Psychoacoustic Impacts Estimation in Manufacturing based on Accelerometer Measurement using Artificial Neural Networks

Minjie Zou[1], Laura Folk[2] and Julien Provost[1]

***Abstract*— In recent years, psychoacoustic impacts of a product has won increasing importance in various design and manufacturing sectors. However, conventional measurement setups based on microphones are expensive and noise-sensitive. This paper proposes a novel method to estimate psychoacoustic parameters from accelerometer measurement by using artificial neural networks. The proposed method has been successfully applied on automotive vehicle interior components which produce nonstationary sounds when operated. In order to develop and tune the proposed method, the operation sounds are first measured by a microphone and an accelerometer simultaneously. Then, static and dynamic psychoacoustic parameters are calculated from the microphone signals according to the auditory model. Finally, the relationship between the psychoacoustic parameters and the accelerometer signals is approximated by feedforward multilayer neural networks. As a result, the performance of the proposed method using artificial neural networks is successfully validated on the existing database.**

I. Introduction

With the development of electric cars, greater importance has been attached to subjective perception of the vehicle interior sound in the automotive industry. Especially at the first contact with the vehicle in a dealership, where the engine is stopped, the sound of operation components is a deciding factor for customers. Previous studies have shown that subjective perception of nonstationary sound is highly correlated with the psychoacoustic parameters, loudness and sharpness [1]–[4]. The microphone and the artificial head are conventional measuring techniques to quantify the perceived sound by a human, but because of their cost and the complexity to build such a test bench, they are not ideal in the manufacturing industry. Furthermore, the result obtained through conventional measuring techniques is highly disturbed by environmental noises. By contrast, the use of accelerometers shows its superiority in price and noise resistance. This is the reason why this paper investigates the feasibility of psychoacoustic analysis using accelerometer-based measurements.

An accelerometer measures structural vibrations by measuring the structural vibration acceleration. Whereas conventional techniques record the sound waves in the air by measuring the sound pressure over time. Most psychoacoustic research has been based on airborne measurement [1]–[4]. Research on the measurement of psychoacoustic parameters with the accelerometer has been less extensive. To the best of our knowledge, only one case can be reported: Moritz *et al.* [5] have worked on the correlation between airborne and structure-borne sound of the motor noise, which is one specific kind of stationary sound. They have found that loudness calculated from both sensors is linearly correlated, but sharpness is not. Their method of linear regression showed its shortage in handling the correlation of sharpness.

From another perspective, Wang et al. [6] illustrated that Artificial Neural Networks (ANNs) are powerful enough to approximate the existing auditory model for the loudness in the air. ANNs have already been widely applied in vehicle acoustics for evaluation [7], [8], classification [9] and recognition [10]. Research in [7] and [8] focused on stationary sound, and results in [9] and [10] were limited to predefined categories or patterns. None of them covers structure-borne measurement. In our research, the measured sound is nonstationary, and the target values i.e. the psychoacoustic parameters have a large range of value and are not categorized.

The goal of this paper is to fill the above-mentioned lack and aims to achieve psychoacoustic parameters of the nonstationary sound based on the accelerometer measurement using ANNs. Correlation between the microphone measurement and accelerometer measurement with reference to loudness and sharpness will be discussed. Compared to existing literature, the method proposed in this paper permits to:

- Provide a noise-resistant measurement of sound perception using structure-borne accelerometers,
- Measure and evaluate the psychoacoustic parameters earlier in manufacturing,
- Show the potential to replace expensive sensors with economical ones and ANN-processing.

The whole process is shown in Fig. 1. The two branches in the flowchart specify the different treatments, respectively for microphone-based and for accelerometer-based measurements. The branch above (Path 1) shows the conventional way to achieve psychoacoustic parameters, which are calculated from microphone signals according to the auditory model [11]. The branch below (Path 2) describes the proposed new way to achieve sound quality estimation from accelerometer-based measurements with the help of ANN-processing. In order to define the behavior of the ANN, the ANN is first trained using both measurements. This training process permits the ANN to learn the relationship between the accelerometer data and the results of the conventional

This work was partially supported by BMW Group.

[1]Minjie Zou and Julien Provost are with Faculty of Mechanical Engineering, Technische Universität München, Garching bei München, Germany, Email: minjie.zou@tum.de; julien.provost@tum.de

[2]Laura Folk is with BMW Group, München, Germany, Email: Laura.Folk@bmw.de

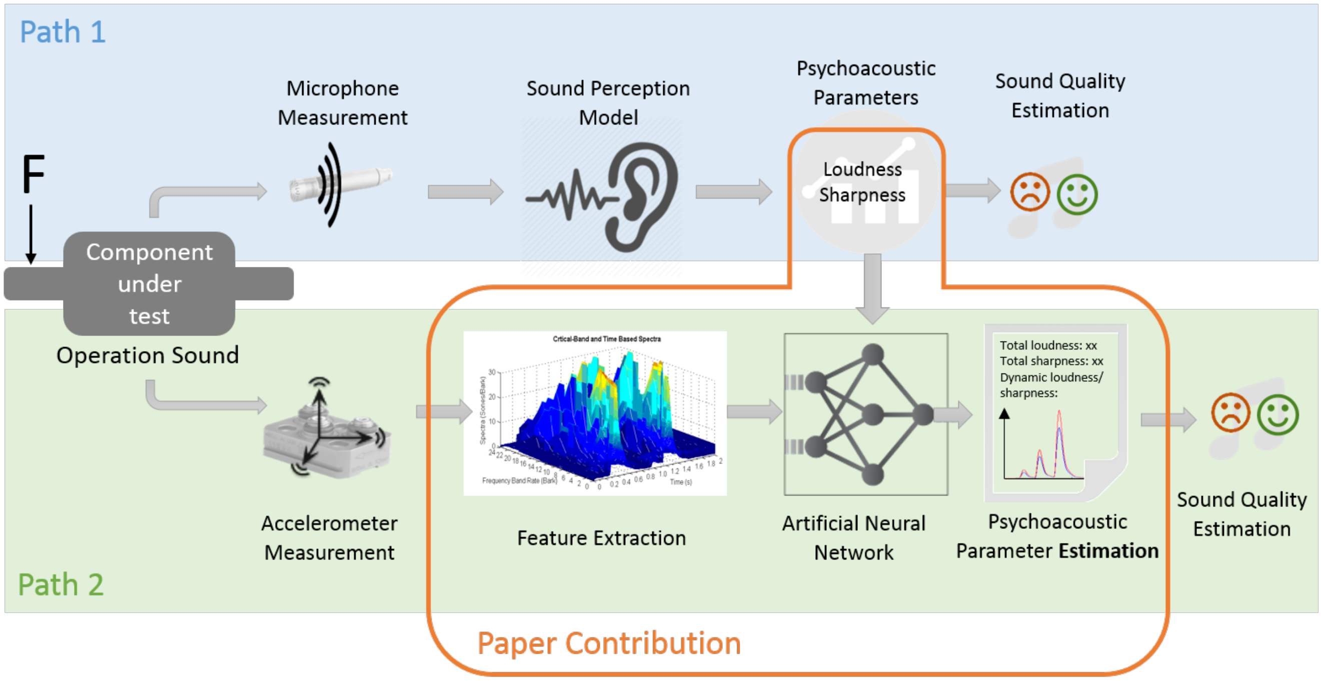

Fig. 1. Flowchart of the proposed method

calculation. Once the ANN has been trained, the branch above can be eliminated, and the sound quality estimation of a new component can be performed according to Path 2. That means, solely accelerometer-based measurement would be needed in further usage.

The remainder of the paper is structured as follows. Section 2 gives out the theoretical background regarding acoustics and ANN techniques. Section 3 describes the methodology of the proposed approach from the experiment, data processing to ANN modeling. Section 4 presents the results of applying the ANN on the measured data. The performance of the ANNs is validated, and the method proposed in this paper is also evaluated from the manufacturing aspects. A conclusion is made in Section 5.

II. Background

A. Psychoacoustics

The study of the human perception of sound is called psychoacoustics. The sensitivity of the human ear changes as a function of frequency. Human hearing combines sound with near frequencies into 24 frequency bands [11]. It is complex to evaluate the perception, because it shows an individual difference. A lot of efforts have been made to use the psychoacoustic parameters to evaluate sound quality subjectively, e.g. loudness, sharpness, rhythm, roughness etc. [11]. It has been found that the sensation of nonstationary sound is highly correlated with the psychoacoustic parameters, loudness and sharpness [1]–[4]. Loudness describes the human perception characteristics regarding the sound magnitude. In this paper it is measured and calculated according to the Zwickers auditory model in ISO 532B in unit of *sone*. Sharpness is a sensation related to the sound spectra density and it influences people's sensory pleasantness. It is highly dependent on the high frequency proportion in the sound. There is no ISO standard to calculate the sharpness. In this study, it is also calculated according to Zwicker [11], which is one of the popular methods. Unit of sharpness is *acum*.

B. Artificial Neural Networks

Artificial neural networks (ANNs) are powerful tools to map the inputs into desired outputs when the input/output relation are complex and nonlinear. An ANN is a data processing system that develops as a generalization of the mathematical model of human cognition [12]. It consists of numerous simple processing units which are interconnected with each other in a structured architecture. Fig. 2 shows a normal feedforward multilayer network, which consists of an input layer, an output layer and one or more hidden layers.

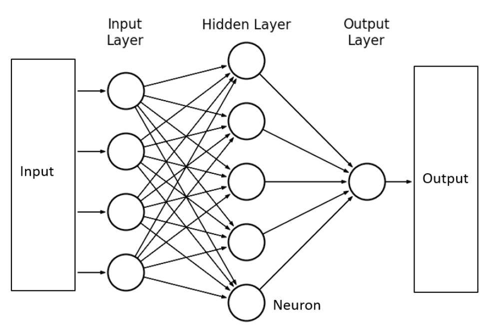

Fig. 2. Architecture of a Feedforward Multilayer Network

A single neuron is shown in Fig. 3, in which the inputs x_i from the previous layer are weighted by w_i, led into a sum block Σ, added with the bias b, processed with an activation function F and finally output in y. Formally, a neuron can be defined as follows:

$$y = F(z) = F(\Sigma_{i=1}^{n} x_i w_i + b), \quad (1)$$

where n is the number of the inputs, and z is the sum of the weighted inputs.

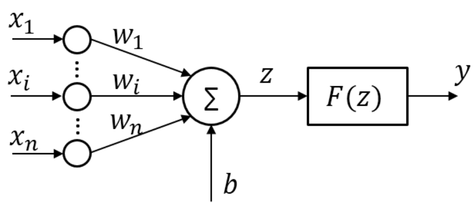

Fig. 3. A Neuron in the ANN

Before the neural network comes into usage it should firstly be trained and validated by a series of training data. In the training phase, network parameters adapt themselves in response to an amount of learning examples from the data pool. The adaptation continues iteratively until the error between the output of the ANN and the desired output lies within tolerance. In this way, the network configures itself and learns the desired input-output relationship.

III. Methodology

This section describes the procedures to obtain the proposed approach. The proposed approach is composed of three main steps: sound measurement, data preparation for ANNs, as well as ANN modeling. Measurements are carried out with a real-time measuring system PAK[1] and all the processing is done with MATLAB. Microphone signals are used to calculate targets of the ANN (see subsection III-B.1), which are specifically the loudness and sharpness. Measured data from the accelerometer are processed to extract features (see subsection III-B.2), which are fed to the ANN as inputs. Based on that, the suitable ANN structure is determined.

A. Measurement protocol

A data pool is required to train the ANN in order to achieve the maximum accuracy. The operation sounds of the same component[2] from different vehicles are recorded in a semi anechoic chamber (see Fig. 4). Both a condenser microphone and a triaxial piezoelectric accelerometer are used in the measurement. The microphone is placed opposite to the component with a constant distance, while the accelerometer is directly fixed to the component surface with wax according to ISO 5348. All the components are operated manually in different parts, which simulates the realistic operating conditions.

Original records of both sensors in the experiments are shown in Fig. 5. Every record lasts 15 seconds with a

[1]http://www.pakbymbbm.com/

[2]For confidential reasons, little specific information can be provided about the component considered in this study. This component is an operating element frequently used by the driver and also manipulated when in the showroom, with the engine stopped

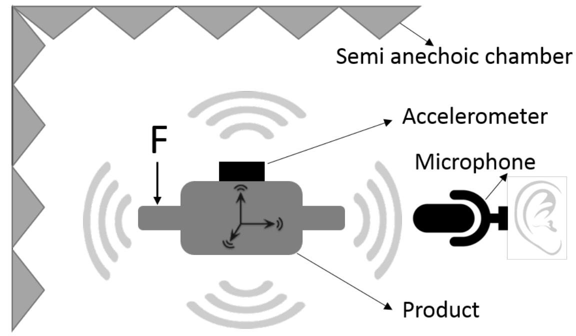

Fig. 4. Measurement Arrangement

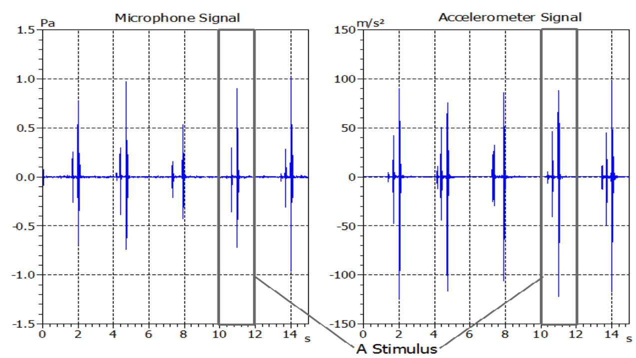

Fig. 5. Measured Data and Stimuli in the Experiment. Here, each experiment consists of 5 stimuli.

sampling rate of 48 kHz. In the 15 seconds, complete operation actions which includes one operation action and one release action are repeated. These actions are called stimuli. Every record contains 5-6 stimuli. In this experiment, every complete stimulus is cut out with a fixed duration of 2 seconds. In total, 11 physical components have been measured. Once the falsely recorded stimuli have been removed, 221 stimuli from the microphone and 221 stimuli from the accelerometer make up the data pool in our research.

B. Data Preparation for the ANN

In order to train ANNs properly and correctly, original measured data are pre-processed to determine the desired inputs and targets. As for the nonstationary sound, dynamic characteristics are as important as the overall evaluation. For this reason, our targets contain two parts: static parameters and parameters which traces the dynamic status. Both parts are selected with reference to loudness and sharpness. Two ANNs are used separately for the different targets, because they are in different complexity degrees. Furthermore, having two separated ANNs makes it more flexible for future users to apply. In order to train the ANNs, the data pool is divided into two parts. Calculated parameters from the microphone, loudness and sharpness, provides the ANN with training targets, and extracted features from the accelerometer are fed into the ANNs as inputs.

1) Psychoacoustic Parameter Calculation: In this step, targets values are prepared for the ANN. Psychoacoustic parameters, loudness and sharpness, are calculated from the microphone signals, which are distinguished into static and dynamic parameters. Static parameters show the statistic evaluation of a measurement, which is represented by two static values i.e. total loudness and total sharpness. They evaluate a stimuli as a whole. Dynamic parameters, which are also called time-varying parameters, are referred to as the time series of instantaneous loudness and instantaneous sharpness specifically in this paper. Considering that the auditory temporal effects have influence when the sound duration is shorter than 200 ms [11], time-varying parameters are acquired by dividing the stimulus into 20 ms frames and calculating parameter values in each frame.

2) Feature Extraction: Feature extraction offers the inputs of the ANN from the accelerometer. The original data are numerous vibration acceleration values measured at a high sampling rate, which are too much information for the neural network. Thus, the input space is reduced by extracting features from the accelerometer signals to allow a better efficiency. Extracted features are expected to be representative for its psychoacoustic impact. Because the accelerometer has a different frequency response from that of the microphone, we analyze the signals in the frequency domain. Corresponding to the two sets of target parameters, input information also falls into the static and the dynamic features. As for the static features, each stimulus is considered as a whole and no time-related factors are in included in the feature. Original signals are filtered with the Butterworth filter into 24 frequency bands from 0 to 15.5 kHz [11]. For each band we get 10 sampling points. Then, the 240 points in spectrum are selected to represent the static feature of the signal. Dynamic features are extracted in a similar way. Instead of processing the whole stimuli, the treatment is carried out using time windowing. A 240-point spectral and temporal matrix is calculated in every 20 ms.

As a result, we obtain two sets of feature data for each stimulus corresponding to the two types of targets. From the static view, every stimulus is featured with a frequency-based vector (240x1). From the dynamic view, it is featured with a time-and-frequency-based 240x100-matrix (see Fig. 6).

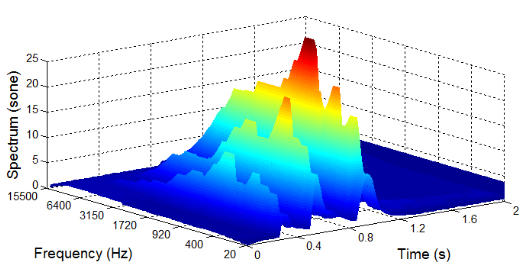

Fig. 6. The Temporal and Spectral Feature of an Accelerometer Stimulus

C. Artificial Neural Network Modeling

In this paper, feedforward multilayer ANNs are used to estimate psychoacoustic parameters from accelerometer signals. Two ANN-models are developed separately for static and dynamic target parameters. ANNs in this research can be regarded as nonlinear function approximator. Previous work showed that the Levenberg-Marquardt algorithm works well with nonlinear function approximation [13]. It offers accurate results and works efficiently for the ANN with a moderate size. This is an iterative process that searches for a local minimum of the multivariate cost function. More information can be found in [14]. Thus, in this study both networks are trained with the Levenberg-Marquardt algorithm.

In both ANNs, a three-layer structure is chosen. For static parameter estimation, ten neurons are used in the hidden layer. In the ANN for dynamic parameters, thirty neurons are used in the hidden layer. The larger number of neurons is required by the higher nonlinearity between inputs and targets. However, when the number is further increased in our experiments, it does not show much improvement in performance. The ANNs use two outputs: the total loudness and total sharpness for the static parameter estimation, and dynamic loudness and sharpness for the dynamic parameter estimation.

In the experiment, available data are divided into three sets: 70% for training, 15% for validation and 15% for test. Training, validation and test data are selected randomly from the whole measured data pool. Some component data may be used only for training, validation or testing, and some can be partly used for several of them.

The ANN configures its network parameters mainly according to the input-output relationship of the training data. Validation data compares the ANN outputs and the targets, and checks the stopping conditions of training. They validate the performance of the ANN in the training process. Test data are not involved in the training process. They are used to evaluate the posteriori performance of the ANN. In other words, the test set checks whether the Path 2 in Fig. 1 could work alone on future data from accelerometer-based measurements only. The division ratio is referred to the previous experience in [13]. A good ANN is expected to provide accurate results in both training and testing processes. A brief overview of the settings is shown in Table I.

TABLE I
CONFIGURATION OF THE ANNS

	ANN For Static Parameters	ANN For Dynamic Parameters
Structure	10 hidden neurons	30 hidden neurons
Training Algorithm	Levenberg-Marquardt	
Data Division	70% Training,15% Validation,15% Test	
Activation Function	Hidden layer: tan-sigmoid , Output layer: linear	
Implementation Platform	Software: Matlab R2013b Hardware: Laptop Intel Core i7-4720HQ CPU @ 2.2GHz, RAM 16GB	

IV. APPLICATION AND RESULTS

A. Results of the ANNs

1) ANN for Static Evaluation: The ANN for static evaluation proves to offer satisfactory results very efficiently. The static parameters, total loudness and total sharpness, are successfully estimated from the selected feature information of accelerometer signals (see Fig. 7 and Fig. 8).

In Fig. 7 and Fig. 8, each circle represents a stimuli, plotted by the target values (total loudness and total sharpness from microphone signals) on the horizontal axis and the ANN outputs on the vertical. The fit line shows the linear regression between the targets and generated outputs. According to the results, both parameters can be acquired at the same time without extra treatment.

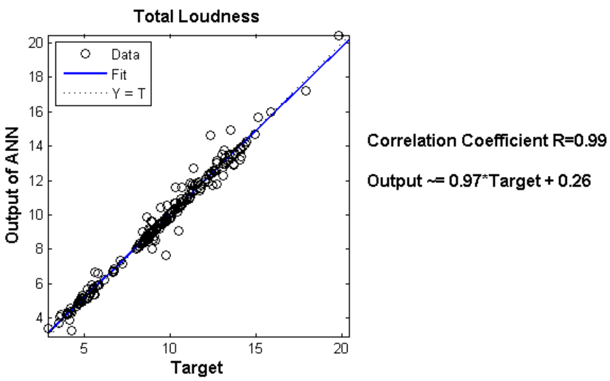

Fig. 7. Results of Total Loudness

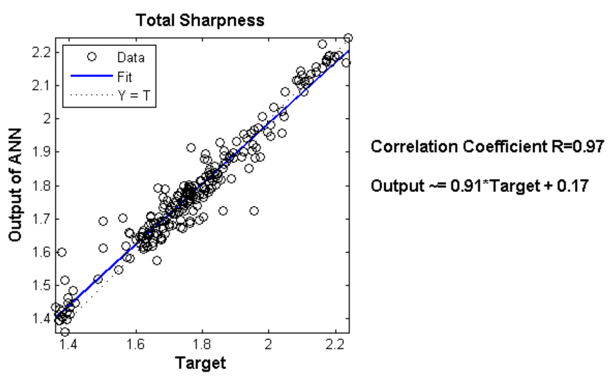

Fig. 8. Results of Total Sharpness

Fig. 9 shows the results of the training and testing process. There is a high correlation R between ANN outputs and targets in the training process, which reaches more than 0.999. It demonstrates that the network is good at finding the input-output relationship within the training data. In the test phase, the correlation reaches more than 0.993. The good results in the test phase proves the generalization capability of the network, which implies that the ANN is robust for untrained data.

2) ANN for Dynamic Evaluation: Correlation of the outputs with desired parameters are good too, with 0.953 and 0.928 respectively for the training and testing process. There are more training data for this ANN since the input features and target values vary with time. Taking account of time-related factors also leads to a higher complexity of the input-output relationship. However, the good performance of the ANN-processing is ensured by increasing the number of neurons in the ANN. Fig. 10 shows the results of the training and testing processes. Each circle represents a 20 ms slice of the stimulus, plotted by the target values on the horizontal axis and the ANN outputs on the vertical.

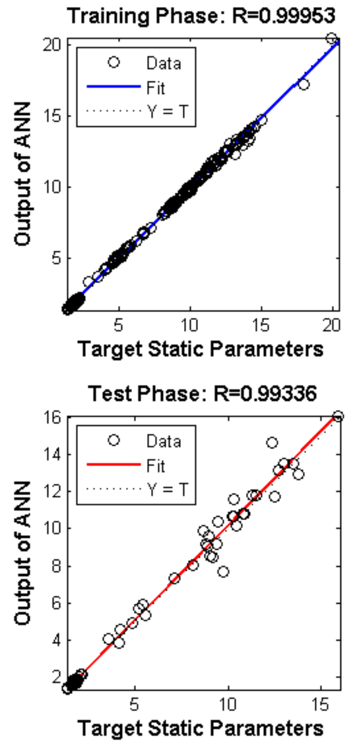

Fig. 9. Training (above) and Test (below) Results for Static Parameters

B. Evaluation of the Method

The estimation of static parameters is accurate. The root-mean-square error (RMSE) of the total loudness estimation is 0.616 sone, and RMSE for total sharpness is 0.082 acum. The maximum error of the total loudness is 2.8 sones, and it is 0.2 acum for total sharpness. Therefore, compared to the target values which mainly varies from 8 to 14 sones for the loudness and from 1.6 to 2 acums for the sharpness, the method to estimate static parameters has been judged robust enough in the given context.

In comparison, even though the ANN correlation factors were good, estimation errors of dynamic parameters are much larger than those of static parameters. RMSE of the instantaneous loudness is 1.272 sones, and RMSE of the instantaneous sharpness is 0.693 acum. The estimation error peaks at 25.4 sones and 2.4 acums respectively for instantaneous loudness and instantaneous sharpness. Though the correlation is high, such large errors are not acceptable in our case. Therefore, the method to estimate dynamic parameters is still not robust enough and needs further investigation.

Because the training process is an independent process, which does not take place during the application phase, the training time is not significant to the manufacturing usage. Instead, the recall time in the application process is worth analyzing. In our experiment, once the network is trained, it can be applied i.e. recalled almost in real time. As for a single 2-sec stimulus which is sampled at 48000 Hz, the ANN

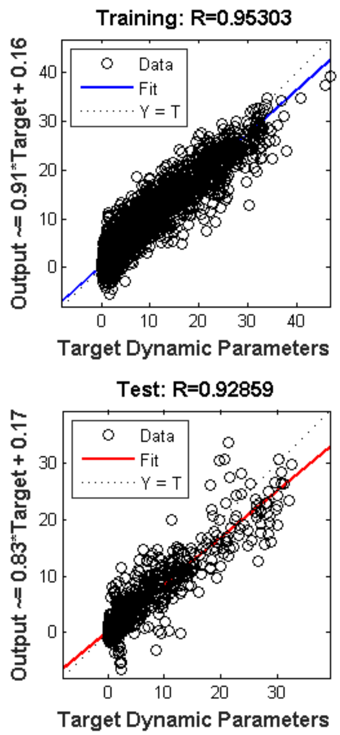

Fig. 10. Training and Test Results for Dynamic Evaluation

takes 0.01 sec in average to generate the static parameters, and it takes 1 sec to generate 100 instantaneous dynamic parameters. Table II shows the average processing time for one stimulus in Matlab, including the time to extract feature and to recall an ANN. On the whole, the processing time of our proposed method is shorter than 2 sec. Thus, one way to improve the robustness of the ANN analysis for dynamic psychoacoustic parameters would be to define it with a higher complexity and to use a larger training set.

TABLE II
PROCESSING TIME OF THE METHOD WITH A SINGLE STIMULUS

	For Static Evaluation	For Dynamic Evaluation
Feature Extraction	0.13 s	0.17 s
ANN Application	0.01 s	1.00 s
Total	0.14 s	1.17 s

V. CONCLUSION

This paper has presented a method to evaluate the correlation between accelerometer-based measurement and conventional microphone-based measurement with special reference to the psychoacoustic parameters. Both static parameters (total loudness, total sharpness) and dynamic ones (instantaneous loudness, instantaneous sharpness) are taken into account with two independent ANNs. The operation sound of the vehicle interior components have been measured and 221 pairs of stimuli were used to train and test the ANNs. Experiment results show that the ANNs are able to estimate psychoacoustic parameters from the accelerometer signals.

The estimation of static psychoacoustic parameters are accurate and efficient. ANN outputs are highly correlated with the target values calculated from the microphone signals. Thus, the proposed using accelerometer can be used as a cost-effective, reproducible and noise-resistant alternative to the conventional microphone-based approach.

However, the estimation of dynamic parameters are not robust enough yet: even though correlation factors and RMSE errors are good, peak errors are still too high.

Future work considers extraction of alternative features from accelerometer signals, or training of the ANNs directly from raw accelerometer signals.

REFERENCES

[1] M. C. Bezat, V. Roussarie, R. Kronland-Martinet, and S. Ystad, "Relations between acoustic parameters and perceptual properties: an approach by regressions tree applied to car door closure sounds," *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 3253–3253, 2008.

[2] A. M. Willemsen and M. D. Rao, "Characterization of sound quality of impulsive sounds using loudness based metric," in *Proceedings of 20th International Congress on Acoustics, Sydney*, vol. 5, 2010, pp. 3397–3404.

[3] M. Höchstetter, P. Seifert, O. Jokisch, J. L. Verhey, and U. Gabbert, "Influence of synthetically varied signal parameters of impulsive vehicle sounds on perceived quality," in *26. Konferenz Elektronische Sprachsignalverarbeitung, 25.-27.03.2015, Eichstätt*, G. Wirsching, Ed., 2015, pp. 264 – 271.

[4] M. Höchstetter, D.-I. F. M. Wackerbauer, J. Verhey, *et al.*, "Psychoacoustic prediction of singular impulsive sounds," *ATZ worldwide*, vol. 117, no. 7-8, pp. 58–63, 2015.

[5] K. Moritz, S. Ochs, J. Schlittenlacher, J. Bös, and T. Melz, "Wahrnehmung von getriebegeräuschen - übertragung psychoakustischer analysen auf körperschallsignale (german)," in *DAGA 2015; 41. Jahrestagung für Akustik*, ser. Paper ID 000057, March 2015, pp. 113–116. [Online]. Available: http://tubiblio.ulb.tu-darmstadt.de/73503/

[6] Y. Wang, C.-M. Lee, H. He, and Y. Tian, "A new intelligent technique for sound quality evaluation of nonstationary vehicle noises," in *Strategic Technology, The 1st International Forum on*. IEEE, 2006, pp. 33–36.

[7] S.-K. Lee, B.-S. Kim, and D.-C. Park, "Objective evaluation of the rumbling sound in passenger cars based on an artificial neural network," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 219, no. 4, pp. 457–469, 2005.

[8] H.-H. Lee and S.-K. Lee, "Objective evaluation of interior noise booming in a passenger car based on sound metrics and artificial neural networks," *Applied ergonomics*, vol. 40, no. 5, pp. 860–869, 2009.

[9] M. Paulraj, A. M. Andrew, and S. Yaacob, "Classification of interior noise comfort level of proton model cars using artificial neural network," in *Advanced Computing and Communication Systems (ICACCS), 2013 International Conference on*. IEEE, 2013, pp. 1–7.

[10] Y. Xing, Y. Wang, L. Shi, H. Guo, and H. Chen, "Sound quality recognition using optimal wavelet-packet transform and artificial neural network methods," *Mechanical Systems and Signal Processing*, vol. 66, pp. 875–892, 2016.

[11] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*. Springer Science & Business Media, 2013, vol. 22.

[12] M. Sivanandam *et al.*, *Introduction to artificial neural networks*. vikas publishing House PVT LTD, 2009.

[13] H. B. Demuth, M. H. Beale, O. De Jess, and M. T. Hagan, "Neural network design," 2014.

[14] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.