Classification of Military Ground Vehicles Using Time Domain Harmonics’ Amplitudes

Peter E. William
University of Nebraska-Lincoln, peter.ezzat@ieee.org

Michael W. Hoffman
University of Nebraska-Lincoln, mhoffman1@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/electricalengineeringfacpub

Part of the Electrical and Computer Engineering Commons
Classification of Military Ground Vehicles Using Time Domain Harmonics’ Amplitudes

Peter E. William, Member, IEEE, and Michael W. Hoffman, Member, IEEE

Abstract—In the context of the United Nations peacekeeping operations, we developed an energy-efficient method for the detection and classification of military vehicles using a group of wireless sensors. The method is adequate for low-power unattended sensors which perform sensing, feature extraction, and classification in a standalone scenario. Harmonics’ amplitudes approximating the harmonic signature of the time domain acoustic signal captured by wireless sensor nodes are estimated for vehicle discrimination. The computational complexity for the time domain features extracted from ground vehicles’ acoustic signals is lower than their equivalent spectral features. Classification is performed using a multilayer feedforward neural network, where discrimination between vehicles depends on their acoustic signature irrespective of their speed or location. Evaluation of the time domain method, through processing of an acoustic data set for heavyweight and lightweight military ground vehicles with comparison to spectral features, shows that time domain harmonics’ amplitudes are simpler to obtain and provide the reliability of the spectral features in both the detection and false alarm rate.

Index Terms—Acoustic emission, decision fusion, harmonic signals, time domain features, vehicle classification.

I. INTRODUCTION

Monitoring cease-fire and peace agreements is one among a range of activities undertaken by the United Nations (UN) and other international actors to maintain peace and security throughout the world [1]. The UN operations have relied almost exclusively on human observers, both military and civilian, for monitoring. It is difficult, if not impossible, to monitor large regions and to maintain a permanent presence in distant locations relying only on human observers [2]. Modern monitoring technologies offer aerospace, ground, and even underground continuous surveillance. Besides the extended range, unmanned monitoring offers the ability to record data and retrieve information which can be very useful for further analysis.

Advances in microelectromechanical systems, signal processing techniques, wireless communication, and low-power hardware modules have enabled the development of low-cost sensor nodes that are capable of performing complex tasks [3]. The purpose of a sensor network is to provide users access to the information of interest from data gathered by spatially distributed sensors. In most monitoring operations, users require only certain aggregate functions of this distributed data. There are several system functions that contribute to the energy profile of the individual sensor node. These functions, shown in Fig. 1, include data acquisition, processing, power mode change, sleep/idle mode, and communication. The sensing power depends on the transducer, and for most passive transducers (acoustic, seismic, infrared, etc.), the amount of consumed power is very reasonable. Communication power depends on the rate, separation, protocol, channel, and SNR. Radios with low bit rates (≤100 Kb/s) are very advantageous in terms of power consumption, which motivates performing the processing in the node instead of simply forwarding the large amount of acquired data to a central location. Low-power sensor node modules commonly employ a low-power microcontroller which is not power efficient in executing a complex algorithm. In [4], the power consumed for a military vehicle classification scheme using a Mica2 sensor node module by Crossbow [5] was estimated. The algorithm extracted features from the spectrum of the acoustic signal. Classification was performed using support vector machines. The power consumed, on the Mica2 testbed, for a fast Fourier transform (FFT) on 512 samples was 23.9 μW, while the power for feature selection and classification was 0.28 μW. This tells us that for detection and classification applications, the majority of the processing power is consumed during the feature extraction phase especially when spectral analysis is used.

The majority of previous ground vehicle detection and classification techniques extract the characteristic features from the frequency representation of the acoustic signal [6]–[14]. The motivation for using frequency domain features is the energy concentration of the acoustic signal, which lies mostly in the range 20–400 Hz. In [8] and [9], consecutive FFT components were selected for generating elements of the feature vector (FV), while in [13], [15], [16], only harmonically related FFT peaks that exceeded the noise level were selected for generating the FV. Harmonically related peaks result in a lower dimension FV, but determining those peaks has two requirements: first,
the effective resolution of the computed FFT components has to be about 1 Hz or less, which raises the overall number of complex multiplications; second, a good estimate of the fundamental frequency (FF) and the number of harmonics should be obtained. Computing the FFT and analyzing the spectrum can be relatively expensive to perform on-board a battery operated sensor node [17].

We previously proposed extracting features from the time domain acoustic signal [18], [19] for vehicle detection and classification. The algorithm was based on approximating the harmonic structure in the higher frequency band (>125 Hz) and was tested only on two heavyweight military vehicles. The previous time domain algorithm was unsuccessful in distinguishing among a wide variety of vehicles since a significant amount of information was lost by neglecting the lower frequency band. In addition, it is more challenging to be able to differentiate not only among large armored tanks but also among lightweight wheeled vehicles which produce lower acoustic sound levels with gasoline engines that have more effective mufflers.

In this paper, a time domain harmonics’ amplitudes (TDHA) method is developed for vehicle classification. The extracted features are the signal energy, the estimated strongest harmonic frequency, and the harmonics’ amplitudes estimated from a template of the acoustic signal. The computational power needed for extracting the time domain features using the TDHA method is less than that required to obtain the spectral features. The TDHA method aims to increase the life time of individual sensor nodes. The TDHA method can also be used in many applications that utilize acoustic/seismic signatures for source identification as in traffic control [20], surveillance [10], [15], [16], health and environmental monitoring, and fault diagnosis [21].

In this paper, the TDHA method was tested on actual acoustic data for different vehicles collected in the context of the Bochum Verification Project (BVP) [6], which investigates the potential of automatic sensor systems for verification of disarmament and peace keeping agreements. The Project was conducted at Institut für Experimentalphysik III, Ruhr-Universität Bochum, Germany, and is continued now at Experimentelle Physik III, Technische Universität Dortmund, Germany.

The paper is organized as follows: Section II describes the acoustic signature of ground vehicles, Section III illustrates the TDHA method applied on ground vehicles’ acoustic emissions, Section IV shows how neural networks are used in this paper for classification of the set of characteristic time domain features, and Section V includes the experimental data processing results followed by the conclusion in Section VI.

II. ACOUSTIC SIGNATURE OF GROUND VEHICLES

The acoustic emissions of ground vehicles contain a wealth of information which can be used for vehicle classification [16]. Ground vehicles have two main sources of acoustic emissions, namely, the engine and the propulsion mechanism [11], [22], which can be modeled as a combination of two main components, a deterministic component and a nondeterministic component. The engine and drive train are responsible for the deterministic component and can be represented using a coupled harmonic signal model. In the spectral domain, the acoustic emission is composed of a family of narrowband harmonic lines with an FF directly related to the engine revolution per minute and other periodic components that arise from the tracks or the interaction between the tires and the road. For tracked vehicles, the track produces an additional series with another FF [7]. Ground vehicles can be sensed using seismic, acoustic, thermal, electric, magnetic, chemical, or optical sensors. Among these sensors, seismic and acoustic sensors are characterized by their passive nature, nonlinear of sight sensing, relatively long range, and low-power requirements and can be used either during day or at night [23]. Acoustic-seismic sensors are now found in various battlefield ground sensors, generally known as unattended ground sensors [24].

Acoustic features have proven to be quite useful in vehicle classification, equivalent seismic features less so [8]. The most significant limitation that affects the performance of acoustic sensors employed for monitoring cease-fire territories is wind noise [16]. While acoustic propagation is dependent on atmospheric conditions, propagation effects over a short range are not very relevant.

The time-frequency responses of sample runs of the acquired acoustic signals from two tracked military vehicles passing by a sensor node are shown in Fig. 2. The harmonic structure and the time-varying nature of the signals are very apparent in this figure. There are significant components up to 400 Hz when the vehicle is approaching the closest point of approach. A coupled harmonic signal accurately models a large portion of the vehicle’s acoustic signature, especially that coming from the engine. The noise and a portion of the target’s acoustic
signature that does not fit into the coupled harmonic model are responsible for the remaining components in the signal spectrum. For ground vehicles, the engine FF of the acoustic signal typically lies within the range (8–20) Hz\(^1\) [10], [11]. According to the FF range for military vehicles, there might be up to 100 distinct harmonics that characterize the signature of the vehicle under investigation.

III. TDHA

Extracting characteristic features from the time domain signal is the core of the TDHA method, which eliminates the need to perform an FFT for the generation of characteristic FVs. Time or frequency domain features are evaluated according to the number of multiplications required for feature extraction, the dimensionality of the FV, and the amount of information carried by the characteristic feature that is crucial for discrimination. In our proposed model, we select only the acoustic deterministic signature of the ground vehicle for identification since it was shown in [13] that most of the vehicle’s acoustic signal energy lies in the harmonically related components.

A. Coupled Harmonic Signal Model

A simple coupled harmonic model that represents a single harmonic signal can be described as [11]

\[
x(t) = \sum_{k=1}^{M} A_k \cos(2\pi f_k t + \varphi_k)
\]

(1)

where \(A_k \geq 0\) and \(\varphi_k\) are the amplitude and phase of the \(k\)th harmonic, respectively; \(f\) is the FF; and \(M\) is the total number of harmonics. It is clear that the number of parameters is \(2M + 1\). Assuming that \(y = x + \epsilon\) is the vector of acquired acoustic sensor data, where \(\epsilon\) is assumed to be the noise vector with a multivariate Gaussian distribution such that \(\epsilon \sim \mathcal{N}(0, \Sigma)\). The acquired signal may also be written in matrix form as

\[
y = Cu + Sv + \epsilon
\]

(2)

where the elements of the \(u\) and \(v\) vectors and the \(C\) and \(S\) matrices are given by

\[
u_k = [A_k \cos \varphi_k]^T \quad \text{and} \quad v_k = [-A_k \sin \varphi_k]^T
\]

\[
c_{nk} = \cos \left(2\pi f_k n + 2\pi f_k (N-1)/2\right) \quad \text{and} \quad s_{nk} = \sin \left(2\pi f_k n + 2\pi f_k (N-1)/2\right)
\]

(3)

where \(k = 1, \ldots, M\); and \(n = 0, \ldots, N-1\). Defining the phase in the middle of the observation window guarantees that \(CTS = 0\) and \(S^TC = 0\) (diagonalizing the information matrix). Under a white noise assumption, the maximum likelihood (ML) estimation is given by the least squares (LS) solution. In our case, the simplified LS solution is given by

\[
\hat{u} = (CTC)^{-1}CTy \quad \text{and} \quad \hat{v} = (STS)^{-1}STy.
\]

(4)

For the single harmonic model, with both the FF and the number of harmonics defined, the costly matrix inversion is reduced to simple scaling since \((CTC)^{-1}\) and \((STS)^{-1}\) produce diagonal matrices. Using the amplitude vector estimates, the coupled harmonic signal amplitude \(A_k\) is simply equal to \(\sqrt{\hat{u}^2 + \hat{v}^2}\). Experimental results in [11] and [12] on real-time recordings showed that the estimated phase does not produce improved performance over the estimated signal amplitude due to propagation effects and it is not used in our approach.

B. Strongest Harmonic Period

Without knowing the FF and the number of harmonics, the complexity in estimating the harmonics’ amplitudes is large. Even when the complexity was discarded, several verification schemes in the literature [6], [7] failed in defining the harmonic line series from the spectrum of the acoustic signal searching among all distinct peaks, while other schemes [11], [12] found difficulty in precisely estimating the FF value, even when the number of harmonics was fixed.

The TDHA approach avoids searching for the actual FF by estimating the strongest harmonic component instead. There are two reasons why we are interested in estimating the strongest harmonic component: first, it requires simple and few computations to be estimated; second, it will be combined with the required spectral resolution \((N_s)\) in determining the fundamental component of the approximated harmonic model for estimating the amplitude parameters. We will call this fundamental component the assumed FF \((\hat{F}_F)\). The time domain signal is processed in the form of short events (windows). Selecting events of short durations \((0.5\ s)\) guarantees that variation in the harmonic structure is minimum.\(^2\) The amplitude of the strongest harmonic component was previously used for normalization [25], in order to mitigate the variation in signature with range (i.e., the separation between sensor node and vehicle). Estimating the strongest harmonic period \((\tau)\) is accomplished through computing the correlation between the first half of the current event time signal and its shifted version. Computing the correlation of the acoustic signal represents a simple and efficient method for defining repeated patterns. The correlation maximum occurs at zero shift, and the distinct correlation peaks are identified to occur at shifts equivalent to the strongest harmonic component period and its multiples. If correlation peaks are located at displacements equal to \(d_1, d_2, d_3, \ldots\) etc., the strongest harmonic period becomes

\[
\tau = \frac{d_1}{F_s}
\]

(5)

where \(F_s\) is the sampling frequency. The estimated period can be refined by considering all peaks or selecting those with amplitude above the noise level as follows:

\[
\tau = \text{mean}\left(\frac{d_1}{F_s}, \frac{d_2}{F_s}, \frac{d_3}{F_s}, \ldots\right).
\]

(6)

\(^1\)In [7], the engine FF for the tracked YPR 765 armored vehicle was 26 Hz at 3000 r/min.

\(^2\)In [6], a window size of \(\leq 0.82\ s\) is recommended such that the harmonic structure of ground vehicles will not change by more than 10% during the same event.
Simple examples of the correlation function are shown in Fig. 3 for different SNRs. The original signal has a harmonic structure with the strongest harmonic period equal to 8.9 ms. The estimated harmonic period using either the first correlation peak \(d_1\) or the second correlation peak \(d_2\) is found to be 8.5 and 9 ms, respectively. The accuracy in estimating \(\tau\) depends on two important parameters, the expected value of the strongest harmonic period and the sampling frequency. As the strongest harmonic period decreases or the sampling frequency decreases, the resolution in estimating the strongest harmonic period also decreases and vice versa. This limitation comes from the fact that we compute the correlation at positive integer shifts only.

The simple harmonic model discussed in Section III-A is defined by two components, namely, the FF and the number of harmonics. The TDHA method uses the estimated strongest harmonic frequency \((\xi = (1/\tau))\) and \(N_s\) in computing \(\tilde{FF}\) such that

\[
\tilde{FF} = \left\lfloor \frac{\xi}{N_s} \right\rfloor \tag{7}
\]

where \(\left\lfloor \xi/N_s \right\rfloor = \min\{L \in \mathbb{Z}|L \geq (\xi/N_s)\}\). The number of harmonics is fixed and is selected such that the generated coupled harmonic model approximates the acoustic signature up to 250 Hz. Note that averaging over nine different vehicles, 96.64% of the energy of the acoustic emission for more than 32 000 detected events\(^3\) was found to be in the range 0–250 Hz. Using (7), the \(\tilde{FF}\) is always an integer divisor of the estimated strongest harmonic frequency with a defined spectral resolution, which guarantees that the generated harmonic model will cover this frequency in addition to others with higher energy concentration.

\(^3\)Each event represents 0.5 s.
Fig. 4 shows the time-frequency response for all detected events for a tank passing through a sensor node with both the actual and estimated strongest harmonic frequency. The acoustic signal was processed in 0.5-s windows, the total number of detected events for this specific run was 64 (32 s). It is clear in Fig. 4 that another harmonic component dominates, most probably that coming from the tracks, when the vehicle is closer to the sensor node. Although the simple estimation of the strongest component slipped for few events, the use of (7) in estimating a fundamental helps in correctly estimating a quite good approximation of the actual signal spectrum. This is seen from the comparison in Fig. 5, between the actual signal spectrum computed with 0.17-Hz resolution and the estimated harmonics' amplitudes using the LS method at four different events \((t_1, t_2, t_3, t_4)\). Table I shows the actual and estimated frequencies for the four events.

The estimated harmonics' amplitudes are likely to have the same envelope as the original signal spectrum, as shown in Fig. 5. Although there was an error in estimating the strongest harmonic component at \(t_4\), the estimated harmonics' amplitudes using the LS method confirm that the strongest harmonic is at 62.95 Hz and not at 96.15 Hz. Although the change in the assumed fundamental is very small, this difference becomes larger for higher harmonic components. It is also shown in Fig. 5 that the estimated amplitudes of all harmonics that are not distinct are close to zero.

### C. Computational Savings

The \(\tilde{F}F\) represents the smallest harmonic component of the coupled harmonic model used for approximating the deterministic acoustic signature. A reduced number of samples equivalent to the period of the \(\tilde{F}F\) will be selected as a template of the original window. The selection is supported
by the validity of the coupled harmonic model, and the fact that harmonic components change slowly [14]. By extracting the characteristic features from the template, we select the signature of this vehicle in the frequency range starting at $F_F$ to $F_F/2$. A simple illustration of the amount of computational savings due to data reduction is shown in Table II for different sampling frequencies and window lengths.

The actual benefit lies in the computational power needed to extract the harmonics’ amplitudes as compared to the FFT components. The number of real multiplications required for estimating the harmonics’ amplitudes in the TDHA method is $O(M \ast NT)$, where $M$ is the number of harmonics and $NT$ is the template width. The number of complex multiplications required to obtain $N$ FFT coefficients is $O(N \log_2 N)$ [26]. Each complex multiplication requires four real multiplications and two real additions. Extracted features using a higher resolution leads to better classification over both methods, but the TDHA method requires a smaller number of multiplications as compared to the FFT method. The number of real multiplications required to estimate the harmonics’ amplitudes compared to the number of real multiplications needed to obtain their equivalent FFT components for different spectral resolutions is shown in Fig. 6. We implemented the proposed feature extraction algorithm in fixed point operation using 16 bit data values with 32 bit computational precision. A fixed point realization of the FFT algorithm, using twiddle factors from [26], for extraction of spectral features is implemented as well for comparison. Both algorithms use lookup tables, namely, twiddle factor tables for the spectral features and sine and cosine tables for the time domain features. On average, the time consumed for extraction of the time domain features is less than 23% of the time required to obtain the spectral features for the same tested signal on the same platform.

### Table II

**Illustration of the Amount of Savings Provided by Using a Template That Corresponds to an $F_F$ Period With $N_s = 5$ Hz**

<table>
<thead>
<tr>
<th>$F_s$ (Hz)</th>
<th>Window length (seconds)</th>
<th>$\text{Template} = \frac{F_F}{F_s}$ (samples)</th>
<th>Percentage of the template size with respect to the original window size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024</td>
<td>0.5</td>
<td>205</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td></td>
<td>20%</td>
</tr>
<tr>
<td>2048</td>
<td>0.5</td>
<td>410</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td></td>
<td>20%</td>
</tr>
</tbody>
</table>

Fig. 6. Real multiplications needed to estimate both the harmonics amplitudes and the FFT components at different $N_s$.

The number of harmonics $M$ is fixed and depends on the spectral resolution and the projected spectral band. After estimating the $M$ harmonics’ amplitudes using the LS method as described in Section III-A, the characteristic FV that will be used in discrimination is constructed from the estimated strongest harmonic frequency, the event energy, and the $M$ estimated harmonics’ amplitudes $(\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_M)$. The length of the FV is $M + 2$.

The performance of vehicle detection and classification systems for peacekeeping operations is highly dependent on many factors as follows: the diversity of vehicle classes; the variety of environments in which peacekeeping operations must operate; the set of selected features; and the classifier. The comparison in [18] between three classifiers [ML, linear vector quantization (LVQ), and artificial neural networks (ANNs)] for military vehicle classification using their acoustic signatures have shown that ANN has the highest performance.

ANNs derive their computational power from their ability to learn combined with the nonlinearity of neurons. Simple feedforward neural networks (FNNs) are typically used in vehicle classification [10], [15], [16], [27]. A three-layer FNN with sigmoid transfer function is utilized, where the number of neurons in the input layer is equivalent to the length of the characteristic FV, and the number of neurons in the output layer is equivalent to the number of vehicle classes. The number of neurons in the hidden layer is usually selected to be sufficiently high so that the network is capable of adapting, as indicated by improved performance. Increasing the number of neurons requires storing more weights and biases and performing more calculations. Hence, a tradeoff exists between increased performance and reduced complexity. For our data set, we use nine output classes, one for each vehicle, and we use feature vectors of length 52. With these constraints, we varied the number of neurons in the hidden layer between 3 and 60, and we found that having fewer than 30 neurons in the hidden layer limited the ability of the ANN to adapt. In addition, there was no significant increase in discrimination capability with more than 50 neurons in the hidden layer. Therefore, the number of neurons chosen for the hidden layer in our simulations is 40. The ANN is trained for 200 epochs, where after each epoch, the weights and biases are updated once using a factor of 1.2 times the error derivative after calculating the ANN output using all
available passes in the entire training data set. The FNN started to adapt to the highly variable data in the training set after just a few epochs and training for 100 epochs or more guarantees a good classification rate. Training up to even 1000 epochs or more did not give any significant increase (or decrease) in the classification performance.

Decision fusion represents an energy-efficient solution that reduces noise effects and increases the reliability in the sensor network by reducing individual classification errors [8], [9], [21]. Decision fusion has been successfully utilized in many vehicle classification approaches, either by fusing decisions from spatial nodes [9] or by fusing decisions produced by the same node at different time instances [18] or for different modalities [27]. In our approach, we selected a simple temporal decision fusion algorithm that combines individual decisions produced by the same node at different time instances. The basic temporal decision fusion algorithm employs voting between individual decisions, which is a simple and efficient method with minimum number of computations and memory requirements [28]. In this scenario, a single decision will be given representing the majority voting of all detected events from a single sensor node during a single run.

V. DATA SET AND RESULTS

The acoustic data of ten different vehicles (Leopard 1, Leopard 2, M48, Jaguar, Wiesel, Fuchs, Hermelin, Unimog, Mercedes-Benz 1017, and Volkswagen van) were recorded by the BVP during verification experiments in 2000 [6], [7]. Table III lists some properties regarding the ground vehicles, and Fig. 7 shows their pictures divided into four main categories. The experimental recordings equipped each sensor station with acoustic and seismic sensors to capture the acoustic and seismic signatures of the moving vehicles along four

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Category</th>
<th>Weight (Kg)</th>
<th>Number of Cylinders</th>
<th>Engine Capacity (cm³)</th>
<th>Transmission</th>
<th>Recorded (minutes)</th>
<th>Detected (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavyweight Tracked</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leopard 1</td>
<td>Main Battle Tank</td>
<td>40200</td>
<td>10</td>
<td>37400</td>
<td>auto</td>
<td>137.26</td>
<td>51.00</td>
</tr>
<tr>
<td>Leopard 2</td>
<td>Main Battle Tank</td>
<td>52000</td>
<td>12</td>
<td>47600</td>
<td>auto</td>
<td>138.30</td>
<td>51.64</td>
</tr>
<tr>
<td>M48</td>
<td>Main Battle Tank</td>
<td>45600</td>
<td>12</td>
<td>29400</td>
<td>auto</td>
<td>110.79</td>
<td>48.32</td>
</tr>
<tr>
<td>Jaguar</td>
<td>Tank Destroyer</td>
<td>23200</td>
<td>8</td>
<td>29900</td>
<td>auto</td>
<td>93.60</td>
<td>41.72</td>
</tr>
<tr>
<td>Heavyweight Wheeled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuchs</td>
<td>Armored General Purpose Carrier</td>
<td>14400</td>
<td>8</td>
<td>12800</td>
<td>auto</td>
<td>202.94</td>
<td>58.71</td>
</tr>
<tr>
<td>Hermelin</td>
<td>Reconnaissance Vehicle</td>
<td>9745</td>
<td>6</td>
<td>5700</td>
<td>manual</td>
<td>131.38</td>
<td>43.99</td>
</tr>
<tr>
<td>Lightweight Tracked</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wiesel</td>
<td>Airborne Weapon carrier</td>
<td>1927</td>
<td>5</td>
<td>2000</td>
<td>auto</td>
<td>142.86</td>
<td>28.11</td>
</tr>
<tr>
<td>Lightweight Wheeled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MB 1017</td>
<td>Medium Truck</td>
<td>6800</td>
<td>6</td>
<td>5700</td>
<td>manual</td>
<td>116.51</td>
<td>16.05</td>
</tr>
<tr>
<td>Unimog</td>
<td>Small Truck</td>
<td>5150</td>
<td>6</td>
<td>5700</td>
<td>manual</td>
<td>85.72</td>
<td>14.18</td>
</tr>
<tr>
<td>VW 70X0C</td>
<td>Van</td>
<td>1713</td>
<td>5</td>
<td>2400</td>
<td>manual</td>
<td>20.73</td>
<td>4.70</td>
</tr>
</tbody>
</table>

Fig. 7. Ten different vehicles.
different lanes (paved and unpaved) as shown in Fig. 8. Only one vehicle at a time was driven past the sensing stations while the data were collected. The road length exceeded 400 m, 200 m on each side of the main sensor line. The TDHA method is employed on the acoustic data since the harmonic structure is more apparent in the acoustic emission of vehicles, and the performance was degraded by incorporating the seismic data. Each run represents one vehicle passing through the sensor field, and the acoustic data are captured using two different sensor stations on different sides of the road. The separation between the two stations is 101.4 m. More than 365 runs (>19.5 Hours) were recorded at variable speeds with different directions and on different surfaces. Fig. 8 shows a layout for the sensor stations’ positions with respect to the four lanes.

The acoustic data set was originally sampled at 20 kHz. Prior to detection and feature extraction, the data were downsampled to 5 kHz. The original recording was triggered manually when the vehicle crosses the 200 m flag to acquire the whole run. In order to determine whether an event has occurred or not, the adaptive constant false alarm rate (CFAR) [9] outputs a decision every 0.5 s, based on the average energy level of the acoustic signal. The CFAR threshold is updated according to the noise variance of the detector output. Evaluating the performance of the CFAR detector is done by measuring any gaps that occur in the detection process.4 On average, the implemented CFAR detector misses less than 7% of total number of detected events. The CFAR detector also detects events that do not have a clear signature, and those events are already included in the classification rate in the presented results. The total number of detected events per run depends on the vehicle itself and its speed. During normal speeds, medium and lightweight vehicles are detected within 50 m from the sensor station, while heavyweight tracked vehicles are detected beyond 100 m. Although the same number of runs was recorded for each vehicle, the variation in the detection range resulted in having the total number of detected events per vehicle range between 1200–5500 events/vehicle. The total number of runs for the Volkswagen van was very small compared to the remaining military vehicles which resulted in having a very small number of generated FVs. ANNs are very sensitive to the size of the training set, and since we were trying to adapt a single network to differentiate among all vehicles, the Volkswagen van was dropped from our simulations. The complete set of detected events used for evaluation includes around 32 000 events.

Discriminating between tracked and wheeled vehicles or heavyweight and lightweight vehicles can be performed using features generated from the acoustic and seismic signal energy or the signal’s peak amplitudes in addition to other statistical parameters. In this paper, we investigate the ability of features extracted from the time domain acoustic signal to correctly classify among nine vehicles5 using a single three-layer FNN. The input layer constructed from 52 neurons and represents the FV, the hidden layer consists of 40 hidden neurons, and the output layer consists of nine neurons (corresponding to nine vehicles). The training (learning) of the network is performed according to the resilient backpropagation (Rprop) algorithm developed from an analysis of the performance of the standard steepest descent algorithm [29]. Rprop is a high-performance learning algorithm which obtains fast convergence of the weights and biases.

Classification results are given in the form of detection, false alarm, and classification rates. The detection rate for each class is the ratio of the number of events correctly classified for the given class to the total number of events in the given class. The false alarm rate for each class is the ratio of the number events for all other classes classified as the given class to the total number of events in the given class. The classification rate is the ratio of the number of events correctly classified for all classes to the total number of events. Military vehicle classification using acoustic features is an application where 80% or better correct classification rate is considered to be excellent [28].

In order to evaluate the discrimination ability of harmonics’ amplitudes as compared to spectral features, FFT components were computed for three different spectral resolutions (5, 10, and 20 Hz for NFFT points = 1024, 512, and 256, respectively, at 5 kHz sampling frequency). NFFT spectral components were computed for each detected event with a duration of 0.5 s using a rectangular window. Three different windows were tested (Hamming, Hanning, and rectangular), and the spectrum obtained using the rectangular window resulted in better classification than the other two windows. The FFT characteristic FV that will be used in discrimination is constructed from the event energy followed by the first 50 FFT coefficients approximating the spectral signature in the low-pass band up to 250, 500, and 1000 Hz, respectively. This results in a FV of length 51. The TDHA method6 was applied for the same set of detected events for the same approximate spectral resolution as the FFT method. To determine the advantage of estimating the fundamental of the approximated harmonic model from the strongest

4A gap is defined as one or more events that are preceded and followed by at least 2 s or more of continuous detection.

5In the BVP data set, one specific vehicle is used for each class-rather than a set of the same type vehicle representing each class.

6The strongest harmonic component was estimated using the first three peaks of the correlation function.
In this setting, the ANN is trained using events that are randomly selected from all available events, and the ANN is tested using all remaining events. Table IV illustrates the average detection, false alarm, and classification rate for the TDHA method and the FFT method for Monte Carlo simulations of 100 iterations. In each iteration, 1/3 of the total number of detected events are randomly selected for network training, and the remaining 2/3 are used for testing. Temporal decision fusion results were also included in Table IV showing the average classification rate among all vehicles when a single decision is produced per run\(^7\) by combining all decisions for a single run of one target crossing a sensor node.

\(^7\)A run represents all detected events from a single recording for a single target.

Selecting the $\overline{F_F}$ proportional to the strongest harmonic component works substantially better than having a constant FF. Table IV shows that for a fixed FF length, the TDHA modeling of the deterministic signature of the acoustic emission of vehicles is at least as effective in vehicle classification as are the equivalent spectral features from the FFT method. The classification rate of lightweight vehicles is less than the equivalent heavyweight vehicles as expected since the harmonic structure is less distinct. Although the variety of vehicles and unequal number of events per vehicle make the classification phase very challenging to the NN, the classification performance for the given FVs is quite good. When comparing the TDHA method with the results in [7], the average classification rate among nine vehicles is 90.38% using the proposed simple TDHA method. While using a more computationally complex approach based on the relative powers of the first 15 harmonics, the authors in [7] achieved 89% for only eight vehicles (the Wiesel harmonic structure could not be evaluated). Although the authors in [7] used the same data set, they used a different classifier (LVQ).

<table>
<thead>
<tr>
<th>Features $\rightarrow$</th>
<th>Harmonics’ amplitudes $\overline{F_F}$ of strongest component $\rightarrow$</th>
<th>Fixed FF</th>
<th>FFT using a rectangular window $\rightarrow$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle $\downarrow$</td>
<td>Resolution $\rightarrow$</td>
<td>$\sim$ 5Hz</td>
<td>$\sim$ 10Hz</td>
</tr>
<tr>
<td>Leopard 1</td>
<td>97.49</td>
<td>95.07</td>
<td>94.55</td>
</tr>
<tr>
<td>Leopard 2</td>
<td>94.15</td>
<td>90.49</td>
<td>81.71</td>
</tr>
<tr>
<td>M48</td>
<td>97.94</td>
<td>96.80</td>
<td>92.46</td>
</tr>
<tr>
<td>Jaguar</td>
<td>92.26</td>
<td>88.52</td>
<td>83.89</td>
</tr>
<tr>
<td>Fuchs</td>
<td>86.49</td>
<td>84.56</td>
<td>77.44</td>
</tr>
<tr>
<td>Hermelin</td>
<td>74.87</td>
<td>73.69</td>
<td>63.77</td>
</tr>
<tr>
<td>Unimog</td>
<td>66.54</td>
<td>62.39</td>
<td>53.14</td>
</tr>
<tr>
<td>MB1017</td>
<td>70.23</td>
<td>62.26</td>
<td>53.62</td>
</tr>
<tr>
<td>Wiesel</td>
<td>84.57</td>
<td>83.08</td>
<td>79.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Features $\rightarrow$</th>
<th>False alarm rate (%) $\rightarrow$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leopard 1</td>
<td>0.39</td>
</tr>
<tr>
<td>Leopard 2</td>
<td>1.89</td>
</tr>
<tr>
<td>M48</td>
<td>0.35</td>
</tr>
<tr>
<td>Jaguar</td>
<td>1.95</td>
</tr>
<tr>
<td>Fuchs</td>
<td>1.26</td>
</tr>
<tr>
<td>Hermelin</td>
<td>1.10</td>
</tr>
<tr>
<td>Unimog</td>
<td>0.64</td>
</tr>
<tr>
<td>MB1017</td>
<td>1.63</td>
</tr>
<tr>
<td>Wiesel</td>
<td>1.61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Features $\rightarrow$</th>
<th>Classification rate (%) $\rightarrow$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leopard 1</td>
<td>90.38</td>
</tr>
<tr>
<td>Leopard 2</td>
<td>92.05</td>
</tr>
</tbody>
</table>

### Table IV

**Detection, False Alarm, and Classification Rate—Random Training and Testing Events**

*component, the harmonics' amplitudes of a harmonic model with a fixed FF at 5 Hz were estimated using the LS method.*

*Classification Results Using Randomly Selected Events:* In this setting, the ANN is trained using events that are randomly selected from all available events, and the ANN is tested using all remaining events. Table IV illustrates the average detection, false alarm, and classification rate for the TDHA method and the FFT method for Monte Carlo simulations of 100 iterations. In each iteration, 1/3 of the total number of detected events are randomly selected for network training, and the remaining 2/3 are used for testing. Temporal decision fusion results were also included in Table IV showing the average classification rate among all vehicles when a single decision is produced per run\(^7\) by combining all decisions for a single run of one target crossing a sensor node.
and have different training/testing sequences, where 70% of their detected events were selected for codebook generation, with 30% used for testing, so the results are not directly comparable.

Classification Results Using Events From One Sensor Station for Training and the Other for Testing: In this setting, the ANN is trained using events that are detected using one of the two sensor stations and is tested using events detected by the other station. Table V illustrates the average detection, false alarm, and classification rate for Monte Carlo simulations of 100 iterations, where in each iteration, the training and testing of the ANN are performed using events that are detected by two separate stations that are 101 m apart. Although the single event classification rate was reduced when testing with independent events, the simpler TDHA approach again performs at least as well as the spectral features. The advantages in applying decision fusion in this more realistic scenario is more distinct since the improvements in the classification rate are at least 7% compared to only 2% when the ANN is trained using randomly selected events from the two sensor stations.

Table VI illustrates the percentage of misclassification for each vehicle utilizing the TDHA method. It is clear that if a vehicle is misclassified, most probably it will be classified as a vehicle with similar profile regarding the track or the weight. The only exception is the tracked lightweight Wiesel, which is evenly misclassified between tracked and wheeled vehicles because of its small engine and lightweight nature compared to the other tracked vehicles. Note that the Wiesel was also dropped from the simulations in [7] because its harmonic structure could not be evaluated.

Classification Results Using Separate Runs for Training and Testing: In this setting, the ANN is trained using events detected by both stations from half of the recorded runs and is tested using events detected by the other half of the runs (and vice versa for cross-validation). For the TDHA approach with a $\tilde{F}\tilde{F} \sim 5$ Hz, the average single event classification rate is 82.77%. When temporal decision fusion is utilized for each
run, the classification rate per run reaches 89.67%. These results are virtually identical to those shown in Table V. This was true for other cases as well.

VI. DISCUSSION AND CONCLUSION

The TDHA method combines the simplicity in estimating the strongest harmonic component with the straightforward estimation of the harmonics’ amplitudes using the LS method applied only on a reduced number of samples in generating the characteristic features. It must be noted that the TDHA method does not intend to estimate the original FF of the acoustic signal, but it intends to best fit a simple harmonic model, built upon estimating the strongest harmonic component, with a defined resolution to a template extracted from the acquired acoustic signal.

In this paper, we have shown the utility of a simple extraction of selective time domain features from the acoustic emission of harmonic signals for single target classification. The proposed TDHA method enables sensor nodes to carry out the detection, extraction, and classification on-board. The proposed algorithm does not require certain configurations of the sensor nodes or the vehicles to be at a certain range from the sensor node or moving with a certain speed. The main idea was based on the validity of the coupled harmonic signal model. Simulation results show that the selection of the $\tilde{\text{FF}}$ for constructing an approximation for the coupled harmonic model, fixing the number of harmonics, and employing a FNN for classification provides a simple and energy-efficient approach with a high classification rate comparable to or better than the spectral features for the same set of detected events. Decision fusion using temporal decisions represents an efficient solution for increasing the reliability of single target classification.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Altmann at Experimentelle Physik III, Technische Universität Dortmund, Germany, and the BVP for providing the acoustic signatures of the vehicle, the anonymous reviewers for many helpful suggestions, and the Department of Electrical Engineering at the University of Nebraska-Lincoln, Lincoln.

REFERENCES


Peter E. William (S’00–M’10) received the B.S. and M.S. degrees in electrical engineering from Assiut University, Assiut, Egypt, in 2001 and 2004, respectively. He is currently working toward the Ph.D. degree in electrical engineering at the University of Nebraska-Lincoln, Lincoln.

His research interests include statistical signal processing, feature extraction, pattern recognition, and machine learning. He has worked on developing low-power feature extraction algorithms for smart sensors for signal identification. Other research interests include data compression, multimodal signal analysis, and spatiotemporal decision fusion.

Michael W. Hoffman (S’84–M’87) received the B.S. degree in electrical engineering from Rice University, Houston, TX, the M.S. degree in electrical engineering from the University of Southern California, Los Angeles, and the Ph.D. degree in electrical engineering from the University of Minnesota, Twin Cities.

From 1985 to 1988, he was a Signal Processing System Engineer in the Space Communications Division, TRW, Inc. In 1993, he joined the University of Nebraska, Lincoln, where he is currently a Professor. His research interests include joint source channel coding, sensor array processing, and low-power mixed-signal processing.