

EXPLORING SOUND SIGNATURE FOR VEHICLE DETECTION AND CLASSIFICATION USING ANN

Jobin George¹, Anila Cyril², Bino I. Koshy³ and Leena Mary⁴

¹Department of EC Engineering, Rajiv Gandhi Institute of Technology, Kottayam, India
jobing4@gmail.com

^{2,3}Department of Civil Engineering, Rajiv Gandhi Institute of Technology, Kottayam, India

anilacyril1110@gmail.com, bino@rit.ac.in

⁴Department of CS and Engineering, Rajiv Gandhi Institute of Technology, Kottayam, India

leena.mary@rit.ac.in

ABSTRACT

This paper attempts to explore the possibility of using sound signatures for vehicle detection and classification purposes. Sound emitted by vehicles are captured for a two lane undivided road carrying moderate traffic. Simultaneous arrival of different types vehicles, overtaking at the study location, sound of horns, random but identifiable back ground noises, continuous high energy noises on the back ground are the different challenges encountered in the data collection. Different features were explored out of which smoothed log energy was found to be useful for automatic vehicle detection by locating peaks. Mel-frequency cepstral coefficients extracted from fixed regions around the detected peaks along with the manual vehicle labels are utilised to train an Artificial Neural Network (ANN). The classifier for four broad classes heavy, medium, light and horns was trained. The ANN classifier developed was able to predict categories well.

KEYWORDS

Traffic characterises, vehicle detection, multilayer feedforward neural network, vehicle classification.

1. INTRODUCTION

Vehicles produce sound when they move. This include engine noise, noise due to tyre - pavement interaction, body rattling, vibrations, horns etc. The sound level also depends on the characteristics of the road and operating conditions. Under same type of operating conditions, similar types of vehicle may emit similar sound, thus suggesting a specific sound signature for each type of vehicle. Vehicle detection and classification have been attempted using many ways. Image processing techniques are used to classify the vehicles under the real time traffic management and for Intelligent Transportation Systems (ITS). Inductive loops based systems are widely used for determination of vehicle counts.

Vehicle detection while it is in motion is a prerequisite for traffic and speed management, classified vehicle count, traffic signal time optimization, gap/ headway measurement and military purposes. Vehicle detection techniques are in a stage of continuous evolution. Traditional methods use intrusive sensors such as inductive loops [1], magnetometers, microloop probes, pneumatic road tubes and piezoelectric cables [2]. Unattended ground magnetometer sensor [3]

was found to be effective to detect the vehicle passing. Since these intrusive methods decrease the pavement life and affect traffic during installation and maintenance, it has been replaced by non intrusive techniques such as video image processing, microwave radar, laser radar, passive infrared, ultrasonic, passive acoustic array and combinations of sensor technologies such as passive infrared and microwave doppler or passive infrared and ultrasonic [2].

Researches on the application of image processing techniques for vehicle detection have proved it to be a perfect substitute for conventional methods. Now-a-days, video monitoring is widely used for signal control and traffic management. Traffic video processed using unsupervised vehicle detection and spatio-temporal tracking is effective in vehicle tracking in complex urban situations [4]. In temporal difference based image processing (also known as Consecutive image subtraction techniques), differences in consecutive image is analysed to detect a moving vehicle and its speed [5]. Gray scale images that are frame differences of a video were effective for detecting vehicles. Using edge detection, resolution reduction and modified morphological transformation techniques when incorporated into processing techniques along with Hough transforms and fuzzy integrals, it was possible to obtain a good vehicle detection system [6].

Processing videos for classification of vehicle detection and classification is computationally intensive and time consuming. Vehicle classification using sound signal have been attempted by researchers previously. Munich [7] compared conventional methods used for speaker recognition, namely, systems based on Mel-frequency Cepstral Coefficients (MFCC) and either Gaussian Mixture Models (GMM) or Hidden Markov Models (HMM), with Bayesian subspace method based on the Short Term Fourier Transform (STFT) of the vehicles' acoustic signature. He found that a probabilistic subspace classifier outperformed conventional MFCC-GMM- and MFCC-HMM-based systems by 50%. Two sets of features were fused for better vehicle classification [8]. One being harmonic components, mainly characterizing engine noise and second is a group of key frequency components, designated to reflect other minor but also important acoustic factors, such as tire friction noise. Fusing these two sets of features provides a more complete description of vehicles' acoustic signatures, and reduces the limitation of relying one particular feature set [8]. Bhave and Rao [9] reported an investigation of acoustic features relating to vehicular traffic in relation to type of vehicle and state of motion in monitoring traffic congestion. Different vehicles, broadly classified into two, three wheelers and heavy vehicle were studied for their acoustic signatures. A source filter model of engine sound was used to derive suitable features. The performance of formant based features is compared with that of Mel-Frequency Cepstral Coefficients (MFCC) via a k-NN classifier on a manually labelled database of traffic sounds.

In this work, we use smoothed logarithmic energy for automatically detecting vehicle by a peak picking algorithm. Spectral characteristics of the sound signature represented using MFCC for vehicles are used to train a classifier for three broad categories of vehicles. While testing, MFCC extracted from automatically detected peaks are applied to the classifier to get the class labels. Remaining part of the paper is organized as follows. Section 2 describes data collection procedure and analysis of the collected data. In Section 3, methodology used for vehicle detection and classification are discussed. Experimental details and results are discussed in Section 4. Section 5 summarizes the study in this paper.

2. DATA COLLECTION AND ANALYSIS

In this section the data collection and analysis is elaborated. Roadside recording was performed on typical working days during several traffic conditions. The initial task pertained to sound recordings without background noise to eliminate the need for filtering. Simultaneous video and sound recordings were performed by Sony Handycam and Zoom H4n audio recorders. Stereo

recordings using two onboard microphones were used for the study. The recordings were about 15-30 minutes continuous duration at a distance of 1.5m from pavement edge. The sites were selected to capture the sound of moving vehicles without much acceleration or gear change. Also the recordings were attempted without capturing undue attention from motorists on clear sunny working days of the week.

2.1. Analysis of Vehicle Data

Collected vehicle data was analysed and studied using wavesurfer. Wavesurfer is a free software tool designed for analysis and labelling of speech data. Here wavesurfer is used for labelling and observing the characteristics of the recorded vehicle data as illustrated in Fig 1.

The vehicle sound recordings were manually labelled for identification of broad categories of vehicles like light vehicles, medium vehicles and heavy vehicles as shown in Fig. 1(a). The light vehicles included two and three wheelers. The medium type vehicles included cars, SUVs and light trucks and the heavy vehicles included buses and trucks. Discernable incidents like horn sound were also labelled. The noise of wind or accidental movement of audio recorder was not labelled as an incident. Labelling was also done with the specific class of vehicles as *hep*, *mep* and *lp* for heavy, medium, and light vehicles respectively as shown in Fig. 1(b). Horns were labelled as *hop*.

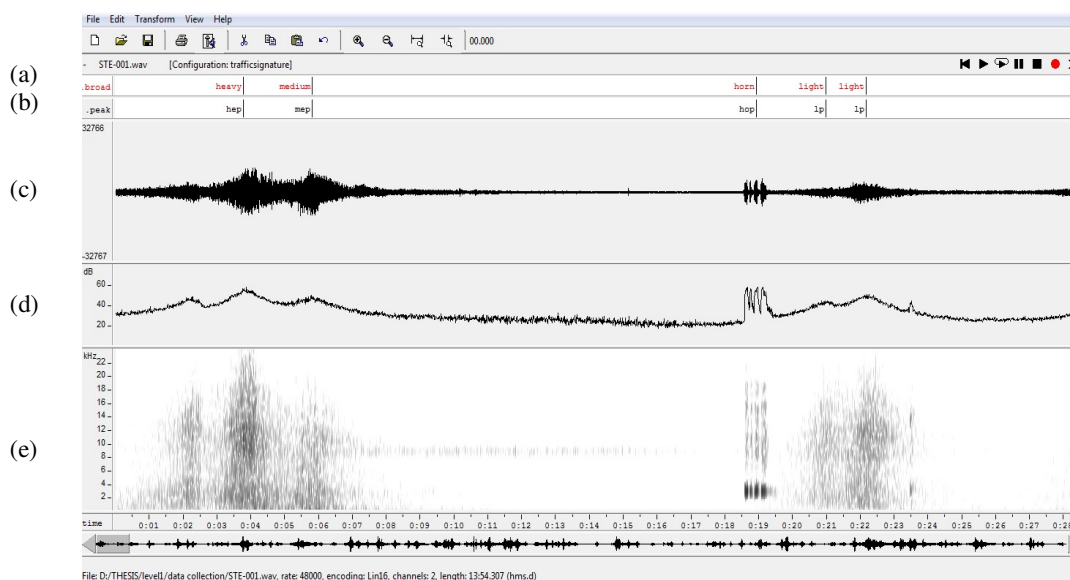


Fig. 1. Visualisation of collected vehicle sound data using wavesurfer showing (a)& (b) broad category labels (c) sound wave form (d) energy contour in dB (e) wideband spectrogram.

Analysing the energy contour shown in Fig. 1(d), it was observed that corresponding to passing of each vehicle, there is an increase in energy. The maximum energy is recorded at the instant vehicle crosses the audio recorder. The wideband spectrogram as in Fig. 1(e) shows presence of high energy, spread at different range of the audio spectrum (0-20KHz). Spectrum corresponding to horn shows a distinct periodicity in energy distribution as shown in Fig. 1(e). The analysis of vehicle audio recordings indicated that detection of energy peaks may give an approximate count of vehicles passing the recorder at a particular point.

2.2. Features for Vehicle Detection

A preliminary study was conducted to identify discriminative features for vehicle detection. Various features explored include the following: (i) short time energy (ii) log energy (iii) smoothed log energy.

(i) Energy

One of the basic short-time analysis functions useful for vehicle sound signals is the short-time energy. The short-time energy, E_n is defined as

$$E_n = \sum_{m=1}^L (x[m]w[n-m])^2 = \sum_{m=1}^L x^2[m]w^2[n-m] \quad (1)$$

Where L is the number of samples of the signal, $w[n-m]$ represents a time shifted window sequence, whose purpose is to select a segment of the sequence $x[m]$ in the neighbourhoods of sample $m = n$. As there are different forms for energy representation, we explored different ways for calculating energy of the sound signal. Fig. 2 (a) and (b) show a sound waveform and corresponding short time energy. It can be observed that the energy is relatively high for the vehicle occurrence compared to the low energy regions corresponding to noises of the background. But it consists of high frequency variations or ripples which may not be suited for vehicle peak detection.

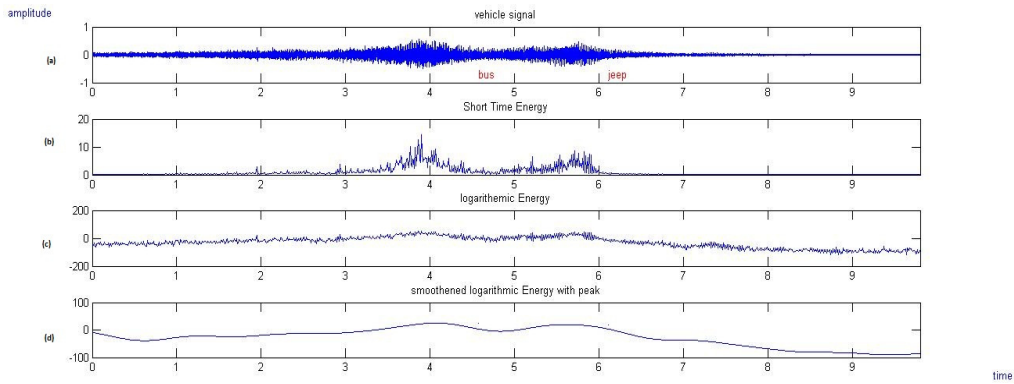


Fig. 2. Features explored for vehicle detection (a) sound waveform with labels (b) short time energy (c) logarithmic energy in dB (d) smoothed logarithmic energy in dB

(ii) Logarithmic energy

In order to reduce high frequency variations and dynamic range of short term energy, the logarithmic energy (in dB) is computed. As shown in Fig 2(c) log energy also contains high frequency ripples/ components. Logarithmic energy in dB is found using $20 \cdot \log(\text{energy})$.

(iii) Smoothed logarithmic energy

Logarithmic energy in dB shown in Fig. 2(c) contains high frequency ripples/components which may create problems while detecting the presence of vehicles using peak detection. Hence high frequency components are removed using appropriately designed low pass filter. An elliptic low pass filter with $R_p = 3$; $R_s = 60$; $W_p = (1000/\text{fs})$; $W_s = (2000/\text{fs})$ is used for this where fs is the sampling rate (48000 samples/second), R_p is the pass band ripple factor, R_s is the stop band ripple factor, W_p is the pass band frequency and W_s is the stop band frequency. Fig 2(d) shows the smoothed logarithmic energy after removing the high frequency components from 2(c).

3. METHODOLOGY FOR VEHICLE DETECTION AND CLASSIFICATION

3.1. Vehicle detection

The peak finding algorithm for vehicle detection is as follows.

1. Start with inspecting every 10 ms window of energy contour in dB, find peak for each window, and inspect the values of neighboring samples (previous 40th sample and next 40th sample in our studies).
2. If both neighbouring point have very low value compared to the peak then identify the considered peak as the potential peak, otherwise discard the peak.
3. Now move to next 10 ms of energy contour and repeat the procedure.
4. Plot the identified potential vehicle peaks.

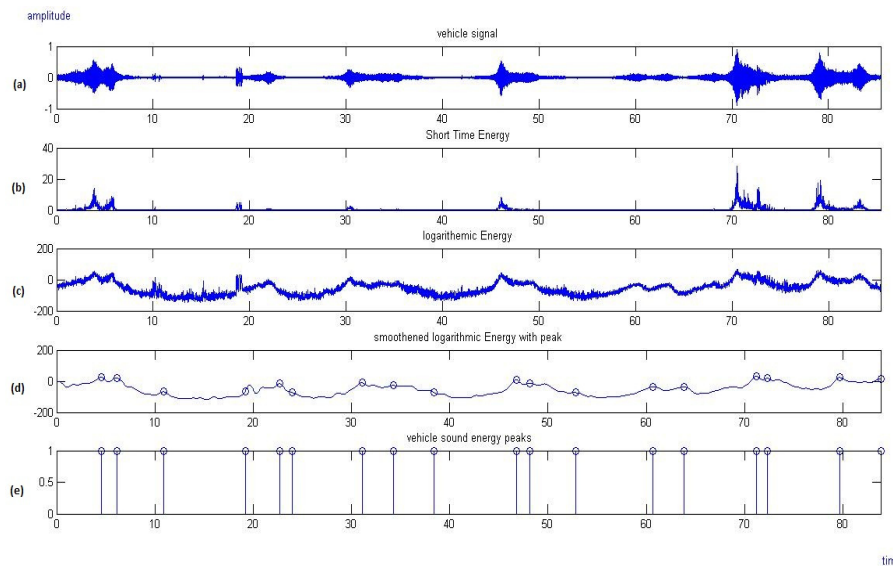


Fig. 3. Methodology for vehicle detection (a) sound waveform (b) short time energy (c) logarithmic energy in dB (d) smoothed logarithmic energy in dB with peaks marked (e) Detected peaks corresponding to vehicles.

3.2. Algorithm for vehicle detection and classification

1. Divide 75% of the collected data for training and remaining for testing.
2. Do manual labelling for training and testing data as illustrated in Fig 1.
3. For each manually labelled peak compute 30 dimensional MFCC for 20ms region around the peak. Perform cepstral mean subtraction for MFCC values. Normalise MFCC value to the range between (0,1).
4. Initialise appropriate MLFFNN structure with random weights.
5. Train MLFFNN using MFCC as input and vehicle label as output using back propagation training.
6. Testing of trained MLFFNN is done as shown in Fig 4. The various steps are:
 - i) Compute the log energy of test data
 - ii) Smooth the log energy using appropriate low pass filter.
 - iii) Detect peaks corresponding to vehicles using peak detection algorithm.
 - iv) Extract MFCC for the region around the detected peak.

- v) Perform cepstral mean subtraction to remove the influence of unwanted background noise
- vi) Normalize the MFCC values to limit its range between (0,1)
- vii) Apply this normalized MFCC as input to trained MLFFNN and find out the output labels.
- viii) Compute the classification accuracy using the manual labels of test data

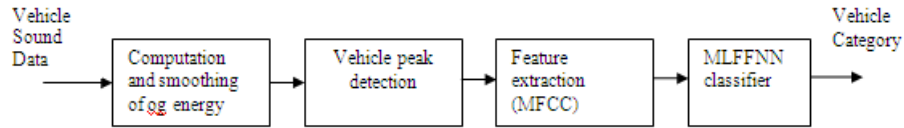


Fig 4. Block schematic illustrating various stages during the test phase of vehicle detection and classification.

4. EXPERIMENTAL RESULTS

Vehicle sound recordings of about 120 minutes are used for demonstrating the effectiveness of the proposed algorithm. Recording for 90 minutes duration is used for training and another recording of duration 30 minutes is used for testing. In the proposed method, MFCC features are used for training and testing purpose. Extracted features and labels along with is used for training the MLFFNN with structure 13L 35N 10N 4N where L denotes neurons with linear activation function, N denotes neurons with nonlinear (sigmoid in this case) activation function, and the numerals denote the number of neurons in each layer. This MLFFNN classifier is then tested using MFCC features derived from regions around automatically detected vehicle peaks in the test data. The classification accuracy is computed with the manual labelling of test data. Results are as shown in table 1. An overall classification of approximately 67% was obtained using this method. Confusion matrix corresponding to test data 1 and 2 are given in Table 2 and 3 respectively, which indicate that maximum confusion has occurred in case of light and medium category of vehicles which is obvious.

Table 1. Performance of proposed vehicle detection and classification algorithm

Particulars of test data	Vehicle classification accuracy (%)
Test data 1	67.4%
Test data 2	66%

Table 2. Confusion Matrix for Vehicle Classes (Data set- No.1)

Model \ Test	Heavy	Medium	Light	Horn
Heavy	7	1	2	0
Medium	3	16	11	2
Light	2	5	34	0
Horn	2	3	0	7

Table 3. Confusion Matrix for Vehicle Classes (Data set- No.2)

Model Test	Heavy	Medium	Light	Horn
Heavy	5	1	2	0
Medium	1	14	4	3
Light	4	10	34	0
Horn	3	1	2	6

5. CONCLUSIONS

The present study has shown the potential of low pass filtered log energy being used for automatic vehicle detection by locating peaks. The ANN classifier developed was able to predict different categories of vehicles. The prediction success shows considerable confusion between the classes of medium and light category of vehicles where similar types of engines are used. The study has not addressed multilane traffic and its effect on the sound sampling due to simultaneous arrivals of small vehicles, overtaking of different classes of vehicles. The effect of acceleration and gear shift on classification need further exploration.

Acknowledgements

The authors would like to thank All India Council of Technical Education (AICTE), India for supporting this study under the Research Promotion Scheme (RPS).

REFERENCES

- [1] Janusz Gajda, Ryszard Sroka, Marek Stencel, Andrzej Wajda & Tadeusz Zeglen, (2001) "A Vehicle Classification Based on Inductive Loop Detectors", *IEEE Instrumentation and Measurement Technology Conference*, pp 460-464.
- [2] Luz Elena Y. Mimbela & Lawrence A. Klein, (2000) *A Summary of Vehicle Detection and Surveillance Technologies used in Intelligent Transportation System*, FHWA.
- [3] Carol T. Christou & Garry M. Jacyna, (2010) "Vehicle Detection and Localization Using Unattended Ground Magnetometer Sensors", *International Conference on Information Fusion* pp 1-8.
- [4] Chengcui Zhang, Shu-Ching Chen, Mei-Ling Shyu & Srinivas Peeta, (2003) "Adaptive Background Learning for Vehicle Detection and Spatio-Temporal Tracking", *International Conference on Information, Communications and Signal Processing (ICICS)*, 2A3.7, pp 1-5.
- [5] L.G.C Wimalaratna & D.U.J. Sonnadara, (2008) "Estimation of the Speeds of Moving Vehicles from Video Sequences", *Proceedings of the Technical Sessions*, 24, Institute of Physics – Sri Lanka, pp 6-12.
- [6] Xiaobo Li, Zhi-Qiang Liu & Ka-Ming Leung, (2002) "Detection of vehicles from traffic scenes using fuzzy integrals", *Pattern Recognition* 35, pp 967–980.
- [7] Mario Munich (2004) "Bayesian Subspace Methods for Acoustic Signature Recognition of Vehicles", *Proceedings of 12th European Signal Processing Conference (EUSIPCO 2004)*, pp 2107-2110.
- [8] Baofeng Guo, Mark Nixon & Thyagaraju Damrala, (2011) "Improving acoustic vehicle classification by information fusion", *Pattern Analysis and Applications*, Volume 15, Issue 1, pp 29-43.
- [9] Nikhil Bhave & Preeti Rao, (2011) "Vehicle Engine Sound Analysis Applied To Traffic Congestion Estimation", *Proceedings of International Symposium on Computer Music Modeling and Retrieval (CMMR) and Frontiers of Research on Speech and Music (FRSM)*.

Authors

Jobin George graduated from Kannur University in Electronics and Communications Engineering in 2010. He is currently doing masters degree in Advanced Communication Engineering and Information Systems at Rajiv Gandhi Institute of Technology, Kottayam, Kerala, India. His area of interest is signal processing and neural network



Anila Cyril graduated in Civil Engineering from College of Engineering, Trichur, Kerala in 2011. She is currently perusing masters degree in Transportation Engineering at Rajiv Gandhi Institute of Technology, Kottayam, Kerala, India. Her area of interest is traffic simulation and modelling.



Bino I. Koshy received his Bachelors degree from Kerala University in 1985. He obtained his M.Tech and Ph.D. from Indian Institute of Technology, Madras, India in 1987 and 2007 respectively. Currently he is working as Professor in Civil Engineering, Rajiv Gandhi Institute of Technology, Kottayam, Kerala, India. His research interests are travel demand modelling, traffic modelling and artificial neural networks. He is a life member of Indian Society for Technical Education and Institute of Urban Transport (India). He is also a Fellow of Institution of Engineers (India).



Leena Mary received her Bachelors degree from Mangalore University in 1988. She obtained her M.Tech from Kerala University and Ph.D. from Indian Institute of Technology, Madras, India. She has 22 years of teaching experience. Currently she is working as Professor in Computer Science and Engineering at Rajiv Gandhi Institute of Technology, Kottayam, Kerala, India. Her research interests are speech processing, speaker forensics, signal processing and neural networks. She has published a book on Extraction and Representation of Prosody for Speaker, Speech and Language Recognition by Springer. She is a member of IEEE and a life member of Indian Society for Technical Education.

