Available online at www.sciencedirect.com

ScienceDirect

Journal of Sound and Vibration 299 (2007) 933–947

JOURNAL OF SOUND AND VIBRATION

www.elsevier.com/locate/jsvi

Sound-quality prediction for nonstationary vehicle interior noise based on wavelet pre-processing neural network model

Y.S. Wang[a,b], C.-M. Lee[b,*], D.-G. Kim[b], Y. Xu[a,b]

[a]School of Automobile and Transportation Engineering, Liaoning Institute of Technology, Jinzhou 121001, PR China
[b]School of Mechanical and Automotive Engineering, University of Ulsan, Ulsan 680-749, Republic of Korea

Received 14 October 2005; received in revised form 2 June 2006; accepted 24 July 2006
Available online 2 October 2006

Abstract

A new concept for sound-quality prediction, the so-called wavelet pre-processing neural network (WT-NN) model, is presented in this paper. Based on interior vehicle noise, the WT-NN sound-quality evaluation model was developed by combining the techniques of wavelet analysis and neural network (NN) classification. A wavelet-based 21-point model for vehicle noise feature extraction was established, as was a NN model. Verification results show that the trained WT-NN models are accurate and effective for sound-quality prediction of nonstationary vehicle noises. Due to its outstanding time–frequency characteristics, the proposed WT-NN model can be used to deal both with stationary and nonstationary signals, and even transient ones. In place of conventional psychoacoustical models, the WT-NN technique is suggested not only to predict, classify, and compare the sound quality (loudness and sharpness) of vehicle interiors, but also to apply to other sound-related fields in engineering.

1. Introduction

Reduction of the ever-increasing noise levels in our environment may improve sound quality and thereby our quality of life. Vehicle noises, which constitute about 40 percent of city environmental noise, have been considered in the past few decades, and vehicle noise control has accordingly become a very active research area. The majority of problems in vehicle acoustics concern acoustic comfort, not hearing damage. To improve acoustic comfort in vehicle designs, researchers should first understand how to evaluate a noise. A large research effort related to sound-quality evaluation (SQE) of vehicle noises has recently been conducted [1–3]. Also, the characteristics of a sound as it is perceived are not exactly the same as the characteristics of the sound being emitted, a phenomenon related to the physics of human hearing process. Thus, many psychoacoustic metrics, such as loudness, sharpness, tonality, roughness, fluctuation strength, pleasantness, etc., which can explain the quantitative relation between acoustical stimuli and human sensations, have also been applied to evaluate vehicle noises [4,5]. It can be observed that the algorithm complexity and time-consumption of the SQE methodology cannot include all human feeling factors and specify all sound

*Corresponding author. Tel.: +82 52 259 2851; fax: +82 52 259 1681.
E-mail addresses: jzwbt@163.com (Y.S. Wang), cmlee@ulsan.ac.kr (C.-M. Lee).

0022-460X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2006.07.034

sensations. In the vehicle SQE, for example, the studies often have to focus on responses to only one or two types of noise, due to the particular features of each type. It is thus both necessary and useful to develop a new, powerful approach for more accurate evaluation of sound qualities.

In SQE engineering, DSP techniques need to be carefully selected according to characteristics of the sound signals of interest. Stationary frequency-based techniques and nonstationary time–frequency techniques have been frequently mentioned in the literature [7,8]. The former is used to highlight nonunique features and therefore is not appropriate for nonstationary sounds. Time–frequency techniques, however, can be used to extract transient features of sounds [9,10]. Short-time Fourier transform (STFT) uses a standard Fourier transform over several types of windows. Wavelet-based techniques can apply a mother wavelet function with either discrete or continuous scales to surmount the fixed time–frequency resolution issue inherent in STFT. In applications, the discrete versions, such as the discrete wavelet transform (DWT) and wavelet packet analysis (WPA), are usually used for reducing the number of calculations to be done, thereby saving computer-running time. Ambiguity functions (AF) and Wigner–Ville distributions (WVD) have better resolution than STFT, but suffer from cross-term interference and produce results with coarser granularity than wavelet techniques do [7]. AF and WVD with excessive transformation durations are obviously unacceptable in the development of real-time monitoring systems. Almost all vehicle noises in actual working cases, such as those generated in acceleration and deceleration processes, are nonstationary. Therefore, based on the above findings, the STFT, DWT, WPA, and CWT techniques, which are suitable for feature extraction of nonstationary signals, may be considered in this study. However, their ability to characterize sounds for special purposes should be further tested in SQE system developments.

The sound classification techniques in common use have been investigated and compared in [11]. The results show that: self-organizing maps and learning vector quantization techniques are complementary to each other, while long-term statistics cannot be applied in combination with nonstationary feature extraction techniques. The Gaussian mixture model may be adopted in the unsupervised classification of musical signals. Neural networks (NNs) and fuzzy identification are widely used in the biomedical field for modeling, data analysis, and diagnostic classification [12–14]. The ability to reproduce arbitrary nonlinear functions of input, as well as the highly parallel and regular structure of the NNs, makes them suitable for complex pattern recognition and classification tasks [15]. Therefore, a NN algorithm is adopted to project the input signal features to the output SQE patterns of vehicle noises in this paper.

Based on the above discussion, it can be concluded that sound perception, as a neural response of humans, is an ambiguous concept: Individual differences in perception among people are as much influenced by personal factors as by noise factors. Thus, it is sometimes impossible to find an exact physical model to describe the annoyance response for all people. Therefore, for a correct assessment of sound quality, intelligent methods such as fuzzy logic and NNs should naturally be considered. Based on vehicle interior noises, a new SQE technique of wavelet pre-processing neural network (WT-NN) is developed in this work. The WT-NN model may be directly used to predict not only nonstationary, but also stationary and even transient sound-quality matrices (SQMs) in vehicle engineering. In view of the application, it may also be used in sound recognition, failure diagnostics of equipments and clinical disease diagnostics in medical treatment, etc.

2. Establishment of vehicle noise database

In this paper, sample vehicle interior noises were prepared using the binaural recording technique. The following data acquisition parameters were used: signal length, 10 s, sampling rate, 22 050 Hz. Inevitably, distortion of the measured sounds by certain additive noises occurred, which came from both ambient background noise and the hardware of the measurement system; therefore, the signal needed to be denoised. Some techniques for white noise suppression in common use, such as the least square, spectral subtraction, matching pursuit methods, and the wavelet threshold method have been used successfully in various applications [16–18]. The wavelet threshold method in particular has proved very powerful in the denoising of a nonstationary signal. Therefore, a DWT-based shrinkage denoising technique was applied. This technique may be performed in three steps: (a) decomposition of the signal, (b) determination of threshold and nonlinear shrinking of the coefficients, and (c) reconstruction of the signal. Mathematically, the soft threshold signal is sign(x) ($|x|-t$) if $|x|>t$, and otherwise is 0, where t denotes the threshold. The candidate threshold rules and other options in the denoising functions in nonlinear shrinking were also carefully investigated. Finally, the selected parameters were: Daubechies wavelet "*db*3," 7 levels, soft universal threshold equal to the root square of $2\log(\text{length}(f))$, assuming the model is basic and with unscaled noise. As an example, a denoised interior signal is shown in Figs. 1 and 2 is the corresponding spectrum. It can be seen that the harmony and white noise components of the sample interior noise are well-controlled. The wavelet shrinkage denoising technique is effective and sufficient for denoising vehicle noises. After signal denoising, a database of vehicle interior noises was established for evaluation by using filtering technique. The parameters of the filters designed by Matlab toolbox are shown in Table 1. Eighteen noise signals were generated by filtering the interior noises from the right "ear" of the dummy head, and 18 from the left "ear".

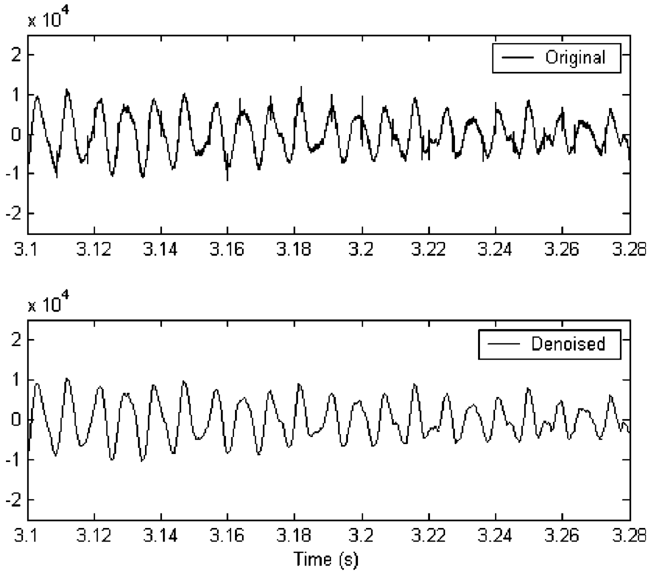

Fig. 1. Comparison of the interior noises before and after the wavelet-denoising model.

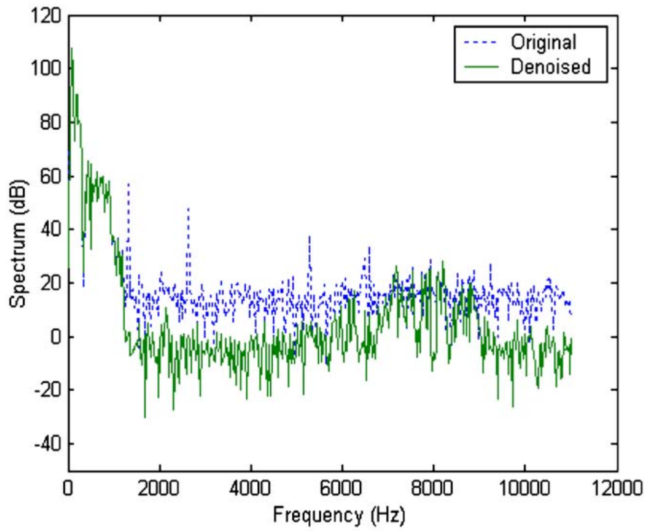

Fig. 2. Comparison of the interior noise spectra before and after the wavelet-denoising model.

Table 1
The parameters of the designed filters

Band number	1	2	3	4	5	6	7	8	9
Sampling rate (Hz)	345	345	345	2756	2756	2756	22 050	22 050	22 050
Center frequency (Hz)	14.1	28.1	56.3	112.5	225	450	900	1800	3600
Bandwidth (Hz)	9.9	19.9	39.8	79.5	159.1	318.2	636.4	1272.8	2545.6

3. Feature extraction of vehicle noises

3.1. Wavelet scalogram analysis

Wavelet analysis (WA) is widely used in various subfields of mathematics, science, and engineering. Due to its outstanding transient trait in time–frequency domain, WA is regarded as an effective approach for nonstationary data processing. WA is the process of decomposing or reconstructing a signal using wavelets, a family of orthogonal functions of type,

$$\Psi_{a,b}(t) = |a|^{-1/2}\psi[(t-b)/a], \quad a, b \in R \quad a \neq 0, \tag{1}$$

generated from a wavelet function $\psi(t)$ by dilation and translation operations, which are governed by the scale factor a and shift factor b, respectively. The CWT of a signal $x(t)$ is defined as

$$W_x(a,b) = |a|^{-1/2}\int_{-\infty}^{+\infty} x(t)\psi^*[(t-b)/a]\,\mathrm{d}t = \langle x(t), \Psi_{a,b}(t)\rangle, \tag{2}$$

where $\psi^*[(t-b)/a]$ is the complex conjugate of $\psi[(t-b)/a]$. DWT is defined by setting $a = a_0^{-j}$, $b = a_0^{-j}kb_0$ $(j, k \in Z)$. If $a_0 = 2$, $b_0 = 1$, Eq. (2) may be rewritten as

$$W_x(a,b) = W_x(2^{-j}, 2^{-j}k) = 2^j \int_{-\infty}^{+\infty} \psi(2^j t - k)x(t)\,\mathrm{d}t. \tag{3}$$

A scalogram may be defined as square of the wavelet decomposition coefficients

$$\mathrm{SCAL}_f(a,b) = \left|W_x(a,b)\right|^2. \tag{4}$$

3.2. Wavelet packet analysis

Superior to STFT, WT giving an equal exponential frequency partition has been regarded as a prominent advantage in signal processing. However, frequency resolution of WT decreases with increasing frequency. This shortcoming protects WT from applications in sound signal analysis, especially for analyzing nonstationary or transient sounds. We want an adjustable resolution, which can be arbitrarily selected according to the characteristics and the study requirements of a signal. WPA provides this property. It can be defined by using the scaling function $\phi(t)$ and the wavelet function $\psi(t)$. Let $u_0(t) = \varphi(t)$, $u_1(t) = \psi(t)$, and define

$$u_{2n}(t) = \sqrt{2}\sum_k h_k u_n(2t-k), \quad u_{2n+1}(t) = \sqrt{2}\sum_k g_k u_n(2t-k), \tag{5a,b}$$

specifies a wavelet packet $\{u_n(t) + |n \in Z_+\}$. An orthogonal basis is given by the $\{u_{j,m,n}(t) = 2^{j/2}u_m(2^j t - m)$ $|j, m \in Z, n \in Z_+\}$ where j, m, and n are the scale, translation, and oscillation parameters. Thus, WPA may be derived as

$$C_m^{j,2n} = \sum_{k=-\infty}^{\infty} h^*_{2m-k} C_k^{j+1,n}, \quad C_m^{j,2n+1} = \sum_{k=-\infty}^{\infty} g^*_{2m-k} C_k^{j+1,n}. \tag{6a,b}$$

Imposing Eq. (6) on the original signal $C^{0,0}$, one may achieve the complete WPA of a signal.

3.3. Twenty-one-point feature extraction models for vehicle noises

For extracting the features of vehicle noises, the above-mentioned WSA methods including CWT, DWT, and the WPA were performed and discussed in this paper. This procedure was performed using the filtering algorithm, i.e. Mallat algorithm in the Matlab toolbox. The parameters selected for CWT calculation are: the Daubechies wavelet with a filter length of two (center frequency $F_c = 0.6667\,\mathrm{Hz}$), scaling factor $a = 1$–300. Figs. 3 and 4 show the CWT results and its corresponding auto-spectrum of an interior noise, respectively. Beyond the FFT, CWT provides the information not only in the time domain, but also in the frequency domain, and therefore in the time–frequency plane. The interior noise in Fig. 3 shows some in-line energy peaks fluctuating along the time axis within the almost-fixed frequency ranges. It can be seen that the energies of the vehicle noises mainly distribute in a low-frequency range (below 1000 Hz) and decrease with increasing

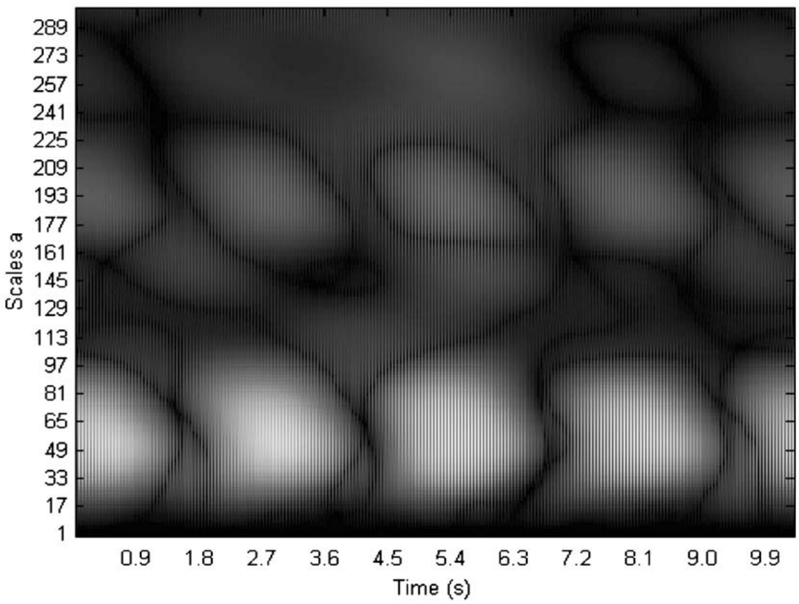

Fig. 3. CWT of an interior noise signal.

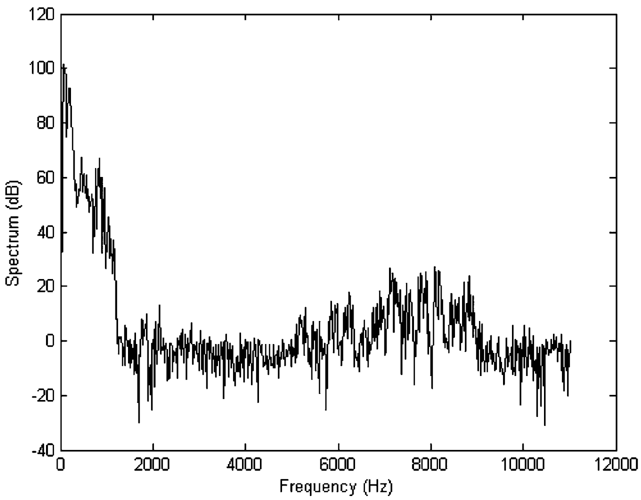

Fig. 4. Auto-spectrum of an interior noise signal.

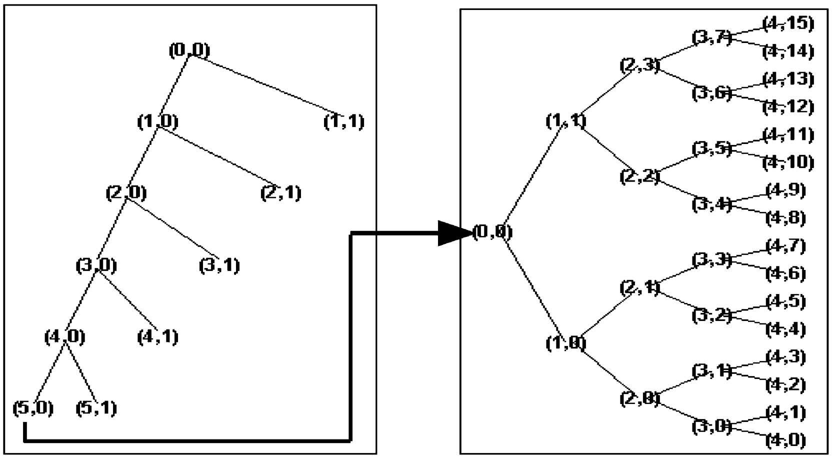

Fig. 5. Twenty-one-point feature extraction model for interior noise analysis.

Table 2
The calculation parameters selected in the 21-point feature extraction model

Analysis types	Discrete wavelet transform (DWT)	Wavelet packet analysis (WPA)
Wavelet functions	Daubechies	Daubechies
Filter lengths	6	4
Cost functions	—	Shannon entropy
Analysis levels	5	4
Order of vanish moments	3	2

frequency. According to the obtained signal energy distributions, a 21-point modeling combined by DWT and WPA algorithms for feature extraction of vehicle interior noise was established and is shown in Fig. 5. Two steps are conducted in this feature extraction model: a five-level DWT and a complete WPA with four levels. The selected parameters in the 21-point model are listed in Table 2. The Shannon entropy cost function is usually used in acoustical engineering, due to its minimal number of coefficients for signal reconstruction with a small mean square error. It is defined as

$$E(s) = -\sum_i s_i^2 \log(s_i^2). \tag{7}$$

Here, s is the sound signal and s_i is decomposed WPA coefficients in. The four-level WPA has $2^4 = 16$ terminal node signals. Using the 21-point model, the feature of the interior noises is extracted, and an example is shown in three dimensions in Fig. 6. Note that each point represents a frequency band and the point numbers are arranged from low to high frequencies. Fig. 7 shows the correction in calculation of the newly proposed 21-point feature extraction model.

It should be mentioned that, to determine a reasonable point number, a set of tests were performed in the Matlab toolbox by adjusting both the DWT and WPA levels from two to ten. The test results suggest that, considering a tradeoff between the refined features of the vehicle noises and the number of computation in the following NN models, the combination of five- level DWT followed by four-level WPA with total 21 points is the most felicitous for vehicle noises, and therefore was finally selected in this paper.

4. Psychoacoustic modeling for vehicle noises

In SQE engineering, psychoacoustical attributes such as loudness, sharpness, roughness, etc., have been considered and generalized in SQMs. The purpose of this study is to predict the SQMs of vehicle noises using

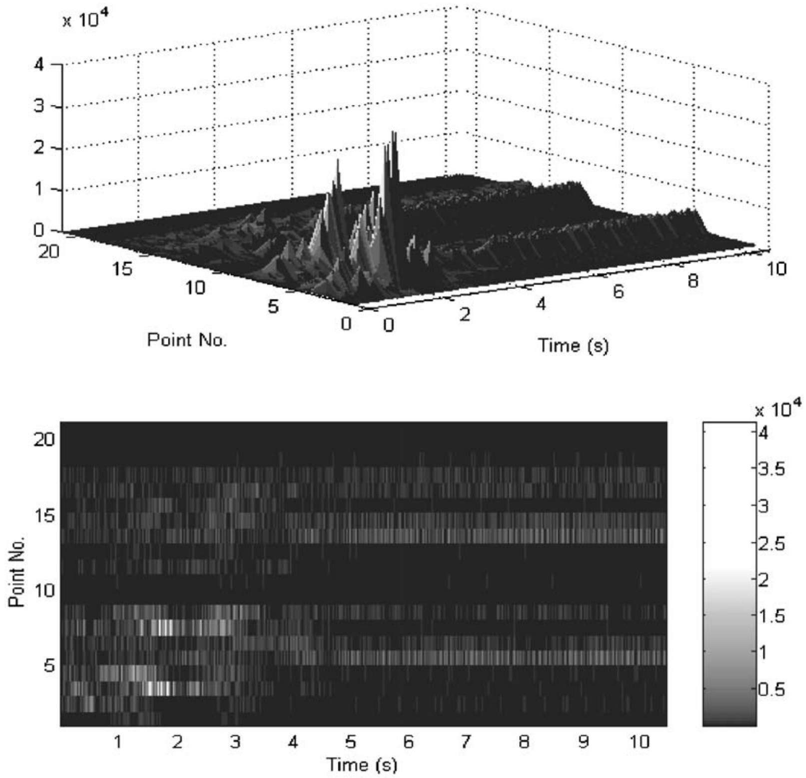

Fig. 6. Feature of the interior noise in time–frequency map extracted by the 21-point model.

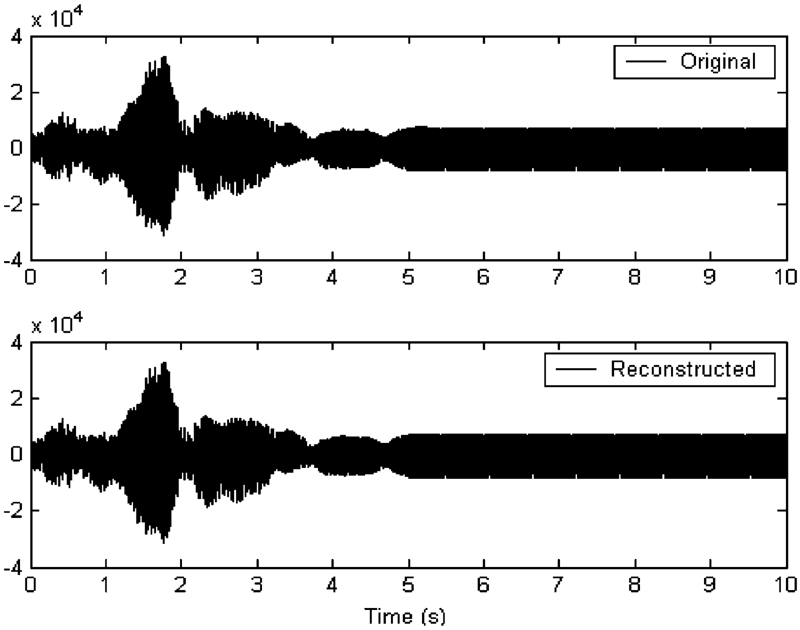

Fig. 7. Verification of the reconstructed interior noises using the 21-point model.

WT-NN models. As the inputs of the NN models, the above WA results provide the time–frequency features of the signals. However, the output SQMs need to be discussed. Here we investigate only the loudness and sharpness, which are related to the SEQ of the vehicle noises.

The measurement procedure for loudness considers the critical-band spectra in human hearing, according to the Zwicker model [19] (DIN 45631, ISO 532B). A specific loudness can be calculated from the dB level for each third-octave band under the assumption that a relative change in loudness is proportional to a relative change in intensity. Masking curves can then be constructed around these levels representing the effects of critical bands. The total loudness is calculated by integrating the specific loudness in the individual critical bands,

$$N = \int_0^{24\mathrm{Bark}} N'(z)\,\mathrm{d}z, \tag{8}$$

where N is the loudness in sones, N' is the specific loudness in sone/Bark, and z is the critical band rate in Bark. A formula for converting frequency values f (in kHz) into bark values z is

$$z(f) = \begin{cases} 11.82\arctan(1.21f), & f \leqslant 1.5, \\ 5\ln(f/1.5) + 12.61, & f > 1.5. \end{cases} \tag{9}$$

Based on the specific loudness, specific sharpness S' can be calculated as,

$$S' = 0.11\,\frac{\int_0^{24\mathrm{Bark}} N' g(z) z\,\mathrm{d}z}{\int_0^{24\mathrm{Bark}} N'\,\mathrm{d}z}, \tag{10}$$

where the weighting function $g(z)$ is determined by psychoacoustics in Eq. (11),

$$g(z) = \begin{cases} 1, & 0 \leqslant z \leqslant 16, \\ 0.066\mathrm{e}^{0.171z}, & 16 < z \leqslant 24. \end{cases} \tag{11}$$

Then total sharpness S in "acum" can be obtained by integrating the specific sharpness

$$S = \int_0^{24\mathrm{Bark}} S'\,\mathrm{d}z. \tag{12}$$

The vehicle noises considered are partly nonstationary; therefore, the temporally involved characteristic of the psychoacoustic parameters in SQM need to be taken into account. For varying sounds, Zwicker's tracked evaluation approach [20] considering temporal effects has been widely accepted, although it is not yet an ISO standard. The Zwicker method allows us to distinguish between unmasked and masked contours in the one-third octave SPL pattern. In this work, the loudness is computed based on the Zwicker model, the selected cutoff frequencies in the designed third-octave filters are 689, 1378, 2756, 5512 and 11 025 Hz, respectively, and the frame size is 46.4 ms, corresponding to 1024 points under the presented sampling rate: 22 050 Hz.

The specific values of loudness and sharpness of the interior noises (signal 1) are shown in Figs. 8 and 9, which indicate an uneven characteristic of loudness and sharpness in the domains of time and critical band rate. The time-varying total loudnesses in Fig. 10 suggest a curtain relationship between the sound loudness and its signal amplitude (energy). The signal components in vehicle noises, which significantly affect the

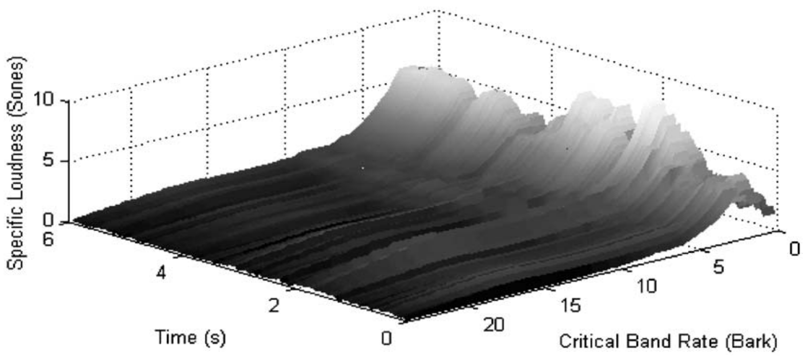

Fig. 8. Specific loudness of an interior noise (signal 1) in time–frequency map.

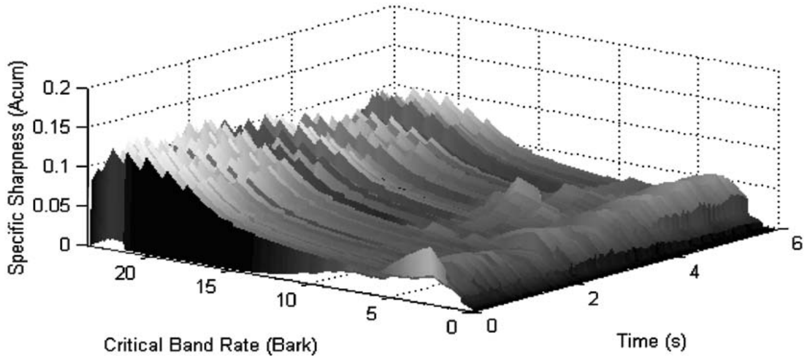

Fig. 9. Specific sharpness of an interior noise (signal 1) in time–frequency map.

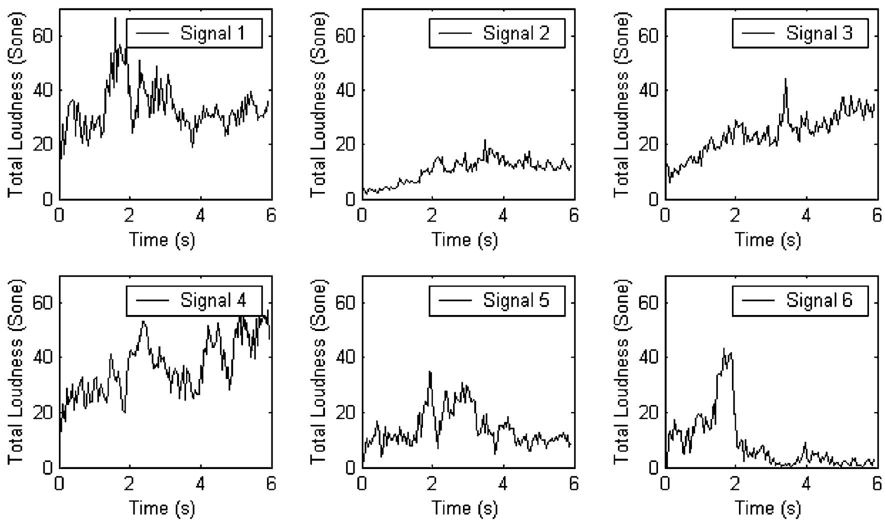

Fig. 10. Temporal total loudness of each interior noise based on Zwicker loudness model.

perceived loudness, mainly lie in the frequency range below 2400 Hz; the sharpness components, however, lie in the high frequency range. Accordingly, these noise components need to be taken into account and reduced in the vehicle designs.

5. Architecture of wavelet pre-processing neural network

5.1. Theory of neural networks

The NN acquires the ability to generalize based on the training data and, if these represent an entire population well, the network can predict outcomes for new, previously unseen data. In many described NNs, one of the most commonly used structures is the multilayer perceptron. In this work, we present an important and much-applied supervised feed-forward NN structure, and a multi-layer perceptron (MLP). A three-layer perceptron, which has proven efficient for representing decision problems, was adopted (Fig. 11) [21]. The sample $\{x_k\}$ is fed to the network and produces an output $\{y\}$. The input pattern $\{x_k\}$ is then propagated through the network in the following way:

$$y_i = f\left(\sum_{j=1}^{M} w_{ij}^{(2)} f\left(\sum_{k=1}^{N} w_{jk}^{(1)} x_k\right)\right), \tag{13}$$

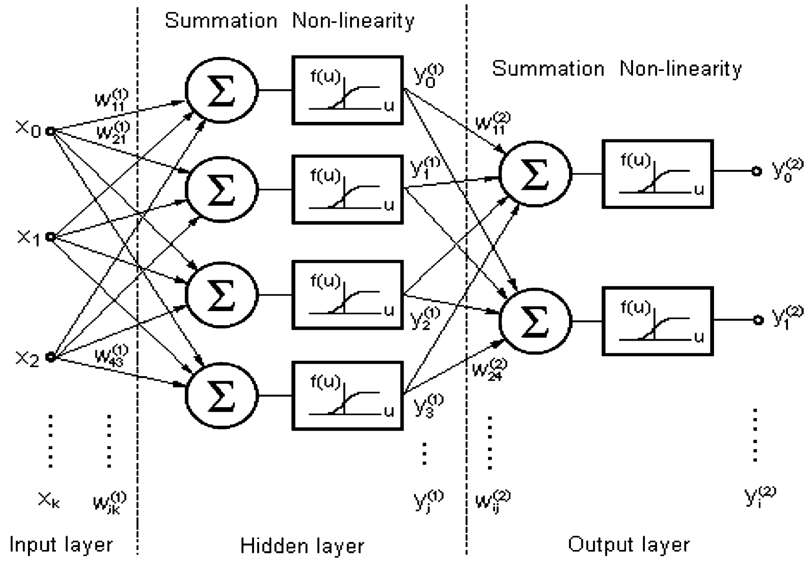

Fig. 11. A simple multi-layer perceptron with a hidden layer.

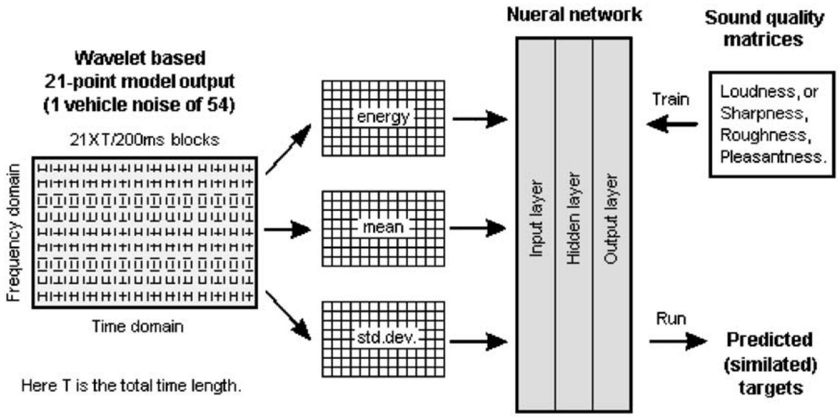

Fig. 12. Schematic presentation of the data inputs and outputs to the neural network.

where y_i denotes the output unit i, and N, M denote the number of input units and hidden units, respectively. $w_{ij}^{(n)}$ is the a from a "source" unit j to a "target" unit i, and n the layer.

To implement this procedure, one needs to calculate the error derivative with respect to weight in order to change the weight by an amount that is proportional to the rate at which the error changes as the weight is changed. The back-propagation algorithm is used here.

5.2. Wavelet pre-processing neural network modeling for SQE

An overall schematic presentation of the WT-NN model is shown in Fig. 12. A vehicle noise is first fed to the wavelet-based preprocessor. Then, the SQM of the interior noise is taken from the psychoacoustical model as the outputs; after training the NN, the sound sensations can be intelligently mimicked and predicted.

To improve performance of the WT-NN, the size of the network input layer must be kept small, and thus reduction of the redundancy in the input data is required. The data reduction is performed in time and scale domains. As the preprocessor of a network, CWT is regarded as too time-consuming and cumbersome. The reduction work has been partially done in the scale domain by using the DWT-based 21-point feature extraction model that we deliberately designed for vehicle noise quality estimation. In the time dimension, the concept of temporal masking in the auditory models is introduced to divide the time-axis of the 21-point output into $T/200$ ms frames, thus creating 21 by $T/200$ ms blocks for each of the noise signals, where T is the signal length and 200 ms is the frame length usually used in psychoacoustics. For acquiring all signal feature information, a strategy called extract certain statistics is applied to the 21-point outputs, in which statistical parameters such as mean, standard deviation, median, and maximum are considered. In this study, only the mean and standard deviation of the wavelet preprocessor output are taken into account. The mean scalogram characterizes the energy of each frame of the vehicle noise, and should be able to detect energy deformations. The standard deviation provides information about the temporal traits of the signal: how the scalogram fluctuates with time. The reduction result of a vehicle noise with 21 points by 50 frames named energy feature unit shown in Fig. 13. Accordingly, the mean and standard deviation units may be easily calculated. Thus, we arrange them in a line forming an extracted feature matrix (EFM) structure for each signal as shown in Fig. 14. For a training signal set with a number of n, there will be n feature matrices. Therefore, the total extracted feature matrix (TEFM) may be fabricated by apposing the feature blocks of the n signals, thus with a size of $n \times 3150$.

The SQM from psychoacoustical models in Fig. 12 as outputs of the NN model may be represented in different ways. In this work, it was decided to train one network for all considered sensations, including both the total and the specific information, using only one type of vehicle noise. For a nonstationary interior noise, by taking the above time-tracked evaluation results from the Zwicker loudness models, the output SQM may be expressed as

$$\mathrm{SQM} = \begin{bmatrix} \mathrm{LOUDESS} & \mathrm{SHARPNESS} \end{bmatrix}^{\mathrm{T}}. \tag{18}$$

Thus, the corresponding total SQM (TSQM) may be represented as Eq. (18) by n signals. Then, the network can be trained to project the input feature matrices to the output evaluation matrices.

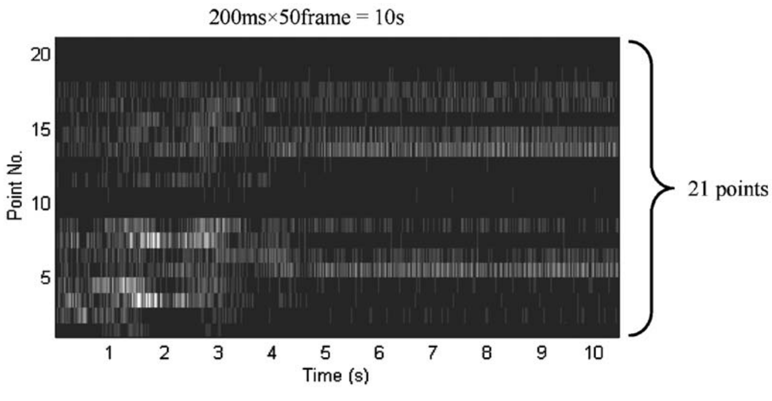

Fig. 13. A feature extraction unit (21 × 50) from the time–frequency map of a signal.

Unit of Energy (21×50)	Unit of Mean (21×50)	Unit of Standard Deviation (21×50)
21 points×50 frames×3 = 3150 blocks		

Fig. 14. The input feature matrix structure of each signal for the neural network.

6. Experimental prediction and verification

As mentioned, the basic NN structure chosen for the current work is a three-layer perceptron with a single hidden layer. The optimal number of hidden neurons is difficult to determine, but typically it should be selected between the number of input and output nodes, and should be much smaller than the number of training samples. We used to modify the network structure by adding or deleting the hidden neurons to improve the network performance, and then determined parameters of the NNs for loudness and sharpness evaluation of the interior noises, see Table 3.

To verify the accuracy of the WT-NN model, the NN shown in Table 3 was trained and simulated. The signals used in the training and simulation procedures are from the left and right "ears" of the dummy head, respectively. The training performance of the WT-NN model is shown in Fig. 15. As seen through 3079 training echoes, i.e. 3079 upgrades of the weightings in each layer of the NN, the sum-squared error approached and finally reached a preset target value of 0.002. All the interior noises in the vehicle noise database were used in the model training. After training, we, respectively, fed the signals from the right "ear" to the Zwicker loudness model and the trained WT-NN model. Figs. 16–18 show the calculated and predicted

Table 3
MLP architecture and training parameters in the NN model for loudness

Architecture	
The NN type	Momentum adaptive learning rate ("TRAINGDX")
The number of layers	3
The number of neuron on the layers	Input: 3150, hidden: 60, output: 50
The initial weights and biases	The Nguyen–Widrow method ("NEWFF")
Activation functions	Log-sigmoid function ("LOGSIG")
Training parameters	
Learning rule	Back-propagation
Adaptive learning rate	Initial: 0.001, increase: 1.05, decrease: 0.7
Momentum constant	0.95
Sum-squared error	0.002

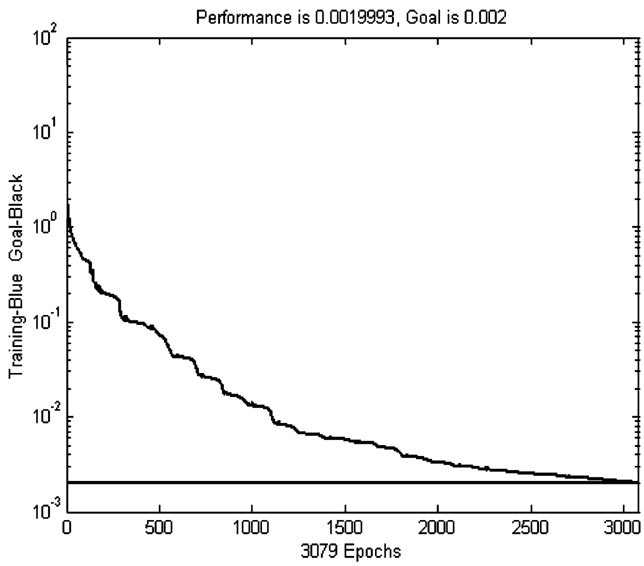

Fig. 15. The NN training performance in the WT-NN model.

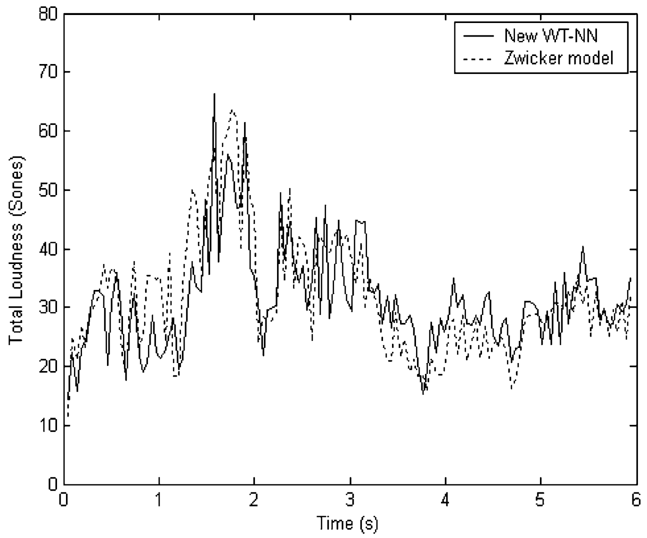

Fig. 16. Total loudness comparison between the Zwicker model and the WT-NN model using an interior noise (right "ear").

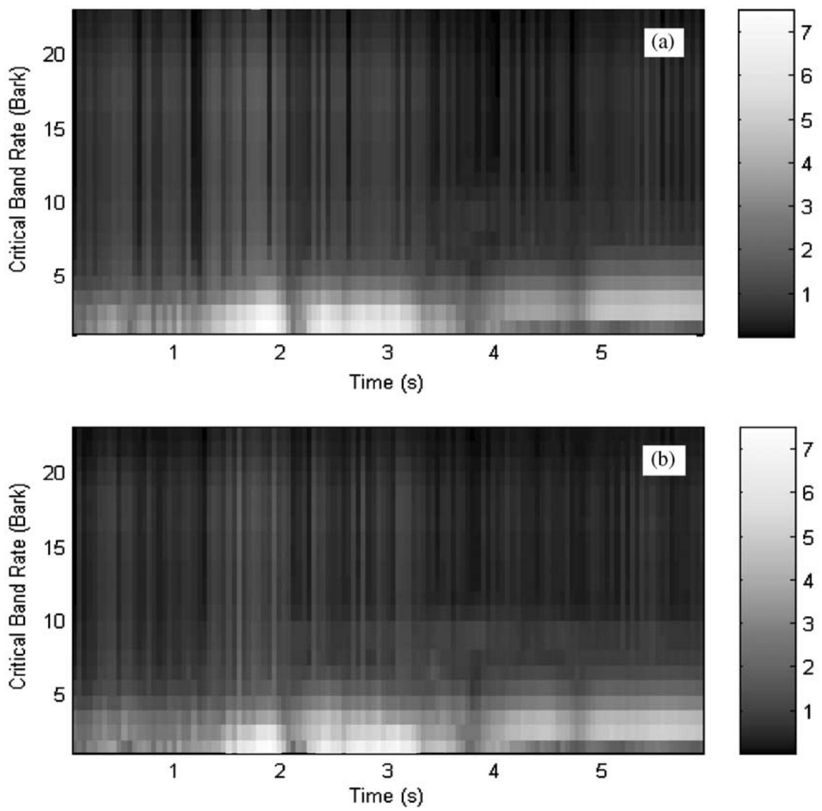

Fig. 17. Specific loudness comparison between (a) the Zwicker model, and (b) the WT-NN model (right "ear").

results of the interior noise samples. It can be seen that the predicted total and specific loudnesses in Figs. 16 and 17 and the specific sharpness in Fig. 18 are consistent with those from the Zwicker models. Following the above validation procedure, we obtained similar comparison results by using the signals from the left "ear" in

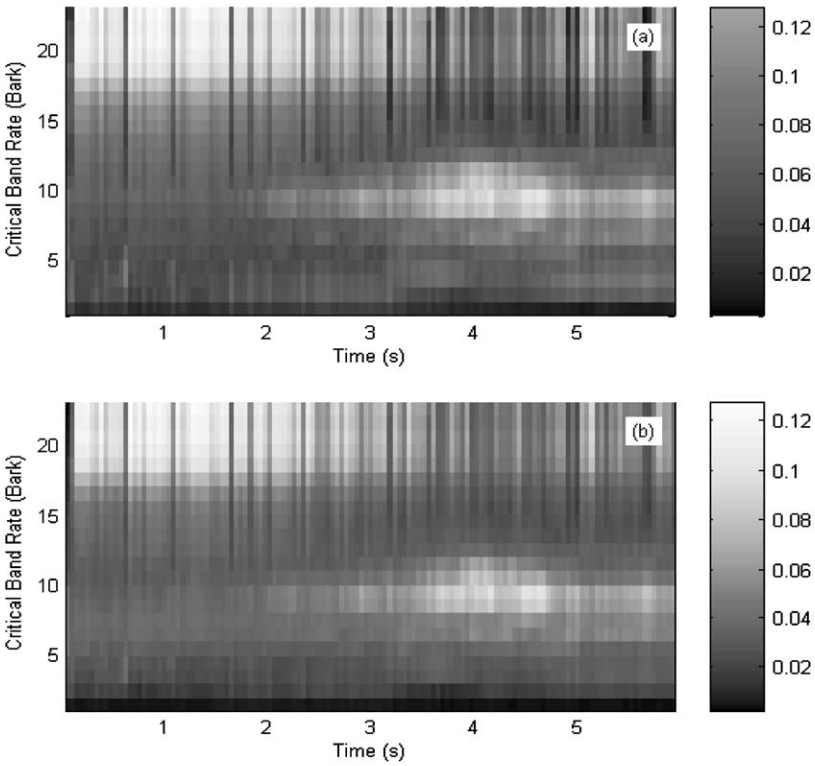

Fig. 18. Specific sharpness comparison between (a) the Zwicker model, and (b) the WT-NN model (right "ear").

the vehicle noise database. Thus, the conclusion may be drawn that the newly proposed WT-NN model is precise enough to predict sound sensations in vehicle SQE engineering.

It may also be found, comparing with the in situ psychoacoustical models, that the WT-NN model can be used to predict different sensations in a target SQM at the same time, whenever it is trained and fixed. In addition, as seen, the WT-NN model deals with the sound sensations separately in the training process; this makes it possible to extend its prediction scope to the SQE parameters in subjective evaluation domain. This point goes beyond any psychoacoustical models, and its benefit might be seen in future research.

7. Summary and conclusions

This paper investigated a new WT-NN SQE technique to predict the sound sensations of vehicle noise. Based on the denoised interior noises, a wavelet-based, 21-point model was developed and applied to extract the noise features. Considering the SQMs from the conventional psychoacoustical models, a three-layer NN model was designed and trained by the adaptive-learning-rate BP algorithm with gradient descent momentum for intelligent identification of the noises. In applications, the WT-NN models can be used to estimate and compare sound quality of vehicles, and may be extended not only to other sound signals, such as speech, music, and machinery failure noise, but also to other engineering fields of signal processing.

It can be concluded that the 21-point model is effective for feature extraction of nonstationary vehicle noises, and the WT-NN predictions are in agreement with those from the psychoacoustical models. This implies that the WT-NN models are accurate enough to map a nonstationary vehicle noise to its SQMs. Another principle has been found that the training time of the WT-NN models is roughly proportional to the values of their training goals. Also, the WT-NN model can deal with any kind of signal, save computation time, and provide any element in SQMs including some subjective sensations, and is a good substitute for complex psychoacoustical models in vehicle SQE engineering.

Acknowledgments

This work was supported by 2005 Research Fund of University of Ulsan, Korea.

References

[1] E. Zwicker, H. Fastl, *Psychoacoustics, Facts and Models*, Springer, Berlin, Germany, 1990.
[2] H. Fastl, The psychoacoustics of sound-quality evaluation, *Acustica* 83 (1997) 754–764.
[3] F.K. Brandl, W. Biermayer, A new tool for the onboard objective assessment of vehicle interior noise quality, *SAE* 1999-01-1695 (1999) 1–11.
[4] H. Murata, H. Tanaka, H. Takada, Y. Ohsasa, Sound quality evaluation of passenger vehicle interior noise, *SAE* 931347 (1993) 675–681.
[5] E. Terhardt, G. Stoll, M. Seewann, Algorithm for extraction of pitch and pitch salience from complex tonal signals, *Journal of the Acoustical Society of America* 71 (3) (1982) 679–688.
[7] L. Cohen, *Time–Frequency Analysis*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.
[8] B. Gold, N. Morgan, *Speech and Audio Signal Processing*, Wiley, New York, 2000.
[9] B. Hubbard, *The World According to Wavelets: The Story of a Mathematical Technique in the Making*, Wellesley, USA, 1995.
[10] Y.S. Wang, C.M. Lee, L.J. Zhang, Wavelet analysis of vehicle nonstationary vibration under correlation four-wheel random excitation, *International Journal of Automotive Technology* 5 (4) (2004) 257–268.
[11] T. Kohonen, *Self-organizing Maps*, Springer, Berlin, Germany, 1997 (Printed in the USA).
[12] T. Villmann, Neural network approaches in medicine—a review of actual developments, in: *Proceedings of the European Symposium on Artificial Neural Networks*, Bruges, Belgium, 2000, pp. 165–176.
[13] F. Gurgen, Neural network based decision making in diagnostic applications, *IEEE Engineering in Medicine and Biology Magazine* 18 (4) (1999) 89–93.
[14] P.J.G. Lisboa, A review of evidence of health benefit from artificial neural networks in medical intervention, *Neural Networks* 15 (1) (2002) 11–39.
[15] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, Wiley, New York, 2001.
[16] R. Le Bouquin, Enhancement of noisy speech signals: application to mobile radio communications, *Speech Communication* 18 (1) (1996) 3–19.
[17] N. Ruiz-Reyes, P. Vera-Candeas, J. Curpián-Alonso, R. Mata-Campos, J.C. Cuevas-Martínez, New matching pursuit-based algorithm for SNR improvement in ultrasonic NDT, *NDT&E International* 38 (6) (2005) 453–458.
[18] Y.Y. Kim, J.-C. Hong, N.-Y. Lee, Frequency response function estimation via a robust wavelet de-noising method, *Journal of Sound and Vibration* 244 (4) (2001) 635–649.
[19] E. Zwicker, H. Fastl, U. Widmann, K. Kurakata, S. Kuwano, S. Namba, Program for calculating loudness according to DIN 45631 (ISO 532B), *Journal of Acoustic Society of Japan (E)* 12 (1) (1991) 39–42.
[20] E. Zwicher, Procedure for calculating loudness of temporally variable sounds, *Journal of the Acoustical Society of America* 62 (1977) 675–682.
[21] T.J. Sejnowski, C.R. Rosenberg, Parallel networks that learn to pronounce English text, *Complex Systems* 1 (1987) 145–168.