Mechanical Systems and Signal Processing 66-67 (2016) 875–892

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Sound quality recognition using optimal wavelet-packet transform and artificial neural network methods

Y.F. Xing, Y.S. Wang*, L. Shi, H. Guo, H. Chen

School of Automotive Engineering, Shanghai University of Engineering Science, No. 333, Longteng Road, Songjiang District, Shanghai 201620, PR China

ARTICLE INFO

Article history:
Received 11 August 2014
Received in revised form
14 December 2014
Accepted 1 May 2015
Available online 30 June 2015

Keywords:
Sound quality recognition
Human auditory perception
Optimal wavelet-packet transform
Artificial neural network
Nonstationary vehicle noise

ABSTRACT

According to the human perceptional characteristics, a method combined by the optimal wavelet-packet transform and artificial neural network, so-called OWPT–ANN model, for psychoacoustical recognition is presented. Comparisons of time–frequency analysis methods are performed, and an OWPT with 21 critical bands is designed for feature extraction of a sound, as is a three-layer back-propagation ANN for sound quality (SQ) recognition. Focusing on the loudness and sharpness, the OWPT–ANN model is applied on vehicle noises under different working conditions. Experimental verifications show that the OWPT can effectively transfer a sound into a time-varying energy pattern as that in the human auditory system. The errors of loudness and sharpness of vehicle noise from the OWPT–ANN are all less than 5%, which suggest a good accuracy of the OWPT–ANN model in SQ recognition. The proposed methodology might be regarded as a promising technique for signal processing in the human-hearing related fields in engineering.

1. Introduction

Sound quality (SQ) refers to the perceptional characteristics of the hearing sensation invoked by a sound. The characteristics of a sound as it is perceived are not exactly the same as those of the sound being emitted, but a phenomenon involved in the physics of human hearing process. It has been found that the human hearing process related not only to the physical characteristics of a sound, but also to the psychological, physiological factors and social practices of a listener (or assessor) is very complex and the human perception of sound is nonlinear. It is sometimes impossible to find an exact physical model to describe the perception response for all people, due to the individual hearing characteristics and listening environments.

The SQ of the noise from a product is of ever-increasing importance when assessing the total quality of the product. For products ranging from cars to hand tools, not only the level but also the quality of its noise is part of what attracts the customers and manufacturers. To recognize the perceiving process of the human auditory system, the psychoacoustics was introduced into SQ studies. Aiming at different sound objects in applications, a large number approaches for sound quality evaluation (SQE) have been proposed and widely discussed by the academic community in the past few decades [1–3]. Some psychoacoustic parameters (metrics), such as A-, B-, C- and D-weighted sound pressure levels (SPLs), loudness, sharpness, roughness, fluctuation strength, tonality, boom, pitch strength, impulsiveness, brightness, annoyance, and pleasantness, etc., were presented to quantitatively relate sound stimuli to human sensations [4,5]. The SQ metrics matrices can only replace subjective

* Corresponding author. Tel.: +86 21 67791149; fax: +86 21 67791152.
E-mail addresses: jzwbt@163.com, wysgcd@sues.edu.cn (Y.S. Wang).

http://dx.doi.org/10.1016/j.ymssp.2015.05.003
0888-3270/

testing if they correlate well with a jury response. The effectiveness of a certain metric is dependent on the meaning of a sound being tested, which sometimes may strongly influence subjective evaluation results of the sound. Thus, only a few SQ metrics were mathematically modeled and standardized. The international standardization organization (ISO) had generalized SQE methods into two classes. The first is the octave-band analysis for weighted SPLs. The second is referring to more appropriate values based on the human sensation of loudness. Typically, the standard ISO 532, which provide the graphic methods for loudness calculation using one-third-octave bands, have been applied in the SQE areas of engine noise, vehicle interior noise and door-slam quality [6–9]. It should be mentioned that the user choice of which SQ metric to apply is specific to individual appliance. Manufacturers who undertake SQ testing often develop their own metrics, as well as making use of the well-known ones. In the automobile industry, typically, additional SQ metrics sometime include squeaks, rattles, buzzes, itches and creaks, which may be assessed using nonstationary loudness analysis [10].

Among in-situ psychoacoustical parameters commonly used in the SQE engineering, it has been found that the weighted SPLs are not perfect methods because the sound masking effects are ignored in their calculation procedure. Based on the previously defined equal-loudness-contours, Zwicker [4] presented a methodology to calculate the specific loudness, thereby the total loudness and sharpness models. The Zwicker loudness model has been applied in the ISO 532B [6], and the sharpness model has not yet been internationally standardized, but was accepted by the German national standard DIN 45692. A loudness model for time-varying sounds proposed by Glaberg and Moore [11] has been used in the SQE of nonstationary noise, and was proved helpful for sound source identification. Many mathematical models for sound roughness calculation were proposed [12–14]. Some improved methods based on the Aures model [12] were developed by considering the sound phase difference and carrier frequencies in the critical bands of human auditory system [15,16]. Hashimoto [17] studied the rumbling noise of a vehicle below 300 Hz and suggested that the rumbling noise has a linear relationship with the boom index under the stationary working conditions. The vehicle rumbling noise was also investigated by Lee, who pointed out that the boom is strongly dependent on the loudness of a sound [18]. As a synthesized metric, the pleasantness (or annoyance) is often defined by combining the sound loudness, sharpness, fluctuation strength, tonality and roughness. Recently, some studies on the weights of these objective parameters of sounds were conducted for synthesis evaluations [19–21]. It has been found that the loudness and sharpness constitute more than 50% of the overall assessment of a sound. For the stationary vehicle noise, specially, this percentage is more than 80% [21,22]. Thus, the loudness and sharpness were mainly considered in vehicle SQE engineering.

To extract feature from the nonstationary signals, some time–frequency analysis techniques, such as the short-time Fourier transform (STFT), wavelet transform (WT), Wigner–Ville distribution (WVD) and the Hilbert–Huang transform (HHT), etc., have been frequently mentioned in literatures [23,24]. Recently, these time–frequency algorithms were introduced into sound feature extraction and fault diagnosis in vehicle engineering [25,26]. The discrete WT-based methods for sound feature extraction and octave analysis were developed and used in SQE of time-varying vehicle noises [27]. It has been found that, comparing with the frequency-based STFT, the WT can be specially designed and provide a SQE filterbank with approximate center frequencies and bandwidths as those of the critical bands in the human auditory system. The SQE methods based on the WVD and the empirical mode decomposition (EMD) in HHT were presented and applied to the door slamming of a car [28,29]. The calculated psychoacoustic parameters show good correlations with those from the subjective evaluations. Due to the complex and nonlinear characteristics of the perception process in the human auditory system, naturally, the intelligent recognition methods should be considered. To classify sound patterns, the artificial neural network (ANN) and the support vector machine (SVM) approaches have been attempted to use in the level classifications of the loudness, sharpness and annoyance indices of vehicle noise [21,26,30,31]. Based on the applications in product SQEs, both the most common-used multiple linear regression and the intelligent ANN approaches were reviewed [32], which suggests that a specially designed ANN is effective and can be used in an intelligent SQE systems, although it sometime shows a shortcoming associated with robust. More recently, Aiming at the integrated satisfaction index of the diesel engine-radiated noise, a new method combined by the SVM and genetic algorithm (GA), so-called GA–SVM, is proposed [33], which gives a good accuracy in prediction of engine noise quality. Additionally, both the GA–SVM and the back-propagation (BP) ANN were proved effective for SQE of engine noise. From the literatures, it can be found that the ANN and the SVM are feasible for SQ recognition of the stationary vehicle interior noises. For the nonstationary noises, however, effectiveness of these methods needs to be further investigated in SQE engineering.

Based on the above discussions, it may be seen that the SQE methods have been developing from the traditional to the intelligent ways. For the intelligent SQEs, specially, the approaches for sound feature extraction and quality pattern recognition need to be carefully selected and designed. Based on some measured vehicle noises, in this presented work, a new SQ recognition technique called OWPT–ANN model is developed by combining the WPT and ANN algorithms and validated by tests. The scheme of this research is shown in Fig. 1. In view of the application, the OWPT–ANN model can be directly used to evaluate SQ of the sample vehicle, and may be extended to other sound-related fields for signal processing in engineering, such as speech recognition, music evaluation and noise-based machine failure diagnosis, etc.

2. Theory background

2.1. Artificial neural network

The ANN, a mathematical model inspired by biological neural networks, is composed of many simple, densely interconnected nonlinear artificial neurons operating in parallel. The feedback network with multilayered perceptrons, which