Vehicle noise pattern recognition by Self-Organizing Maps

Stefan Twieg stefan.twieg@volkswagen.de
Volkswagen AG/Group Research
Wolfsburg, Germany

Holger Opfer opfer.holger@vgj.co.jp
Volkswagen Group Japan
VTT, Tokyo, Japan

Helmut Beikirch helmut.beikirch@uni-rostock.de
Faculty of Computer Science and
Electrical Engineering
University Rostock, Germany

Abstract

Interior vehicle acoustics are in close connection with our quality opinion. The noise in vehicle interior is complex and can be considered as a sum of various sound emission sources. A nice sounding vehicle is objective of the development departments for car acoustics. In the process of manufacturing the targets for a qualitatively high-valuable sound must be maintained. However, it is possible that production errors lead to a deviation from the wanted vehicle interior sound. This will result in customer complaints where for example a rattling or squeak refers to a worn-out or defective component. Also in this case, of course, the vehicle interior noise does not fulfill the targets of the process of development. For both cases there is currently no possibility for automated analysis of the vehicle interior noise. In this paper an approach for automated analysis of vehicle interior noise by means of neural algorithms is investigated. The presented system analyses microphone signals from car interior measured at real environmental conditions. This is in contrast to well known techniques, as e.g. acoustic engine test bench. Self-Organizing Maps combined with feature selection algorithms are used for acoustic pattern recognition. The presented system can be used in production process as well as a standalone ECU in car.

Keywords: signal classification, SOM, car noise, pattern recognition.

1. MOTIVATION

Car interior noise is a sum of different sound emission sources, e.g. engine, wind, chassis or suspension. Furthermore, it is a challenging task to examine objective assessment criteria for a sound. In production process acoustic and vibration technology is used e.g. for analysis at engine test bench. Engines and test benches are tested in [1] to recognize damages of failure parts. Suppliers are also checking their parts before delivery by similar methods [2], [3] to achieve a reliable and correct product. Commonly, the acoustic analysis is done in defined environmental conditions to get no interference with unknown noises. In this case quite simple mathematic is used to detect and to interpret changes. This method is in contrast to the method shown in this paper where an undefined environment influences the acoustic conditions. For assessment of car interior sound by algorithms the time signal which is recorded by microphone in car interior has to be fragmented in different mathematic parameters. The challenge is to find parameters which are corresponding to failures or changes of car part sounds and not to acoustic changes in environment. Thus the parameters have to be explicit relevant for sounds of interest. To analyze these relevant parameter knowledge storage is required to build up experience of parameter behavior. Therefore a neural net by Kohonen [4], the self-organizing map (SOM), models an input space where similar patterns can be categorized. For the described purpose the Kohonen-map is combined with parameter relevance detection algorithm. As a result the developed system assesses continuously the microphone signals, e.g. in two categories like “error-free” and “error” sound. In future the described method can be used for applications of sound assessment at production-line end (at test-bench or test-drive) or in car combined with hands-free kit.

2. MODEL-LAYOUT FOR PATTERN RECOGNITION

An adequate amount of data sets is the basis for acoustic pattern recognition. For this purpose different noticeable sounds are stored in database. They were recorded at different conditions like:

- Car type,
- Driving situation (longitudinal and transversal dynamic),
- Car parts (new vs. used and damaged),
- Environment (weather, pavement).

In the following analysis one car type was chosen. For this one 1430 measurements with a length of 15 to 30 seconds are stored in database. In Figure 1, you can see the model for datasets analysis and pattern recognition. As next step the one (e.g. rms, sound pressure level) or multidimensional (e.g. spectrum, DCT, MFCC) parameters are calculated. However, psycho-acoustic (e.g. loudness, roughness) parameters are calculated too. In the next step the focal point is

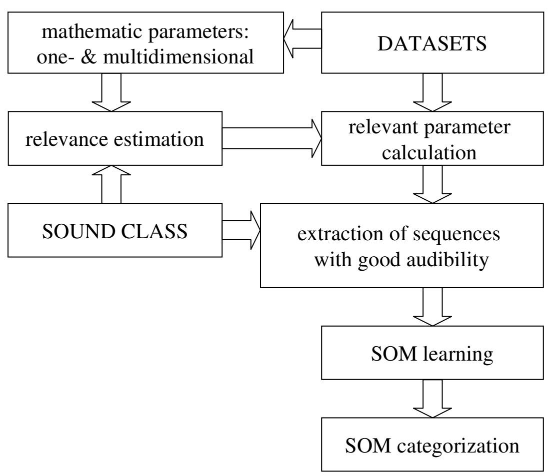

FIGURE 1: Model-design for classification of acoustic in car recordings.

relevance detection to identify parameters which are corresponding to noticeable sounds caused by wear (e.g. steering rod ball joint) or damage. The relevance detection algorithm has to identify the connection between:

- the sound classes (e.g. “error-free” and “error” sounds),
- sound classes and parameters.

Each parameter is rated and ordered by rank information. The parameters with best relevance are used as input features for SOM. In the next step the datasets with best audibility of defined sound classes are used for training process. There are two defined sound classes “error” and “error-free” sound. For more detailed information it is also possible to define more sub-sound-classes, e.g. different noise emitting parts. The sound class representing sequences are extracted from database, whereas a huge variance of recording situations is very important. After learning process the weights of the neural net are adopted. Now the SOM represents the input space. In labeling process the sound class information is added. Thereafter it is possible to classify unknown test datasets and to get a feedback of associated sound class.

In the following the described method is combined in a so-called SOM-Agent which will monitor and assess microphone signals. The knowledge – the neural weights – of the input space and the sound classes – from labeling process – are transferred to the agent. Relevant parameters are calculated for the sampled audio data and will be used as input features for the SOM-Agent. Identification of sound class active neurons will follow. In Figure 2 the created SOM-Agent characteristic curve for one recording of 11 seconds is shown. The 2 defined sound classes and an “unknown” class – in case of no active neuron – are represented by an output value in range 0 to 2 on y-axis. The x-axis represents time. Statistics are also computed for the three classes.

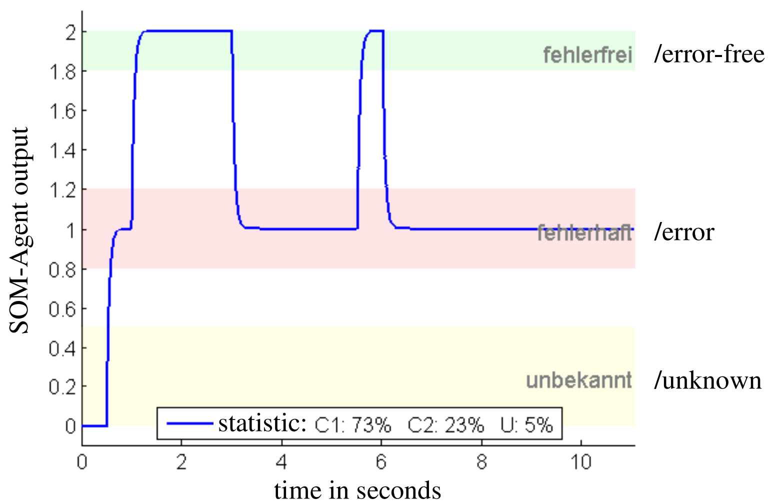

FIGURE 2: Autonomous detection of noticeable acoustic changes in car interior. The SOM-Agent characteristic curve (blue) toggles between three sound classes “error-free”, “error” and “unknown” sound, e.g. if output value is “1” the microphone recording was made in a car with damaged or worn-out parts.

3. FUNDAMENTALS

As described in previous section the shown method consists of two main components. First one is identification and calculation of sound class relevant parameters. For the second step these are the input features for a self-organizing map which will categorize them in defined sound classes.

3.1 Relevance Identification

Several methods exist for analyzing and calculation of feature influence, feature relevance, feature (subset) selection or feature extraction. Reduction of feature space to the essential is main target of several classification tasks. Anyway, the connectivity between the parameters and classes is determined. Therefore the correlation coefficient [5] – as linear method – or the information gain [6], [7] – as non-linear method – can be used. In Equation 1 the calculation rule for correlation coefficient, also known as normalized covariance is given. Thus, the linear connection between the vectorial values x and y is estimated.

$$r(x, y) = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}} \quad (1)$$

On the other hand the information gain (IG, Equation 2 at [6]) is calculated by entropy H and represents a non-linear probabilistic method. The information gain returns the contribution of one attribute to class decision making.

$$IG(S) = H(S) - \sum_{v \in Values(A)} \frac{|S_V|}{|S|} H(S_V) \quad (2)$$

Whereas S is the set and A represents the class information of S. S_V is the subset of the selected class.

Further known filter methods (parameter selection by pre-processing) are for example: Mutual Information, Fisher Discriminant, RELIEF, PCA, ICA, FCBF (Fast Correlation Based Filtering, [8], [9]). As you can see in [8], [9] the FCBF algorithm solves the two main aspects of feature subset selection. First one is in our case the decision whether mathematic parameter (feature) is relevant or not for a acoustic condition (sound class). In second step the algorithm determines whether a relevant parameter is redundant.

The symmetrical uncertainty (SU, Equation 3) is used for determination of nonlinear connection between the two sources x and y. It is based on information gain but it is normalized on interval [0,1].

$$SU(x, y) = 2\left[\frac{IG(x \mid y)}{H(x) + H(y)}\right] \quad (3)$$

In correlation analysis for a dataset the relevance and redundancy are assessed by this formula. If we are assuming that the values of one source contain same information like second source the SU output will be 1. If SU output is 0 the sources are independent of each other. The symmetrical uncertainty is used in FCBF algorithm as measure for correlation between two sources. In this algorithm two main aspects are shown.

First one is the SU calculation for the feature-class-SU:

- Serves as measure for correlation between each feature and the class information.
- For each feature the SU to all classes is calculated and stored in a list.
- For this list a threshold determines if features SU is too low. A feature is deleted if threshold is undercut.
- For each class the residual features are ordered in descending order and stored in a final list.

In second process the redundancy examination occurs by feature-feature-SU.

- Serves as a measure for correlation between each combination of the listed features.
- For each feature the SU is calculated to all other features.
- Comparison of feature-feature-SU and feature-class-SU.
- Features with very high SU are deleted.
- The remaining features are representing the best features in dataset.

In most cases the efficiency of the algorithm is shown by a classifier. For this case the recognition rate is the measure of assess. Anyway, as well known no algorithm will be the best or worst solution for all datasets. In most cases algorithms are optimized for all-round use or for dataset. As you can see in analysis (Figure 3, [9]) of UCI library [10] datasets the FCBF algorithm has best performance for *CoIL2000* (compared to three reduction methods). For *Lung-Cancer* and *Splice* dataset on the contrary all methods have nearly same recognition rates.

Title	FCBF$_{(log)}$	FCBF$_{(0)}$		Full Set		ReliefF		CFS-SF		FOCUS-SF	
	Acc	Acc	p-Val	Acc	p-Val	Acc	p-Val	Acc	p-Val	Acc	p-Val
Lung-cancer	83.33	86.67	0.34	78.33	0.34	84.17	0.85	86.67	0.34	87.5	0.46
Promoters	93.27	93.27	1	91.55	0.55	87.82	0.25	95.18	0.17	90.45	0.40
Splice	93.95	96.14	$\mathbf{0.00}^{+}$	95.52	$\mathbf{0.00}^{+}$	91.32	$\mathbf{0.00}^{-}$	93.54	0.24	94.36	$\mathbf{0.08}^{+}$
USCensus90	97.94	97.88	0.19	93.49	$\mathbf{0.00}^{-}$	97.97	0.17	97.99	0.65	97.87	0.44
CoIL2000	93.94	93.92	0.34	78.68	$\mathbf{0.00}^{-}$	93.89	0.66	92.92	$\mathbf{0.01}^{-}$	83.22	$\mathbf{0.00}^{-}$
Chemical	71.91	67.73	$\mathbf{0.02}^{-}$	60.90	$\mathbf{0.00}^{-}$	71.26	0.77	70.51	0.35	66.35	$\mathbf{0.00}^{-}$
Musk2	84.59	84.59	1	84.78	0.51	84.59	1	64.87	$\mathbf{0.00}^{-}$	83.53	$\mathbf{0.01}^{-}$
Arrhythmia	67.48	65.73	0.45	60.88	$\mathbf{0.01}^{-}$	55.79	$\mathbf{0.00}^{-}$	69.05	0.45	69.06	0.56
Isolet	50.06	83.33	$\mathbf{0.00}^{+}$	84.10	$\mathbf{0.00}^{+}$	60.90	$\mathbf{0.00}^{+}$	87.31	$\mathbf{0.00}^{+}$	71.03	$\mathbf{0.00}^{+}$
Multi-feat	95.9	95.65	0.50	94.1	$\mathbf{0.01}^{-}$	67.65	$\mathbf{0.00}^{-}$	96.15	0.64	93.7	$\mathbf{0.02}^{-}$
L/W/T	**-**	**1/2/7**		**5/2/3**		**3/1/6**		**2/1/7**		**4/2/4**	

Figure 3: Comparison of recognition rates of FCBF, ReliefF, CFS-SF and FOCUS-SF algorithm. ([9], “Acc records 10-fold cross-validation accuracy rate (%) and *p*-Val records the probability associated with a paired two-tailed t-Test.”).

3.2 Self-Organizing-Map

The neurons of a self organizing map [4] are represented by a multidimensional weight vector. If the neural net is adopted the weight vectors are already adjusted (self organized without supervise) to the input space (defined by the features/parameters). Field of application is the recognition of clusters in high dimensional data. Each input node has a connection by weights to each neuron of the map. One way of visualization is the component plane where the weights are shown as contour or surface plot. However, more detailed information for SOM-visualization (e.g. U-Matrix) methods is discussed in [11].

Representation of the n SOM-neurons $w_1, \ldots, w_n$ is done by the weight vectors $\vec{w}_1, \ldots, \vec{w}_n$. The number of elements in $\vec{w}$ is equal to the number m of input features $x_1, \ldots, x_m$. The initialization of the weights is done randomly or by prior input space knowledge. Adjustment and allocation of neurons to the input space is made in different ways, e.g. as line, two- or three-dimensional map.

The adaptation is done by a learning rule with defined step size. The interconnected neurons are adjusted by means of following equation:

$$\vec{w}_k(n+1) = \vec{w}_k(n) + \alpha \cdot (\vec{x}(n) - \vec{w}_k(n)) \quad (4)$$

In Equation 4 the calculation rule (from [4]) for adaptation of weight vectors $\vec{w}_k$ for selected neuron k is shown. The step size α is chosen in interval $0 < \alpha < 1$. There is no general rule for determination of α. If *batch learning* is used the input vector $\vec{x}$ with m features is randomly selected out of dataset. If we have a closer look to the learning process the adjustment of the weights results in a movement of the neuron in the direction of the training vector. The selection of the trained neuron occurs by a distance function, e.g. Euclid distance. The neuron which has lowest distance to input vector $\vec{x}$ is adopted. However, the algorithm has to be updated in the way that the connected neurons are covering the complete input space to get a map which is topology preservative. For this purpose a neighborhood function φ defines a radius for the selected winner neuron which has lowest distance to input vector. Thus, neurons inside the radius were adopted by the input vector, too. Furthermore a distance function $d(w_i, w_k)$ allows a division in different neighborhood-distances R for a weighted neuron adaptation. Symbolically neurons which are closest to the winner neuron have the distance $R = 1$ and neurons in next radius have a distance of $R = 2$ and so on. The degree of adaptation which defines the amount of movement for the winning neuron in the direction of input vector $\vec{x}$ depends on the neighborhood function $\varphi(w_i, w_k) \leq R$ with a range of values in [0,1]. The neurons w_i with a small distance will be adapted much more than neurons with a big distance. Also the design of the neighborhood function can be varied depending on computational resources.

- Computationally efficient Stair-Neighborhood function

$$\varphi(w_i, w_k) = \begin{cases} \left(\frac{1}{d(w_i, w_k)} \right); & d(w_i, w_k) \leq R \\ 0; & sonst \end{cases} \quad (5)$$

- Computationally intensive Gaussian-Neighborhood function

$$\varphi(w_i, w_k) = \begin{cases} e^{-\left(\frac{d(w_i, w_k)^2}{2\sigma^2} \right)}; & d(w_i, w_k) \leq R \\ 0; & sonst \end{cases} \quad (6)$$

In comparison Equation 6 realizes a more detailed and non-linear neighborhood-function than Equation 5. Finally, after insertion in Equation 4 the learning rule is shown updated in Equation 7.

$$\vec{w}_i(n+1) = \vec{w}_i(n) + \alpha \cdot \varphi(w_i, w_k) \cdot (\vec{x}(n) - \vec{w}_i(n)) \quad (7)$$

Commonly the adaptation process is done by a defined amount of training cycles. The adaptation process shown in Equation 4 and 7 can be optimized by a fast adjustment at the beginning and a slow one at the end of training. This behavior can be realized by a time-varying step-size $\alpha(t)$.

$$\alpha(t) = \alpha\left(1 - \frac{t}{T_E}\right) \quad (8)$$

By means of the fixed ratio between actual time t to the number of cycles T_E in Equation 8 a linear function is given. Non-linear step-size functions are also possible alternatives. Also it is possible to modify the neighborhood functions (Equation 5 and 6) with a time descending factor. For interpretation of the self-organizing map two main aspects were analyzed:

- Characteristics of SOM-neuron-weights in dependence of input vector.
- Class assignment (sound classes) of neurons with a maximum reaction (smallest distance) at defined stimulation (a.k.a. labeling process).

4. METHOD DESCRIPTION AND EVALUATION

As a basis for the training of the SOM-Agent (see Figure 2) a noise database with different measurements is used from which the mathematical properties are calculated. For training, 11 different noise types were defined which come from worn-out coupling rods, wheel bearings, suspension strut mountings and other running gear components from the front of a defined motor vehicle type. Noises which originate from the normal sound of a series vehicle without worn-out components represent a further noise type. In sum there are 129 measurements - with a length between 10 and 30 seconds – for which finally two training classes are defined:
“C1”: error
“C2”: error-free
The time constant for the window length for which the parameters are calculated and averaged is $250ms$. For each of the 129 datasets several parameter-sets are calculated and marked as class one or two. For the dimension reduction and relevance determination for the features the FCBF-algorithm was selected. From the 17 available parameters the parameters shown in Table 1 are determined as relevant by the FCBF-algorithm described in section C1. The training process runs T cycles, in which a randomly selected parameter-set is presented to the network. In next step all measurements are used for evaluation by SOM-Agent. Therefore all 129 acoustic records are

Roughness (Time)
Speed
Sharpness (Time)
Spectral Centroid
Yaw Rate
Loudness (Time)

TABLE 1: The 6 most relevant parameters are estimated by means of FCBF-algorithm.

presented one after the other and the characteristic curve shown in figure 2 is calculated. The parameters must be estimated for the acoustic signals continuously. Thus, these time continuous parameters equal time variable data rows where each column set is an input vector of self-organizing map. Through determination of the active neurons in dependence of the class the acoustic time signals can be assessed and equipped with class information ("C1" or "C2"). For reliable condition recognition a buffer (e.g. ring-buffer with 12000 samples) is filled with these data rows and is analyzed periodically. For this purpose a subset of the stored data in buffer is used to estimate the class. Depending on distribution of active neurons the following output is calculated:

- SOM-Agent output $s \in N$ in interval $s = [0,1,2]$, where
  - $s = 1$ corresponds with error-noise class "C1" and
  - $s = 2$ corresponds with error-free-noise class "C2".
- SOM-output $s = 0$ if input data didn't activate any neuron. That's why Zero corresponds to an unknown state "U".

For visualization a low pass filtered curve is generated from signal s:

$$s_{LP}(n) = \left(1 - \frac{1}{L}\right) s_{TP}(n-1) + \frac{1}{L} \cdot s(n) \quad (9)$$

A statistical evaluation is done afterwards either for the values of s or s_{LP}. In the case of strong fluctuating curves of s the curve of s_{LP} will generate an averaged value over the window length characterized by L. However diffuse values (e.g. $s = 1.5$) are not used for statistic analysis, that's why thresholds are implemented (have a look at colored areas in Figure 2). These thresholds are optimized for a clear assessment between two classes. The probability of occurrence for the classes (in the case of use of s_{LP} within the threshold regions) is estimated over the length of time signals. In the legend of Figure 2 the probability of occurrence is shown for the analyzed record. As a result class "C1" was determined by SOM-Agent for 70% of analyzed time.

For training process and parameterization of self-organizing map the obviously noticeable sequences are extracted out of 129 measurements. Hence a dataset with 2456 patterns with 17 parameters is generated. The 6 most relevant parameters were presented to 2 different self-organizing maps with 169 or 1000 neurons and 6 inputs. The Gaussian function described in Equation 6 was used as a neighborhood function. Neurons within the neighborhood radius R (defines the width of the bell curve) are adopted by following rule (out of Equation 7, 8, 9):

$$\vec{w}_i(n+1) = \vec{w}_i(n) + \alpha\left(1 - \frac{t}{T_E}\right) \cdot e^{-\left(\frac{d(w_i, w_k)}{R}\right)^2} \cdot \left(\vec{x}(n) - \vec{w}_i(n)\right) \quad (10)$$

As distance function, the Euclidean distance is used:

$$d(w_i, w_k) = \|w_i - w_k\| = \sqrt{\sum_{x=1}^{N} (w_{ix} - w_{kx})^2} \quad (11)$$

In figure 4 the evaluation of the recognition rates is shown for the SOM1-agent (169 neurons). On x-axis the acoustic error conditions (class "C1") in the range 1 to 11 are reproduced. Number 12 is the acoustic original state without any worn-out or damaged part (class "C2"). On y-axis the averaged probabilities of occurrence are reproduced.

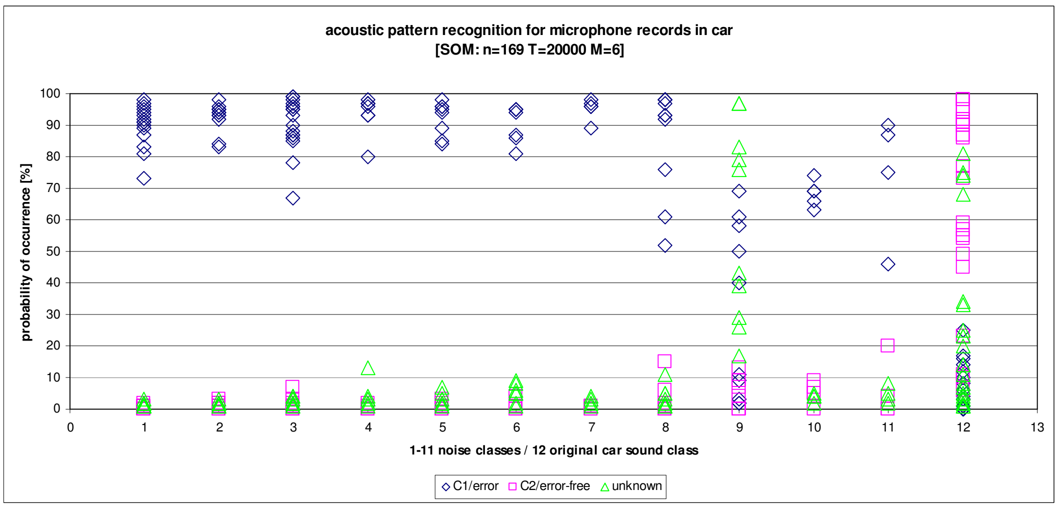

Figure 4: SOM1-Agent statistic: A calculated probability of occurrence for 129 measurements, SOM with 169 neurons and 6 parameters is used. Legend defines whether the analyzed measurements are assessed as "error", "error-free" or "unknown" condition.

The results of SOM1-Agent (Figure 4) show that primarily the noise classes 1-8 are assessed much better than 9-11. Just in the case of noise type 9 many measurements are classified as "unknown". Also many of the original sound records (class 12) are assessed as unknown condition (some records are assessed for more than 70% as unknown). This behavior shows that no neuron for one of the defined classes "error" or "error-free" was activated. This bad performance could be caused by too little records for this noise type. Anyway, the noise of a worn-out drive shaft is only noticeable for a few situations when the car is accelerated or high steering angle at low speed occurs. This is in contrast to steering rods where noticeable noise is emitted in many driving situations. Definitely all records without any noticeable noise are not considered for the database but there are still records where the noise occur e.g. for only one time in 10 seconds. However, for this case the SOM-Agent shall assess the moment where noise is noticeable as "error" class and all other as "error-free".

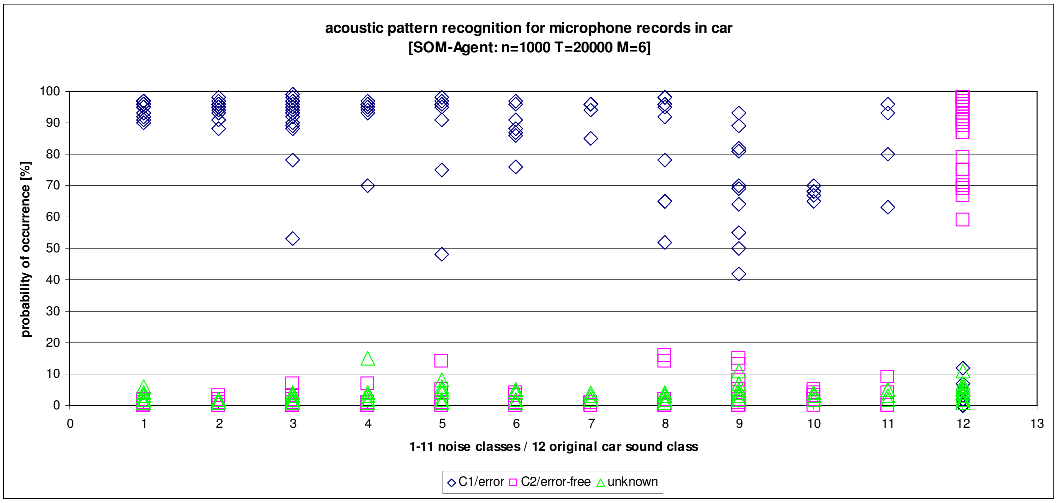

Figure 5: SOM2-Agent statistic: A calculated probability of occurrence for 129 measurements, SOM with 1000 neurons and 6 parameters is used. Legend defines whether the analyzed measurements are assessed as "error", "error-free" or "unknown" condition.

In comparison the SOM2-Agent (Figure 5) assesses much less time domain data (from 129 measurements) as "unknown". Just the original sound records are classified to more than 50% of the time for the correct category "C2" ("error-free"). This is a result of the increased amount of neurons (1000), the corresponding expansion of the input space and the more detailed information of the trained input space. In contrast to SOM1-Agent, 6 times more neurons are used.

5. CONSLUSION & FUTURE WORK

The SOM2-Agent produces a reliable statement of the matching category for the analyzed 129 acoustic in-car records. In comparison to SOM1-Agent (Figure 4 and 5) much less time-parts are declared as "unknown". All car interior time signals of an error-free car are classified for more

class	*C1 [%]*	*C2 [%]*	*U [%]*
1	95,0	0,3	2,4
2	94,6	1,3	1,6
3	90,9	1,1	2,1
4	92,0	1,3	4,4
5	87,0	2,9	3,6
6	88,7	1,7	3,7
7	93,4	0,6	2,6
8	83,5	3,4	2,4
9	69,5	4,6	4,8
10	67,6	2,8	3,0
11	83,0	3,3	3,8
12	1,8	87,3	3,4

Table 2: Averaged probabilities of occurrence (SOM2-Agent) for 129 measurements and for different error noise (class 1-11) and error-free (class 12) records.

than 50% of time to the right category "C2". The probability of occurrence of the other two categories is near less than 10%. Anyway, noises of worn-out components are assessed at least for more than 40% of analyzed time to the right category/class. So it is important to make an decision by means of the relations between all statistic outputs "C1", "C2" and "U". For example, if

"C2" is much higher than "C1" it is likely that the analyzed signal contains no information for a worn-out part.
In Table 2 all averaged probabilities of occurrence of SOM2-Agent are listed. The noise types of the class 9 and 10 are recognized worst. All other noise types emitted from defect or worn-out components are recognized as a "error"-sound ("C1") correctly via at least 83 % of the analyzed time. The "error-free"-sound ("C2") - reproduced through class 12 - is evaluated above 87 % of the analyzed time correctly by the agent. However, the average shows one more time that the introduced method works well but nevertheless further improvements will be focused on the outliers. The presented method seems to be good for pattern recognition in acoustic records due to combination of traditional pattern recognition, signal processing and statistic analyze. However, it is understood that the results will be proof for the analyzed environment conditions (e.g. car type, microphone type, measurement procedure). Furthermore, the analysis of the influence of different car types on the features and pattern recognition is necessary in further studies. It was shown that the number of the neurons has a decisive influence on the accuracy and reliability of the SOM-agent. It has to be examined also by means of further measurements how far the number of neurons, adaptation accuracy and over-fitting influence the training and assessment process.

6. REFERENCES

[1] B. Orend, I. Meyer, *Schadensfrüherkennung mittels Körperschall*. MTZ, 2009

[2] R. Jahnke, *Beitrag zur akustischen Qualitätsprüfung im Bereich industrieller Fertigung*. TU Berlin, 2001

[3] K. Stepper, *Ein Beitrag zur akustischen Güteprüfung von Komponenten der Kfz.- und Automatisierungstechnik*. TU Berlin, 1999

[4] T. Kohonen, *Self-organizing maps*, Springer Verlag, 2001

[5] S. Twieg, H. Opfer, H. Beikirch, *Correlation based method for acoustic condition recognition*. 1st IEEE Germany Student Conference, Hamburg University of Technology, 2009

[6] K.-M. Würzner, *Textkategorisierung: Entscheidungsbäume*. Universität Bielefeld, 2003

[7] A.-W. Moore, *Information Gain*. Carnegie Mellon University Pittsburgh, 2003

[8] L. Yu, H. Liu, *Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution.* ICML, 2003

[9] L. Yu, H. Liu, I. Guyon, *Efficient Feature Selection via Analysis of Relevance and Redundancy.* Journal of Machine Learning *Research*, pp. 1205-1224, 2004

[10] A. Asuncion, D. Newman, *UCI Machine Learning Repository*, University of California, Irvine, School of Information and Computer Sciences, 2007

[11] T. Fincke, V. Lobo, F. Baco, *Visualizing self-organizing maps with GIS*, GI Days 2008 Münster, 2008

[12] J. Biesiada, W. Duch, *Feature Selection for High-Dimensional Data: A Kolmogorov-Smirnov Correlation-Based Filter*, Advances in Soft Computing, Springer Verlag, pp. 95-104, 2005

[13] O. Chapelle, *Feature selection for Support Vector Machines*, Max Planck Institute for Biological Cybernetics, Tübingen, 2005

[14] M. A. Hall, L. A. Smith, *Feature subset selection: a correlation based filter approach*. International Conference on Neural Information Processing and Intelligent Information Systems. © Springer, pp. 855-858, 1997

[15] Zhipeng Li & Fuqiang Liu, Cellular Automaton Based Simulation for Signalized Street. International Journal of Engineering and Interdisciplinary Mathematics, vol. 1, no. 1, pp. 11-22, 2009

[16] Niemann, H. (1988). *Klassifikation von Mustern*. Springer Verlag, Berlin, Heidelberg, New York, Tokyo, 1988