PAN
ARCHIVES OF ACOUSTICS
Vol. **46**, No. 1, pp. 55–65 (2021)
DOI: 10.24425/aoa.2021.136560



## Research Paper

# Development of a Sound Quality Evaluation Model Based on an Optimal Analytic Wavelet Transform and an Artificial Neural Network

Mehdi POURSEIEDREZAEI[(1)], Ali LOGHMANI[(2)*], Mehdi KESHMIRI[(2)]

[(1)] *Mechanical Engineering Group, Pardis College*
*Isfahan University of Technology*
Isfahan 84156-83111, Iran; e-mail: poorseied@yahoo.com

[(2)] *Department of Mechanical Engineering*
*Isfahan University of Technology*
Isfahan 84156-83111, Iran; e-mail: mehdik@iut.ac.ir
*Corresponding Author e-mail: a.loghmani@iut.ac.ir

(*received March 6, 2020; accepted October 2, 2020*)

The purpose of this study was to develop a sound quality model for real time active sound quality control systems. The model is based on an optimal analytic wavelet transform (OAWT) used along with a back propagation neural network (BPNN) in which the initial weights and thresholds are determined by particle swarm optimisation (PSO). In the model the input signal is decomposed into 24 critical bands to extract a feature matrix, based on energy, mean, and standard deviation indices of the sub signal scalogram obtained by OAWT. The feature matrix is fed into the neural network input to determine the psychoacoustic parameters used for sound quality evaluation. The results of the study show that the present model is in good agreement with psychoacoustic models of sound quality metrics and enables evaluation of the quality of sound at a lower computational cost than the existing models.

**Keywords:** analytic wavelet transform (AWT), sound quality evaluation (SQE), psychoacoustic metrics, back propagation neural network (BPNN).

## 1. Introduction

Sound quality is a perceptual reaction that reflects the degree of the listener's satisfaction with a given sound. As the physical characteristics of acoustic signals does not represent the perceived attributes of sound, sound quality is often evaluated by groups of listeners in jury tests, which is a time consuming task (Wang *et al.*, 2014).

Sound quality can also be evaluated with the use of sound quality (SQ) metrics based on psychoacoustic models (Blauert, Jekosch, 1998; Lyon, 2000; Hafke-Dys *et al.*, 2016; Pleban, 2014). The main purpose of SQ metrics is to replace jury tests with an acoustic measurement that would provide an accurate prediction of sound quality judgements made by human listeners. In recent studies several SQ metrics, such as loudness (Fastl, Zwicker, 2007; Klonari *et al.*, 2011), sharpness (Leite *et al.*, 2008; Wang *et al.*, 2007), roughness (Aures, 1985a; Miśkiewicz *et al.*, 2007; Szczepańska-Antosik, 2008; Vencovský, 2016), and tonality (Aures, 1985b; Cuddy *et al.*, 2007; Terhardt *et al.*, 1982) were developed and used for sound quality evaluation (Carletti, 2013; Pleban, 2010). The calculations of individual SQ metrics have been used to compute combined, overall measures of sound quality, such as pleasantness and unbiased annoyance (Kaczmarek, Preis, 2010).

The complex nature of auditory system signal processing causes various difficulties and limitations in the calculation of SQ metrics, therefore various so called intelligent methods, based on artificial neural networks, were proposed (Chen *et al.*, 2015; Maleczek, 2008). A previous study (Pourseiedrezaei *et al.*, 2019) has shown that a combination of an optimised artificial neural network (ANN) and a wavelet packet transform (WPT) can be used for reliable prediction of sound quality at a low computational cost.

Most of previous SQ studies that employed artificial neural networks (Huang *et al.*, 2015; 2017; Lee,

Lee, 2009) estimated the overall sound quality by feeding individual SQ metrics, such as loudness, sharpness, roughness, and tonality into the network. As the main computational load is the calculation of those metrics, the use of an artificial neural network does not reduce the amount of calculations. It also should be noted that previous models had various limitations in sound quality evaluation. The model developed by Xing *et al.* (2016) predicted loudness and sharpness but did not estimate roughness and tonality. Pourseiedrezaei *et al.* (2019) have shown that there was only weak correlation between the energy matrices at the inputs of the neural network and the prediction of roughness and tonality at the outputs. Thus, two other indices, i.e., the mean and standard deviation of each subsignal were added to the inputs of the ANN to estimate all the sound quality metrics.

The aim of the present study was to develop a model for the prediction of sound pleasantness that could be implemented in real time active sound quality control (ASQC) systems, such as those used in neonatal intensive care units (NICU). It was assumed that the model would have the possibly lowest computational load while meeting the calculation accuracy requirements for real time active noise control (ANC) systems (Kuo, Morgan, 1996). In the present study an analytical wavelet transform (AWT) was designed to create a filter bank corresponding to the critical band model of signal processing in the auditory system. The modified model, called OAWT-BPNN (optimised analytical wavelet transform – back propagation neural network), is a combination of an optimised AWT and a back propagation neural network (BPNN).

## 2. Background theory

### *2.1. Psychoacoustic metrics*

Pleasantness is predicted from the calculations of loudness, sharpness, roughness, and tonality. A detailed description of those metrics is available in the literature (Fastl, Zwicker, 2007). Pleasantness is calculated in MATLAB using Eq. (1) from Fastl and Zwicker (2007) and implemented in the LabVIEW Sound and Vibration Toolkit (NI-Tutorial-1526, 2013; Technical note, 2015):

$$P = e^{-0.55R} e^{-0.113S} \left(1.24 - e^{-2.2T}\right) e^{-(0.023L)^2}, \quad (1)$$

where $P$ denotes pleasantness, $L$ – loudness, $S$ – sharpness, and $T$ – tonality.

### *2.2. Artificial neural network*

In most applications the back propagation (BP) algorithm has been used to train the neural network. In the present algorithm the weights and thresholds of the network are computed through the gradient of the error. If the ANN output does not reach the desired value the computation process moves backwards to adjust the weights with the BP algorithm. Any continuous nonlinear function can be estimated by a three layer back propagation neural network (Hecht-Nielsen, 1992).

To minimise the training error the transfer functions $f_1$ and $f_2$ are selected as logsig and purelin, respectively (Pourseiedrezaei *et al.*, 2019). The training error is computed as the mean square error (MSE) between the predicted and desired outputs, defined by Eqs (2a) and (2b).

$$E = \frac{1}{2} \sum_{k=1}^{q} \frac{E_k}{(q \cdot m)}, \quad (2a)$$

in which

$$E_k = \sum_{i=1}^{m} (y_i^k - d_i^k)^2, \quad (2b)$$

where $q$ is the number of training samples, $m$ is the number of outputs, $y_i^k$ and $d_i^k$ are the actual and desired outputs of the $i$-th node for the $k$-th training sample. The formula for updating the weights between network layers is (Fausett, 1994):

$$w^p = (1 - \alpha) w^{p-1} - \eta \left( \frac{\partial E}{\partial w} \right), \quad (3)$$

where $w$ is the weight between network layers, $p$ denotes the number of training samples, $\eta$ is the learning rate, and $\alpha$ $(0 \le \alpha \le 1)$ is the momentum coefficient.

One of the most effective parameters in the neural network performance is selecting the initial weights and thresholds (Jaddi, Abdullah, 2018; Zhang *et al.*, 2007). A particle swarm optimisation (PSO) algorithm is used to optimise the initial values of weights and thresholds (Pourseiedrezaei *et al.*, 2019). The particle swarm is created in an $N$-dimensional search space, using Eq. (4)

$$N = n_i n_h + n_h n_o + n_{h_{\text{bias}}} + n_{o_{\text{bias}}}, \quad (4)$$

where $n_i$, $n_h$, and $n_o$ are the number of neurons at the input and the numbers of hidden and output layers of the neural network, respectively, and $n_{h_{\text{bias}}}$ and $n_{o_{\text{bias}}}$ are the numbers of biases in the hidden and output layers. The fitness function is the mean square error (MSE) calculated from Eq. (2).

### *2.3. Analytic wavelet transform (AWT)*

Owing to its fine time-frequency resolution AWT is an efficient way for analysing non-stationary signals. The selection of a suitable mother wavelet is the main point in the wavelet decomposition. The mother wavelet of the AWT is defined as:

$$\psi(t) = g(t) e^{j\eta t}, \quad (5)$$

where $j^2 = -1$, $\eta$ is a parameter related to the frequency and $g(t)$ is a real value function, usually a Gaussian:

$$g(t) = \frac{1}{(\sigma^2\pi)^{1/4}} e^{-t^2/2\sigma^2} \tag{6}$$

where $\sigma$ determines the shape of the Gaussian function. The wavelet transformation of sound signal $f(t)$ is defined as:

$$\begin{aligned} W(u) &= \int_{-\infty}^{+\infty} \frac{1}{s} f(t) g\left(\frac{t-u}{s}\right) e^{-j\,\eta(t-u/s)}\,\mathrm{d}t \\ &= \mathrm{conv}\,(f(t), h_s(t)), \end{aligned} \tag{7}$$

$$h_s(t) = \frac{1}{s} g\left(\frac{1}{s}\right) e^{j\,\eta(t/s)},$$

where $s$ is the scale factor, and $u$ is the shift factor. The Fourier transforms of $g(t)$ and $\psi(t)$ are described as follows:

$$\widehat{g}(\omega) = (4\pi\sigma^2)^{1/4} e^{-\sigma^2\omega^2}, \qquad \widehat{\psi}(\omega) = \widehat{g}(\omega - \eta). \tag{8}$$

## 3. Design of an optimal wavelet analytic transform (OAWT) for nonstationary signal decomposition

To decompose the signal into 24 critical bands the wavelet parameters ($\eta$, $\sigma$, $s$) should be tuned in accordance with the critical bands. The frequency parameters of the mother wavelet: the lower limit frequency, the upper limit frequency, and the centre frequency correspond exactly to those used in Zwicker's critical band model (Fastl, Zwicker, 2007).The centre angular frequency $\omega_c$, based on Eq. (8) is:

$$\omega_c = 2\pi f_c = \frac{\eta}{s}, \tag{9}$$

where $f_c$ is the band's centre frequency, and $\eta$ and $s$ denote the frequency shifting factor and the scale factor, respectively.

Similarly, the lower and the upper limits for angular frequencies are defined as:

$$\omega_l = 2\pi f_l, \qquad \omega_u = 2\pi f_u. \tag{10}$$

The energy difference between the centre and lower limit frequencies of the wavelet is expressed as:

$$\Delta dB_l = 20\left|\log_{10}\frac{\widehat{g}(\eta - s\omega_l)}{\widehat{g}(\eta - s\omega_c)}\right| = 10\frac{\sigma^2\eta^2\left(1-\frac{\omega_l}{\omega_c}\right)^2}{\log_e 10}. \tag{11}$$

Similarly, for the energy difference between the centre frequency and the upper limit frequency of wavelet:

$$\Delta dB_u = 20\left|\log_{10}\frac{\widehat{g}(\eta - s\omega_u)}{\widehat{g}(\eta - s\omega_c)}\right| = 10\frac{\sigma^2\eta^2\left(1-\frac{\omega_u}{\omega_c}\right)^2}{\log_e 10}. \tag{12}$$

The energies of the upper and lower limit frequencies are half of that of the centre frequency and are equal to $\Delta dB_l = \Delta dB_u = 3$ dB therefore:

$$\omega_c - \omega_l = \omega_u - \omega_c. \tag{13}$$

Accordingly,

$$\left(1-\frac{\omega_l}{\omega_c}\right)^2 = \left(\frac{\omega_c-\omega_l}{\omega_c}\right)^2 = \left(\frac{\omega_u-\omega_c}{\omega_c}\right)^2 = \left(1-\frac{\omega_u}{\omega_c}\right)^2. \tag{14}$$

Considering Eqs (11)–(14), we may write:

$$\sigma^2\eta^2 = \frac{3\cdot\log_e 10}{10\left(1-\frac{\omega_u}{\omega_c}\right)^2} = \frac{3\cdot\log_e 10}{10\left(1-\frac{\omega_l}{\omega_c}\right)^2}. \tag{15}$$

Having $\omega_c$, $\omega_l$ and $\omega_u$ values for each critical band, Eq. (15) is equal to a constant value that is represented by $\Gamma$. Assuming that $\sigma = 1$ (Zhu, Kim, 2006):

$$\eta = \sqrt{\Gamma}. \tag{16}$$

Based on Eqs (9) and (15), the transform parameters are obtained for each of the 24 critical bands. By choosing the OAWT parameters, as shown in Table 1, a suitable sound quality evaluation model is obtained.

Table 1. Optimal mother wavelet parameters based on 24 critical bands, calculated from Eqs (9) and (15).

| Wavelet | Frequency band [Hz] | Scale factor [$s$] | Frequency shifting factor ($\eta$) |
|---|---|---|---|
| 1 | 0–100 | 0.002646 | 0.831129 |
| 2 | 100–200 | 0.002646 | 2.493387 |
| 3 | 200–300 | 0.002646 | 4.155645 |
| 4 | 300–400 | 0.002646 | 5.817903 |
| 5 | 400–510 | 0.002405 | 6.875704 |
| 6 | 510–630 | 0.002205 | 7.895726 |
| 7 | 630–770 | 0.001890 | 8.311291 |
| 8 | 770–920 | 0.001764 | 9.364054 |
| 9 | 920–1080 | 0.001653 | 10.389113 |
| 10 | 1080–1270 | 0.001392 | 10.279754 |
| 11 | 1270–1480 | 0.001260 | 10.883833 |
| 12 | 1480–1720 | 0.001102 | 11.081721 |
| 13 | 1720–2000 | 0.000945 | 11.042143 |
| 14 | 2000–2320 | 0.000827 | 11.220242 |
| 15 | 2320–2700 | 0.000696 | 10.979652 |
| 16 | 2700–3150 | 0.000588 | 10.804678 |
| 17 | 3150–3700 | 0.000481 | 10.351335 |
| 18 | 3700–4400 | 0.000378 | 9.617351 |
| 19 | 4400–5300 | 0.000294 | 8.957724 |
| 20 | 5300–6400 | 0.000241 | 8.840191 |
| 21 | 6400–7700 | 0.000204 | 9.014554 |
| 22 | 7700–9500 | 0.000147 | 7.941900 |
| 23 | 9500–12000 | 0.000106 | 7.147710 |
| 24 | 12000–15500 | 0.000076 | 6.530180 |

### 3.1. Feature extraction based on OAWT

OAWT is a kind of wavelet transformation with adjustable independent resolution in individual frequency bands, therefore it can be used for decomposing the signal into the frequency bands corresponding to those of the human hearing system, described in the critical band model (Fastl, Zwicker, 2007). Figure 1 shows an example of signal decomposition with the use of OAWT.

The model presented in (Xing *et al.*, 2016) estimates only the loudness and sharpness of sound and cannot be used for the prediction of roughness and tonality (see Pourseiedrezaei *et al.*, 2019). To estimate roughness and tonality two other statistical indices are added in the present model which are the mean and the standard deviation of the OAWT output scalogram used as the ANN input. The sound is analysed in the time and in the frequency domains. The temporal masking effects in the human auditory system are reflected by setting the sound resolution to 50 ms in the time domain and the frequency masking is taken into account by setting the frequency interval to 24 critical bands. Thus, each sound signal is partitioned into $24 \times T/50$ ms blocks where $T$ is signal duration and 50 ms is the frame length commonly used in psychoacoustics (Fastl, Zwicker, 2007).

A block diagram of the OAWT model for extracting the sound features is shown in Fig. 2. The sound signal is first divided into $M$ frames of 50 ms, then each frame is divided into 24 sub-signals by the OAWT. The energy value $E_i$ for each subsignal is obtained as:

$$E_i = \sum_t [a_i(t)]^2 \, \Delta t, \tag{17}$$

where $a_i(t)$ and $\Delta t$ are the amplitude of the $i$-th subsignal and the time interval of $a_i(t)$.

The mean and standard deviation of the scalogram is calculated for each sub signal from Eqs (18) and (19):

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Scal}_i, \tag{18}$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\mathrm{Scal}_i - \mu)^2, \tag{19}$$

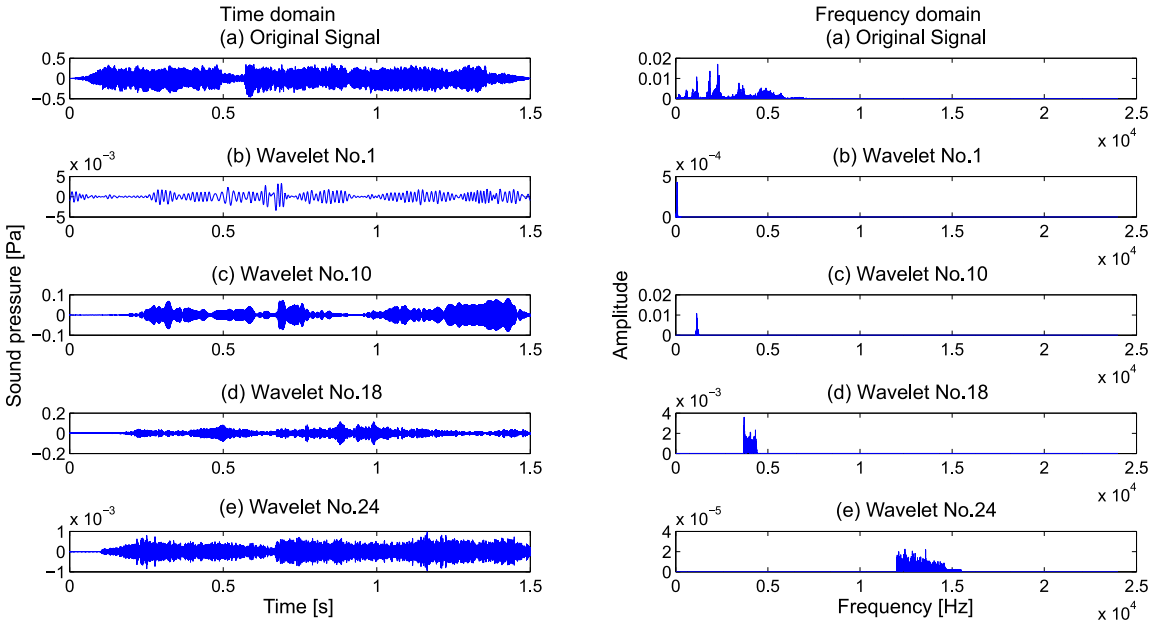


Fig. 1. Example of signal decomposition with the use of OAWT.

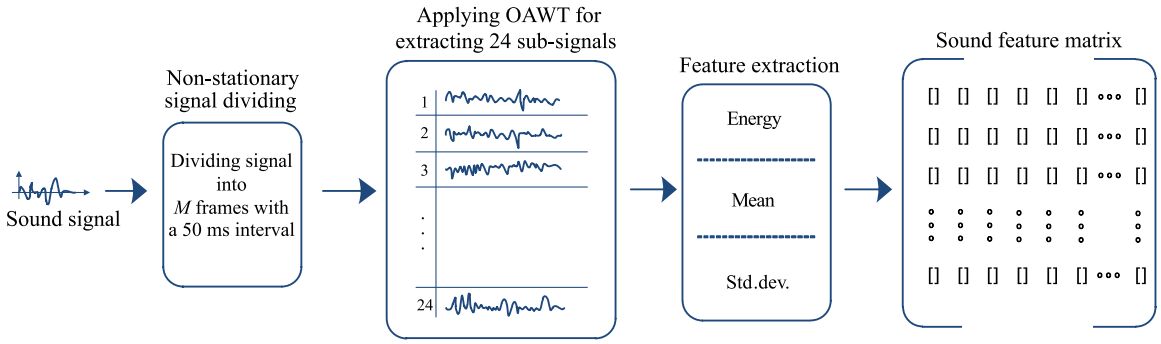


Fig. 2. Block diagram of the OAWT model for the extraction of the sound feature matrix.

where the scalogram of the signal is contained in $\mathrm{Scal}_1$ through $\mathrm{Scal}_N$, $i$ is an index from 1 to $N$, $N$ is the number of samples, $\mu$ is the mean scalogram, and $\sigma$ is the standard deviation of the scalogram.

The total feature matrix, extracted by juxtaposing the feature blocks of $n$ signals with a size of $n \times (24 \times 3 \times T/50)$ is fed to the BPNN. The outputs of the BPNN, which are the sound quality matrices (SQM), are expressed as:

$$\mathrm{SQM} = [\mathrm{Loudness\ \ Sharpness\ \ Roughness\ \ Tonality}]^{\mathrm{T}}. \tag{20}$$

### 3.2. Development of the model for sound quality prediction

In order to verify the model a sound database was selected to incorporate the model into an ASQC system for neonatal intensive care units (NICU). One of the criteria for NICU ranking is the quality of the acoustic environment (Dunn *et al.*, 2013). Noise is produced in NICU facilities by equipment and human activity, and the noise levels can be considerably increased by baby crying (Olbrych, 2010).

In this study the Oxford vocal (OxVoc) sounds database was used for extracting the sound indices. The OxVoc is a collection of natural affective vocal sounds from infants and adults (Parsons *et al.*, 2014). The database consists of 173 nonverbal sounds representing various emotional states. Non-verbal vocalisations are useful for recognising the psychological features, as such sounds do not involve any individual characteristics and do not pose the problem of authenticity. To predict the sound quality the extracted sound features must be mapped against the corresponding psychoacoustic metrics. For this purpose a BPNN has been employed with a PSO algorithm used to determine the BPNN initial weights and threshold. The successive steps in predicting the objective psychoacoustic parameters are as follows (Fig. 3):

*Step 1:* Identification of the input and output nodes with the consideration of three characteristics, including energy, mean, and standard deviation scalogram of the output sub signals of the OAWT. The number of neurons in the input layer is $3 \times 24 = 72$. The number of output layer neurons is 4 to represent loudness, sharpness, roughness, and tonality. The pleasantness index is obtained from Eq. (1).

*Step 2:* Selection of the hidden layer neurons: the number of neurons in the hidden layer is set to 12 (Pourseiedrezaei *et al.*, 2019).

*Step 3:* Determination of $f_1$ and $f_2$, which are the network transfer functions in the hidden layer and the output layer. Logarithmic sigmoid and linear functions are used for $f_1$ and $f_2$, respectively (Pourseiedrezaei *et al.*, 2019).

*Step 4:* Optimisation of the initial weights and thresholds of the BPNN using the PSO algorithm.

*Step 5:* Training of the BPNN.

*Step 6:* Computation of the hidden layer and the output layer values.

*Step 7:* Determination of the MSE of the network (Eq. (2)).

*Step 8:* Updating of the weights and the thresholds of the neural network (Eq. (3)).

*Step 9:* If the obtained MSE is less than a predetermined value, the BP algorithm stops.

### 3.3. Architectural design of the PSO-BPNN

It should be noted that 70%, 15%, and 15% of the sound samples were randomly used for network training, validation, and testing the neural network. An accurate prediction model can only be obtained when the network training parameters are determined correctly. In this study a three layer neural network was chosen to predict the SQ metrics. The neural network consists of 72 neurons in the input layer, corresponding to the sound feature vector in the form of $X = [x_1, x_2, ..., x_{72}]^{\mathrm{T}}$ and four neurons in the output layer in the form of $Y = [y_1, y_2, y_3, y_4]^{\mathrm{T}}$ which corresponds to loudness, sharpness, roughness, and tonality. The structure of the BPNN was presented in a previous study (Pourseiedrezaei *et al.*, 2019). Table 2 gives a summary of the optimal parameters.

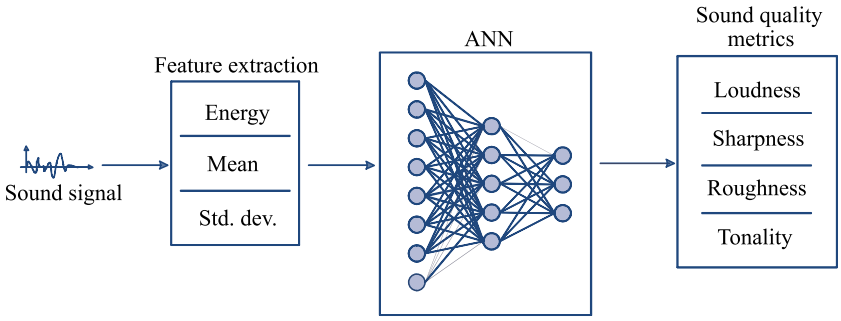


Fig. 3. Block diagram of a OAWT-BPNN model for predicting psychoacoustic metrics (Pourseiedrezaei *et al.*, 2019).

Table 2. Optimal parameters for the PSO-BPNN model.

| Type of algorithm | Parameters | Value |
|---|---|---|
| BP | Number of the hidden layers | 1 |
| | Number of neurons in the input layer | 72 |
| | Number of neurons in the hidden layer | 12 |
| | Number of neurons in the output layer | 4 |
| | Transfer function of the input-hidden layer | Logsig |
| | Transfer function of the hidden-output layer | Purelin |
| | Training function | Levenberg-Marquardt |
| | Momentum factor | 0.9 |
| | Learning rate | 0.5 |
| | Training target of MSE | 0.001 |
| | Testing performance | MSE |
| PSO | Population size of PSO | 200 |
| | Max generation | 100 |
| | Acceleration factors | 2 |
| | Inertial factor | 0.7 to 0.4 |
| | Particle dimension | 928 |

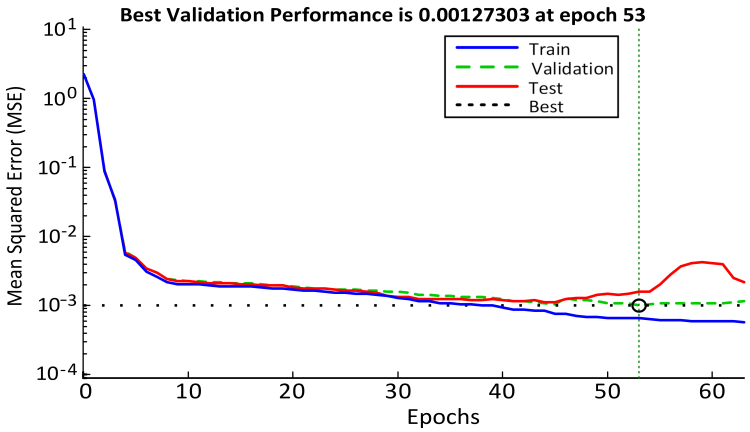


Fig. 4. Training performance of the OAWT-BPNN model.

Figure 4 shows a performance curve of the trained PSO-BPNN model. The abscissa is the number of epochs and the ordinate is the network prediction error. The data indicate that the validation and testing sets are rapidly reduced and the network becomes stable at the 53rd repetition and converges to an error of 0.001, which means that the network parameters were well chosen for sound quality estimation of non-verbal sounds.

Before network training all data should be normalised to a $[-1, 1]$ interval in order to eliminate the magnitude effect of the data and secure the model against a high prediction error. Normalisation is made using Eq. (21):

$$x_i' = (x_i - x_{\min})/(x_{\max} - x_{\min}), \tag{21}$$

where $x_i'$ is the normalised value and $x_{\min}$ and $x_{\max}$ are the minimum and maximum values.

## 4. Results, analysis, and discussion

Figure 5 shows – in individual panels for loudness, sharpness, roughness, and tonality – the normalised error between the calculations made with the use of psychoacoustic models (Zwicker, Fastl, 2007) and values predicted from the OAWT-BPNN model. It is apparent in Fig. 5 that the values predicted form the OAWT-BPNN model are in good agreement with the psychoacoustic model.

Figure 6 shows the regression plots of the OAWT-BPNN model and the corresponding psychoacoustic model. All the R-squared coefficient values given in in-

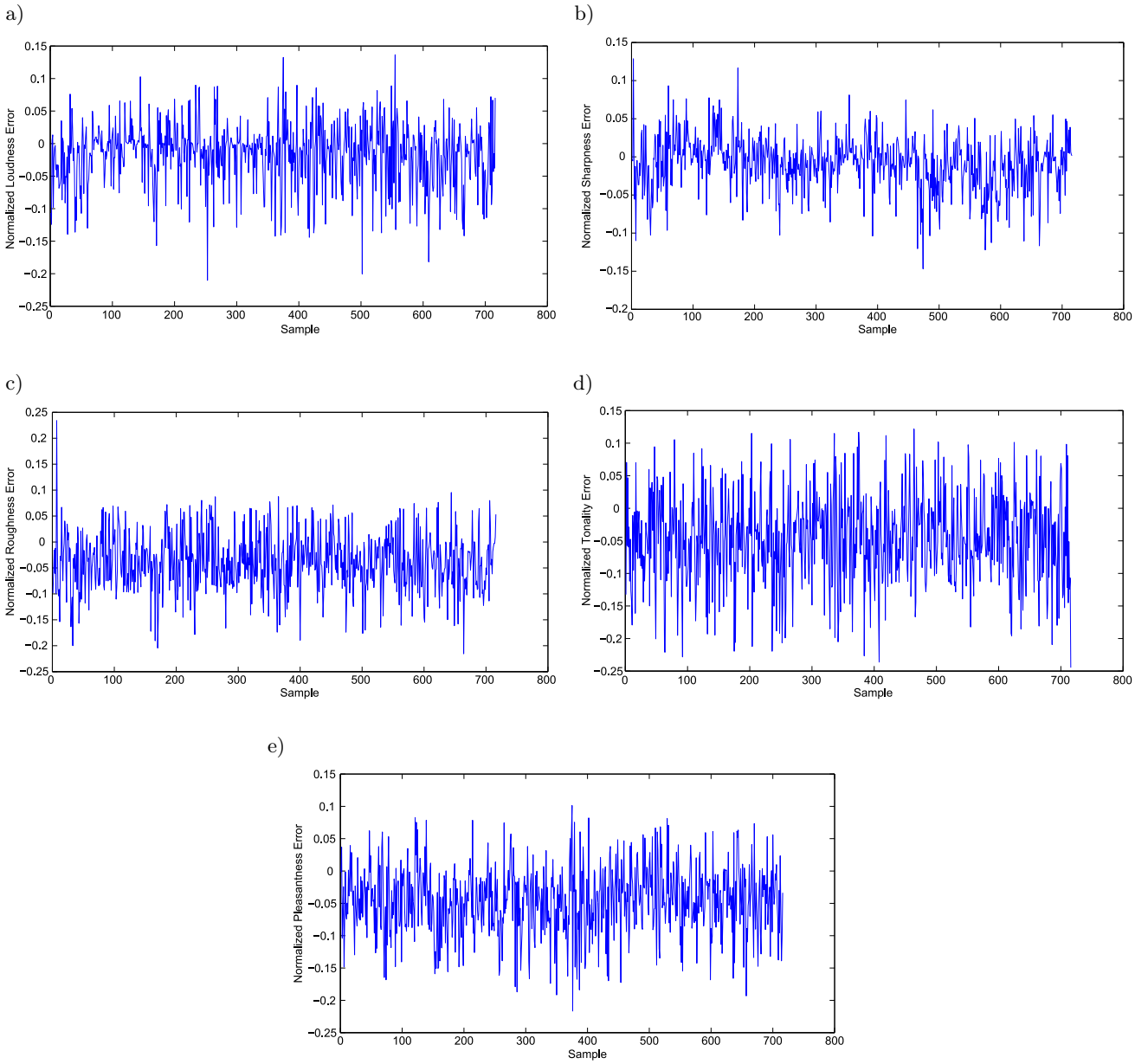


Fig. 5. Normalised error between the calculations made with the use of psychoacoustic models (Fastl, Zwicker, 2007) and the OAWT-BPNN model. Individual panels show the data for: a) normalised loudness, b) normalised sharpness, c) normalised roughness, d) normalised tonality, e) normalised pleasantness.

dividual panels in Fig. 6 are greater than 0.9, which indicates that the OAWT-BPNN model predicts the psychoacoustic characteristics of sound with a high accuracy.

Figure 7 shows the RMS prediction error for the present OAWT-BPNN model and the WPT-BPNN model described in a previous paper (Pourseiedrezaei *et al.*, 2019). The data indicate that the present model is more accurate than the previous one.

To further verify the present model a feature matrix for a sound sample (crying) was fed into the trained OAWT-BPNN algorithm for predicting loudness, sharpness, roughness, tonality, and pleasantness. Figure 8 shows the difference (error) between the outputs of the OAWT-BPNN model and the WPT-BPNN model calculated relative to the predictions made with the use of Zwicker's model (Zwicker, Fastl, 2007). The data plotted in individual panels in Fig. 8 indicate that the OAWT-BPNN model yields lower error than the previous WPT-BPNN model.

Table 3 shows a comparison of computation load in the calculation of the psychoacoustic indices with the use of psychoacoustical models (Fastl, Zwicker, 2007) implemented in commercial software (LabVIEW) (NI-Tutorial-1526, 2013; Technical note, 2015), and with the use of the WPT-BPNN and the OAWT-

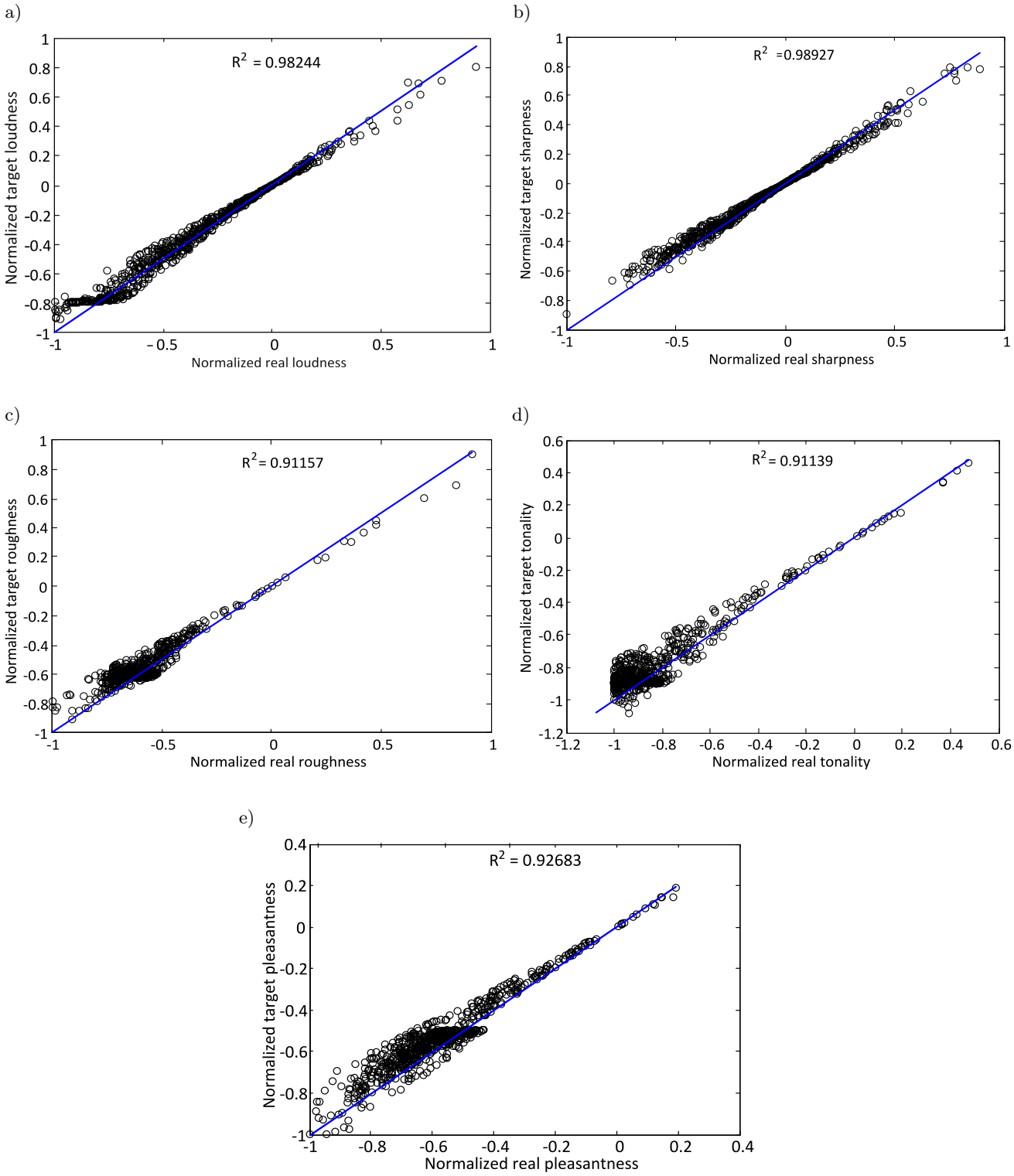


Fig. 6. Regression plots of the outputs of the proposed model with the corresponding psychoacoustic models outputs for normalised parameters: a) loudness, b) sharpness, c) roughness, d) tonality, e) pleasantness.

BPNN models. The data show the time used to calculate the loudness, sharpness, roughness, tonality, and pleasantness of a randomly chosen sample with the use of a computer with an Intel Core 2 Duo 2.66GHz processor and RAM 4 GB RAM. The data indicate that the OAWT-BPNN requires much less computational load than the other algorithms compared in Table 3.

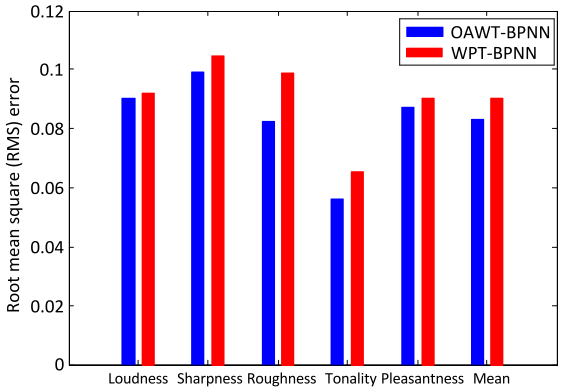


Fig. 7. RMS error of the OAWT-BPNN model and WPT-BPNN model.

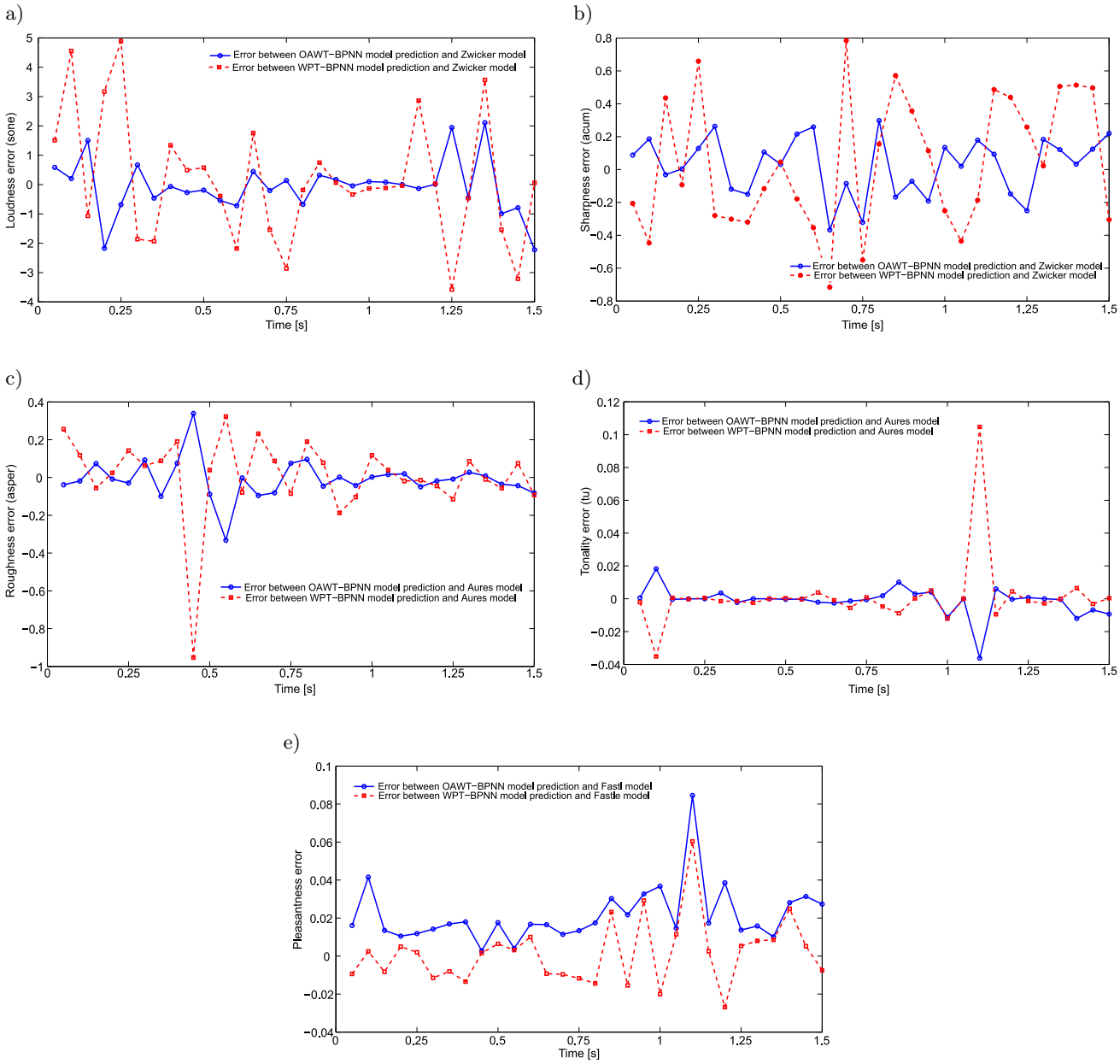


Fig. 8. Difference between the outputs of the WPT-BPNN the OAWT-BPNN models calculated relative to the predictions made with the use of Zwicker's psychoacoustic model (Fastl, Zwicker, 2007). The individual panels show the data for loudness, sharpness, roughness, tonality, and pleasantness.

Table 3. Comparison of computational time used for the calculation of psychoacoustic metrics.

| | MATLAB Code | LabVIEW Sound and Vibration Toolkit | WPT-BPNN | OAWT-ANN |
|---|---|---|---|---|
| Computational time [s] | 1.6659 | 1.5609 | 0.7141 | 0.0895 |

## 5. Conclusions

In this paper a modified intelligent model, developed by combining the OAWT and PSO-BPNN models, was proposed for the determination of the sound quality metrics for using in a real time SQE system. The results show that the OAWT-BPNN model can accurately predict the loudness, sharpness, roughness, tonality, and pleasantness indices. An advantage of the new model is in its higher accuracy of prediction of psychoacoustic metrics and much lower computational load, in comparison to previous models.

The model was verified in the present study with the use of samples of non-verbal vocalisations. The results show that the model is suitable for sound quality evaluation in real time active sound quality control systems, such as those used in neonatal intensive care units.

## References


1. Aures W. (1985), Method for calculating auditory roughness [in German: Ein Berechnungsverfahren der Rauhigkeit], *Acta Acustica united with Acustica*, **58**(5): 268–281.
2. Aures W. (1985), Calculation method for the sensory euphony of any sound signals [in German: Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale], *Acta Acustica united with Acustica*, **59**(2): 130–141.
3. Blauert J., Jekosch U. (1998), Product-sound quality: A new aspect of machinery noise, *Archives of Acoustics*, **23**(1): 105–124.
4. Chen K., Paurobally R., Pan J., Qiu X. (2015), Improving active control of fan noise with automatic spectral reshaping for reference signal, *Applied Acoustics*, **87**: 142–152, doi: 10.1016/j.apacoust.2014.07.003.
5. Cuddy L.L., Russo F.A., Galembo A. (2007), Tonality of Low-Frequency Synthesized Piano Tones, *Archives of Acoustics*, **32**(4): 541–550.
6. Dunn M.S., Erickson D., Avenue H., Gregory S. (2013), Recommended standards for newborn ICU design, eighth edition, *Journal of Perinatology*, **33**(Suppl. 1): S2–S16, doi: 10.1038/jp.2013.10.
7. Fastl H., Zwicker E. (2007), *Psychoacoustics: facts and models*, Springer, Berlin, Germany, 3rd ed., doi: 10.1007/978-3-540-68888-4.
8. Fausett L. (1994), *Fundamentals of Neural Networks*, Englewood Cliffs, NJ: Prentice-Hall.
9. Hafke-Dys H., Preis A., Kaczmarek T., Biniakowski A., Kleka P. (2016), Noise annoyance caused by amplitude modulated sounds resembling the main characteristics of temporal wind turbine noise, *Archives of Acoustics*, **41**(2): 221–232, doi: 10.1515/aoa-2016-0022.
10. Hecht-Nielsen R. (1992), Theory of the Backpropagation Neural Network, [in:] *Neural networks for perception*, *Vol. 2: Computation, Learning, Architectures*, H. Wechsler [Ed.] Harcourt Brace & Co., Orlando, FL, USA, doi: 10.5555/140639.140643.
11. Huang H.B., Li R.X., Huang X.R., Yang M.L., Ding W.P. (2015), Sound quality evaluation of vehicle suspension shock absorber rattling noise based on the Wigner-Ville distribution, *Applied Acoustics*, **100**: 18–25, doi: 10.1016/j.apacoust.2015.06.018.
12. Huang H.B., Li R.X., Yang M.L., Lim T.C., Ding W.P. (2017), Evaluation of vehicle interior sound quality using a continuous restricted Boltzmann machine-based DBN, *Mechanical Systems and Signal Processing*, **84**(Part A): 245–267, doi: 10.1016/j.ymssp.2016.07.014.
13. Jaddi N.S., Abdullah S. (2018), Optimization of neural network using kidney-inspired algorithm with control of filtration rate and chaotic map for real-world rainfall forecasting, *Engineering Applications of Artificial Intelligence*, **67**: 246–259, doi: 10.1016/j.engappai.2017.09.012.
14. Kaczmarek T., Preis A. (2010), Annoyance of time-varying road-traffic noise, *Archives of Acoustics*, **35**(3): 383–393, doi: 10.2478/v10168-010-0032-2.
15. Klonari D., Pastiadis K., Papadelis G., Papanikolao G. (2011), Loudness assessment of musical tones equalized in a-weighted level, *Archives of Acoustics*, **36**(2): 239–250, doi: 10.2478/v10168-011-0019-7.
16. Kuo S.M., Morgan D. (1996), *Active Noise Control Systems: Algorithms and DSP Implementations*, John Wiley & Sons, Inc., New York, NY, USA.
17. Lee H.H., Lee S.K. (2009), Objective evaluation of interior noise booming in a passenger car based on sound metrics and artificial neural networks, *Applied Ergonomics*, **40**(5): 860–869, doi: 10.1016/j.apergo.2008.11.006.
18. Leite R.P., Paul S., Gerges S.N.Y. (2008), A sound quality-based investigation of the HVAC system noise of an automobile model, *Applied Acoustics*, **70**: 1–10, doi: 10.1016/j.apacoust.2008.06.010.
19. Lyon R. (2000), *Designing for product sound quality, Mechanical Engineering*, Marcel Dekker, Inc.: New York, Basel.
20. Maleczek S. (2008), Testing the sound quality of acoustic vacuum tube amplifier, *Archives of Acoustics*, **33**(4 suppl.): 135–140.
21. Miśkiewicz A., Rogala T., Szczepańska-Antosik J. (2007), Perceived roughness of two simultaneous

harmonic complex tones, *Archives of Acoustics*, **32**(3): 737–748.

22. NI-Tutorial-1526 (2013), *White paper on Measurement of Sound Quality*.
23. Olbrych S. (2010), *Noise pollution in the NICU*, Case West. Reserv. Univ.
24. Parsons C.E., Young K.S., Craske M.G., Stein A.L., Kringelbach M.L. (2014), Introducing the oxford vocal (OxVoc) sounds database: A validated set of non-acted affective sounds from human infants, adults, and domestic animals, *Frontiers in Psychology*, **5**: 562, doi: 10.3389/fpsyg.2014.00562.
25. Pleban D. (2014), Definition and measure of the sound quality of the machine, *Archives of Acoustics*, **39**(1): 17–23, doi:10.2478/aoa-2014-0003.
26. Pourseiedrezaei M., Loghmani A., Keshmiri M. (2019), Prediction of psychoacoustic metrics using combination of wavelet packet transform and an optimized artificial neural network, *Archives of Acoustics*, **44**(3): 561–573, doi: 10.24425/aoa.2019.129271**.**
27. Szczepańska-Antosik J. (2008), Roughness of two simultaneous harmonic complex tones in various pitch registers, *Archives of Acoustics*, **33**(1): 73–78.
28. Technical note (2015), *An introduction to sound quality testing*, Manchester, UK: Acoustic Research Centre, School of Computing, Science and Engineering, University of Salford.
29. Terhardt E., Stoll G., Seewann M. (1982), Algorithm for extraction of pitch and pitch salience from complex tonal signals, *The Journal of the Acoustical Society of America*, **71**(3): 679–688, doi: 10.1121/1.387544.
30. Vencovský V. (2016), Roughness prediction based on a model of cochlear hydrodynamics, *Archives of Acoustics*, **41**(2): 189–201, doi: 10.1515/aoa-2016-0019.
31. Wang Y.S., Lee C.M., Kim D.G., Xu Y. (2007), Sound-quality prediction for nonstationary vehicle interior noise based on wavelet pre-processing neural network model, *Journal of Sound and Vibration*, **299**: 933–947, doi: 10.1016/j.jsv.2006.07.034.
32. Wang Y.S., Shen G.Q., Xing Y.F. (2014), A sound quality model for objective synthesis evaluation of vehicle interior noise based on artificial neural network, *Mechanical Systems and Signal Processing*, **45**: 255–266, doi: 10.1016/j.ymssp.2013.11.001
33. Xing Y.F.F., Wang Y.S.S., Shi L., Guo H., Chen H. (2016), Sound quality recognition using optimal wavelet-packet transform and artificial neural network methods, *Mechanical Systems and Signal Processing*, **66–67**: 875–892, doi: 10.1016/j.ymssp.2015.05.003
34. Zhang J.R., Zhang J., Lok T.M., Lyu M.R. (2007), A hybrid particle swarm optimization-back-propagation algorithm for feedforward neural network training, *Applied Mathematics and Computation*, **185**(2): 1026–1037, doi: 10.1016/j.amc.2006.07.025.
35. Zhu X., Kim J. (2006), Application of analytic wavelet transform to analysis of highly impulsive noises, *Journal of Sound and Vibration*, **294**(4): 841–855, doi: 10.1016/j.jsv.2005.12.034.