VEHICLE ENGINE SOUND ANALYSIS APPLIED TO TRAFFIC CONGESTION ESTIMATION

Nikhil Bhave and Preeti Rao
Department of Electrical Engineering,
Indian Institute of Bombay, Mumbai-76, India
Email: nikhilbhavev@gmail.com, prao@ee.iitb.ac.in

Abstract

An investigation of acoustic features relating to vehicular traffic on roadways is reported. Computable features that relate to the type of vehicle and state of motion can be useful in monitoring traffic congestion. In the present work, different vehicles, broadly classified into two, three wheelers and heavy vehicle, are studied for their acoustic signatures. A source filter model of engine sound is 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.

Key words - Vehicle classification, acoustic sensing, vehicle activity detection, vehicle source filter model.

1. Introduction

The often annoying and unpleasant sound produced by a moving vehicle may be usefully applied order to help in reducing chaotic road conditions especially on major roads in mega cities. There have been attempts involving processing sound aimed at detecting the congestion state of traffic so that prospective road users can be alerted in time and helped with identifying less congested alternate routes. This is an important goal of many futuristic intelligent transport systems. Various non-acoustic based congestion detection of traffic has been attempted in the past including camera sensing with image processing. Acoustics based solutions traditionally used the honking sound of the vehicles to estimate its speed and hence the state of traffic [1]. This solution was based on exploiting the Doppler Effect. The honking sound was captured using two sensors placed with certain distance between them. The frequencies of a single honk recorded simultaneously by the two sensors were used in estimating the speed of vehicle.

While the presence of honking can be used to indicate the condition of traffic, it is not reliable due to the possibility of restrictions on honking in certain zones. We consider other non-honk acoustics of the vehicle to look for novel solutions to automatic detection of the state of the traffic. Vehicles moving on the road create sounds that are distinctive of the state of motion as well as vehicle type, often easily identifiable. Acoustic features extracted from traffic noise thus have the potential of providing important clues to traffic conditions. From the knowledge of human sound perception, such acoustic cues are expected to be prominent in the short-time magnitude frequency spectrum of the vehicle sound.

Spectral domain representations have been explored in past work on vehicle recognition [2]. The spectrum of the vehicle sound however is influenced by the structure of the vehicle body as well as its state of motion (static, moving, accelerating, etc.). A wavelet based method [3], uses a sixth order spline wavelet transform to detect arrival of vehicles of arbitrary type when other noises are present by analysis of acoustic signatures against an existing database of recorded and processed acoustic signals. To ensure least possible false matching, a training database is constructed using the distribution of energies among blocks of wavelet packet coefficients with a procedure for random search for a near-optimal footprint. Among spectral based methods, [4] extracts the fundamental frequency associated with the engine of the vehicle, and establishes a relation between the fundamental frequency, number of cylinders of the engines and their RPM. It identifies the role of this fundamental frequency and the harmonic spectral structure in classification of vehicles. Vehicle classification algorithm using back propagation described in [5] is another spectral domain based method. A one third-octave filter bands is used for getting the important signatures from the emanated noise. The extracted features are associated with the type and distance of the moving vehicle. A fusion of harmonic features which correspond to engine noise and key features which correspond to tires friction noise have also been used in vehicle identification [6].

In this paper we propose to use the parameters of a source filter model for vehicle detection. The system is described and its performance evaluated on an available labelled database.

2. Proposed source-filter modelling

The sources of sound in a vehicle are many like bumping and friction of vehicle tyres with ground, wind effects, etc., but the primary contributor is its engine. Hence a basic knowledge about the engine mechanics is helpful in order to understand the nature of sound generated. There are three basic process that are occur periodically viz. compression, combustion and exhaust. The combustion and exhaust give rise to periodic sound bursts. Later this impulsive and periodic sound source that is generated at engine due to combustion, and more so because of exhaust that is blown out at high pressure goes through one or more tube like structures like exhaust pipe. The sound, so filtered by the components corresponding to the body parts of vehicle and the surroundings, eventually reaches the sensor as depicted in Figure 1.

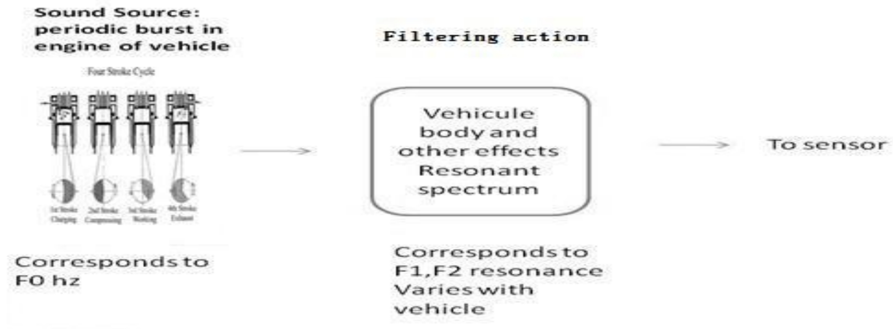


Figure 1. Proposed source-filter model

The estimation of parameters associated with the vehicle source filter model (VSFM) can help to detect the type of vehicle and its state of motion. The fundamental frequency of vibration of F0 is the parameter associated with the source and it corresponds to the revving of the engine, it does depends on the type of the engine of the vehicle and number of cylinder [4], thus it gives an idea about the state of the motion of the vehicle. The resonant frequency due to the tube structures in vehicles corresponds to the formants which represent the filter characteristics. They vary with vehicle type as confirmed by a study of spectrograms of different vehicles. Hence the first two formants are considered as features for vehicle type classification. An alternate set of features, Mel-frequency Cepstral Coefficients (MFCC) are also investigated. MFCCs offer a compact representation of the spectral envelope of the sound and form the basis of speech recognition systems. The performance of MFCC features can form a benchmark for the proposed source-filter model features of this work.

3. Traffic noise database

Several road side recording were performed at different roads and during different conditions (morning, noon and evening). The recordings were of about 15-30 min duration. The segments in these recording were clipped such that the clip had initial background noise followed by a recognizable vehicle sound in foreground. This extraction was done manually by listening to the recordings and viewing the waveform and spectrogram. The extracted clips were labelled to belong to particular class among two wheeler, three-wheeler and heavy vehicle. The data set collected had 300 sound clips sampled at 5 kHz, 100 of each type with clip duration varying from 2 to 8 seconds ensuring they had predominant sound of the vehicle. The state of the motion for vehicles varies across the clip.

4. Vehicle classification

It forms a typical case of pattern recognition wherein k-nearest neighbour based algorithm was used for classification. A pre-processing to enhance spectral and temporal features against the background noise was first implemented followed by vehicle classification.

4.1 Pre-processing

Spectral subtraction [7] was implemented on the traffic data to make the target clip free of background traffic hum so that the spectrum primarily carries the enhanced spectral signature of the vehicle in the foreground. The signal sampled at 5 kHz is assumed to comprise of sound from vehicle in foreground and the overall noise due to the traffic in background. These sounds are assumed to be statistically independent. The signal was first windowed using Hamming window of 24 ms duration with an overlap of 12ms. The power spectrum was obtained using 128-point FFT. The frame was then classified into either noisy frame or frame with vehicle signature based on the spectral variance. The weighted estimate of noise was then subtracted from the power spectrum of the frame. The noise spectrum is estimated and its values are updated after every next detected noise frame .The initial estimate of noise is obtained from the first few frames assumed to contain only the background. With every noise frame detected the statistics of noise are updated. The 'α' depends on SNR of the signal [7] while spectral floor factor 'β' is kept constant at 0.0005.The result of the processing of a two wheeler clip are shown in Figures 2 and 3.

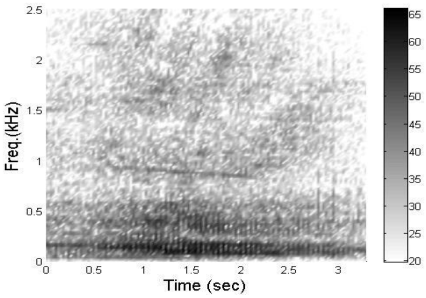


Figure 2. Spectrogram of a two wheeler sound with noise

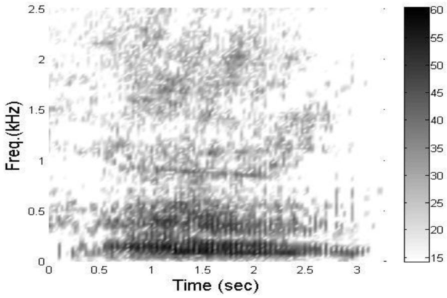


Figure 3. Spectrogram after pre-processing

4.2 Vehicle detection

The pre-processed traffic audio signal is windowed with 180 ms duration Hamming window. For each windowed segment a feature vector is extracted. This feature vector and its class information from all the data clip forms the input to k-NN algorithm. The segment of the test clip is classified based on the 5 nearest neighbours, among the other pre-labelled training data set as two, three wheeler or heavy vehicle. A clip was recognised to consist of a given class based on the mode of the classes' presents in it. The following two types of feature vector were evaluated.

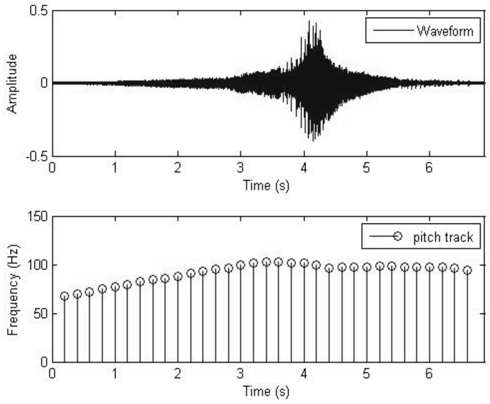
Formant features: For each window formants are extracted using LPC analysis [7]. In the extraction and tracking of the formants the filter in VSFM is modelled using an all pole filter whose order is chosen as 2+(Fs/1000), spectral observation showed existence of at least one resonant frequency per 1000 Hz , here the order is 7. Only the first two prominent formants form the feature vector to classify the segment.

MFCC features: MFCC's have become a standard in pattern recognition problems primarily in acoustics since they relate to perceptual dimensions of the signal. In the use of these as feature vector for classification of vehicle, for each windowed segment 1024 point FFT is obtained with appropriate zero padding. The first 512 power spectrum coefficients are Mel scale filtered. Mel filters are a set of 24 triangular band pass filters, spanning the 0-2500 Hz region, with their positions equally spaced along the Mel frequency axis [7]. Amplitude of the spectrum resulting from application of discrete cosine transform on the log Mel-band powers gives the MFCC coefficients. First 20 of these coefficients form the feature vector.

5. Experimental results

5.1 Tracking of F0

Autocorrelation based method [7] generally used to extract periodicity of a signal or pitch in case of speech signal was used to extract the F0 values. It was observed that for accelerating vehicles F0 steadily increases and decreases for decelerating vehicle while it remains constant for vehicle moving at near constant speed.



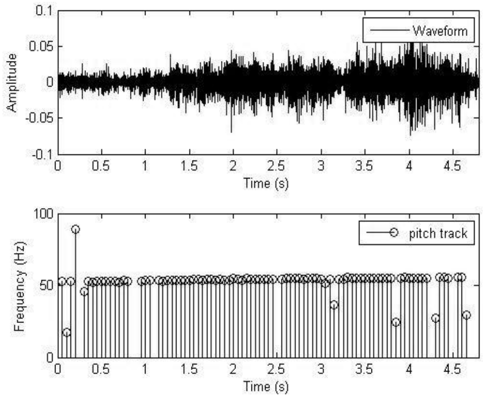


Figure 3. F0 tracking for initially accelerating vehicle

Figure 5. F0 tracking for vehicle moving at constant speed

5.2 Results of vehicle detection

The extracted formants F1 and F2 from each segments of all the clips forms the feature vector in first method while MFCC formed the feature vector in the alternate method, segments of these clips are labelled to belong to one of the three classes. The k-NN algorithm with k=5, uses this labelled training data information to recognise the type of vehicle in a given traffic clip. For testing purpose one of the clips from the data set is used as test clip while others form the training set. Each frame of the clip is classified into one of the three types. The clip itself is then classified based on the majority class of its frames. The percentage of clips classified correctly is tabulated below. The classification matrices for both the methods are provided in Tables 3 and 4.

Vehicle type	F1 (Hz)		F2 (Hz)	
	Avg.	Std. dev.	Avg.	Std.dev.
Three wheeler	584	232	1604	301
Heavy Vehicle	281	106	1084	271
Two wheeler	351	191	1286	318

Table 1. Formants F1, F2 (average and std. dev.) for different vehicles

Vehicle type	Total no. of clips	No. of clips recognised correctly	
		Formants based method	MFCC based method
Two Wheeler	100	92	96
Three Wheeler	100	69	80
Heavy Vehicle	100	50	68

Table 2. Classification accuracy for formant and MFCC based method

6. Conclusion

The classification matrix for MFCC based method show that the two and three wheelers where recognised more accurately while the low percentage of accuracy of detection of heavy vehicle is because many of them were detected to be of two wheeler type. This is because of similar spectrographic structure of few heavy vehicles and two wheelers. The low accuracy in detecting

Classified / Actual	Two-wheeler	Three-wheeler	Heavy vehicle
Two wheeler	92	3	5
Three-wheeler	24	69	7
Heavy vehicle	41	9	50

Table 3. Confusion matrix for formant based method

Classified / Actual	Two-wheeler	Three-wheeler	Heavy vehicle
Two-wheeler	96	1	3
Three-wheeler	17	80	3
Heavy vehicle	30	2	68

Table 4. Confusion matrix for MFCC based method

heavy vehicle could be attributed to the fact that it included in its cluster sound from various type of vehicle which includes cars, tempo, tractors and buses. Many or often all two wheelers do have good silencers also heavy vehicles like cars are designed by manufactures to have low noise performance, this curbing of sound emitted leads in similar spectral signatures. This led to misclassification of many heavy vehicles as two wheelers.

Future works should be rendered in the direction wherein we identify a feature which along with MFCC's distinguishes between these types of vehicle more accurately, further improvements can be brought in by creating more classes or sub classes among heavy vehicles.

7. Acknowledgements

We would like to thank Rijurekha Sen for motivating the problem, and Rijurekha Sen with Arpita Agrawal for help with database creation.

References

[1] R. Sen, B. Raman and P. Sharma, "Horn-Ok-Please", The 8th Annual International Conference on Mobile Systems, Applications and Services, Mobisys'10, pp. 15-18, 2010.

[2] H. Wu, M. Siegel and P. Khosla "Vehicle sound signature recognition by frequency vector principal component analysis", IEEE Trans. Instrum. Meas., vol. 48, pp. 1005-1009, 1999.

[3] A. Averbuch, V. Zheludev, N. Rabin, Schclar, "Wavelet based acoustic detection of moving vehicles", Multidimens. Syst. Signal Process. Vol. 20, pp. 55–80, 2009.

[4] S. Kozhisseri and M. Bikdash, "Spectral Features for the Classification of Civilian Vehicles using Acoustic Sensors", in proceedings of the IEEE workshop on Computational Intelligence in Vehicles and Vehicular Systems, pp. 93-100, 2009.

[5] N. A. Rahim, Paulraj M P, Abdul H.A, S. Sundararaj, "Moving Vehicle Noise Classification using Back propagation Algorithm", in proceedings of the 6th International Colloquium on Signal Processing & Its Applications (CSPA), pp 1-6, 2010.

[6] B. Guo, M. S. Nixon, and T. R. Damarla, "Acoustic information fusion for ground vehicle classification", in proceedings of the 11th International Conference of Information Fusion, Cologne, Germany, pp. 1-7, 2008.

[7] X. Huang, et.al. "Spoken Languge Processing," Princeton Hall, 2001.

[8] J. J. Wei, Q. Xuan , X. Han , Y. Zhong-hu, "An Approach of Automatic Vehicle Classification by Acoustic Wave based on PCA-RBF", in proceedings of the 6th international conference on Information Engineering and Computer Science, pp 1-4, 2009.

[9] J. J. Wei, Q. Xuan , X. Han , Y. Zhong-hu, "An Approach of Passive Vehicle Type Recognition by Acoustic Signal Based on SVM", in proceedings of the 3rd International Conference on Genetic and Evolutionary Computing, pp 545-548, 2009.