 324 (2009) 1124-1140. doi: 10.1016/j.jsv.2009.02.034.
- [30] ISO, 532B Acoustics - Method for Calculating Loudness Level, International Organization for Standardization, Geneva, 1975.
- [31] H. Liu, J. Zhang, P. Guo, F. Bi, H. Yu, G. Ni, Sound quality prediction for engine-radiated noise, *Mech. Syst. Signal Process.* 56–57 (2015) 277-287. doi: 10.1016/j.ymssp.2014.10.005.
- [32] E. Parizet, E. Guyader, V. Nosulenko, Analysis of car door closing sound quality, *Appl. Acoustics* 69 (2008) 12-22. doi: 10.1016/j.apacoust.2006.09.004.
- [33] H.H. Lee, H.W. Kim, S.K. Lee, Sound Quality evaluation for the axle gear noise in the vehicle, *J. Acoust. Soc. Am.* 123 (2008) 3260. doi: 10.1121/1.2933563.
- [34] H.B. Huang, R.X. Li, W.P. Ding, M.L. Yang, H.L. Zhu, Research on the rig test for identifying abnormal noise of suspension shock absorber, *J. Shock Vibration* 34 (2015) 191-196.
- [35] L. Egab, X. Wang, Objective evaluation of interior trim effects on sound quality and noise reduction of a coupled plate cavity system, *Mech. Syst. Signal Process.* 70–71 (2016) 919-931. doi: 10.1016/j.ymssp.2015.09.035.
- [36] L. Zou, X.L. Tang, H.S. Yu, T. Zhang, J.W. Zhang, Analysis on the torsional vibration and noise of transmission system hybrid electric cars, automobile, *Engineering* 36 (2014) 21-27.
- [37] K. Soobbarayen, S. Besset. A simplified approach for the calculation of acoustic emission in the case of friction-induced noise and vibration, *Mech. Syst. Signal Process.* 50 (2015)

732-756.

- [38] Z.M. Xu, Y. Zhang, J.L. Liu, Y.S. He, Q. Yuan, Sound quality analysis of wiper system noise in cars, automobile, Engineering 36 (2014) 1009-1013.
- [39] 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. doi: 10.3901/JME.2011.24.091.
- [40] K. Manimala, K. Selvi, R. Ahila, Optimization techniques for improving power quality data mining using wavelet packet based support vector machine, Neurocomputing 77 (2012) 36-47. doi: 10.1016/j.neucom.2011.08.010.
- [41] G.R. Ibrahim, A. Albarbar, Comparison between Wigner–Ville distribution and empirical mode decomposition vibration-based techniques for helical gearbox monitoring, P. I. Mech. Eng. C-J Mec. 225 (2011) 1833-1846.
- [42] G.B. Huang, H. Lee, M.E. Learned. Learning hierarchical representations for face verification with convolutional deep belief networks, computer vision and pattern recognition (CVPR), in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2518-2525. doi: 10.1109/CVPR.2012.6247968.
- [43] Z. Ling, L. Deng, D. Yu, Modeling spectral envelopes using restricted Boltzmann machines and Deep belief networks for statistical parametric speech synthesis, IEEE. Transactions of the on Audio, Speech, and Language Processing, 21 (2013) pp. 2129-2139.
- [44] 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, Signal Processing Magazine, 29 (2012) 82-97.
- [45] P. Tamilselvan, P. Wang, Failure diagnosis using deep belief learning based health state

- classification, *Reliab. Eng. Syst. Saf.* 115 (2013) 124-135. doi: 10.1016/j.ress.2013.02.022.
- [46] 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. doi: 10.1016/j.eswa.2013.12.026.
- [47] Y. Tang, Deep learning using linear support vector machines, *Arxiv Preprint Arxiv* 1306 (2013) 0239.
- [48] E.M. Schmidt, Y.E. Kim. Learning emotion-based acoustic features with deep belief networks, 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011, pp. 65-68. doi: 10.1109/ASPAA.2011.6082328.
- [49] C. Juang, C. Cheng, T. Chen, Speech detection in noisy environments by wavelet energy-based recurrent neural fuzzy network, *Expert Syst. Appl.* 36 (2009) 321-332. doi: 10.1016/j.eswa.2007.10.028.
- [50] R. Moshrefi, M.G. Mahjani, M. Jafarian, Application of wavelet entropy in analysis of electrochemical noise for corrosion type identification, *Electrochem. Commun.* 48 (2014) 49-51. doi: 10.1016/j.elecom.2014.08.005.
- [51] L.X. Zhang, J. Wu, Deep belief networks based voice activity detection, *IEEE Trans. Audio Speech Language Process.* 21 (2013) 697-710.
- [52] J. Wu, C. Huang, An engine fault diagnosis system using intake manifold pressure signal and Wigner–Ville distribution technique, *Expert Syst. Appl.* 38 (2011) 536-544. doi: 10.1016/j.eswa.2010.06.099.
- [53] J. Liu, W. Bao, L. Shi, B. Zuo, W. Gao. General regression neural network for prediction of sound absorption coefficients of sandwich structure nonwoven absorbers, *Appl. Acoustics* 76

(2014) pp. 128-137.

- [54] J. Wu, P. Chiang, Application of Wigner–Ville distribution and probability neural network for scooter engine fault diagnosis, *Expert Syst. Appl.* 36 (2009) 2187-2199. doi: 10.1016/j.eswa.2007.12.012.
- [55] C. Chang, C. Lin, LIBSVM: A library for support vector machines, *ACM Trans. Intell. Syst. Technol.* 2 (2011) 1–27. doi: 10.1145/1961189.1961199.