Vehicle Acoustic Classification in Netted Sensor Systems Using Gaussian Mixture Models

Burhan F. Necioğlu, Carol T. Christou, E. Bryan George, Garry M. Jacyna The MITRE Corporation, 7515 Colshire Dr., McLean, VA 22102-7508, USA

ABSTRACT

Acoustic vehicle classification is a difficult problem due to the non-stationary nature of the signals, and especially the lack of strong harmonic structure for most civilian vehicles with highly muffled exhausts. Acoustic signatures will also vary largely depending on speed, acceleration, gear position, and even the aspect angle of the sensor. The problem becomes more complicated when the deployed acoustic sensors have less than ideal characteristics, in terms of both the frequency response of the transducers, and hardware capabilities which determine the resolution and dynamic range. In a hierarchical network topology, less capable Tier 1 sensors can be tasked with reasonably sophisticated signal processing and classification algorithms, reducing energy-expensive communications with the upper layers. However, at Tier 2, more sophisticated classification algorithms exceeding the Tier 1 sensor/processor capabilities can be deployed. The focus of this paper is the investigation of a Gaussian mixture model (GMM) based classification approach for these upper nodes. The use of GMMs is motivated by their ability to model arbitrary distributions, which is very relevant in the case of motor vehicles with varying operation modes and engines. Tier 1 sensors acquire the acoustic signal and transmit computed feature vectors up to Tier 2 processors for maximum-likelihood classification using GMMs. In a binary classification task of light-vs-heavy vehicles, the GMM based approach achieves 7% equal error rate, providing an approximate error reduction of 49% over Tier 1 only approaches.

Keywords: Acoustics, feature extraction, Gaussian mixture models, distributed classification.

1. INTRODUCTION

Vehicle detection and classification has been one of the mainly emphasized tasks of an internally-funded research program in netted sensors recently undertaken by the MITRE Corporation, with potential application areas that include border monitoring, combat vehicle identification, and urban warfare. The focus of the first-year effort has been an acoustics-only based approach, the preliminary results of which are reported here.

Acoustic vehicle classification is a difficult problem due to the non-stationary nature of the signals, the lack of strong harmonic structure for most civilian vehicles brought about by their highly muffled exhausts, and the usually dominant tire noise. A vehicle's acoustic signature will vary largely depending on its speed, acceleration, gear position and engine speed, the road conditions and texture, and even the aspect angle of the sensor itself. Usually, features extracted from the acoustic waveform are not fully adequate for clear vehicle type discrimination.^{1, 2} The problem becomes more complicated within the paradigm of networked sensors, with significant size and energy constraints on the lower tier sensor/processor nodes. The deployed acoustic sensors will usually have less than ideal characteristics, in terms of both the frequency response of the transducers, and the hardware capabilities which determine the resolution, dynamic range, and sampling interval of the captured acoustic signals.

In a hierarchical network topology, the less capable sensors at the "Tier 1" level can be tasked with reasonably sophisticated signal processing and classification algorithms, like the linearly weighted discriminator (LWD), reducing energy-expensive communications with the upper layers.³ However, at the "Tier 2" level, where information from the lower tier "trickles up", and where there is substantially more computing power, more sophisticated classification algorithms can be deployed, which would certainly exceed the capabilities of the "Tier 1" sensors/processor combinations. One such sophisticated approach that is very suitable for these "Tier 2" processor nodes is to use a maximum likelihood (ML) classifier based upon Gaussian mixture models (GMM),

For further information: E-mail: {necioglu,christou,bgeorge,gjacyna}@mitre.org, Telephone: 1-703-883-6865



Figure 1. Map of traffic circle and leading roads, and placement of mote/sensor combinations.

and this has been the focus of the investigation presented here. The use of GMMs is motivated by their ability to model arbitrary distributions, which is very relevant in the case of motor vehicles which have varying engine types, sizes, configurations, and operating modes. Components of vehicular noise (engine, muffler, tire) have differing contributions to the overall acoustic waveform, depending on the vehicle speed, gear and throttle positions, type of road surface, and so forth. With adequate training data, GMMs have the potential to represent variations within each class which are caused by factors such as the number of cylinders, different muffler characteristics, gear position/engine speed, vehicle speed and acceleration/deceleration.

Naturally, this approach requires a data set that is representative of the problem at hand to train the GMM parameters for the considered vehicle classes. Therefore a data collection effort was an integral part of this investigation and is described in Section 2. In Section 3, feature extraction on the remote sensor is described, followed by Section 4 on ML classification using GMMs. Finally, the experimental results are presented and discussed in Section 5, with summary and conclusions given in Section 6.

2. DATA COLLECTION

The performance of statistical model-based automatic classification algorithms is largely dependent on the availability of an adequate amount of training data. Lacking any available suitable corpora of vehicle acoustic data, an effort was undertaken to collect real data as part of this initiative. The main goal was to record vehicles using sensors that would actually be deployed, so that more realistic data could be utilized for the investigation of the classification problem.

The wireless remote sensors utilized in this investigation comprised Crossbow[®] Technology Inc. MICA2 (MPR400/MPR410) processor/radio boards ("motes") mated to MTS310 general-purpose multi-sensor boards with an on-board microphone, a light detector, a temperature sensor, an accelerometer and a magnetometer. The mote/sensor board combinations were enclosed in plastic food containers, mounted to the lid. Holes were drilled at the bottom of the containers and polyester quilt batting was used to act as windscreen. The inside of the containers were lined with foam to make the characteristics approach an anechoic chamber. The containers were fitted with four legs in each corner, so that they would sit about two inches off the ground. This packaging allowed for a non-directional acoustic sensor configuration. Due to the limitations of the hardware, the data acquisition was restricted to 8-bit PCM at a sampling rate of 3750 Hz, with 125 ms long data-frames extracted every 250 ms. The mote radios were too slow to ex-filtrate this output, and as a result, each mote/sensor combination was connected to a Crossbow MIB510 programming and serial interface board using a 51-position ribbon cable through a slit cut on the side of the container. Each interface board was in turn connected via its serial port to an x86 notebook computer running GNU/Linux. The software running on the motes implemented an adaptive detection algorithm and sent out data frames through the interface board when a "vehicle" was detected. Each detection segment was written to a separate file on the notebook computer hard drive, with time-stamp information inserted into the filename.

The data collections were performed in September 2004 around the MITRE campus in McLean, Virginia. Specifically, a traffic circle and three roads leading to it were selected for the placement of six motes (Figure 1). This area sees a mix of light and heavy vehicles, going up-hill and down-hill. Two of the roads leading to the circle also have stop-signs. The vehicles decelerate when approaching the circle, generally come to a full-stop where there is a stop-sign, and accelerate leaving the circle. This provided for a good variety of vehicle and engine speed combinations. The mote/acoustic sensor combinations were also individually calibrated to equalize their sensitivity. The whole duration of the data collection was captured on video by a consumer digital camera placed in a sixth-floor office overlooking the area. During post-processing of the collected data, the time-stamped audio segments were manually categorized into vehicle types, using this accompanying video footage.

The end-product of this data collection effort was 2640 segments of vehicle audio data, representing a total of about 146 minutes (about 73 minutes of actual captured acoustic waveform due to the 50% duty cycle mote data exfiltration process). The vehicles included a variety of cars, small, medium and large sport-utility vehicles (SUV), minivans, light trucks, full-size/delivery vans, medium and large diesel trucks and buses, and several examples of motorcycles. For this study, it was decided to focus on the two-class classification problem of "light"-vs-"heavy" vehicles, using data recorded by the sensor which was closest to each passing vehicle. Cars, SUVs, minivans and light trucks were lumped into the "light" category, whereas medium and and heavy diesel trucks and buses were designated as "heavy". This "closest point" sub-set of the whole collection data set consisted of 1123 "light" vehicle segments (almost 32 minutes actual), and 65 "heavy" vehicle segments (about 4 minutes of actual). Furthermore, the "closest point" data set was split into two disjoint partitions at the mid-point of the data collection time window, and both partitions included data captured at all six locations. These partitions were used alternatively as class model parameter training sets and test sets. Table 1 summarizes the vehicle segment statistics for the data sub-set used in this investigation.

3. FEATURE EXTRACTION

In a netted sensor environment, the Tier 1 sensors would be tasked with reasonably sophisticated signal processing and classification algorithms, reducing energy-expensive communications with the upper layers. As such, the time-waveform associated with acoustic events detected by the remote sensor would not be communicated out. However, features extracted by the Tier 1 sensors might be conveyed upwards, where they can be utilized by more sophisticated and more computationally expensive algorithms.

The features should be simple enough to be computed with the limited resources of the Tier 1 sensors, yet they should capture the essential information necessary for the classification task at hand. Motor vehicles are complicated systems, and a combination of different sources contribute to their acoustic signals, which include the engine, its fans, the exhaust system, the transmission system components, as well as the interaction between the tires and the road surface.^{4–6} The harmonics of the engine and exhaust noise depend on the number of cylinders, piston transient settling times, piston ping amplitudes, engine speed, and other quantities which may

Set	Number of Segments	Segment Duration (in seconds)			
		Average	σ	Minimum	Maximum
Partition 1 "light"	559	3.46	2.36	0.25	14.26
Partition 2 "light"	564	3.36	2.24	0.25	12.01
Partitions 1&2 "light"	1123	3.41	2.30	0.25	14.26
Partition 1 "heavy"	32	7.75	2.48	3.50	12.26
Partition 2 "heavy"	33	6.42	2.92	0.75	10.76
Partitions 1&2 "heavy"	65	7.07	2.77	0.75	12.26
All	1188	3.61	2.47	0.25	14.26

Table 1. Segment statistics for the vehicle acoustic data sub-set of "closest point" recordings and its partitions.

vary significantly with vehicle size and operation mode.⁷ However, these harmonics might also be masked, by considerable tire noise which is a function of vehicle speed, tire size and tread pattern, and the road conditions.^{8,9}

In light of all this, feature vectors formed by the output energy levels of a generalized parametric non-linear filter-bank form were investigated. The parameters were defined as the number of filters in the bank (n_b) and the "stretch factor" (k_s) . To keep the computations on the sensor processor simple, non-overlapping, rectangular filters were considered. For any given filter-bank implementation, the Nyquist bandwidth is divided into $n_b + 1$ unequal width sections with bandwidths $b_0, b_1, \ldots, b_{n_b}$, such that for $i = 1, \ldots, n_b$, $b_i = b_{i-1} \cdot k_s$, and filter 0 corresponding to b_0 which contains the DC term is dropped, yielding an n_b dimensional feature vector. (Of course, in actual implementation, this definition can only be a general guide subject to quantization because of the discrete frequencies of the DFT, and to satisfy $N_F/2 = \sum_{i=0}^{n_b} b_i$, where N_F is the DFT size.) Then, for each analysis frame x(n) of duration $N_a \leq N_F$, the mean energy of filter i becomes element i of the feature vector \vec{v} , such that,

$$v_i = \frac{1}{b_i} \sum_{\substack{k=1+\sum_{j=0}^{i-1} b_j \\ j=0}}^{b_i + \sum_{j=0}^{i-1} b_j} |X(k)|^2, \quad i = 1, \dots, n_b,$$
(1)

where X(k) is the complex DFT value at discrete frequency k. The energy normalized variation of \vec{v} is given by

$$\vec{v}_n = \frac{N_F/2}{\sum_{k=0}^{N_F/2-1} |X(k)|^2} \vec{v}.$$
(2)

As an example, given the sampling frequency of 3750Hz, frame duration of 0.125 ms (corresponding to 469 samples), $N_F = 512$, $n_b = 15$ and $k_s = 1.25$, the rectangular filter center frequencies and bandwidths are given in Table 2.

4. CLASSIFICATION

As stated in Section 3, for the problem of acoustic vehicle classification, the feature space will encompass substantial variability, even only considering the features extracted from the *same vehicle*, at different road and engine speeds, at acceleration, or at deceleration. Therefore, it is quite possible that the even the distribution of feature vectors from a single vehicle will have multiple modes. Considering the fact that the classification problem considered here has two broad classes with an even larger variation of vehicle sizes, types, engine types and displacements, cylinder numbers, tire sizes and tread patterns, and so forth, the ability of GMMs to model arbitrary distributions with multiple modes makes them all the more appealing. (The GMM approach has been widely used in pattern recognition problems, especially in speech and speaker recognition,^{10–12} and is a well established technique.)

Filter $\#$	f_c (Hz)	Bandwidth (Hz)	Bandwidth ratio to previous
1	29.30	14.65	_
2	51.27	21.97	1.50
3	80.57	29.30	1.33
4	117.19	36.62	1.25
5	161.14	43.95	1.20
6	219.73	58.59	1.33
7	292.97	73.24	1.25
8	388.19	95.22	1.30
9	505.38	117.19	1.23
10	651.86	146.48	1.25
11	834.97	183.11	1.25
12	1062.02	227.05	1.24
13	1347.66	285.64	1.26
14	1706.55	358.89	1.26
15	1794.44	161.13	0.45

Table 2. Non-overlapping rectangular filter center frequencies and bandwidths for the sampling frequency of 3750Hz, $N_F = 512$, $n_b = 15$ and stretch factor $k_s = 1.25$.

The probability density function for a multivariate random variable \vec{v} with a Gaussian mixture distribution is given by:

$$P(\vec{v}) = \sum_{i=1}^{M} w_i N(v, \mu_i, C_i), \qquad \vec{v} \in \mathbf{R}^n$$
(3)

where $\Lambda_i = \{w_i, \mu_i, C_i\}$ are the parameters for the *i*th mixture component, and represent the mixture weight, mean vector, and the covariance matrix, respectively. While there is no closed-from solution for estimation of these parameters, the iterative Expectation-Maximization (EM) algorithm¹³ is the most commonly used method.

For this investigation, estimation (training) of the GMM parameters for each class was performed using a MATLAB PDF estimation toolbox for Gaussian mixtures.¹⁴ Once the class model parameters are estimated, the maximum-likelihood classification of an observation vector \vec{x} is performed according to:

$$\log(P(\vec{x}|\text{Class}_0)) - \log(P(\vec{x}|\text{Class}_1)) \stackrel{0}{\underset{1}{\geq}} T \tag{4}$$

where T is a threshold that determines the operating point of the binary classifier on the detection error trade-off (DET) curve^{*} In the case of multiple observation vectors $\vec{x_1}, \ldots, \vec{x_p}$, as would happen for a detected acoustic segment that is longer than a single analysis frame, with the assumption of independent and identically distributed observations, the same test would become:

$$\frac{1}{p} \sum_{i=1}^{p} \left(\log(P(\vec{x_i} | \text{Class}_0)) - \log(P(\vec{x_i} | \text{Class}_1)) \right) \overset{0}{\underset{i=1}{\overset{>}{_{\sim}}} T$$
(5)

using an average log-likelihood measure over the incoming feature vectors.

The alternative classifier that represents a solution running at the Tier 1 nodes, is the closed-form LWD algorithm, the details of which are given in Reference 3. In summary, the hypothesis test for the LWD classifier

^{*}The DET curve is similar to the receiver operating characteristic curve (ROC).

with parameters $\vec{\gamma}$ and β is given by:

$$\frac{1}{p}\sum_{i=1}^{p}(\vec{x_i}'\cdot\vec{\gamma}+\beta)\underset{1}{\stackrel{0}{\geq}}T$$
(6)

for a segment of observation vectors $\vec{x_1}, \ldots, \vec{x_p}$, and decision threshold T.

The performance comparison of the GMM based classifier for the Tier 2 nodes, and the LWD algorithm for the Tier 1 nodes is the subject of the next section.

5. EXPERIMENTAL RESULTS

Using the two class data set of "light" and "heavy" vehicles (Table 1) a variety of classification experiments were performed by varying the parameters of the non-linear filterbanks to obtain a variety of front-end features, for the Tier 1 LWD classifier, and for the Tier 2 GMM classifier.

For every experiment undertaken, model parameters were trained on one partition, tested on the other, and the DET curve was established for the "light"-vs-"heavy" vehicle classification task. This was repeated by exchanging the training and test partitions. The average equal error rate (EER) operating point from the two runs was used for the evaluation and comparison of various classification schemes.

A number of experiments were run by varying the number of filterbanks (n_b) between 8 to 17, stretch factors between 1 (linearly spaced filterbank) to 1.5. Further variation was introduced by using normalized feature vectors (\vec{v}_n) as given in Eq. (2) so that only the information present in the relative frequency content of the feature vector would be used in classification. These variations in features were implemented using both GMM based and LWD based classifiers. For the GMM classifier case, the feature vectors were subjected to a logtransformation, and the minimum standard deviation allowed for the mixture components during training was varied between 0.1, 0.05, and 0.01, and the rest of the estimation parameters were left at their default values. Table 3 lists a sub-set of the results for 20 out of a total of 560 experiments performed using these variations, including the best EER figures for the varied parameters and classification schemes.[†]

The best overall performance of $7.0\pm1.5\%$ average EER was achieved by using a GMM classifier and a filter bank with 15-bands and a stretch factor $k_s = 1.25$ (Table 2), using un-normalized log-energy terms (see Figure 2 for the DET curve). With this feature set configuration, the "light" class converged to have 16 mixture components (modes) training on Partition 1 and 11 mixture components training on Partition 2 data, whereas the "heavy" class converged to have four mixture components training on Partition 1 data, and three mixture components training on Partition 2 data. Reducing the number of filters by 33% to 10 and using a stretch factor $k_s = 1.1$, with un-normalized log-energy terms, the EER slightly increases to $7.7\pm1.5\%$ which is only an insignificant difference. Using normalized log-energy terms, the 16-element filterbank with a stretch factor $k_s = 1.20$ achieves an average EER of $11.4\pm1.8\%$ (see Figure 3 for the DET curve). Using the same type of energy terms, reducing the number of filters to 11 and using a stretch factor $k_s = 1.25$, the increase in EER to 11.6 ± 1.8 is insignificant.

In comparison, almost all of the LWD classifier configurations achieved a best average EER figure of $13.8\pm2.0\%$, using un-normalized filter RMS-energy terms (see Figure 4 for the DET curve). However, when the RMS-energy terms were *normalized*, the average EER figures for the LWD classifier increased significantly to between 30–40%. The DET curve for the best configuration among these is given in Figure 5.

It is observed that by applying the logarithm function to the filter energies, better performance is achieved using the GMM based classifiers compared with the case of using raw RMS-energy terms. This probably can be best explained by the fact that although GMMs can model arbitrary distributions, the highly skewed distributions of non-negative filter energy outputs still cause some under-fitting problems, and the logarithm function helps by decreasing this skew in the distributions, making them easier to fit with a Gaussian mixture model. However, for the LWD classifier, using the un-normalized raw RMS-energy values always gives better average EER figures than using the un-normalized log-energy values, consistent with the derivation.³

[†]The confidence intervals given in Table 3 are estimates obtained by approximating the binomial distribution with a normal distribution

Test Part.1	Test Part.2	Overall	Type	n_b	k_s	$\log()$	normalized	σ_{\min}
6.2 ± 1.9	7.9 ± 2.2	7.0 ± 1.5	GMM	15	1.25	Yes	No	0.1
$6.2{\pm}1.9$	$9.1{\pm}2.3$	$7.7{\pm}1.5$	GMM	10	1.10	Yes	No	0.05
$6.2{\pm}1.9$	$9.1{\pm}2.3$	$7.7{\pm}1.5$	GMM	16	1.20	Yes	No	0.01
$7.4{\pm}2.1$	$10.7 {\pm} 2.5$	$9.1{\pm}1.6$	GMM	16	1.17	Yes	No	0.1
$7.4{\pm}2.1$	$12.1{\pm}2.6$	$9.8{\pm}1.7$	GMM	11	1.25	Yes	No	0.1
$10.6{\pm}2.5$	$12.1{\pm}2.6$	$11.4{\pm}1.8$	GMM	16	1.20	Yes	Yes	0.01
$9.4{\pm}2.4$	$13.7 {\pm} 2.8$	$11.6{\pm}1.8$	GMM	11	1.25	Yes	Yes	0.05
$9.4{\pm}2.4$	18.2 ± 3.1	$13.8{\pm}2.0$	GMM	16	1.17	Yes	Yes	0.1
$12.5 {\pm} 2.7$	$16.4 {\pm} 3.0$	$14.4{\pm}2.0$	GMM	15	1.25	Yes	Yes	0.1
$12.5{\pm}2.7$	18.2 ± 3.1	$15.3 {\pm} 2.0$	GMM	10	1.10	Yes	Yes	0.01
$12.5{\pm}2.7$	$15.2 {\pm} 2.9$	$13.8{\pm}2.0$	LWD	8	1.10	No	No	n/a
$12.5{\pm}2.7$	$15.2{\pm}2.9$	$13.8{\pm}2.0$	LWD	10	1.10	No	No	n/a
$12.5{\pm}2.7$	$15.2{\pm}2.9$	$13.8{\pm}2.0$	LWD	11	1.25	No	No	n/a
$12.5 {\pm} 2.7$	$15.2 {\pm} 2.9$	$13.8{\pm}2.0$	LWD	15	1.25	No	No	n/a
$12.5 {\pm} 2.7$	$15.2 {\pm} 2.9$	$13.8{\pm}2.0$	LWD	16	1.17	No	No	n/a
$12.5 {\pm} 2.7$	$15.2 {\pm} 2.9$	$13.8{\pm}2.0$	LWD	16	1.20	No	No	n/a
$26.4{\pm}3.6$	$33.4{\pm}3.8$	$29.9{\pm}2.6$	LWD	16	1.17	No	Yes	n/a
$25.0{\pm}3.5$	$34.9{\pm}3.8$	$30.0{\pm}2.6$	LWD	10	1.10	No	Yes	n/a
32.1 ± 3.8	$33.4{\pm}3.8$	$32.8 {\pm} 2.7$	LWD	15	1.25	No	Yes	n/a
$37.5 {\pm} 3.9$	$38.6{\pm}3.9$	$38.1 {\pm} 2.8$	LWD	16	1.20	No	Yes	n/a
$34.4{\pm}3.8$	$44.2 {\pm} 4.0$	$39.3{\pm}2.8$	LWD	11	1.25	No	Yes	n/a

Table 3. Percent equal error rate (EER) figures with estimated 95% confidence intervals for the "light"-vs-"heavy" vehicle binary classification task, with two-partition jack-knifing (train on Partition 2, test on Partition 1, and vice versa).

6. CONCLUSIONS

These experiments suggest that in a binary classification task of "light"-vs-"heavy" vehicles, the GMM based approach can reduce the EER by approximately 49% compared to the LWD classifier with the best configuration. GMM based classifiers are computationally more expensive and therefore more suitable for Tier 2 processor nodes, whereas LWD classifiers can be implemented at the Tier 1 sensor/processor level. However, the GMM based classifiers degrade more gracefully when the feature vectors are normalized with respect to energy, and this is an important detail, given that employing un-normalized features requires calibration of the sensors, which might be impractical in actual deployment systems. The results also suggest that employing GMMs, result in better utilization of the information present in the frequency content, especially when considering the energy-normalized feature sets.

Since both the LWD and GMM classification schemes provide a confidence measure associated with the classification decision that could be fed upward in a netted sensing environment, and utilized in collaborative approaches, both have the potential to be very useful in tracking and classifying targets in a sensor field.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Dr. Paul M. Baggenstoss of the Naval Undersea Warfare Center, Newport, RI, for the MATLAB PDF estimation toolboxes for Gaussian mixtures and HMMs he makes available on-line at http://www.npt.nuwc.navy.mil/csf. The authors would also like to thank Dr. Brian Flanagan



Figure 2. The DET curve for the GMM classifier with a filter bank of 15-bands and a stretch factor of 1.25, using un-normalized log-energy terms.

and Dr. Kenneth Parker of The MITRE Corporation for their help in programming the MICA2 processor boards, and Ms. Naomi Kolatch of The MITRE Corporation for her help in the data collection and annotation process.

REFERENCES

- G. S. Anderson, C. S. Lee, and G. G. Fleming, "FHWA Traffic Noise Model, Version 1.0: User's Guide," Tech. Rep. FHWA-PD-96-009 and DOTVNTSC-FHWA-98-1, John A. Volpe National Transportation Systems Center, Acoustics Facility, Cambridge, MA, January 1998.
- C. W. Menge, C. J. Rossano, G. S. Anderson, and C. J. Bajdek, "FHWA Traffic Noise Model, Version 1.0: Technical Manual," Tech. Rep. FHWA-PD-96-010 and DOT-VNTSC-FHWA-98-2, John A. Volpe National Transportation Systems Center, Acoustics Facility, Cambridge, MA, February 1998.
- G. M. Jacyna, C. T. Christou, B. E. George, and B. F. Necioğlu, "Netted Sensors-Based Vehicle Acoustic Classification At Tier 1 Nodes," in SPIE Defense and Security Symposium, Proceedings of SPIE 5796, (Orlando, FL), March 2005.
- C. T. Christou and G. M. Jacyna, "Simulation of Vehicle Acoustics in Support of Netted Sensor Research and Development," in *SPIE Defense and Security Symposium*, Proceedings of SPIE 5796, (Orlando, FL), March 2005.



Figure 3. The DET curve for the GMM classifier with a filter bank of 16-bands and a stretch factor of 1.20, using *normalized* log-energy terms.

- 5. M. Allman-Ward, M. P. Balaam, and R. Williams, "Source Decomposition for Vehicle Sound Simulation Simulation and Test Centre Report," tech. rep., MSX International, Milbrook, England.
- A. Crewe et al, "Sound Decomposition-A Key to Improved Sound Simulation, paper no. 2003-01-1423," tech. rep., Society of Automotive Engineers, 2003.
- M. Bodden and R. Heinrichs, "Analysis of the Time Structure of Gear Rattle," in *Proceedings of Internoise* 99, pp. 1273–1278, (Fort Lauderdale, FL), 1999.
- U. Sandberg and J. A. Ejsmont, "Noise Emission, Friction and Rolling Resistance of Car Tires–Summary of An Experimental Study," in *Proceedings of The 2000 National Conference On Noise Control Engineering*, (Newport Beach, CA), December 3-5, 2000.
- U. Sandberg, "The Multi-Coincidence Peak around 1000 Hz in Tyre/Road Noise Spectra," in *Proceedings* of Euronoise Conference, (Naples, Italy), 2003.
- L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of The IEEE 77, pp. 257–286, Feb. 1989.
- D. A. Reynolds and R. C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Transactions on Speech and Audio Processing* 3, pp. 72–83, Jan. 1995.
- J. P. Campbell, Jr., "Speaker Recognition: A Tutorial," Proceedings of The IEEE 85, pp. 1437–1462, Sept. 1997.



Figure 4. The DET curve for the LWD classifier with a filter bank of 8-bands and a stretch factor of 1.10, using un-normalized RMS-energy terms.

- A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Royal Stat. Society 39-1, pp. 1–38, 1977.
- 14. P. M. Baggenstoss, "Statistical Modeling Using Gaussian Mixtures and HMMs with MATLAB," in http://www.npt.nuwc.navy.mil/Csf/htmldoc/pdf, Naval Undersea Warfare Center, Newport, RI.



Figure 5. The DET curve for the LWD classifier with a filter bank of 16-bands and a stretch factor of 1.17, using *normalized* RMS-energy terms.