CLASSIFICATION OF NON-SPEECH ACOUSTIC SIGNALS USING STRUCTURE MODELS

C. Tschöpe and D. Hentschel

Fraunhofer Institute for Nondestructive Testing
Dresden, Germany

M. Wolff, M. Eichner and R. Hoffmann

Dresden University of Technology,
Germany

ABSTRACT

Non-speech acoustic signals are widely used as the input of systems for non-destructive testing. In this rapidly growing field, the signals have an increasing complexity leading to the fact that powerful models are required. Methods like DTW and HMM, which are established in speech recognition, have been successfully used but are not sufficient in all cases. We propose the application of generalized structured Markov graphs (SMG). We describe a task independent structure learning technique which automatically adapts the models to the structure of the test signals. We demonstrate that our solution outperforms hand-tuned HMM structures in terms of class discrimination by two case studies using data from real applications.

1. INTRODUCTION

The recognition of acoustic signals is widely concentrated on, but not limited to, speech signals. Other biometric data coming from humans or animals, but also signals emitted by machines, vehicles, etc., must be analyzed in many cases. Among these tasks, the non-destructive acoustic analysis plays an important role. It is based on data recorded during the normal operation of a device or in special test runs, resp.

The classification of these data is frequently performed using statistical parameters of features which are calculated from the waveforms in time domain or from its transforms in spectral or wavelet domains. Such approaches are sufficient in many applications of non-destructive analysis and lead to useful results. They need, however, special knowledge on the processes generating the acoustic signals and cannot be applied in more complicated cases which are characterized by a sequential structure of the signal.

It can be easily demonstrated that the structural classification methods developed for speech recognition can be applied successfully to structured non-speech signals, too. In some cases a simple DTW recognizer can be used [1]. In more complicated cases, the concept of hidden Markov models (HMM) has been applied [2, 3, 4, 5]. Because of the well-defined sequential structure of a speech signal, hidden Markov modeling normally uses simple left-right graphs having nodes (states) that emit feature vectors following a mixed Gaussian probability density function.

It must be noted, however, that signals appearing with more sophisticated problems of non-destructive testing do not have a simple left-right structure in general. Therefore, it is necessary to reveal the structure of acoustic signals by automatically learning graphs of a rather general structure.

This paper applies our former work on a general structure training for speech signals [6] to non-speech signals. This solution has the following benefits: *(a)* it is a universal approach, *(b)* it is entirely data driven and *(c)* it requires only few prior knowledge on the underlying process. The proposed method offers a wide application range in the fields of process integrated non-destructive testing (PINT), health monitoring and life cycle prediction.

The paper describes the sequential model, the training and the classification process. We also describe two case studies: acoustic quality assessment of gear wheels and health monitoring of microfluidic valves. The results of the case studies show a clear improvement of the performance of our method compared to the HMM approach.

2. TRAINING AND CLASSIFICATION PROCEDURE

2.1. Stochastic Markov Graphs

As already mentioned, our acoustic models are structured as Markov graphs which generalize HMM's (figure 1 shows an illustration). In contrast to HMMs, the graph topology becomes an essential object of the training. Such graphs are sometimes called stochastic Markov graphs (SMG, [7]).

A stochastic Markov graph

$$\mathcal{G} = \left\{ V, E, \{\mathcal{N}\}, \nu^{(V)}, \pi^{(E)} \right\} \tag{1}$$

consists of a node (or state) set V and a directed edge set $E \subseteq V \times V$. Through the map $\nu^{(V)} : V \rightarrow \{\mathcal{N}\}$ a single n-dimensional Gaussian probability density function $\mathcal{N}_i(\underline{\mu}_i, \Sigma_i) \in \{\mathcal{N}\}$, which defines a certain area in the secondary feature space (see 2.2), is assigned to each node. Because of the arbitrary topology of a SMG, there is no need to associate mixture Gaussians with the nodes. These may easily be expressed by an appropriate graph structure. As in conventional HMMs, each edge carries a transition probability $\pi^{(E)} : E \rightarrow \mathbb{R}$.

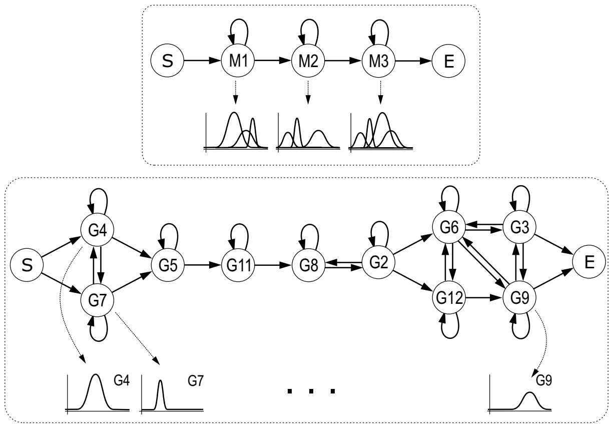

Fig. 1. 3 state HMM with 3.3 Gaussian mixtures per state (top) versus 10 state SMG (bottom). The Gaussians of both models comprise the same number of parameters. The figures correspond to the models *B1* and *E* in section 3.1.

2.2. Feature Extraction and Transformation

Our classification tasks include acoustic and supersonic signals in a frequency range from 100 Hz to 250 kHz. We use a straightforward FFT based feature extraction. In a subsequent feature transformation step we apply a principal component analysis (PCA) to the primary FFT feature vectors thus significantly reducing the feature dimensionality. The secondary feature vectors produced by the feature transformation should at most comprise 25 vector components to allow a robust estimation of Gaussian parameters with a reasonable amount of training data.

2.3. Training

This section gives a brief overview on the SMG training procedure. A detailed description and the theoretical foundation can be found in [7, 8].

Training of stochastic Markov graphs includes the estimation of the *acoustical* parameters (mean vectors $\underline{\mu}_i$ and covariance matrices Σ_i) of the Gaussians $\mathcal{N}_i$ and the determination of the *graph topologies* $\{V, E, \pi^{(E)}\}$. The training procedure performs these two tasks simultaneously (see figure 2 for a block diagram).

In a first stage one SMG model for each class m of observations is created. The models are initialized by creating a left-to-right HMM topology with K nodes, assigning Gaussians to the nodes and initializing their parameters with a sufficiently large sample of observed feature vectors (see 2.2).

The model training follows the k-means clustering scheme. First, we iteratively refine the parameters of the Gaussians using the Viterbi algorithm. We also estimate the transition probabilities $\pi^{(E)}$ between the nodes of the SMGs. After the Viterbi training has converged, we clean the SMGs removing all edges whose transition probabilities fall below

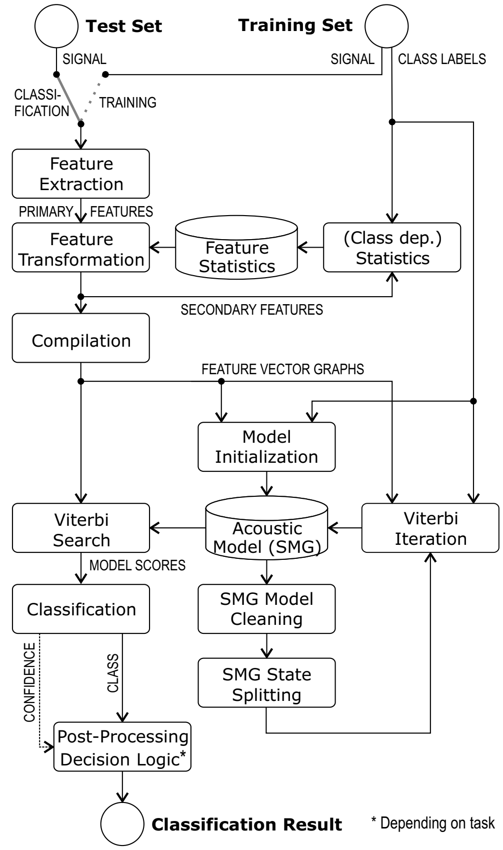

Fig. 2. Flowchart of a structural classification system with feature extraction (top, 2.2), training (right hand side, 2.3) and classification (left hand side, 2.4). Eventual post processing and decision making steps depend on the actual classification task.

a given threshold. Then we split all Gaussians (and the respective SMG nodes) along the axis of their greatest standard deviations. The new nodes inherit the transitions to all predecessors and successors of the original node.

We repeat the process of Viterbi training, cleaning and state splitting until *(a)* a given number of SMG nodes is reached or *(b)* a given number of nodes is removed during the cleaning. The latter strategy is motivated by the assumption, that the model should not overfit the training set.

2.4. Classification

Classification with SMGs works exactly the same way as classification with HMMs. The only – technical – difference lies in the arbitrary graph topology of the SMGs. For decoding we build a linear graph $\mathcal{X}$ from the observation feature vector sequence $\underline{X}$ which assigns exactly one feature vector to each node of the graph. Then we match this graph on the SMG models $\mathcal{G}_m$ of all M classes of observations. As local distance measure between a feature vector and a Gaussian we use the Mahalanobis distance (log-likelihood):

$$LL(\underline{x}_i|\mathcal{N}(v_{k,m})) = (\underline{x}_i - \underline{\mu}_{k,m})^{\mathrm{T}} \Sigma_{k,m}^{-1} (\underline{x}_i - \underline{\mu}_{k,m}) \qquad (2)$$

where $\mathcal{N}(v_{k,m}) = (\underline{\mu}_{k,m}, \Sigma_{k,m})$ denotes the Gaussian associated with node v_k of SMG m.

A Viterbi search determines the path $\mathcal{U}_m^*$ through the SMG $\mathcal{G}_m$ which maximizes the log-likelihood sum for the observation feature vector sequence:

$$\mathcal{U}_m^* = \arg\max_{\mathcal{U} \in \mathcal{G}_m} \left[\sum_{i=1}^{|\mathcal{X}|} LL(\underline{x}_i|\mathcal{N}(u_i)) \right] \qquad (3)$$

where $\mathcal{N}(u_i)$ denotes the Gaussian associated with the node $u_i \in \mathcal{U}$ and $|\mathcal{X}|$ denotes the length of the feature vector sequence. We write $LL^*(\underline{X}|\mathcal{G}_m)$ for the log-likelihood of the best path in model $\mathcal{G}_m$ given the feature sequence $\underline{X}$.

In case of a multi-class model ($M > 1$) we now choose the model m^* giving the greatest log-likelihood for the observation as recognition result:

$$m^* = \arg\max_{\mathcal{G}_m} LL(\underline{X}|\mathcal{G}_m) \qquad (4)$$

Further we use the log-likelihood as a confidence measure which allows us to determine how similar the observation and the selected class are.

3. CASE STUDIES

We tested the described classification approach in two tasks, one from the field of process integrated non-destructive testing and one from the field of health monitoring.

3.1. Process Integrated Nondestructive Testing

The first task consisted of evaluating the quality of sinter elements (gear wheels for passenger car transmissions) and of identifying errors, particularly cracks. Finally a good-bad-decision had to be made. The mechanical characteristics of sinter-metallurgically manufactured elements can be measured by their vibration response. Different influences such as cracks, material inhomogeneities or geometrical deviations change the oscillation behavior.

In the experiment the gear wheels were excited with a defined mechanical pulse from a transmitter. We used the pulse responses of the inspected items to an excitation with a 250 kHz SINC-function which were received through two different ultrasonic sensors as test signals. The signal was analyzed as described in 2.2.

In our experiment 620 gear wheels were inspected in this way. We used 20 test signals (10 good and 10 bad) as training set for the feature statistics and 160 test signals for the training of a single-class model of "good" items. The classification experiments were carried out on the remaining 450 test signals. We conducted 3 classification experiments: In the first two (baseline) experiments we used HMMs with mixture Gaussians (*B1*: 3 states, 3.3 Gaussians in average per mixture; *B2*: 10 states, 1 Gaussian per mixture). Due to the limited amount of available training data the HMMs had to be hand-tuned to achieve an optimal trade-off between number of states and number of Gaussians per mixture. In the third experiment (*E*) we used a SMG model with 10 states which was automatically adapted to the structure of the test signals using the method described in 1. Figure 1 shows the topology of this SMG.

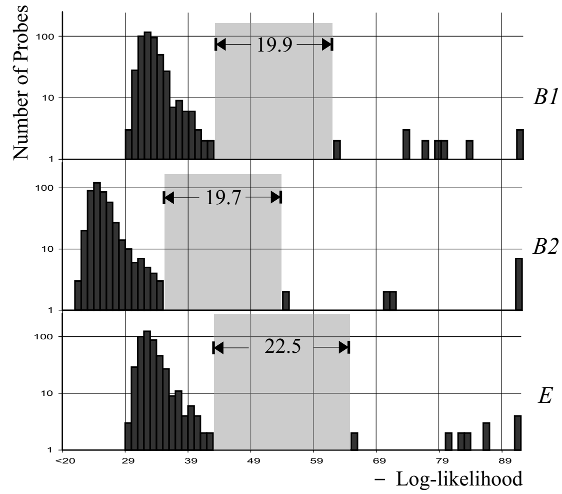

Fig. 3. Histograms of log-likelihoods for the three tested models. All models separate bad probes correctly from good ones. However, the SMG model *E* has a noticeably greater discrimination margin (gray area).

As we only model the "good" items we define a classification threshold of the log-likelihood sum according to equation (3). Probes falling below this threshold are classified as "bad". All gear wheels were also inspected allowing us determine the classification error rate of the models. All three experiments worked without any classification errors. However, experiment *E* achieved a significantly higher (16% relative) discrimination, in terms of log-likelihood difference, between the good and bad samples with approximately the same means and standard deviations of the log-likelihoods of the good samples (table 1).

Table 1. Results of PINT experiment.

Log-likelihood LL	*B1*	*B2*	*E*
Threshold good–bad	-57.0	-50.0	-59.0
Classification error	0.0 %	0.0 %	0.0 %
Good class minimum	-41.7	-34.3	-42.2
Good class average	-32.6	-24.9	-32.6
Good class std. deviation	1.9	2.1	1.9
Bad probes maximum	-61.7	-54.0	-64.7
Discrimination Margin	20.0	19.7	22.5

3.2. Health Monitoring

In the second task we accomplished a life span analysis of pneumatic valves. We attached eight new valves, which were opened and closed with a frequency of approximately 1 Hz cyclically, on a board. Compressed air with a pressure of 0.5 bar served as flow medium. The long time investigation took place until the valves did not work correctly any longer.

Mechanical changes, which occur during the aging process, do effect the switch noise of the valves. Therefore, a sound recording of the valve during its switching process can be used to evaluate the valve's condition. The goal of the experiment is a remaining useful life estimation in order to detect a forthcoming failure.

On behalf of a well-known manufacturer of valves a piezoelectric sensor system has been developed at our institute. The sensor element records the impact sound of the valve and consists of a Piezo ring which is, as a washer, wedged under the retaining screw of the valve head. Miniaturized hybrid electronics performing signal preprocessing is situated directly besides this sensor. A CompactPCI computer with integrated DSP board serves as control unit and realizes control, signal processing and data acquisition.

The manufacturer guarantees a lifetime of about 10 million switching cycles. The long time investigation has been run since the first of August 2003. We recorded one test signal every 2500 switching cycles. From the 4000 recorded test signals we used the first 600 to train individual single-class models of the "new"-conditions of all valves in the test. The remaining 3400 test signals were used to test the models. We conducted similar classification experiments

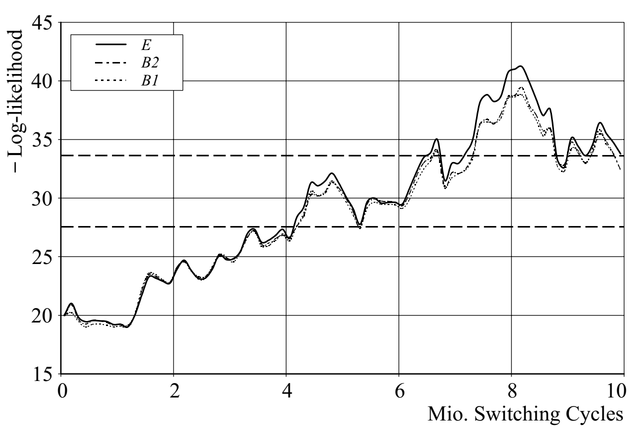

Fig. 4. Log-likelihoods of good models given the test signals of one selected valve with increasing number of switching cycles.

as described in 3.1: *B1*: HMM with 10 states and 16 Gaussians per mixture; *B2*: HMM with 20 states and 2 Gaussians per mixture. As in the first task, the number of states and Gaussian mixtures of the HMM's were hand-tuned. By the SMG training procedure (*E*) we obtained a model with 37 states which is comparable to *B1* and *B2* in terms of the number of Gaussian parameters. As figure 4 shows, all three classificators again separated the new from the used state of the valves well. The peak of the log-likelihood value around 8 million switching cycles indicates approaching the end of the valve's lifetime. Also in this task the SMG model performs better than the HMMs even though the improvement is here not as significant as in the PINT experiment. However, it must be noted that the good performance of the HMMs was achieved by hand-tuning of their structure while the SMG structure was obtained by our automatic procedure without any hand tuning.

4. CONCLUSION

We successfully applied an SMG based classification technique to different types of acoustic and supersonic signals from the field of non-destructive testing. We showed that a structured stochastic model of the signals can be automatically learned from a test set by a task independent procedure and that the obtained models outperform hand-tuned HMMs. In our future work we will apply the introduced classification technique to a greater variety of tasks.

5. REFERENCES

[1] P. Holstein, M. Koch, D. Hirschfeld, R. Hoffmann, D. Bader, and K. Augsburg, "A strategy for signal recognition under adverse conditions," in *Proc. 32nd Conf. Internoise*, 2003, Jeju, Korea.

[2] D. Zhang, Y. Zeng, X. Zhou, and Cheng Y., "The pattern recognition of non-destructive testing based on HMM," in *Proc. 4th World Congress on Intelligent Control and Automation (Cat. No.02EX527)*, 2002, vol. 3, pp. 2198–2202, Piscataway, NJ, USA.

[3] P. Baruah and R. B. Chinnam, "HMMs for diagnostics and prognostics in machining processes," in *Proc. 57th Meeting of the Society for Machinery Failure Prevention Technology*, 2003, pp. 389–398, Virginia Beach, USA.

[4] A. R. Taylor and S. R. Duncan, "A comparison of techniques for monitoring process faults," in *Proc. Conf. Control Systems*, 2002, pp. 323–327, Stockholm, SE.

[5] H. Y. K. Lau, "A hidden Markov model-based assembly contact recognition system," *Mechatronics*, vol. 13(8-9), pp. 1001–1023, 2003, ISSN 0957-4158.

[6] M. Eichner, M. Wolff, and R. Hoffmann, "A unified approach for speech synthesis and speech recognition using stochastic Markov graphs," in *Proc. 6th Int. Conf. Spoken Language Processing (ICSLP)*, 2000, vol. 1, pp. 701–704, Beijing, PR China.

[7] F. Wolfertstetter and G. Ruske, "Structured Markov models for speech recognition," in *Proc. ICASSP*, 1995, pp. 544–547, Detroit.

[8] M. Eichner, S. Ohnewald, M. Wolff, and R. Hoffmann, "Speech synthesis using Stochastic Markov Graphs," in *Proc. ICASSP*, May 5–7 2001, Salt Lake City, UT, USA.