Vehicular Mechanical Condition Determination Using Acoustic Signal

Ms. Minal Bhandarkar[1] ,Mrs.Tejashri Waykole[2]

[1,2]*Department Of Electronics And Telecommunication,G.S.Moze College of Engineering, Balewadi ,Pune,University of Pune,Pune, India*

Abstract: *In this paper we will determine the mechanical condition of vehicle by using acoustic or sound signals. Here we will determine wheather mechanical condition of vehicle is good or bad. As well as we can estimate the traffic density on road also. Cumulative sound signals consist of various noises coming from various part of vehicles which includes rotational parts, vibrations in the engine, friction between the tires and the road, exhausted parts of vehicles, gears, etc. Noise signals are tire noise, engine noise, engine-idling noise, occasional honks, and air turbulence noise of multiple vehicles. These noise signals contains spectral content which are different from each other, therefore we can determine the mechanical condition of vehicle. For example, under a good condition of vehicle, the vehicles typically having the smooth audio and very less engine ideling noise. Here we uses SVM and ANN classifiers.*

Index terms: *Artifitial Neural-Network, Support Vector Machine, Signal processing, Noise signal recognition.*

I.   Introduction

Now a daysDensity of traffic on roads and highways hasbeen increasing constantly due to motorization, urbanization, and population growth. Usually Traffic congestion reduces the efficiency of the transportation infrastructure of a city; it also increases travel time, fuel consumption, and air pollution; and leads to increased user frustration and fatigue in developing countries like India and South Asia. As number of vehicles are constantly increasing, it becomes necessary to facilitate effective control of traffic flows in urban areas[1,16]. Especially in rush hours, if there is a poor control at traffic signals it may result in a long time trafficjam causing a hours of delays in traffic flows and also CO_2 emission[2].In developed countries intelligent transportation system (ITS) plays a vital role for traffic congestion where traffic flow should be orderly and lane driven and mostly homogeneous. For this purpose they have used multiple sensors such as magnetic loop detectors, speed guns and video cameras. But the cost of installation and maintenance is very high during their life cycle. Therefore over the past decade, researchers have been developing several traffic monitoring technologies based on laser, ultrasound, radar, video and audio signals. Video image processing is very natural and impressive technique for the traffic. However the computer vision technology is not still adequate to handlevariable and poor lightening condition and visual occlusions. In order to overcome such drawbacks, some researchers provide excellent references which provides robust traffic monitoring techniques using video signals, ranging from vehicle tracking to vehicle occlusion handling.In [3], Cucchiaradetectedvehicles in urban traffic scenes by means of rule-based reasoning on visual data. In [4],Kamijointroduced a hidden Markov model -based computer-vision technique. It detects accidents and other events such as reckless driving at road intersections which further causes accidents. However, these techniques do not directly indicates the problem of average-speed estimation. In [5], Coifman proposed and feature-based computer vision technique for vehicle tracking. He uses some of the "corner" features of the vehicles, which are used to track them and then estimate traffic parameters such as average speed and volume of the vehicles. They provided impressive results on free way traffic. However, it is not clear if this tracking technique could still work in the chaotic and nonlane-driven city traffic conditions and the probable conditions are heavy, medium, and slow traffic.Such traffic conditions are very common in developing countries like India and South Asia and this is the focus of this paper. As the sensors such as magnetic loop detectors[6]are widely used for traffic monitoring in developed countries but theyStill have very high cost of installation and operation and maintenance as well as they required traffic flow to be orderly, lane driven and mostly homogeneous. All these conditions of magnetic loop detector not met at all in developing countries like South Asia and India, where the traffic is highly chaotic, non-lane driven and heterogeneous. Therefore, traffic monitoring is an even severe problem in developing countries, where the ITS systems such as loop detectors and computer-vision-based tracking techniques are not so effective. In such situations, the roadside acoustic signal seems to be a good approach for traffic monitoring due to its inherent low cost of installation and operation. Therefore, several researchers have been developed various traffic monitoring techniques based on audio signals. Doppler frequency shift provides a theoretical description of single vehicle speed. from this description it can be Assume that distance to the closest point of approach is known the solution can accommodate any line of arrival of the vehicle with respect to the microphone. Therefore he got the solution

for speed estimation of single vehicle is more applicable as compared to several vehicles. In presence of several vehicles the interference of noise signals is combined with acoustic waveforms[7], [8].Sensing techniques based on passive sound detection are reported in [9], [10]. These techniques utilizes array of microphones to detect the sound or acoustic waves generated by road side vehicles and are capableof monitoring traffic conditions with lane-by-lane and vehicle-by-vehicle basis in a multilane carriageway. Here he usedCorrelation based algorithm [9], [10], [11],It extractskey data which reflects the road traffic conditions, for exa.the speed and density of vehicles. S. Chen introduces multilane traffic sensing concept based on passive sound which is digitized and further processed by an on-site computerwhich also uses a correlation based algorithm. This system having low cost of installation, maintenance and operation, safe passive detection, immunity to adverse weather conditions, and competitive manufacturing cost. The system performs well for free flow traffic but for congested traffic it is difficult to achieve best performance [11].

Valcarce.introduces the differential time delays to estimate the speed in order to achieve the above target which uses a Pair of omnidirectional microphones. It was used and technique is based on maximum likelihood principle. It directly estimates car speed without any assumptions on the acoustic signal emitted by the vehicle [12]. Mohan uses the smartphone features and basic honk signals which is used to estimate vehicles speed [13]. Simple Doppler frequency shift computations are used to estimate speed. Sen introduces Doppler frequency shift rule in which he had assumed that vehicle is moving in same direction as the straight line connecting vehicle to the microphone [14]. However, In presence of multiple honking vehicles it is not clear that how two microphone distinct honk emitted by same vehicle. The experimental setup covers at two roads, on single lane, one way and three lane bidirectional. Jien Kato introduces a method for traffic density estimation based on recognition of temporal variations that appear on the power signals in accordance with vehicle passthrough a reference point [15]. Here he used Hidden Marko model for observation of local temporal variations over small periods of time which are extracted by wavelet transformation. Experimental results show good accuracy for detection of passage of vehicles.VivekTyagi classified traffic density state as free flowing, Medium flow and Jammed which is on speed of 0-10km/hr, 20-40km/hr and above 40km/hr respectively. They consider short term spectral envelope features of cumulative acoustic signal, and then class conditional probability distribution is modelled on one of the three broad traffic density state (i.e. free flow, medium flow or heavy flow). He uses Experimental setup ofomnidirectional microphone placed at about 1.5 m height and cumulative acoustic signal is recorded at 16000 Hz sampling frequency. He uses a Bayes classifier to classify traffic density state which results in ~ 90% of accuracy, which is then improved by using discriminative classifier such as RBF-SVM [16]. Compare with the existing computer vision and traffic monitoring system in [16], [3] and [4] this technique is independent of light condition and visual occlusion and works well for developing geographies which is explained in techniques of [13], [14].

We begin with an overview of past work used for traffic monitoring. In Section II, We provide a description of the proposed work. In Section III, we explained implementation details. Finally, In Section IV the experimental setup and the classification results are provided. And the main findings are summarized in Section V.
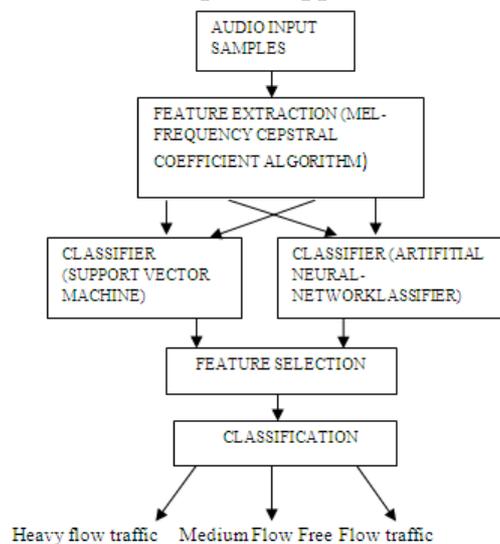
II. Proposed Approach

Fig 1. Flow Chart of proposed approach

As shown in Fig.1 we will use cumulative acoustic signal which can be used to estimate the traffic density as free flow, medium flow, and heavy flow. In order to get the more accuracy we will use Support Vector Machine (SVM) and Artifitial Neural- Network classifier(ANN). In this paper, first we acquire input samples of acoustic signal of each traffic state and mechanical condition. After that we will extract feature vectors using Mel-Frequency Cepstral-Coefficient(MFCC) algorithm. We will train the classifier with these feature vectors, this process occurs in training phase. After that we will take another input sample for testing and compare this signals with database which is acquired during training phase and this phase is called testing phase.

1. *Feature selection algorithm*
- Use Guassian distribution and describe only one fuzzy classification rule for each class.
- Set = 0.5, for i= 1,2,...,K and j= 1,2,...,D, where K is the number of classes and D is the number of features.
- Set the number of selected features (L).
- Train the neuro-fuzzy classifier with LHs. In training, $0 \le P_{ij} \ge 1$
- For i=1 to K. Find the *j*th feature that satisfies the maximum p value for the *i*th class. Then take the *j*th feature into the individual discriminative features set.
- The (L–K) features, which have the biggest P value which are selected as common discriminative features.
- There are L discriminative features. The new training Xnew and testing data are created by the selected features from the original data.

2. *Classification algorithm*
- Set the number of fuzzy rules (V) for every class.
- Then the total fuzzy rules are U= where U is the number of fuzzy rules.
- Set = 1, for i=1,2,...,U and j=1,2,...,D.
- Determine the initial value of nonlinear parameters of ANFC–LH by using K-means clustering.
- Train the ANFC–LH with Xnew training set. In training, value should be equal to or bigger than zero for every feature and fuzzy rules.
- Obtain the training and testing classification results.

III. Impletation Details

For this experimental set up first of all we have to collect the audio signals for different traffic states so, we have collected 20 samples of each traffic state that is high flow, medium flow and free flow traffic. As well as we collected 60 samples of bike, car and truck for determining the mechanical condition of vehicle. Here we have taken only three vehicles namely bike, car and truck. Out of all signals 66% of signals are used for training phase in order to train the classifier and 44% of signals are used for testing phase Each sample is of500ms long. We applied these 66% of signals to the feature extractor to get some unique feature vectors. feature vectors of each signals are somewhat different but somewhere it may be equal. As there are no definite parameters for feature vectors therefore we got these feature vectors after examining all 66% of signals. Mel-Frequency Cepstrum Coefficient algorithm is used for frequency extraction.

1. *Mfcc feature vectors*

The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel-scale, in short it uses a nonlinear frequency scale which approximates the behaviour of the auditory system more closely than the linearly spaced frequency bands used in the normal cepstrum. Additionally, these coefficients are robust and reliable to variations according to recording conditions. Therefore, we decided to use MFCC for feature extraction.
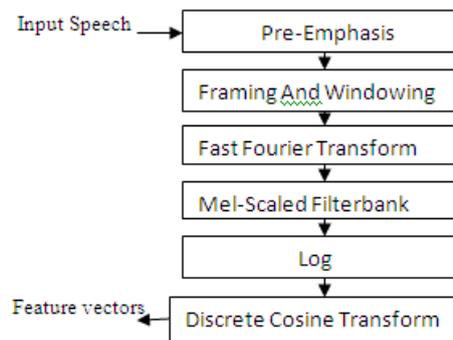

Fig 2. Flow Chart of MFCC

Pre-emphasis phase emphasizes higher frequencies. The pre-emphasis is a process of passing the signal through a filter. It is designed to increase, within a band of frequencies, the magnitude of some frequencies (higher) with respect to the magnitude of the others frequencies (lower) in order to improve the overall SNR. y[n]= x[n]-αx[n-1], α € (0.9, 1)

Typically, speech is a non-stationary signal; therefore, The acquired signal is assumed to be stationary within a short time interval. The input acoustic signal is segmented in frames of 20~40 ms with overlap of 1/3~1/2 of the frame size. In order to keep the continuity of the first and the last points in the frame, typically each frame has to be multiplied with a hamming window. Equation of hamming window is as follows,

$$W[n] = \begin{cases} 0.54 - 0.46 \ \cos\frac{2\pi n}{N}, 0 \le n \le N \\ \qquad 0, \ Otherwise \end{cases}$$

Where N is frame size.

$$Y[n]= X[n] * W[n]$$

Where Y[n] = Output signal, X[n] = Input signal

Fast Fourier Transform (FFT) is used to compute the DFT. It converts each frame of N samples from time domain into frequency domain. The computation of the FFT-based spectrum as follow,

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N} , 0 \le k \le N$$

Where N is the frame size in samples, x[n] is the input acoustic signal, and. X[k] is the corresponding FFT-based spectrum.

The frequencies range in FFT spectrum is wide and acoustic signal does not follow the linear scale. Each filter's magnitude frequency response is triangular in shape and is equal to unity at the Centre frequency and decrease linearly to zero. We then multiply the absolute magnitude of the DFT samples by the triangular frequency responses of 24 mel-filters that have logarithmically increasing bandwidth and cover a frequency range of 0-8kHz in our experiments. Each filter output is sum of its filtered spectral components. Its equation is as follows:

F (Mel) = 2595 * log 10 [1+f/700]

The ith Mel-filter bank energy ($M_{FB}(i)$) is obtained as

$$(M_{FB}(i)) = (Mel_i(k)) * \left| x(k) \right|^2 , \ k € (0, N/2)$$

Where $(Mel_i(k))$ is the triangular frequency response of the ith Mel-filter. These 24 Mel-filter bank energies are then transformed into 13 MFCC using DCT.

This is the process to convert the log Mel spectrum into time domain using DCT. The result of the conversion is called Mel Frequency Cepstral Coefficient. The set of coefficient is called acoustic vectors.

$$C_j = \sum_{i=1}^{24} (MFB(i)) \sqrt{\frac{2}{24}} cos(\pi j \frac{i-0.5}{24}) , j € (0, 12)$$

The acoustic signal and the frames changes, such as the slope of a formant at its transitions. Therefore, there is need to add features related to the change in cepstral features over time. Ex. 13 feature (12 cepstral features plus energy).

$$Energy = \sum X^2 [t]$$

Where X[t] = signal

2. Classifiers

All feature vectors generated by MFCC are provided to the classifiers. In this section first of all we will train the classifiers using these feature vectors so that it can compare the testing signal accurately. So, we decided to use the Support Vector Machine and Artifitial Neural Network.

A. Support vector machine

There are two types of SVM, first is the linear SVM which is the simplest Support Vector Machine. It uses linear decision boundary. But in case of non-separable data set,this SVM wont work effectively. Second is the Non-linear SVM. Non-linear SVM employs a non-linear decision function to classify the training data set by mapping the non-separable data points to higher dimension space where these data points become separable. SVMs build a hyperplane which divides samples such that samples of one class are all on one side of the hyperplane, and samples of the other class are all on theother side. However, data is rarely linearly separable. Thus, SVM project the data in a high-dimensional space, where the classification problem may be linearly separable, and then find the linear hyperplanes that separate the various classes with a high margin. Therefore SVM is proved to be excellent classifiers for diverse pattern recognition applications such as, object detection, speaker recognition etc. In general, a non-linear SVM projects the d-dimensional input features to some high-dimensional space H through a non-linear transform Φ.

Φ:R^d →H

The SVM composes the transformation Φ and the dot product in the higher dimensional space into a single kernel function k, which computes the dot product of two vectors when they are transformed into the higher space.
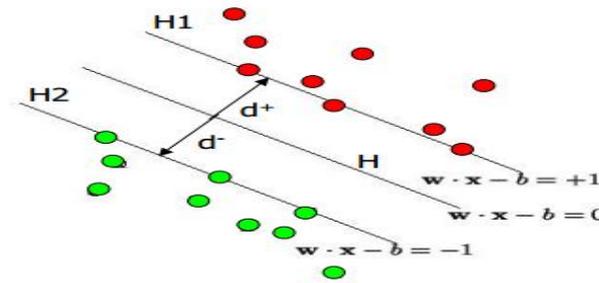

Fig 3.Hyperplane through two linearly separable classes

Consider the input data in the form (X_i, Y_i), where vectors
are in real-valued multi-dimensional space H and are the class labels. So any hyperplane is defined as
{ ∈ H | <w, > + b=0}   w ∈ H, b ∈ R
Hyperplane is defined such that
X_i.w + b ≥ +1 when = +1
X_i . w + b ≤ -1 when = -1
And minimize ||w||2
H1 and H2 are the planes:
H1: . w + b = +1
H2: . w + b = -1
The points on the planes H1 and H2 are the support vectors. d+ = the shortest distance to the closest positive point
d- = the shortest distance to the closest negative point
The margin of a separating hyperplane is d+ + d-.

B.   *Artifitial neural network*

An artificial neural network is composed of many artificial neurons that are linked together according to a specific network architecture. The objective of the neural network is to transform the inputs into meaningful outputs.It is machine learning approach that models human brain and consists of a number of artificial neurons.Neuron in ANNs tend to have fewer connections than biological neurons.Each neuron in ANN receives a number of inputs. An activation function is applied to these inputs which results in activation level of neuron (output value of the neuron).Knowledge about the learning task is given in the form of examples called training examples.
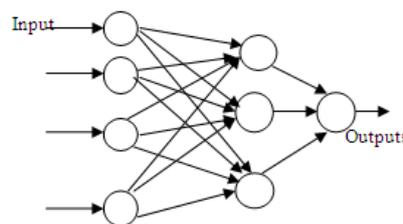

Fig 4. Structure if ANN

IV.   Results And Discussion

1.   *Mechanical Condition Of Vehicle*

For Mechanical Condition determination, we have taken 60 samles for truck, bike and car. Out of 60 samples 66% samples are used for training phase and 44% samples are used for testing phase. We have converted these samples into 'Wav. File' and then we have converted these signals into frames of 200μsec. But when we made frames there are chances of occurring discontinuities so to minimize it we have shifted these frames by 50 μsec and then we convolute it with hamming window and then we get the FFT signals. Then these signals are applied to the mel filter bank which gives 13 coefficients. After applying MFCC to the sample we got following graph.
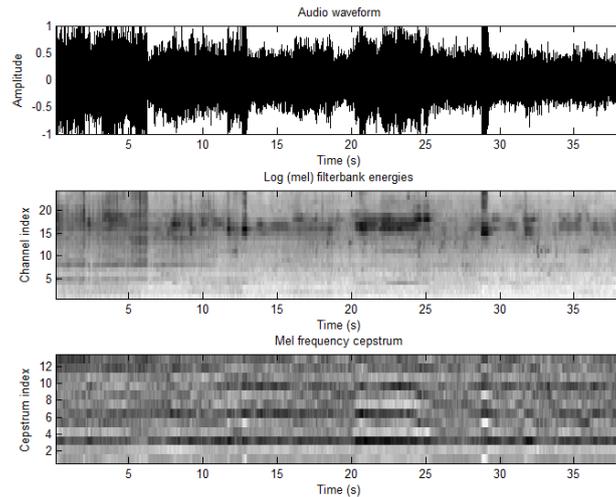
Now these coefficients are applied to the classifiers to train them and we got following results. As shown in table 1, all the features vectors are different from each other.In table no.2 we can see, the training inputs and testing inputs. These testing inputs are compare with training inputs and we got accuracy of actual state, and we got following results. Fig.(5) shows graph forgood bike and bad bike

Table 1 : Accuracy of good bike

Input		Output	Accuracy of three vehicles					
Itrfin	Itrfin	Ans	good bike	bad bike	good car	bad car	good truck	bad truck
1	2	1	0	100	0	0	0	0
2	2	2	0	100	0	0	0	0
3	2	3	0	100	0	0	0	0
4	2	4	0	100	0	0	0	0
5	2	5	0	100	0	0	0	0
6	2	6	0	100	0	0	0	0

Table 2 : Accuracy of bad bike

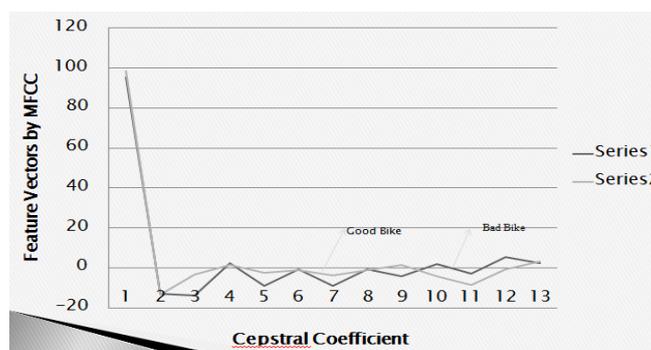| Input		Output	Accuracy of three vehicles					
Itrfin	Itrfin	Ans	good bike	bad bike	good car	bad car	good truck	bad truck
1	2	1	0	100	0	0	0	0
2	2	2	0	100	0	0	0	0
3	2	3	0	100	0	0	0	0
4	2	4	0	100	0	0	0	0
5	2	5	0	100	0	0	0	0
6	2	6	0	100	0	0	0	0

Fig 5 : difference between bad bike and good bike

V. Conclusion

We have come to this conclusion by reviewing the above literatures that we can estimate three probable conditions of traffic that is free flow, medium flow and heavy flow traffic and mechanical condition of vehicles that is good and bad. For concluding this we are expected the MFCC features for classification using SVM and ANN Classifier. From MFCC we get different feature vectors we will make its database and train the classifier. Then with this database we compare the new signals that is testing signals.After analysing it we get accuracies of three traffic states and good and bad vehicles. As we are using here recording system for acquiring input traffic signals itsindependent of lighting conditions and visual occlusion problems which occurs in computer-vision or tracking based technology. As well as our technique is low cost and having easy operations and hence this technique can be used in developing countries where chaotic and non-lane driven traffic are very common.

Acknowledgment

References

[1]. Chen Xiao-feng, Shi Zhong-ke and Zhao Kai, "Research on an Intelligent Traffic Signal Controller," Intelligent Transportation Systems, 2003. Proceedings, Volume 1, Pg. 884-887, 2003 IEEE.

[2]. Chunxiao LI and Shigeru SHIMAMOTO, "A Real Time Traffic Light Control Scheme for Reducing Vehicles CO2 Emissions," The 8th Annual IEEE Consumer Communications and Networking Conference - Emerging and Innovative Consumer Technologies and Applications.

[3]. R. Cucchiara, M. Piccardi, and P. Mello, "Image analysis and rule based reasoning for a traffic monitoring system," *IEEE Trans. Intell. Transp.Syst.*, vol. 1, no. 2, pp. 119–130, Jun. 2000.

[4]. S. Kamijo, Y. Matsushita, and K. Ikeuchi, "Traffic monitoring and accident detection at intersections," *IEEE Trans. Intell. Transp.Syst.*,vol.1,no.2,pp.108–118, Jun. 2000.

[5]. B. Coifman, D. Beymer, P. McLauchlan, and J. Malik, "A real-time computer vision system for vehicle tracking and traffic surveillance," *Transp. Res. C, Emerging Technol.*, vol. 6, no. 4, pp. 271–288, Aug. 1998.

[6]. D. I. Robertson and R. D. Bretherton, "Optimizing networks of traffic signals in real time—The SCOOT method," *IEEE Trans. Veh. Technol.*, vol. 40, no. 1, pp. 11–Type equation here.15, Feb. 1991.

[7]. B. G. Quinn, "Doppler speed and range estimation using frequency and amplitude estimates," J. Acoust. Soc. Amer., vol. 98, no. 5, pp. 2560–2566, Nov.1996.

[8]. C. Couvreur and Y. Bresler, "Doppler-based motion estimation for wide-band sources from single passive sensor measurements," in Proc. IEEE ICASSP, Apr.1997,pp.21–24.

[9]. S. Chen, Z. P. Sun, and B. Bridge, "Automatic traffic monitoring by intelligent sound detection," Proc. IEEE Intelligent Transportation Systems Conf., Nov. 1997.

[10]. S. Chen and Z. P. Sun, "Traffic sensing by passive sound detection," in Proc. Sensors and Their Applications VIII Conf., Glasgow, Scotland, U.K., Sept. 1997.

[11]. S. Chen, Z. Sun, and Bryan Bridge, "Traffic Monitoring Using Digital Sound Field Mapping," IEEE Transactions on vehicular technology,vol.50,no. 6, pp.1582-1589, Nov 2001

[12]. R. Lopez-Valcarce, C. Mosquera, and R. Perez-Gonzalez, "Estimation of road vehicle speed using two omnidirectional microphones: A maximum likelihood approach," EURASIP J. Appl. Signal Process., pp. 1059–1077, 2004.

[13]. P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphone," in Proc. SenSys, Nov. 2008, pp. 323–336.

[14]. R. Sen, B. Raman, and P. Sharma, "Horn-ok-please," in Proc. ACM MobiSys, San Francisco, CA, 2010, pp. 137–150

[15]. J. Kato, "An Attempt to Acquire Traffic Density by Using Road Traffic Sound," Active Media Tech., pp. 353-358, IEEE 2005.

[16]. VivekTyagi, "Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics," IEEE Transactions on Intelligent Transportation Systems. 2012.