Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.jzus.zju.edu.cn; www.springerlink.com
E-mail: jzus@zju.edu.cn

A comprehensive evaluation method of diesel engine sound quality based on paired comparison, uniform design sampling, and improved analytic hierarchy process*

Wen-yun HE, Jun-hong ZHANG†‡, Jian WANG
(*State Key Laboratory of Engines, Tianjin University, Tianjin 300072, China*)
†E-mail: zhangjh@tju.edu.cn

Received Jan. 15, 2016; Revision accepted Nov. 9, 2016; Crosschecked June 16, 2017

Abstract: The paired comparison (PC) is an easy and effective way to distinguish the nuances among sound stimuli of diesel engine in sound quality subjective evaluation. However, the PC method is inefficient for the subjective evaluation of large quantities of stimulus or to rank the results of overall perception of sound stimuli. To overcome the shortcoming of the PC method, this paper presents a comprehensive evaluation method of diesel engine sound quality based on the PC method, uniform design sampling (UDS), and the improved analytic hierarchy process (AHP). An hierarchical tree of comprehensive evaluation models of diesel engine sound quality is constructed and simplified through UDS to ensure that the sound sample size is in the application range of the PC method or the grouped paired-wise comparison (GPC) method. An improved AHP is developed by introducing a new method calculating the relative weight based on evaluators' preference order instead of the traditional 1–9 scale method, which reduces subjective liberty and difficulty of the consistency of the judgment matrix meeting the requirement, and the requirements of evaluators' specialized knowledge. With three diesel engines' noise stimuli as example, a realization of sound quality comprehensive evaluation based on UDS-AHP method with the GPC method is introduced. The results show that the proposed subjective comprehensive evaluation method can accurately provide the ranks of the sound stimuli over all subjective perception with a significant decrease of the evaluation workload.

Key words: Sound quality; Comprehensive evaluation; Diesel engine; Paired comparison (PC); Uniform design sampling (UDS); Analytic hierarchy process (AHP)
http://dx.doi.org/10.1631/jzus.A1600025 **CLC number:** TK421.6

1 Introduction

A measurement of sound quality takes human hearing characteristics into account and quantitatively reflects the subjective feeling discrepancy for different noises. It has become the focus of current diesel engine noise research (Genuit, 2008; Zhang *et al.*, 2012). Sound quality subjective evaluation is the key link in sound quality study. It directly reflects the subjective perception ranks of the sound sample. The common subjective evaluation methods mainly include rank order, the rating scale (Ronacher *et al.*, 1999), the semantic differential (Chouard and Hmpel, 1999), and the paired comparison (PC) method (David, 1966; Chouard and Weber, 1994). The PC method is widely recognized and applied in diesel engine sound quality evaluation. It can be carried out easily and can be employed to discriminate the nuances of sound stimuli with no difficulty. In the current study, sound stimuli lasting 4 s each were diesel engine noises recorded 1 m away from the engine surface through the noise test carried out with HEAD

---

‡ Corresponding author

* Project supported by the National High-Tech R&D Program of China (No. 2016YFD0700701)

ORCID: Wen-yun HE, http://orcid.org/0000-0001-7085-0199 acoustic digital artificial head in the semi-anechoic chamber, and played back with a professional soundcard and professional headphone in a quiet listening room after screening and pre-processing.

If N sound stimuli are submitted to human listeners, subjective sound quality evaluation based on the complete PC method involves N^2 comparisons. Therefore, the time consumption of each complete PC evaluation is proportional to the square of the number of sound stimuli. With the increase of the sample size, the evaluation time will increase dramatically. It will lead to the impatience and fatigue of the evaluator during a long term evaluation test. Therefore, a large sample evaluation is hard to carry out. Due to the deficiency of traditional PC, a lot of research has been done on how to improve it. Chouard and Weber (1994) analyzed the influence of the ordering of noise in sequences on the PC evaluation result and hold that it could be performed with a semi-matrix. Mao *et al.* (2006) presented a grouped paired-wise comparison (GPC) method, which divided the all the evaluation sound stimuli into Z correlative test groups. With the reference link stimuli configured in each group, the final evaluation results of all stimuli are reconstructed. Huang *et al.* (2008) proposed an adaptive grouped paired comparison (AGPC) design to avoid the reduction of the precision with reference samples, which is adaptive instead of pre-selected. Some researchers put forward the classified paired comparison (CPC) using the clustering analysis such as a K-means cluster (Jiao *et al.*, 2005), free sorting (Parizet and Koehl, 2011), and a hierarchical cluster (Lemaitre *et al.*, 2002; Ling *et al.*, 2006). Although GPC and CPC are two effective methods which can reduce the time requirement and difficulties of the subjective comparison test to some degree, the large sample evaluation or the complex multivariate analysis is sometimes beyond the scope of their capabilities.

Uniform design (UD) (Fang and Ma, 2001; Fang *et al.*, 2003a; 2003b; Huang *et al.*, 2007) based on a quasi-Monte-Carlo algorithm is a “space filling” experimental design. It was developed from an orthogonal experimental design by Fang and Wang to solve the problem of missile design. UD allocates experimental points uniformly scattered on the experimental domain and enables every point to be better represented. These representational points are not tidy and contrasting, so it can reduce the number of the experimental points. Uniform design sampling (UDS) is a statistical sampling method which is a development of UD and an improvement for Latin hypercube sampling (LHS) (McKay *et al.*, 1979). It is especially suitable for multi-factor and multi-level experiment design. Zhang and Wang (1996) studied the properties of UDS and applied it to computations of numerical integrals, and compared it with the Monte-Carlo method, UD, and LHS. The results showed that UDS was more powerful than the other methods. Jing *et al.* (2005) proposed a novel yield estimation and optimization method based on UDS. Compared with the popular Monte-Carlo sampling method, this new method needed only a few circuit simulations to offer a valuable estimation and was immune to the number of statistical variables. Furthermore, owing to a simple algorithm to generate samples, the UDS method added no computational complexity. Zhou *et al.* (2012) redesigned the crossover operation in a genetic algorithm based on UDS. It proved that the bias of samples in a UDS point set was $O(n^{-1+\varepsilon})$, the same as a good point set. The sample in the UDS point set was a random uniform distribution and had better representativeness.

Actually, there are two purposes of diesel engine sound quality analysis. They are sound stimuli sequencing and products sequencing. That is to say, it is aimed at judging the sound stimuli or products with some measurable parameters for sound quality research. Previous studies mainly focus on the former. The relationships between the objective characteristics of the sound signal and the subjective feeling were studied. In a practical project, the comprehensive evaluation of a certain product sound quality and the sequencing of different products are what we need. However, evaluation results of different sound stimuli are mutually independent, and the effect of the diesel engine characteristics (such as the brand, the running condition, the noise test position) on the sound quality and the comprehensive annoyance evaluation of the diesel engine noise could not be obtained directly by the common methods. To solve this problem, the analytic hierarchy process (AHP) (Satty, 2008; Xu and Wei, 1999) is applied in sound quality evaluation. AHP is a multi-criteria decision-making method which integrates qualitative analysis with quantitative analysis to improve the reliability and validity of the decision. It can integrate clearly many different effect factors through the construction of a hierarchical model. Liu *et al.* (2014) proved that AHP was instructive and useful for the subway's sound quality evaluation. The different operating conditions, passenger's attitudes, and test positions were considered in the research. Su *et al.* (2012) combined AHP with fuzzy comprehensive evaluation based on a semantic differential to rank the overall perception of vehicle interior sound. AHP is an applicable and simple method for the comprehensive evaluation of sound quality.

In this study, a comprehensive evaluation method of sound quality is proposed focusing on solving the multi-factor, multi-condition problem, and reducing time and difficulty of a subjective jury test at the same time. The comprehensive evaluation model of diesel engine sound quality is developed based on hierarchy analysis, UDS, and subjective perception analysis. An improved AHP, using a new method to determine the relative weight based on evaluators' preference order instead of a 1–9 scale method, is put forward to reduce the subjective induced error and the difficulty of the consistency of the judgment matrix meeting the requirement. The main effect factors, the experimental domain, and a suitable number of levels for each factor are first determined. Then the hierarchy tree is constructed and simplified through UDS to ensure that the sound sample size is in the application range of GPC. Finally, subjective assessments are carried out and the results of comprehensive evaluation are obtained by multiple regression analysis.

2 Methods

In this section, a comprehensive evaluation method of sound quality based on PC, UDS, and improved AHP (UDS-AHP) is proposed. It is an incomplete jury test method for precise and comprehensive evaluation of sound quality with a small number of sound stimuli. The detailed steps of the UDS-AHP can be described as shown in Fig. 1.

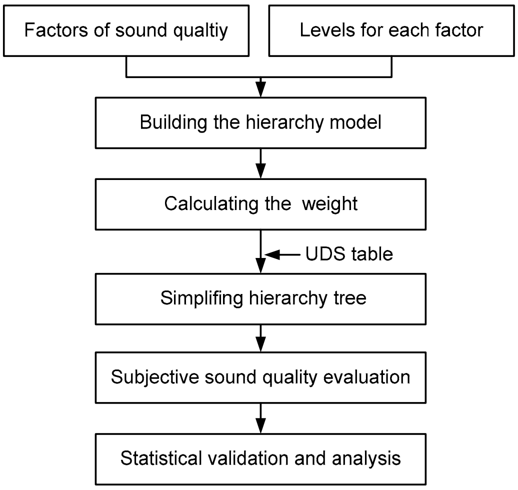

Fig. 1 Flowchart diagram of UDS-AHP

2.1 Uniform design sampling

In the following, the definition of UD will first be given. Suppose that ρ is the set of n points in the domain D, and let $M(\rho)$ be a measure of the nonuniformity of ρ. UD is to seek a set ρ that minimizes $M(\rho)$. In other words, it is to maximize the uniformity through searching all possible n points in D. Generally, the domain D is taken as $C^m=[0, 1)^m$, a unit hypercube in the m-dimensional space, or its subdomain.

An $n\times m$ matrix $\boldsymbol{U}(n, q_1, \cdots, q_m)$, where each column is a permutation of $\{1, 2, \cdots, q_j\}$ ($j=1, 2, \cdots, m$), is called a U-type design, which provides an n-run experimental design for m factors, with the jth factor having q_j levels. Let $\boldsymbol{\mu}(n, q_1, \cdots, q_m)$ denote the set of all $\boldsymbol{U}(n, q_1, \cdots, q_m)$ designs. When some factors have the same levels, the U-type design is denoted by $\boldsymbol{U}(n, q_1^{r1}, \cdots, q_m^{rl})$ and the set of the U-type design is denoted by $\boldsymbol{\mu}(n, q_1^{r1}, \cdots, q_m^{rl})$ for convenience, where $\Sigma_{j=1}^{l} r_j=m$. In particular, the U-type design is denoted by $\boldsymbol{U}(n; q^m)$ when it meets $q_1=q_2\cdots=q_m$. With the mapping function $f=(f_1, f_2, \cdots, f_m)$, where $f_j:1\rightarrow(2l-1)/(2q_j)$, $l=1, 2, \cdots, q_j, j=1, 2, \cdots, m$, and the n rows of a U-type design $\boldsymbol{U}(n, q_1, \cdots, q_m)$ are linearly transformed into n points of C^m. Therefore, the set $\boldsymbol{\mu}(n, q_1, \cdots, q_m)$ can be considered as a subset of C^m. A design of $\boldsymbol{\mu}(n, q_1, \cdots, q_m)$ with the smallest discrepancy $M(\rho)$ is the uniform design denoted by $\boldsymbol{U}_n(q_1, q_2, \cdots, q_m)$.

Clearly it is necessary to define a quantity of discrepancy from uniformity. Generally, the "wrap-around L_2-discrepancy" and the "centered L_2-discrepancy" are widely used, and these have been regarded as reasonable and practicable (Hickernell, 1998a; 1998b; Fang *et al.*, 2003a). A design (a set of n points) on C^m can be expressed by a matrix $\boldsymbol{X}$ of n rows and m columns with each row representing a point and each column representing a dimension or variable. Let $\boldsymbol{x}_i=(x_{i1}, x_{i2}, \cdots, x_{ik})$ denote the ith row of $\boldsymbol{X}$, $i=1, 2, \cdots, n$. Then, the square of the centered L_2-discrepancy for $\boldsymbol{X}$, $\mathrm{CD}_2(\boldsymbol{X})$, can be calculated as

$$\begin{aligned}\left[\mathrm{CD}_2(\boldsymbol{X})\right]^2 = &\left(\frac{13}{12}\right)^m - \frac{2}{n}\sum_{i=1}^{n}\prod_{l=1}^{m}\left(1+\frac{1}{2}\left|x_{il}-\frac{1}{2}\right|\right.\\ &\left.-\frac{1}{2}\left|x_{il}-\frac{1}{2}\right|^2\right)+\frac{1}{n^2}\sum_{i,j=1}^{n}\prod_{l=1}^{m}\left(\frac{1}{2}\left|x_{il}-\frac{1}{2}\right|\right.\\ &\left.+\frac{1}{2}\left|x_{jl}-\frac{1}{2}\right|-\frac{1}{2}\left|x_{il}-x_{jl}\right|+1\right),\end{aligned} \tag{1}$$

and the square of the wrap-around L_2-discrepancy for $\boldsymbol{X}$, $\mathrm{WD}_2(\boldsymbol{X})$, can be calculated as

$$\begin{aligned}\left[\mathrm{WD}_2(\boldsymbol{X})\right]^2 = &-\left(\frac{4}{3}\right)^m + \frac{1}{n^2}\sum_{i,j=1}^{n}\prod_{l=1}^{m}\Big[1.5\\ &-\left|x_{il}-x_{jl}\right|\left(1-\left|x_{il}-x_{jl}\right|\right)\Big].\end{aligned} \tag{2}$$

UDS is based on the random sampling of UD. It has n^m sampling points and each group has n sample points. This sampling proceeds through the following steps:

Step 1. For a given $n\in\mathbb{N}$, $m\in\mathbb{N}$, choose the suitable uniform design $(n, h_1, \cdots, h_m)$. Then, the points set of UD, $R_n=\{\boldsymbol{c}_i=(c_{i1}, c_{i2}, \cdots, c_{ij}), i=1, 2, \cdots, n\}$, can be obtained, where

$$c_{ij}=\left\{\frac{ih_j-0.5}{n}\right\},\quad i=1,2,\cdots,n,\quad j=1,2,\cdots,m. \tag{3}$$

Step 2. Let $\eta_1, \eta_2, \cdots, \eta_m$ be i.i.d. samples of multinomial distribution:

$$\begin{bmatrix} 0 & 1 & \cdots & n-1 \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \end{bmatrix}. \tag{4}$$

Step 3. Let υ_{ij}, which is independent of η_j, be samples uniformly distributed on interval $[-1/2, 1/2]$. Then, the points set $P_n=\{\boldsymbol{x}_i=(x_{i1}, x_{i2}, \cdots, x_{ij}), i=1, 2, \cdots, n\}$ can be defined as

$$x_{ij}=\left\{\frac{ih_j+\eta_j-0.5}{n}\right\}+\frac{\upsilon_{ij}}{n},\quad i=1,2,\cdots,n,\quad j=1,2,\cdots,m, \tag{5}$$

where $\{x\}$ denotes the fractional part of x, can be called samples of UDS. Fig. 2 shows an example of UDS when $n=5$, $m=2$, in which five samples (when $\eta_1=0$ and $\eta_2=0, 1, 2, 3, 4$) are represented by different symbols. It can be observed that each group is uniformly scattered on the experimental region.

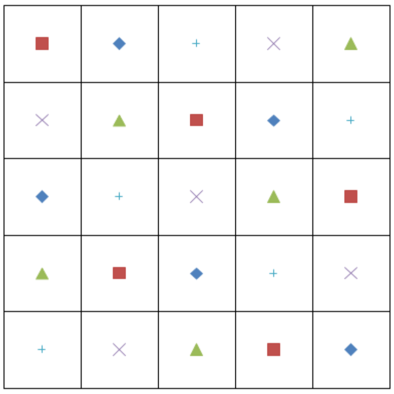

Fig. 2 Five samples by UDS (η_1=0) when n=5, m=2

2.2 Building the hierarchy model

First, the factors are chosen and a suitable number of levels for each factor are determined. Then the hierarchy model is established with a hierarchy analyses method, where the primary factor is placed at the top of the hierarchical tree and the criteria that have direct impacts on the factor are organized into the following layers.

2.3 Calculating the relative weight and the combination weight

The 1–9 scale method based on PC, where 1 means the comparison pair has the same degree of concern, while 9 means one of the pair is far more important than the other, is widely adopted to determine the factor weight. However, because of its large scale among the stages and complete dependence on the evaluators' subjective judgment ability, the 1–9 scale method is not suitable for comprehensive evaluation of diesel engine sound quality, the factors of which usually have little difference in the important degree. It is difficult to make the consistency of the judgment matrix meet the requirement. Thus, a new method based on PC to determine the relative weight is proposed here. Through pair-wise comparisons among the elements at the same level of the hierarchy tree, a series of importance order can be obtained. Then, the preference matrix can be obtained based on the Bradlley-Tery model (Bradley and Terry, 1952).

$$\Pi_{ij} = \frac{\exp\left(\vartheta_i - \vartheta_j\right)}{1+\exp\left(\vartheta_i - \vartheta_j\right)}, \tag{6}$$

where ϑ_i is the merit value of the ith element. The relative weight is defined as

$$w_i = \frac{b_i}{\sum_{i=1}^{n} b_i}, \tag{7}$$

where $b_i = \sum_{j=1}^{m} a_{ij}$, $a_{ij} = \Pi_{ij} \Big/ \sum_{i=1}^{n} \Pi_{ij}$.

The relative weight w_i is calculated based on evaluators' preference order instead of the preference degree from the probability point of view. Any evaluators, who can be the common consumer with simple training, have the same effect on w_i.

The weight vector of the element on level k to the general objective is $\boldsymbol{W}^k$ and the weight vector of the element on level k+1 to the element Ω_i^k is w_i^{k+1}, which is determined by the method mentioned above. They are given as follows:

$$\boldsymbol{W}^k = \left(W_1^k, W_2^k, \cdots, W_{n_k}^k\right)^{\mathrm{T}}, \tag{8}$$

$$w_i^{k+1} = \left(w_{i1}^{k+1}, w_{i2}^{k+1}, \cdots, w_{ij}^{k+1}\right),\ j=1,2,\cdots,n_k, \tag{9}$$

$$\boldsymbol{P}^{k+1} = \left(w_1^{k+1}, w_2^{k+1}, \cdots, w_{n_{k+1}}^{k+1}\right)^{\mathrm{T}}, \tag{10}$$

where $\boldsymbol{P}^{k+1}$ is a $n_{k+1}\times n_k$ matrix. The weight vector of all the elements on level k+1 to the general objective $\boldsymbol{W}^{k+1}$ is given as

$$\boldsymbol{W}^{k+1} = \frac{\boldsymbol{P}^{k+1}\boldsymbol{W}^k}{\sum \boldsymbol{P}^{k+1}\boldsymbol{W}^k} = \left(\boldsymbol{W}_1^{k+1}, \boldsymbol{W}_2^{k+1}, \cdots, \boldsymbol{W}_{n_{k+1}}^{k+1}\right)^{\mathrm{T}}. \tag{11}$$

2.4 Improving the hierarchy tree by uniform design sampling

The hierarchy tree is simplified with a suitable UDS table related to the number of factors and levels. This is in order to reduce the number of stimuli. When the number of tests is determined, the discrepancy of the UDS table increases with the increment of the number of factors. Therefore, the number of tests is at least two times larger than the number of factors in practical applications in order to improve the accuracy of the test.

2.5 Subjective sound quality evaluation

Sound stimuli are diesel engine noise events, which are determined according to the improved hierarchy tree and recorded binaurally with special acoustics digital devices. The recorded original sound stimuli are pre-processed before subjective sound quality evaluation, including length adjusting and loudness editing of the stimuli. Meanwhile, a systematic training procedure for evaluators is carried out before the final jury test.

Subjective sound quality evaluation of diesel engine noise is usually performed on a PC. In the normal PC, sound stimuli are played in pairs by way of permutation and combination, and evaluators are then asked to make relative choice decisions between the presented pairs. This procedure makes it possible for evaluators to discriminate between nuances of the pairs. The total time T_{PC} of the complete PC test including the comparison of the same event can be expressed as follows:

$$T_{\mathrm{PC}} = RN^2 t, \tag{12}$$

where R represents the time of repetition, N is the total number of stimuli to be evaluated, and t is the comparison time required for each sound pair.

The results of many jury tests in sound quality show that the proper duration for each listening test, to help ensure the reliability of subjective test, should be kept in the range of 20–30 min, and never longer than 45 min. For the case of subjective evaluation of large quantity of stimuli, GPC is applicable. It divided the entire sound stimuli into several groups requiring evaluation and the test duration of each group lies in the range of 20–30 min, which is a proper time for the listener to make a credible judgment. Each group is evaluated separately. The final value ζ of sound stimuli can be reconstructed by

$$\zeta_{IJ} = \frac{\alpha}{V_{1J} - V_{2J}}\left(V_{IJ} - V_{1J}\right) + \beta, \tag{13}$$

where J is the identity number of group, I is the serial number of stimuli in the group, V_{1J} and V_{2J} denote the original scores of two reference-link stimuli given by test in group J, respectively, α is the scale factor used for adjusting the evaluating range of final results, and β is a bias factor used for adjusting the neutral point of final results.

Supposing Z groups are divided for the total stimuli number N, the total test time T_{GPC} for all N stimuli can be reduced by a factor of Z, shown as

$$T_{\mathrm{GPC}} = RN^2 t / Z. \tag{14}$$

2.6 Statistical validation and analysis for subjective evaluation data

In the statistical analysis of PC data, weighted factor of consistency (Mao *et al*., 2005) is applied to comprehensive assessment of judging errors. During the determination of the weighted factor of consistency, several factors are considered, namely misjudging rates in the same sound event comparison, the different replay order comparison, and the circular triads. The weight factor of consistency ξ can be expressed as follows:

$$\xi = 1 - \frac{\sum \varepsilon_i E_i}{\sum E_i}, \tag{15}$$

where E_i is the utmost possible times of the ith kind of misjudging, and ε_i is the actual misjudging rate of the ith kind of misjudging.

According to the subjective sound quality evaluation data, the comprehensive annoyance SQ of the diesel engine noise and the determination of these factors' influence on sound quality can be obtained with statistical methods, such as regression analysis and the genetic algorithm:

$$\mathrm{SQ} = f\left(x_1, x_2, \cdots, x_m\right). \tag{16}$$

3 Experiments and application

In this section, three diesel engines' noise stimuli are adopted as test sounds to verify the feasibility of the UDS-AHP method. Implementation and procedure of UDS-AHP is presented, and comparisons of UDS-AHP with AHP are introduced to investigate the suitability of UDS-AHP.

3.1 Hierarchy model and weight

The main influencing factors of sound quality of diesel engine noise, including the brand, the speed, the load, and the noise test position, are adopted to build the hierarchy tree of the comprehensive evaluation model of diesel engine sound quality with a suitable number of levels for each factor, as shown in Fig. 3. These three diesel engines referred to are of the same type. They are produced for a certain type of construction machinery by different manufacturers. It can be assumed that consumers' preferences for the three brands are the same, and these three diesel engines have the same relative weight of the noise test position, the speed, and the load. Table 1 shows the preference matrix and the relative weight of the three engines of different brands. The analyses of the preference orders of the noise test position, the speed, and the load are performed by the complete PC test: 60 consumers of different genders, ages, and education levels are chosen to complete the comparison tasks. The results are validated according to Eq. (15) to examine the consistency of each consumer. Based on the preference order, the preference matrix and the relative weight of the noise test position, the speed, and the load are obtained according to Eqs. (6) and (7). Table 2 shows the preference matrix and the relative weight of five noise test positions. Table 3 shows the preference matrix and the relative weight of five speeds. Table 4 shows the preference matrix and the relative weight of five loads.

3.2 UDS table and the improved hierarchy tree

In the established model shown in Fig. 3, a subjective sound quality evaluation procedure with 375 stimuli is carried out. The large workload makes it impractical and almost impossible to adopt the PC method, though this model has been simplified in line with practical.

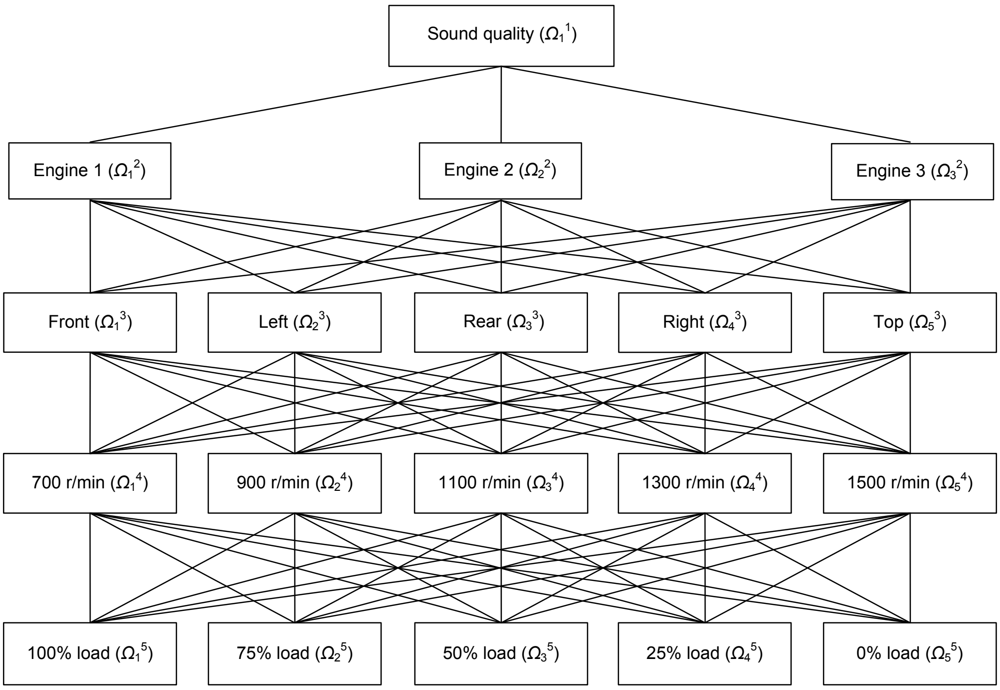

Fig. 3 Hierarchy tree of comprehensive evaluation model of diesel engine sound quality

Table 1 Preference matrix and the relative weight of three engines with different brands

Brand	Engine 1	Engine 2	Engine 3	Relative weight
Engine 1	0.5	0.5	0.5	0.33
Engine 2	0.5	0.5	0.5	0.33
Engine 3	0.5	0.5	0.5	0.33

Table 2 Preference matrix and the relative weight of five noise test positions

Test position	Front	Left	Rear	Right	Top	Relative weight
Front	0.50	0.78	0.93	0.93	0.97	0.37
Left	0.22	0.50	0.79	0.79	0.90	0.26
Rear	0.07	0.21	0.50	0.50	0.70	0.14
Right	0.07	0.21	0.50	0.50	0.70	0.14
Top	0.03	0.10	0.30	0.30	0.50	0.08

Table 3 Preference matrix and the relative weight of five speeds

Speed (r/min)	700	900	1100	1300	1500	Relative weight
700	0.50	0.99	0.99	0.99	0.73	0.41
900	0.01	0.50	0.50	0.50	0.03	0.09
1100	0.01	0.50	0.50	0.50	0.03	0.09
1300	0.01	0.50	0.50	0.50	0.03	0.09
1500	0.27	0.97	0.97	0.97	0.50	0.31

Table 4 Preference matrix and the relative weight of five loads

Load	100%	75%	50%	25%	0%	Relative weight
100%	0.50	1.00	1.00	1.00	1.00	0.46
75%	0.00	0.50	0.55	0.66	0.43	0.14
50%	0.00	0.45	0.50	0.62	0.38	0.13
25%	0.00	0.34	0.38	0.50	0.28	0.10
0%	0.00	0.57	0.62	0.72	0.50	0.16

The noise test position, speed, and load are chosen as the three factors of UDS. For the sake of convenience, levels for each factor are represented by different integers, as shown in Table 5. For the accuracy of the test, the $U_{15}(5^3)$ table shown in Table 6 is used to provide a 15-run experimental design for each brand of diesel engine. Table 7 shows the detailed information of the 15 experimental points of each diesel engine. From Eqs. (1) and (2), the square of the centered L_2-discrepancy of the $U_{15}(5^3)$ table is 0.1147, and the square of the wrap-around L_2-discrepancy of the $U_{15}(5^3)$ table is 0.2003. Fig. 4 shows the distribution chart of experimental points. From Fig. 4, it can be observed that experimental points are uniformly scattered in the 5×5×5 cubic grid.

Fig. 5 shows the improved hierarchy tree of the comprehensive evaluation model. In this improved model, there is a subjective sound quality evaluation procedure with 45 stimuli instead of 375 stimuli. The workload is decreased effectively and the jury test can be performed by GPC. Table 8 shows the relative weight of 15 experimental points. It is calculated by

$$\mathrm{rw}_{\mathrm{EP}}(i,j,k)=\mathrm{RW}_{\mathrm{Position}}(i)\times\mathrm{RW}_{\mathrm{Speed}}(j)\times\mathrm{RW}_{\mathrm{Load}}(k),\tag{17}$$

$$\mathrm{RW}_{\mathrm{EP}}(i,j,k)=\frac{\mathrm{rw}_{\mathrm{EP}}(i,j,k)}{\sum\mathrm{rw}_{\mathrm{EP}}(i,j,k)},\tag{18}$$

Table 5 Corresponding relation between levels for each factor and integers

Code	Test position	Speed (r/min)	Load
1	Front	700	100%
2	Left	900	75%
3	Rear	1100	50%
4	Right	1300	25%
5	Top	1500	0%

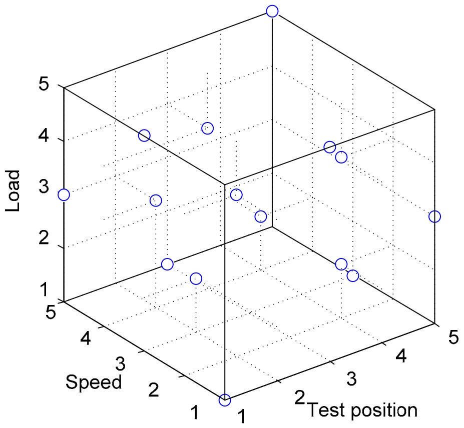

Fig. 4 Distribution chart of experimental points

Table 6 $U_{15}(5^3)$ table

No.	Factor			No.	Factor			No.	Factor		
	Ω^3	Ω^4	Ω^5		Ω^3	Ω^4	Ω^5		Ω^3	Ω^4	Ω^5
1	1	1	1	6	2	4	3	11	4	2	4
2	1	3	5	7	3	1	5	12	4	4	2
3	1	5	3	8	3	4	4	13	5	1	3
4	2	2	4	9	3	5	1	14	5	3	1
5	2	3	2	10	4	2	2	15	5	5	5

Table 7 Detailed information of the 15 experimental points of each diesel engine

No.	Test position-speed-load	No.	Test position-speed-load
1	Front-700-100%	9	Rear-1500-100%
2	Front-1100-0%	10	Right-900-75%
3	Front-1500-50%	11	Right-900-25%
4	Left-900-25%	12	Right-1300-75%
5	Left-1100-75%	13	Top-700-50%
6	Left-1300-50%	14	Top-1100-100%
7	Rear-700-0%	15	Top-1500-0%
8	Rear-1300-25%		

where $RW_{EP}(i, j, k)$ is the relative weight of an experimental point with i-noise-test-position, j-speed, and k-load; $RW_{Position}(i)$ is the relative weight of i-noise-test-position; $RW_{Speed}(j)$ is the relative weight of j-speed; $RW_{Load}(k)$ is the relative weight of k-load.

3.3 Sound quality jury test through GPC

The noise test was carried out with HEAD acoustic digital artificial head in a semi-anechoic chamber. In accordance with the international standard (ISO 6798:1995) and taking previous research as precedent, five positions (1 m away from the engine surface) of all three brands of diesel engine, as shown in Figs. 6 and 7, were used to arrange the microphones. During the measuring process, the intake and exhaust noises were eliminated, and the cooling fan was removed.

After screening and pre-processing of the recorded original sound stimuli, 45 stimuli of 4 s' length are obtained for the jury test of the evaluation model shown in Fig. 5. The jury test is implemented on a PC-based psychoacoustic system in the semi-anechoic chamber. The metric for sound quality

Table 8 Relative weight of 15 experimental points

No.	Relative weight	No.	Relative weight	No.	Relative weight
1	0.48	6	0.02	11	0.01
2	0.10	7	0.14	12	0.01
3	0.04	8	0.01	13	0.02
4	0.02	9	0.06	14	0.03
5	0.02	10	0.01	15	0.03

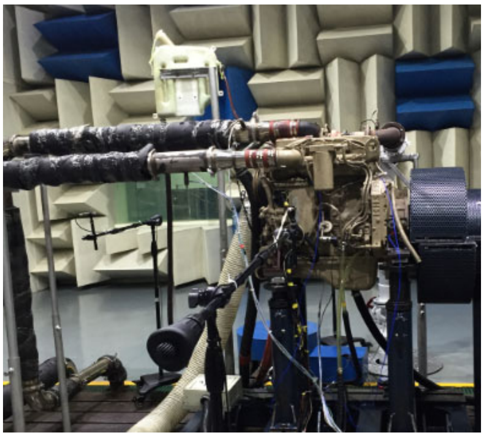

Fig. 7 Diesel engine noise test site

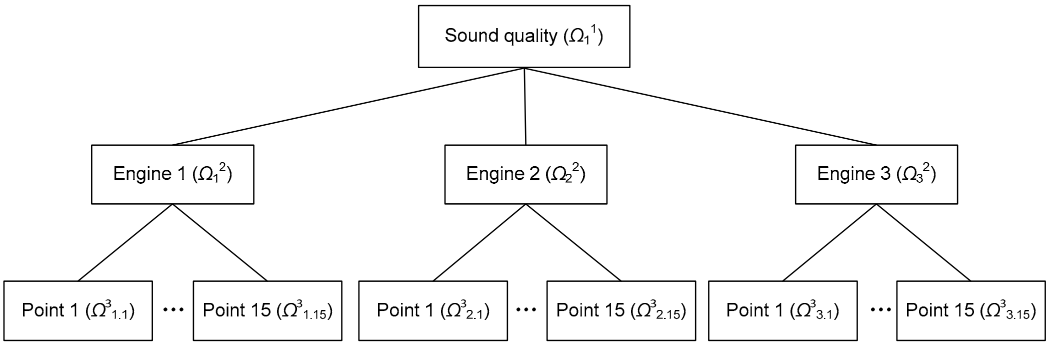

Fig. 5 Improved hierarchy tree of comprehensive evaluation model of diesel engine sound quality

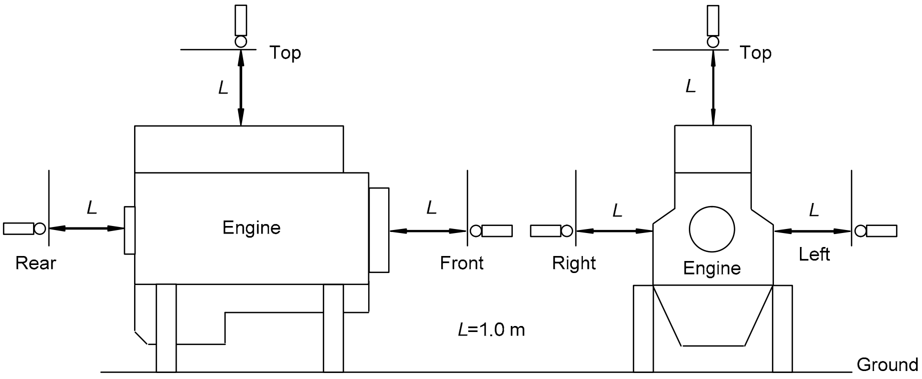

Fig. 6 Arrangement of microphones

evaluation is sensory pleasantness. For two compared stimuli, the annoying one scores 0 point and the other scores 1 point. Moreover, if it is unable to determine which one is better, the two score 0.5 points. The sensory pleasantness was the sum of the score of each comparison. The higher of the sensory pleasantness indicates the higher sound quality of sound stimuli. Twenty-five experienced evaluators who have long been involved in engine or engine related researches are chosen to complete the evaluation tasks. In accordance with the time rule of PC, the entire 45 sound stimuli are divided into three groups. For further reduction of the jury test duration, comparison of same sound event and different playback order comparisons are partially reduced.

3.4 Results and discussion

Subjective test data from each group are validated according to the weighted factor of a consistency criterion to judge the credibility of each evaluator. In the subjective sound quality test, misjudging does usually occur. An abnormal emotion or physiological characteristic of some evaluators may affect consistency during the subjective test. Data with a low weighted factor of consistency should be eliminated before the analysis. It is well recognized that removing about 10% of data from misjudging evaluators has no perceptible effect on the objective interpretation of the test (Mao *et al.*, 2005).

After removing evaluators' data with a weak consistency, the final results of the subjective test are obtained. By taking the average of the score of each noise test position, the evaluation results of the noise test position are ranked. The evaluation results of the speed and the load can be analyzed by the same method. Fig. 8 shows the evaluation result comparison of three diesel engines based on five noise test positions. It can be observed that the best sound quality of these three diesel engines is the noise of the rear-end, and the sound quality of different noise test positions is different, and the influence rule of sound quality by noise test position is also different for these three diesel engines. The evaluation result comparison of the factor is instructive and useful for a targeted acoustic design.

Table 9 shows the values assigned to five noise test positions of three diesel engines based on the sequencings of noise sound quality of different positions. It assumes that the influence rule of sound quality by noise test position is uniformly linear. This assumption is not necessary or very reasonable but is only for the convenience of regression analysis.

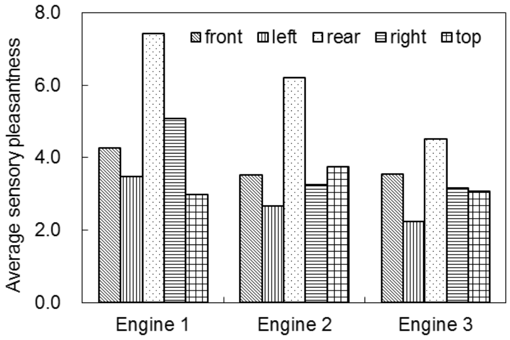

Fig. 8 Evaluation result comparison of three diesel engines based on five noise test positions

Table 9 Values assigned to five noise test positions of three diesel engines

Brand	Value				
	Front	Left	Rear	Right	Top
Engine 1	3	4	1	2	5
Engine 2	3	5	1	4	2
Engine 3	2	5	1	3	4

Looking at engine 1, the correlations between noise test position, engine speed, engine load and the results of subjective sound quality evaluation are shown in Figs. 9–11. It can be observed that the sensory pleasantness of sound stimuli is in inverse proportion to engine speed and engine load. Average sensory pleasantness (ASP) goes down as the speed or load of the engine increases. Data points are not uniformly scattered in the near field of the fitting line in Figs. 9–11. It indicates that there is no significant linear correlation between sensory pleasantness and a single factor of noise test position, speed, and load. None of the three factors has an independent linear effect on the sound quality of engine noise. There are interactional relations among these factors.

In the multivariate linear regression analysis model expressed below, the ASP is regarded as a function and the noise test position, speed, load of diesel engine 1 are viewed as independent variables. Using the SPSS statistical analysis application software, the optimum regression equation is deduced by the least squares method. Table 10 shows the results of analysis of the regression equation's variance (Rawlings *et al.*, 1998).

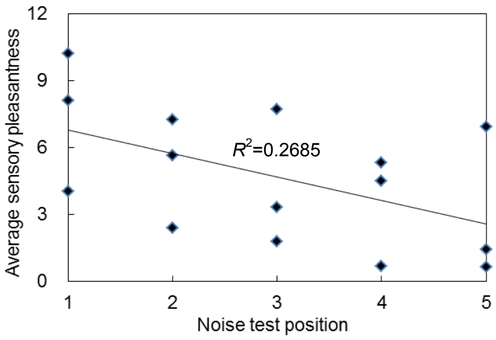

Fig. 9 Correlations between the noise test position and the results of subjective sound quality evaluation

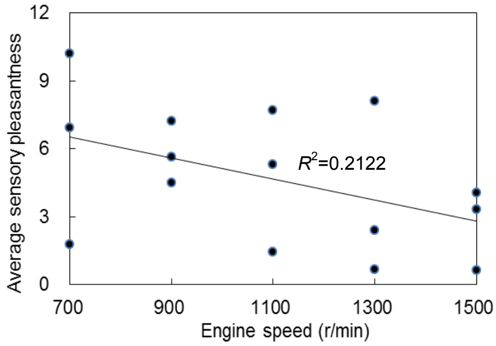

Fig. 10 Correlations between the engine speed and the results of subjective sound quality evaluation

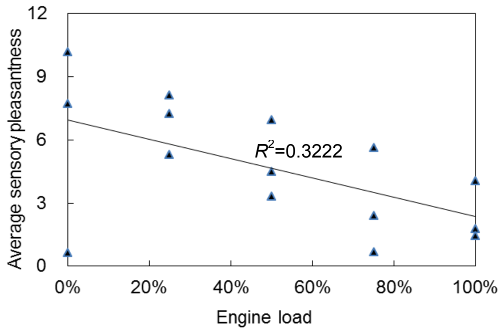

Fig. 11 Correlations between the noise engine load and the results of subjective sound quality evaluation

The results show that the regression equation passes the F-test (F=13.229>F(0.01)=5.42) at a high level of significance,

$$\mathrm{ASP} = 15.0920 - 0.0047x - 4.3402y - 1.0456z, \quad (19)$$

where x is the speed, y is the load, and z is the position.

As shown in Table 11, the scores of the sound stimuli in the subjective test are weighted according to the relative weight shown in Table 8, and then the results of comprehensive evaluation of three diesel engines' sound quality are obtained using the improved hierarchy tree of the comprehensive evaluation model. Fig. 12 shows the radar chart of the regularized results of three diesel engines' sound quality. Sound stimuli sequencing and products sequencing can clearly be observed in Fig. 12. The sound stimulus of No. 7 has the highest ASP for each brand of diesel engine. The results of comprehensive evaluation of three diesel engines sound quality are 4.65, 3.86, and 3.38, respectively, that is, the products sequencing of sound quality is diesel engine 1>diesel engine 2>diesel engine 3.

The comparison with jury test by seven grade rating scale method with the same evaluators is carried out using the entire 375 stimuli involved in the hierarchy tree shown in Fig. 3. The results of sound quality, which are acquired by adopting a weighted calculation on the hierarchy tree shown in Fig. 3, are 3.89, 3.49, and 3.28. Fig. 13 shows the relationship of the regularized results of comprehensive evaluation of the hierarchy tree and that of the improved hierarchy tree, where the ordinate is the regularized result of the hierarchy tree (AHP method), and the abscissa is the regularized result of the improved hierarchy tree (UDS-AHP method). The coefficient R^2 between the two methods is 0.9978. It is proved that data acquired by UDS-AHP method is credible.

Table 10 Analysis of variance table

Source	Sum of squares	Degree of freedom	Mean square	F-statistic	P-value
Due to model	96.5578	3	32.1859	13.2219	0.0006
Residual	26.7772	11	2.4343	–	–
Total	123.3350	14	8.8096	–	–

Table 11 Results of comprehensive evaluation of three diesel engines' sound quality

Sound stimuli	Score in the subjective test			Weight	Weighted result		
	Engine 1	Engine 2	Engine 3		Engine 1	Engine 2	Engine 3
No. 1	1.8	0.8	1.1	0.48	0.849	0.393	0.528
No. 2	7.7	7.2	6.4	0.10	0.790	0.735	0.652
No. 3	3.3	2.6	3.2	0.04	0.121	0.094	0.116
No. 4	4.5	5.3	4.6	0.02	0.101	0.118	0.104
No. 5	5.3	2.8	2.1	0.02	0.111	0.058	0.043
No. 6	0.7	0.0	0.0	0.02	0.011	0.000	0.000
No. 7	14.2	13.5	10.8	0.14	1.948	1.849	1.479
No. 8	8.1	6.6	5.5	0.01	0.070	0.057	0.048
No. 9	4.0	2.6	1.2	0.06	0.255	0.161	0.078
No. 10	5.6	1.1	1.5	0.01	0.068	0.013	0.018
No. 11	7.2	7.1	5.9	0.01	0.088	0.086	0.071
No. 12	2.4	1.6	2.2	0.01	0.021	0.014	0.019
No. 13	6.9	7.0	7.4	0.02	0.158	0.160	0.167
No. 14	1.4	3.6	1.1	0.03	0.042	0.104	0.031
No. 15	0.6	0.7	0.8	0.03	0.017	0.019	0.022
Sum					4.65	3.86	3.38

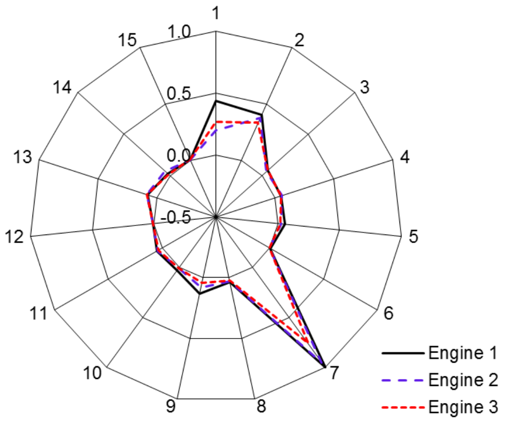

Fig. 12 Radar chart of the regularized results of three diesel engines sound quality

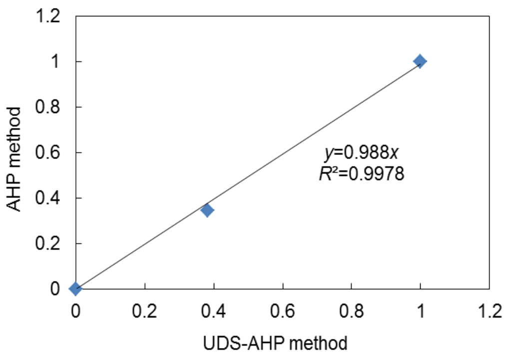

Fig. 13 Correlation between AHP method and UDS-AHP method

4 Conclusions

A comprehensive evaluation of a diesel engine sound quality method based on PC method, UDS, and the improved AHP is proposed and applied in a subjective evaluation procedure. The proposed method is instructive and useful for the acoustic design and rides comfort improvement in the vehicle's sound quality evaluation. Some conclusions can be drawn as follows.

An improved AHP with a new method calculating the relative weight based on evaluators' preference order instead of preference degree is proposed to reduce subjective liberty and difficulty of the consistency of the judgment matrix meeting the requirement. Meanwhile, the requirements for evaluators' specialized knowledge are reduced. The evaluator is not a specialist, but any common consumer.

A hierarchy tree is developed consisting of five levels including sound quality, engine brand, noise test position, engine speed, and engine load, and then the hierarchy tree is simplified into a three-level tree by UDS. Compared to the original one, the improved hierarchy tree has a great advantage on workload and consistency. It is also convenient for the improved hierarchy model to adopt the PC method or the GPC method, which makes it easy to discriminate nuances of diesel engine sound stimuli. Sound stimuli sequencing and products sequencing of three diesel engines noise sound quality are obtained.

Results show that there is a high correlation between the regularized results of comprehensive evaluation of the hierarchy tree and that of the improved hierarchy tree, which indicates that the UDS-AHP method has good agreement with the AHP method.

References

Bradley, R.A., Terry, M.E., 1952. Rank analysis of incomplete block designs: the method of paired comparisons. *Biometrika*, **39**(3-4):324-345.
http://dx.doi.org/10.2307/2334029

Chouard, N., Weber, R., 1994. On the influence of the ordering of noise in sequences for paired comparison judgment utilization. INTER-NOISE and NOISE-CON Congress and Conference Proceedings, **6**(1):1089-1092.

Chouard, N., Hmpel, T., 1999. A semantic differential design especially developed for the evaluation of interior car sounds. *The Journal of the Acoustical Society of America*, **105**(2):1280.
http://dx.doi.org/10.1121/1.426109

David, H.A., 1966. The method of paired comparisons. *Biometrics*, **22**(3):636.
http://dx.doi.org/10.2307/2528197

Fang, K.T., Ma, C.X., 2001. Orthogonal and Uniform Experimental Design. Science Press, China (in Chinese).

Fang, K.T., Lu, X., Winker, P., 2003a. Lower bounds for centered and wrap-around L2-discrepancies and construction of uniform designs by threshold accepting. *Journal of Complexity*, **19**(5):692-711.
http://dx.doi.org/10.1016/S0885-064X(03)00067-0

Fang, K.T., Lin, D.K.J., Liu, M.Q., 2003b. Optimal mixed level supersaturated design. *Metrika*, **58**(3):279-291.
http://dx.doi.org/10.1007/s001840300266

Genuit, K., 2008. Product sound quality of vehicle noise–a permanent challenge for NVH measurement technologies. *SAE Technical Paper*, No. 2008-36-0517.
http://dx.doi.org/10.4271/2008-36-0517

Hickernell, F.J., 1998a. A generalized discrepancy and quadrature error bound. *Mathematics of Computation*, **67**(221):299-322.
http://dx.doi.org/10.1090/S0025-5718-98-00894-1

Hickernell, F.J., 1998b. Lattice Rules: How Well Do They Measure Up. Springer New York, **38**:109-166.
http://dx.doi.org/10.1007/978-1-4612-1702-2_3

Huang, C.M., Lee, Y.J., Lin, D.J.K., *et al.*, 2007. Model selection for support vector machines via uniform design. *Computational Statistics & Data Analysis*, **52**(1): 335-346.
http://dx.doi.org/10.1016/j.csda.2007.02.013

Huang, Y., Chen, K.A., Yan, L., *et al.*, 2008. Adaptive grouped paired comparison: theory and selection of seeds. *Chinese Journal of Acoustics*, **4**(3):298-309 (in Chinese).
http://dx.doi.org/10.15949/j.cnki.0217-9776.2009.04.003

Jiao, F.L., Liu, K., Mao, D.X., 2005. Subjective assessment of car interior sound quality: applicability of nonmetric multidimensional scaling to sound quality research. *Acta Acustica*, **30**(6):521-529 (in Chinese).
http://dx.doi.org/10.15949/j.cnki.0371-0025.2005.06.007

Jing, M.E., Hao, Y., Zhang, J.F., *et al.*, 2005. Efficient parametric yield optimization of VLSI circuit by uniform design sampling method. *Microelectronics Reliability*, **45**(1):155-162.
http://dx.doi.org/j.microrel.2004.06.001

Lemaitre, G., Susini, P., Winsberg, S., *et al.*, 2002. Perception of the timbre of car horns. Proceedings of Forum Acusticum.

Ling, M., Milner, B., Smith, D., 2006. Acoustic environment classification. *ACM Transactions on Speech & Language Processing*, **3**(2):1-22.
http://dx.doi.org/10.1145/1149290.1149292

Liu, N.N., Wang, Y.S., Ju, J., 2014. Fuzzy comprehensive method for subjective noise annoyance evaluation of subway's interior noise. *Noise and Vibration Control*, **34**(3):101-106 (in Chinese).
http://dx.doi.org/10.3969/j.issn.1006-1335.2014.03.021

Mao, D.X., Yu, W.Z., Wang, Z.M., *et al.*, 2005. Statistical validation and criterion for paired comparison data in sound quality evaluation. *Acta Acustica*, **30**(5):468-472 (in Chinese).
http://dx.doi.org/10.15949/j.cnki.0371-0025.2005.05.014

Mao, D.X., Gao, Y.L., Zhou, Y.W., *et al.*, 2006. Grouped pair-wise comparison for subjective sound quality evaluation. *Chinese Journal of Acoustics*, **25**(3):267-276 (in Chinese).
http://dx.doi.org/10.15949/j.cnki.0217-9776.2006.03.008

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in analysis of output from a computer codes. *Technometrics*, **21**(1):239-245.
http://dx.doi.org/10.1080/00401706.1979.10489755

Parizet, E., Koehl, V., 2011. Application of free sorting tasks to sound quality experiments. *Applied Acoustics*, **73**(1): 61-65.
http://dx.doi.org/10.1016/j.apacoust.2011.07.007

Rawlings, J.O., Pantula, S.G., Dickey, D.A., 1998. Applied regression analysis: a research tool. *Technometrics*, **41**(1):82.
http://dx.doi.org/10.2307/1271013

Ronacher, A., Hussain, M., Biermayer, W., *et al.*, 1999. Evaluating vehicle interior noise quality under transient driving conditions. *SAE Technical Paper*, No. 1999-01-1683.
http://dx.doi.org/10.4271/1999-01-1683

Satty, T.L., 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, **1**(1):83-98.

Su, L., Wang, D.F., Jiang, J.G., *et al*., 2012. Fuzzy comprehensive evaluation of vehicle interior sound quality based on semantic differential method. *Journal of Jilin University (Engineering and Technology Edition)*, **42**(2): 309-315 (in Chinese). http://dx.doi.org/10.13229/j.cnki.jdxbgxb2012.02.044

Xu, Z.S., Wei, C.P., 1999. A consistency improving method in the analytic hierarchy process. *European Journal of Operational Research*, **116**(2):443-449. http://dx.doi.org/10.1016/S0377-2217(98)00109-X

Zhang, J.H., Liu, H., Bi, F.R., *et al*., 2012. Psychoacoustic study on contribution of fan noise to engine noise. *Chinese Journal of Mechanical Engineering*, **25**(4):809-815. http://dx.doi.org/10.3901/CJME.2012.04.809

Zhang, R.C., Wang, Z.J., 1996. Uniform design sampling and its fine properties. *Chinese Journal of Applied Probability & Statistics*, **12**(4):337-347.

Zhou, B., Yao, H.L., Shi, M.H., *et al*., 2012. An new immune genetic algorithm based on uniform design sampling. *Knowledge & Information Systems*, **31**(2):389-403. http://dx.doi.org/10.1007/s10115-011-0476-3

中文概要

题　目：基于 PC，UDS 和改进 AHP 的柴油机噪声品质综合评价方法

目　的：提出柴油机噪声品质综合评价方法，解决产品总体噪声品质择优排序问题。

创新点：在噪声品质综合评价模型的研究中，将均匀设计抽样与层次分析法相结合，提出一种考虑多因素、多水平的噪声品质主观综合评价 GPC-UDS-AHP 方法，通过对少量的声样本进行主观评价得到声品质综合评价结果；提出基于偏好序和 Bradley-Terry 模型的权重计算方法，避免了 1-9 标度法主观性强和一致性难以满足要求的问题。

方　法：针对传统的成对比较法难以在声品质主观评价试验中获得评价对象整体感受择优排序的不足，结合均设计抽样法“均匀分散”的特点，提出一种基于分组成对比较法的柴油机声品质主观综合评价方法，确定适合于柴油机声品质主观评价的结构及层次划分，提出基于偏好序和 Bradley-Terry 模型的各级指标权重计算方法。以采集到的某 3 款柴油机共 375 个噪声样本为评价对象，分别利用该方法和等级评分法计算得出 3 款柴油机的声品质评价综合值。

结　论：提出的柴油机声品质主观综合评价 GPC-UDS-AHP 方法可在有效地维持成对比较主观评价试验实施范围的情况下，准确地得到柴油机声品质整体主观感受评价得分的择优排序，可用于实施不同柴油机或 N 次改进试验的声品质比较和评判。

关键词：柴油机；声品质；均匀设计抽样；成对比较法；层次分析法