RI 11010, 6 June 2011 Computer Science

IBM Research Report

Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics

IBM Research Report No. RI 11010
6th June 2011

Vivek Tyagi, Shivkumar Kalyanaraman, Raghuram Krishnapuram
IBM Research Division
IBM India Research Lab
Plot 4, Block-C, Vasant Kunj
New Delhi - 110070, India.

IBM Research Division
Almaden - Austin - Beijing - Delhi - Haifa - T.J. Watson - Tokyo - Zurich

Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics

Vivek Tyagi, *Member, IEEE,* Shivkumar Kalyanarman, *Fellow, IEEE* and Raghuram Krishnapuram, *Fellow, IEEE,*
IBM India Research Laboratory, India
Email: vivetyag@in.ibm.com, shivkumar-k@in.ibm.com, kraghura@in.ibm.com

***Abstract*—In this work, we propose to use a statistical framework that uses the cumulative acoustic signal from a roadside installed single microphone, to classify the vehicular traffic density state. A typical cumulative acoustic signal on a road segment is composed of several noise signals such as the tire noise, engine noise, engine idling noise, occasional honks and the air turbulence noise of the multiple vehicles. The occurrence and mixture weightings of these multiple vehicles' noise signals are determined by the prevalent traffic density conditions on the road segment. For example, in a free flowing traffic, the vehicles would typically be moving with the medium to high speeds thereby producing mainly the tire noise and the air-turbulence noise and less of the engine idling noise and the honks. While for a slow moving congested traffic, the cumulative signal will be largely dominated by the engine idling noise and the honks; the air turbulence and the tire noises will be inconspicuous. Further, these various noise signals have spectral content that are very different from each other and hence can be used to discriminate between the different traffic density states that lead to them. In this work, we extract the short term spectral envelope features of the cumulative acoustic signals, and model their class-conditional probability distributions, conditioned on one of the three broad traffic density states namely *Jammed(0-10KpH), Medium-Flow(10-40Kph) and Free-Flow(40Kph and above)* traffic. While, these states are coarse measures of the average traffic speed, they nevertheless, can provide useful traffic density information in the often chaotic and *non-lane* driven traffic conditions of the developing geographies where the other techniques (magnetic loop detectors) are inapplicable. Based on these learned distributions, we use a Bayes classifier to classify the acoustic signal segments spanning a duration of $5s$ to $30s$, which resulted in a high classification accuracy of $(\sim 95\%)$. Using a discriminative classifier such as Support Vector Machine (SVM), resulted in further classification accuracy gains over the Bayes classifier.**

I. Introduction

Road traffic congestion is a severe problem worldwide. Traffic congestion has been increasing worldwide due to the increased motorization, urbanization and the population growth. Traffic Congestion reduces the efficiency of the transportation infrastructure of a city and increases the travel time, fuel consumption, air pollution and leads to an increased user frustration and fatigue.

In the developed countries, several intelligent transportation systems (ITS) solutions have been proposed that rely on the fact that the traffic is lane driven and multiple sensors which generally are expensive can routinely be used. Magnetic loop sensors, speed guns and video cameras are some of the examples of these sensors. Currently, the most widely used traffic sensor in the developed countries is the magnetic loop detector [19]. However, the installation/maintenance cost of these sensors, significantly adds to the high operational expense of these devices during their life-cycles. The associated cost involved in the earth work during loop laying and traffic closure further limits their widespread use. Therefore, in the past one decade researchers have been developing several non-intrusive traffic monitoring technologies based on the laser, ultrasound, radar, video and audio signals.

Video image processing seems like a natural sensing modality for the traffic. In [16] Cucchiara et. al have proposed an approach for detecting vehicles in urban traffic scenes by the means of rule-based reasoning on visual-data. In [17] Kamijo et. al. have proposed a hidden Markov model (HMM) based computer-vision technique to detect accidents and other events such as reckless driving at the road intersections. However, these techniques do not directly address the problem of average speed/speed-range estimation. In [18], Coifman et. al. have proposed an extensive feature based computer-vision technique for the vehicle tracking. They use "corner" features of the vehicles, which are being driven in the lanes, to track them and then estimate the traffic parameters such as the average speed and the volume. They obtained impressive results on the free-way traffic, where more that 80% vehicles were traveling within the speed range of $50-70mph$ $(80-110Kph)$[18]. These speeds ensured somewhat reasonable headroom between the vehicles, which leads to good tracking as the vehicles are not linked to each other. However, it is not clear if such a tracking technique could still work in the chaotic and *non-lane* driven city traffic conditions with the extremely varied speed ranges of $0-10Kph$, $10-40Kph$ and above $40Kph$. Such traffic conditions are very common in the cities of the developing geographies (India and South Asia) and are the focus of this paper.

In this paper, we study the utility of the cumulative roadside acoustic signal to estimate the traffic density classes (average traffic speed range) on a city's road segment. In particular, its independence to the lighting conditions/visual occlusion, low signal bit rate as compared to video signal and the relatively inexpensive cost of the sensors(microphones, $2-5$\$), make audio an attractive sensing modality.

The cumulative road acoustic signal is composed of varying mixture of several noise signals (tire noise, engine idling noise, engine block noise, air-turbulence noise and the occasional honks) and their mixture weightings in the cumulative signal varies depending on the traffic density conditions[2]. For example, in the free flowing traffic conditions, vehicles tend to move with medium to high speeds and hence their cumulative

acoustic signal is dominated by the tire noise and the air-turbulence noise[2], [3]. Whereas, for a heavily congested traffic, the road-side cumulative acoustic signal gets larger contributions from the engine idling noise and the honks; the air turbulence and the tire noises are almost negligible as none of the vehicles can move at high speeds[2]. Interestingly, these various noises have varied spectral contents and hence can be used to discriminate between the various traffic density states which contribute to them. Therefore, in this work, we extract the spectral features of the roadside acoustic signal. We model the class conditional probability distributions of these feature vectors conditioned on one of the three broad traffic density states namely *Jammed, Medium Flow and Free Flow* traffic using the Gaussian Mixture Models (GMM). The GMM parameters are trained on about 2.5 hours of road-side acoustic data that has been hand-labeled as belonging to one of the three traffic density states. Based on these distributions, a Bayes classifier is used for the traffic density state classification resulting in about 95% accuracy when $20-30s$ of audio signal evidence is presented. We further experimented with a Support Vector Machine (SVM) radial basis function (RBF) kernel classifier[20]. In recent $10-15$ years, SVMs have proved to be excellent classifiers owing to their maximum margin discriminative training[20], [21]. The classification accuracies for our task improved further with the use of a RBF-SVM classifier.

We begin with an overview of the past work in the use of audio modality for the traffic monitoring in the Section II. We provide a description of the various noise signals in the cumulative acoustic signal in Section III followed by their spectral analysis and the GMM based modeling in Section IV. Finally, the experimental setup and the classification results are provided in Section V and the main findings are summarized in Section VI.

II. Audio Related work in Traffic monitoring

While, the use of the magnetic loops, has been the most widely used approach for the traffic monitoring, they still have a very high cost of installation and maintenance (several thousands of dollars). Further they require the traffic flow to be orderly, lane-driven and mostly homogeneous (primarily four wheeled vehicles and multi wheeled trucks).

However, these conditions are not at all met in the developing regions such as India and South Asia/South East Asia which typically have a highly chaotic and non-lane driven traffic. The wide variety of vehicles from two wheeled motorcycles to three-wheeled auto-rickshaws to four-wheeled vehicles to multi-wheeled buses and trucks make it a highly heterogeneous traffic which does not follow lane driving. Therefore, the traffic monitoring is an even severe problem in the developing countries that renders most of the above mentioned ITS systems such as loop detectors and computer vision based tracking techniques ineffective. In this background, roadside acoustic signal seems like a viable modality for the traffic monitoring due to its inherent low cost of installation and operation. Therefore, several researchers have developed various traffic monitoring and a single vehicle profile extraction techniques based on the audio modality.

In [12], [13] a theoretical description of a single point source's (single vehicle) speed estimation is provided using wave propagation effects (Doppler frequency shifts). While this solution is elegant and can accommodate any line of arrival of the vehicle with respect to the microphone (assuming the distance to the closest point of approach (CPA) is known), the description applies only to a single vehicle's acoustic waveform. When several vehicles are present, the interference of their combined acoustic waveforms will render this solution in-applicable for their speed estimation.

In[1] the authors have extensively used the wave propagation physics of a spatial sound source (a vehicle) to estimate a single vehicle's speed, engine RPM, the number of cylinders and its length and width based on its acoustical wave patterns as the vehicle drives by the acoustic sensor. Using a dipole source, they also account for the interference effect of the various tires of a vehicle to arrive at a "unique" feature vector (fingerprint) of each vehicle type (such as Ford F150, Chevy Impala, Honda Accord, Nissan Maxima and VW Passat etc). However their technique is applicable only when there is single vehicle traveling on the road and its vehicle type has to be recognized. Therefore, it cannot be applied for the traffic density state estimation where there are multiple vehicles traveling and producing a cumulative acoustic signal rather than just a single vehicle's acoustic signal.

In [15], the authors have used a combination of the smartphone features such as the accelerometers and a basic honk signal detection followed by a simple Doppler frequency shift computation to arrive at a vehicle's speed estimate. In [11], the authors have presented an approach for detecting a crash-risk level using the computing power and the microphones of mobile devices that can be used to alert the user in advance of an approaching vehicle to avoid a crash.

Building on the work in [15], the authors in [14] have used a simplified Doppler frequency shift rule which implicitly assumes that the vehicle is moving in the same direction as the straight line connecting the vehicle to a microphone. They propose to use the vehicles' honks acoustic signal in their simplified Doppler frequency shift estimation. However, in the presence of multiple vehicles honking (either concurrently or with significant time overlap), and the other background noises present (cumulative engine noise, idling noise and tire noise of several vehicles), it is not clear how the two microphones can robustly detect a honk emitted by the same vehicle. In the presence of multiple overlapping honks and other background vehicular traffic noises, indeed it is a very challenging problem to identify and extract an acoustic signal sub-component that corresponds to the same vehicle's honk at the two microphones. In [14], the authors have used various time based thresholds and a very basic Discrete Fourier transform (DFT) sample energy comparison with a threshold to detect the honks and to correspond them at the two microphones. In general, the spectrum of a honk signal is not a pure sinusoid tone, but instead has several harmonics in the $1-4KHz$ range with a bandwidth of about $100Hz$ around each harmonic. Furthermore, in the presence of the other narrow-band signals (the other vehicles' honk signals in the time vicinity of the honk signal that is being sought) and the other wide-band

noise components (the engine noise, engine idling noise and tire noise) in the cumulative signal, it is clear that a simple threshold based technique cannot robustly detect the same particular honk at the two microphones which are spaced by 30 meters. Further, their experimental setup covers traffic condition on two very busy roads (one with a single lane one way traffic and the other with three lanes bi-directional traffic) with speeds between $0-20Kph$. At the speeds around $40Kph$ and $60Kph$ and above, the probability of the vehicles honking goes down drastically and would be even further low for the event that the vehicles honk in a specific 20 meters stretch which lies in between the two microphones that are separated by 30 meters. The authors in [14] call it " the honking zone of interest" and it is an important requirement for their algorithm.

Given these limitations, we propose to use the entire cumulative roadside acoustic signal rather than just detecting the honk signal. In the next section, we will show, through the spectral analysis, that the cumulative acoustic signal has important discriminative information present in its spectro-temporal plane that allows us to directly build simple statistical classifiers to classify between three broad traffic density states that correspond to an increasing range of speeds, namely $(0, 10)kph$, $(10, 40)Kph$ and above $40Kph$. We denote them as *Jammed*, *Medium-Flow*, and *Free-Flow* traffic density states respectively. Some representative traffic density pictures are shown in the Figs. 1, 2, 3.

A. The need for the traffic density measure

One of the main characteristic of the city traffic in the developing geographies (especially South Asia/India) is that it usually does not move in the lanes even if they are explicitly marked on the roads. Frequent lane changing is very common and hence a lane based *volume* measure (number of vehicles passing a point of a lane per hour) does not seem like an appropriate measure in such conditions. The entire road width, with all the lanes combined, becomes one continuous carriageway. Therefore, we have decided to use the measure of traffic density (*Jammed*, *Medium-Flow*, and *Free-Flow* corresponding to an increasing range of speeds $(0, 10)Kph$, $(10, 40)Kph$ and above $40Kph$ respectively) instead of the two distinct *volume* and *speed* measures, as has been used in some of the traffic monitoring work in the developed geographies[18]. As can be noted from the Figs. 1, 2, 3, our measure of traffic density captures both the measures of *speed* and road-aggregate *volume* in one measure of the traffic density. We have also observed that the *speed* and *volume* measures are sort of inversely related to each other and are not two independent measures. This is especially true for the non-homogeneous and non-lane driven traffic in the cities of South-Asia/India where the fast moving traffic implicitly implies that the volume is low and vice-versa. These observations led to our choice of the above proposed traffic density class based measurements. The experimental results in the later sections are based on this measure.

III. Cumulative Roadside Acoustic Signal

A typical vehicle produces various noises (sounds) depending on its speed, load and its mechanical condition. In general, the vehicle produced noises can be categorized as,

A. Engine noise:

An internal combustion engine's noise consists of a deterministic harmonic train and a stochastic noise component due to the air intake[2]. The deterministic harmonic train is produced by the fuel combustion in the engine cylinders and its lowest harmonic tone is the called the cylinder fire rate f_0 which in turn is a function of the engine's rounds per minute (RPM). In summary, the engine noise varies with the speed and the acceleration of the vehicle[1], [2]. When a vehicle is stationary, though with its ignition on (as is often the case in the congested traffic), it produces a distinct engine idling noise. Whereas, a medium to fast moving vehicle produces different engine noise due to the higher cylinder fire rate.

B. Tire noise:

This is the noise produced by a vehicle's rolling tire as a result of its interaction with the road surface. The tire noise is the main source of a vehicle's total noise at speeds higher than 50kph (kilometers per hour)[3], [4]. It has two components: vibrational noise and air noise [4], [5]. The vibrational noise is caused due to the contact between the tire treads and the road surface texture and its spectral energy lies between 100Hz to 1000Hz. On the other hand the air noise is produced due to the air being sucked in the rubber blocks of the tire and is dominant in the frequency range of 1000Hz to 3000Hz. In the direction of the vehicle's movement, the tire and the road surface form a geometrical structure that amplifies the noise generated by the tire/road-surface interaction. This effect is called the *horn* effect and has a directional pattern. This effect leads to strong tire noise components in the range $(600 - 2000)$Hz[5], [6]. Therefore, the tire noise which becomes dominant in cumulative acoustic signal around and above the speeds of 50kph[3] can provide useful information about the traffic density state.

C. Air Turbulence Noise:

As a vehicle moves at the medium to high speeds, it induces air-turbulence. This noise is due to the air flow generated by the boundary layer of the vehicle and is most prominent immediately after a vehicle has driven past the microphone as a distinctive *drive-by-noise* or *whoosh* sound[7], [8].

D. Exhaust Noise:

This noise is produced by the entire exhaust apparatus that connects the engine combustion compartment through exhaust tubes to the exhaust muffler (silencer) present at the rear-end of the vehicle. Unlike the engine noise, the exhaust noise increase as the vehicle load increases[9].

IV. Short Term Spectral Features and Gaussian Mixture Model (GMM) based modeling

The various traffic density states induce different cumulative acoustic signals[1]. For instance, a fast moving traffic's

Fig. 1. Typical *Free-flowing* traffic (above $40Kph$) in New Delhi, India.

Fig. 2. Typical *Medium-flow* traffic $(10, 40)Kph$ in New Delhi, India.

acoustic signal, primarily consists of the tire-noise and the air-turbulence noise. Whereas, a jammed traffic state will be marked by mostly the engine idling noises and the occasional honks. All these different noises have very different spectral envelopes and hence can be used as cues to discriminate between the three traffic density states. To verify this hypothesis, we examined the spectrogram of the different traffic state's cumulative acoustic signals. In Fig. 4, 5, 6, we show the spectrogram of 120s of acoustic signal, recorded at one microphone installed at the roadside, for the three different traffic density classes (states) namely, *Free-Flow* (above $40Kph$), *Medium-Flow* (speeds of $(10, 40)Kph$) and the *Jammed* traffic (speeds of $(0, 10)Kph$). These spectrogram clearly show the various discriminative acoustic features that characterize these different traffic conditions. Specifically,

- For the *Free-Flowing* traffic in Fig. 4, we only see the wide-band drive-by noise and the air-turbulence noise of the vehicles. We note that there are very few, if any, honks. In particular, there are no honk signals in this 2 min long signal.
- In the *Medium-flow* traffic (Fig. 5), we can notice some wide-band drive-by noise, some honk signals and some concentration of the spectral energy in the low frequency ranges $(0, 0.1)$ of the normalized frequency or equivalently $(0, 800)Hz$.
- Whereas, the *Jammed* traffic condition in Fig. 6 has almost no wide-band drive-by engine noise or air-turbulence noise and is dominated by several honk signals. We note the several harmonics of the honk signals and they range from $(2, 6)KHz$. Further the jammed condition is marked by the high concentration of the spectral energy in a very narrow and low frequency range around $(0, 0.05)$ in the normalized frequency or equivalently $(0, 400)Hz$, which is due to the engine-idling.

We note that the above characterization of the various traffic density states are only indicative in nature. As will become clear in the next sections, we do not propose to use a rule-based technique where a certain noise type is explicitly detected and then the classification is based on that.

A. Mel-Frequency Cepstral Coefficients (MFCC) Features

Encouraged by the presence of these several discriminating features in the spectro-temporal plane of the cumulative acoustic signal, we decided to use the spectral envelope based features in our statistical classifier. As is well known, the

Fig. 3. Typical *Jammed* traffic $(0, 10)Kph$ in New Delhi, India.

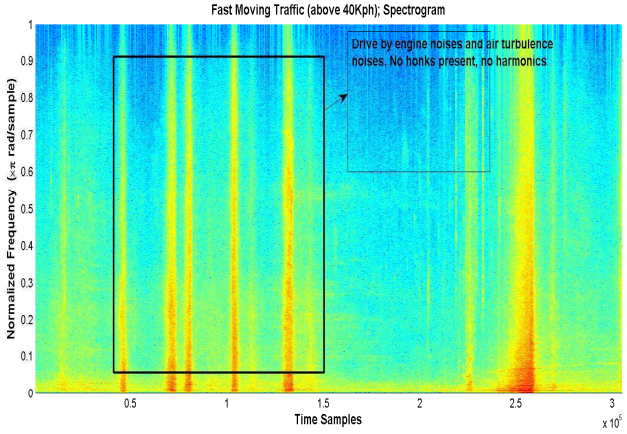

Fig. 4. Spectrogram of the free-flowing traffic (above $40Kph$)

first $10-15$ cepstral coefficients of a signal's short time spectrum, succinctly capture the smooth spectral envelope information[29]. Over the years, Mel-Frequency cepstral coefficients (MFCC)[30], which are the DCT coefficients of a Mel-filter smoothed logarithmic power spectrum, have been applied very successfully as the acoustic features in speech recognition[30], speaker recognition[28] and music recognition. Therefore, in this work we decided to use the MFCC features and their first and second order time-derivatives to characterize the spectro-temporal plane of the acoustic signals. Unlike the speech signal, where a phoneme can change over $10ms$ to $50ms$, the traffic density states evolve over longer time scales. Due to the physical constraints, the traffic density state could change from one to another (jammed to medium flow to free flow) over at least $5-30$ minutes duration and not over, say, every minute. Therefore, we decided to use the long primary analysis windows of the typical size $100-500ms$ to obtain the spectral envelope, followed by the MFCC feature computation. The primary analysis windows were shifted by a shift-size of $50-100ms$ to obtain a series of feature vectors. This feature extraction scheme is illustrated in the Fig. 7. Assuming $x(n)$ to be a short-term windowed acoustic signal, we take its N point DFT, $X(k)$. The acoustic signal has a bandwidth of $8KHz$.

$$X(k) = \sum_{i=0}^{N-1} x(i) \exp\left(\frac{-j2\pi ki}{N}\right) \tag{1}$$

We then multiply the absolute magnitude of the DFT samples by the tri-angular frequency responses of the 24 Mel-filters that have logarithmically increasing band-width[30], and cover a frequency range of $0-8KHz$ in our experiments. These filters are illustrated in Fig. 8. The first $N/2$ absolute magnitude DFT samples correspond to the $0-8kHz$ frequency range. The i^{th} Mel-filter bank energy ($M_{FB}(i)$) is obtained as,

$$M_{FB}(i) = Mel_i(k) \times \log|X(k)|,\ k \in (0, N/2) \tag{2}$$

where, $Mel_i(k)$ is the tri-angular shaped frequency response of the i^{th} Mel-filter. These 24 Mel-filter bank energies are then transformed into a 13 dimensional Mel Frequency Cepstral coefficients (MFCC), c_j, using a discrete Cosine transform (DCT).

$$c_j = \sum_{i=1}^{24} M_{FB}(i) \sqrt{\frac{2}{24}} \cos(\pi j \frac{i-0.5}{24}),\ j \in (0, 12)$$

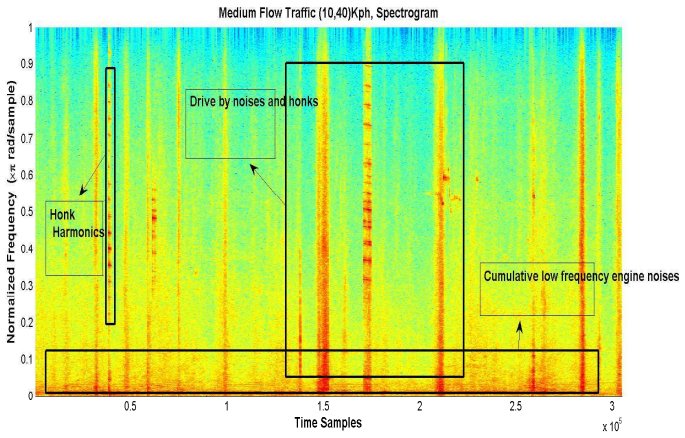

Fig. 5. Spectrogram of the medium-flowing traffic $((10, 40)Kph)$

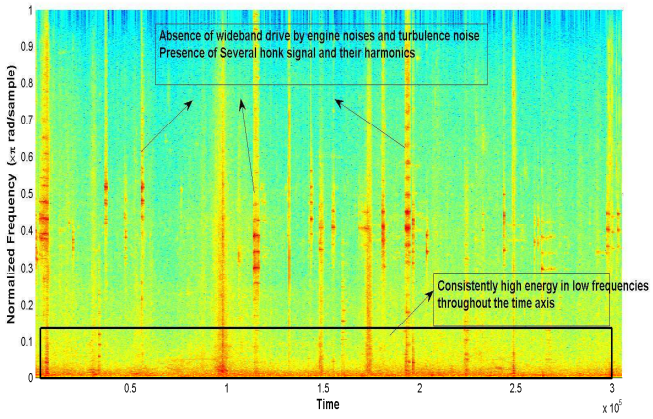

Fig. 6. Spectrogram of the jammed traffic $((0, 10)Kph)$

B. Gaussian Mixture Modeling

Parametric density models such as Gaussian Mixture Models (GMMs) have been very successfully used in diverse pattern recognition problems such as speech/speaker/music recognition[32], [33], [28] and image recognition. They owe their success and popularity to the two important properties. First, given the appropriate number of parameters (mixture components) and a large number of training samples, the GMM based density tends to converge to the true density of the random variable (training samples) under the maximum likelihood criterion[26], [27]. Second, there exists simple and efficient algorithm based on the Expectation Maximization (EM) technique[34], [35] to estimate the maximum likelihood (ML) parameters of a GMM given the training samples.

Therefore, we decided to model the class-conditional probability densities of the feature vectors (MFCCs) conditioned on the traffic density state using the Gaussian Mixture Models (GMMs). Given a sequence of feature vectors $x_{i=1}^{T}$ that belong to a particular traffic density state $'j'$, where $j \in \{$*Jammed, Medium-Flow, Free-Flow*$\}$, we assume the feature vectors are independent of each other given the class (traffic density) labels. Therefore, we have,

$$p(x_{i=1}^{T}|j) = \prod_{i=1}^{T} p(x_i|j) \tag{3}$$

where, $p(x_i|j)$ is modeled by a GMM with $'M'$ mixture components as,

$$p(x_i|j) = \sum_{m=1}^{M} \mathcal{N}(x_i|\mu_j^m, \Sigma_j^m) \tag{4}$$

where, $\mathcal{N}(|\mu, \Sigma)$ denotes a multi-variate Gaussian density with the mean vector μ and the covariance matrix Σ. Standard EM algorithm[35] is employed to estimate the mean vectors μ_j^m and the covariance matrices Σ_j^m of the Gaussian mixture for all values of $'j'$ and $'m'$. Therefore, after performing the training, one has good estimates of the class conditional probability densities of the feature vectors conditioned on the various classes.

C. Traffic Density State Classification using Bayes Rule

Assume $x_{i=1}^{T}$ is the sequence of the feature vectors that we are asked to classify as belonging to one of the traffic density

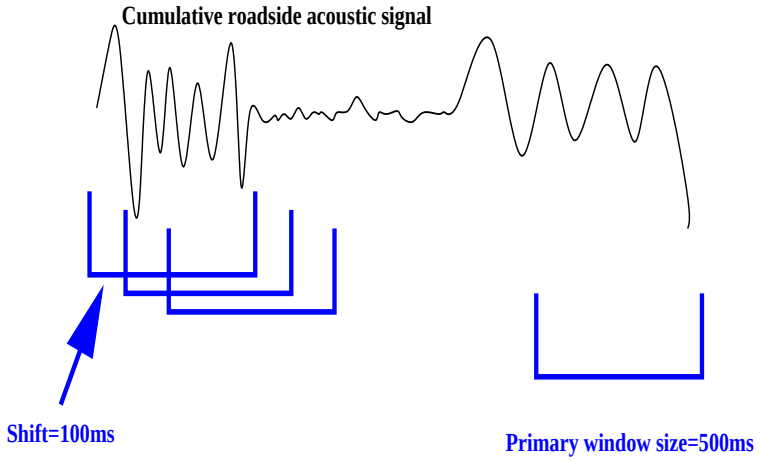

Fig. 7. Blocking of the cumulative acoustic signal into primary windows of size=500ms and shifted by 100ms to obtain a sequence of MFCC feature vectors for an acoustic signal segment.

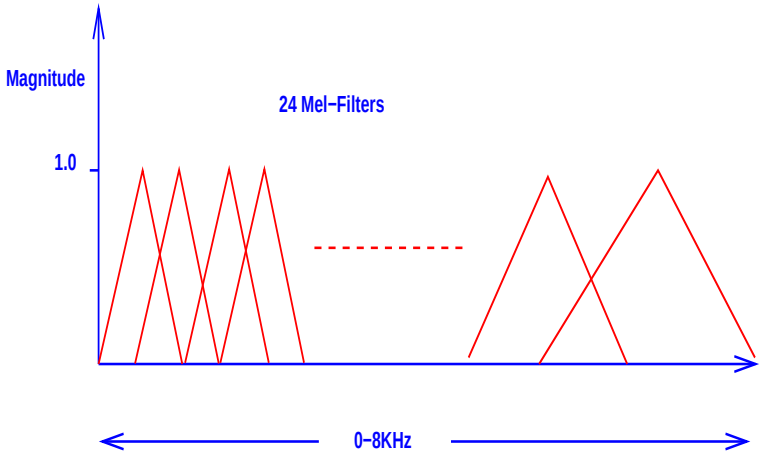

Fig. 8. Tri-angular shaped frequency response of the 24 Mel-filters covering a range of $0 - 8KHz$.

classes. Here $'T'$ is a suitable time duration over which the traffic density can be safely assumed to belong to one class. As per the Bayes classification, we assign it to the class $'j_0'$ such that it maximizes the posterior probability $p(j_0|x_{i=1}^T)$,

$$\begin{aligned} j_0 &= \arg\max_j p(j|x_{i=1}^T) \\ &= \arg\max_j \frac{p(x_{i=1}^T|j)p(j)}{p(x_{i=1}^T)} \end{aligned} \tag{5}$$

Assuming equal prior probabilities and ignoring the denominator term ($p(x_{i=1}^T)$) which remains constant to all the classes, we arrive at a decision rule directly in terms of the class-conditional densities ($p(x_{i=1}^T|j)$); density functions that we have already learned over the training samples,

$$\begin{aligned} j_0 &= \arg\max_j p(x_{i=1}^T|j) \\ &= \arg\max_j \prod_{i=1}^{T} (\sum_{m=1}^{M} \mathcal{N}(x_i|\mu_j^m, \Sigma_j^m)) \end{aligned} \tag{6}$$

Indeed, the above formalism where the GMMs are used to model the class-conditional distributions have been used very successfully in very diverse pattern recognition problems – ranging from speaker recognition[28] to speech recognition[32], [33]. The basic premise remains that if the distribution of a random variable is quasi-stationary, then it can be well modeled by a GMM. In the case of speaker recognition, the speech feature vectors (MFCCs) correspond to a single speaker, and hence can be assumed to be quasi-stationary. Similarly in speech recognition, while the entire speech signal is non-stationary and hence is modeled by a hidden Markov Model (HMM)[32], [33], the distribution of each hidden state (typically a sub-state of a phoneme) is quasi-stationary and is modeled by a GMM.

In the above discussion, we have assumed that the distribution of the cumulative acoustic feature vectors (MFCCs) given a traffic density state (*Jammed, Medium-Flow or Free-Flow*) is quasi-stationary and therefore can be reasonably modeled by a GMM of appropriate mixture components. The appropriateness of this assumption is evaluated through the traffic density state classification experiments that we report in the next section.

V. Experimental Results

We have collected about three hours of cumulative roadside acoustics data from the New Delhi South District area. The data was collected from a roadside installed, omni-directional microphone at 16Khz sampling frequency. This data covered three broad traffic density classes and was collected over about 10 data collection sessions from about 6 different road segments. The labeling of the data was done by a human assessment of the prevailing traffic density state. We further partitioned the data into two independent sets – one for training the GMM parameters of the three classes (traffic density states) and other for the classification experiments based on the learned distributions as in (6). The training set covered three traffic density states (classes) equally and consisted of 90 minutes of audio data. Similarly, the test set was of duration 80 minutes and covered the three classes almost equally.

MFCC features and their first and second order time derivatives (an estimate of the modulation spectrum[31]) were computed over a primary analysis window of varying size from $40ms$ to $500ms$ and shifted by segments of $20ms$ to 100ms. Our intent was to analyze the effect of the primary window size and the shift on the classification accuracies. Four MFCC feature sets were computed with varying configuration as noted below.

- **MFCC**40: Primary window size$=40$ms, shift20ms.
- **MFCC**100: Primary window size$=100$ms, shift$=50$ms.
- **MFCC**200: Primary window size$=200$ms, shift$=100$ms.
- **MFCC**500: Primary window size$=500$ms, shift$=100$ms.

All the feature vectors consisted of first 13 MFCC coefficients (including the zeroth cepstral coefficient c_0, which is equivalent to the log energy of the short-term signal) and their first order and second order time derivatives computed over the 10 adjoining frames. This led to a 39 dimensional feature vector per frame. We trained three GMM models, one for each of the traffic classes on the 1.5 hours of the total training data. Each GMM was initialized as a single Gaussian and the number of mixture components were gradually incremented in steps of 2 to finally train 11 component Gaussian density for each class. Standard EM algorithm was used for the parameter estimation.

Test set consisted of 1.5 hours of data that was completely disjoint from the training set. Our first experiment was on the classification of a single frame of MFCC. Each frame's likelihood was evaluated using the GMM of the three classes and it was assigned to the traffic density class with the highest likelihood. Tab. I, shows the classification accuracy of a single frame (single feature vector) based on its GMM likelihood as computed in (6). We make a couple of observations based on this table,

First, even on the basis of a single feature vector ($T = 1$), we were able to correctly classify the traffic classes with an accuracy in the range $60-70\%$. This experiment supported our hypothesis that the cumulative acoustic signal (and its MFCC feature) does contain discriminative cues about the underlying traffic density condition, and which are learnt by the GMM probability distributions. As a result, it led to an accuracy in the range $60-70\%$ (with a random decision rule, the accuracy of correct classification for a 3 class problem will be 33.33%).

Second, the classification accuracy of the *Medium-Flow* class is below that of both the *Jammed* and the *Free-Flow* traffic density classes. This is due to the fact the medium flow traffic does exhibits the characteristics of both the jammed and the free-flow traffic at different times as its average speed range $(10, 40)Kph$ is in between those of other two traffic classes. This phenomenon is also evident in the spectrogram in Figs. 4, 5, 6, where the medium-flow traffic's spectrogram exhibits "honk-harmonics" and low frequency noise (though with much broader bandwidth as compared to the jammed traffic). At some instances, the medium-flow traffic also exhibits the "drive-by-noise" that is predominantly present in the free-flow traffic.

Third, the classification accuracy of all the three classes increases as the primary analysis window size is increased from $40ms$ to $500ms$. This is due to the fact that the traffic density class (state) is a slow changing physical process owing to its inherent physical constraints (speeds of the vehicles are bounded between $(0, 100)kph$ and hence the density of the traffic cannot change at an arbitrarily high rate). On an average, the traffic density class (state) on a particular road segment can be expected to change from, say, jammed to medium-flow and medium-flow to free-flow on a time-scale of $(5-30)mins$ or even higher. This is unlike the speech signals where a phoneme (a basic classification/recognition unit) can change over $(20-80)ms$ time-period[30]. Therefore a primary analysis window of size $200ms$ or $500ms$ and a window shift of $100ms$ seems to be a reasonable choice.

A. Using multiple contiguous feature frames

So far, our classification was based on a single frame of feature vector that at most covered $100ms$ of the signal. As the traffic density class could only change over the time periods of $(5-30)mins$ or even higher, we can use significantly larger number of contiguous feature vector frames, to perform the classification. Therefore in our next set of experiments we used several contiguous feature frames in the range ($T = 5, 30s$) – time periods over which the traffic density class can be safely assumed to belong to a single class. As explained in (6), the class that maximized the likelihood of the T feature frames, was recognized as the underlying traffic class. This decision was then compared with the ground truth to arrive as the classification accuracy.

In Tables II, III, IV, V, we present the classification accuracy as the number feature frames were increased covering $T = 5s, 10s, 15s, 20s, 25s, 30s$ of the cumulative signal, while different sizes of the primary analysis window and the shift sizes were used. These tables show two important trends. First, the overall accuracy improved for all the feature sets as the observation time period ($'T'$) was increased. Second, the primary window size of $200ms$ and $500ms$ with a shift of $100ms$ seemed to be a good design choice for this classification task. As the observation time period $'T'$ increased, it provided us more evidence to classify the cumulative acoustic signal and hence benefited the classification accuracy. As the

TABLE I
SINGLE FRAME LEVEL CLASSIFICATION ACCURACIES OF THE VARIOUS TRAFFIC DENSITY CLASSES. THE PRIMARY ANALYSIS WINDOW SIZE IS VARIED IN THE RANGE 40MS, 100MS, 200MS, 500MS WHILE THE WINDOW SHIFT SIZE IS VARIED IN THE RANGE 20MS, 50MS, 100MS, 100MS RESPECTIVELY.

Traffic Class	MFCC40	MFCC100	MFCC200	MFCC500
Jammed	61.9	69.1	70.92	73.3
Medium-Flow	52.0	58.6	62.6	63.9
Free-Flow	58.2	66.0	69.30	71.4

TABLE II
CLASSIFICATION ACCURACIES OF THE VARIOUS TRAFFIC CLASSES BASED ON MULTIPLE CONTIGUOUS FRAMES COVERING $T = 5s, 10s, 15s, 20s, 25s, 30s$ OF THE CUMULATIVE ACOUSTIC SIGNAL. FEATURE: MFCC40 (PRIMARY WINDOW=$40ms$, SHIFT=$20ms$)

Traffic Class	**T=5s**	**T=10s**	**T=15s**	**T=20s**	**T=25s**	**T=30s**
Jammed	91.8	95.3	96.5	97.4	99.9	100.0
Medium-Flow	79.1	85.8	88.9	91.1	92.9	94.4
Free-Flow	77.8	87.9	92.7	94.1	94.6	95.6

TABLE III
CLASSIFICATION ACCURACIES OF THE VARIOUS TRAFFIC CLASSES BASED ON MULTIPLE CONTIGUOUS FRAMES COVERING $T = 5s, 10s, 15s, 20s, 25s, 30s$ OF THE CUMULATIVE ACOUSTIC SIGNAL. FEATURE: MFCC100 (PRIMARY WINDOW=$100ms$, SHIFT=$50ms$)

Traffic Class	**T=5s**	**T=10s**	**T=15s**	**T=20s**	**T=25s**	**T=30s**
Jammed	95.8	97.5	99.5	100.0	100.0	100.0
Medium-Flow	79.4	85.9	88.0	91.4	92.4	95.9
Free-Flow	88.8	95.9	97.9	98.3	98.7	99.4

TABLE IV
CLASSIFICATION ACCURACIES OF THE VARIOUS TRAFFIC CLASSES BASED ON MULTIPLE CONTIGUOUS FRAMES COVERING $T = 5s, 10s, 15s, 20s, 25s, 30s$ OF THE CUMULATIVE ACOUSTIC SIGNAL. FEATURE: MFCC200 (PRIMARY WINDOW=$200ms$, SHIFT=$100ms$)

Traffic Class	**T=5s**	**T=10s**	**T=15s**	**T=20s**	**T=25s**	**T=30s**
Jammed	96.1	99.0	100.0	100.0	100.0	100.0
Medium-Flow	81.3	88.0	91.3	93.6	96.2	96.8
Free-Flow	89.5	95.0	96.5	97.2	97.8	97.8

traffic density condition can only change over the time periods of $5 - 30mins$ or even higher, an observation time period of $T = 30s$ that results in highest classification accuracies, was still a fairly quick response time.

B. SVM based classification

While Bayes rule is the optimal classifier in terms of minimizing the error rates, its optimality depends on the availability of the "true" estimate of the class-conditional distributions as in (6). In practice, the "true" distributions are never known and we only estimate them using a family of distributions (for example GMMs). On the other hand, Support Vector Machines (SVM) are discriminative classifiers that do not require the knowledge of the class-conditional distributions – they only focus on the classification. They 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[20], [21]. Therefore, they have proved to be excellent classifiers for diverse pattern recognition applications such as handwritten digit recognition[22], object detection[23], speaker recognition[24] and many others. In particular, a non-linear SVM projects the $'d'$ dimensional input feature x_i (MFCC in this case) to some high dimensional space $\mathcal{H}$, through a non-linear transform Φ.

$$\Phi : \mathbf{R}^d \mapsto \mathcal{H} \tag{7}$$

It then seeks a hyperplane (which is normal to the weight vector $'w'$) in the high-dimensional space $\mathcal{H}$ such that, the margin $\frac{1}{|w|}$ between the two classes (labeled by $y_i = +1, -1$) is maximized[20], [21] and the following constraints are met.

$$\begin{aligned} &w.\Phi(x_i) + b \geq +1 \text{ for } y_i = +1 \\ &w.\Phi(x_i) + b \leq -1 \text{ for } y_i = -1 \\ &\text{and minimize } ||w||^2 \end{aligned}$$

The is easily extended for the multi-class problems[20]. The above minimization problem is a quadratic program with constraints and its solution involves only the dot products between the transformed features taking the form $\Phi(x_i).\Phi(x_j)$. And, these dot products can be efficiently found using kernel functions, without even explicitly knowing the transformation $\Phi(.)$[20], [21]. This so-called *kernel-trick* significantly simplifies the training procedure and brings down the computational requirements[20], [21]. It has been shown that the optimal weight vector can be expressed as[20],

$$w = \sum_{k=1}^{N_{sv}} \alpha_k y_k \Phi(s_k) \tag{8}$$

where, s_i are the so-called $'N'_{sv}$ support vectors, and α_i their coefficients – all values found during the optimization phase (training). In the test phase, a test vector x is assigned to one

TABLE V
CLASSIFICATION ACCURACIES OF THE VARIOUS TRAFFIC CLASSES BASED ON MULTIPLE CONTIGUOUS FRAMES COVERING $T = 5s, 10s, 15s, 20s, 25s, 30s$ OF THE CUMULATIVE ACOUSTIC SIGNAL. FEATURE: MFCC500 (PRIMARY WINDOW=$500ms$, SHIFT=$100ms$)

Traffic Class	**T=5s**	**T=10s**	**T=15s**	**T=20s**	**T=25s**	**T=30s**
Jammed	95.4	98.1	99.7	100.0	100.0	100.0
Medium-Flow	81.0	89.7	93.1	95.6	96.6	97.7
Free-Flow	90.1	96.4	98.3	98.5	98.8	100.0

TABLE VI
SINGLE FRAME LEVEL CLASSIFICATION ACCURACIES OF THE VARIOUS TRAFFIC DENSITY CLASSES USING SVM. THE GMM-BAYES CLASSIFIER ACCURACIES ARE ALSO PRESENTED FOR COMPARISON. THE FEATURE WINDOW SIZE IS $500ms$.

Traffic Class	SVM	GMM-Bayes
Jammed	73.0	73.3
Medium-Flow	80.7	63.9
Free-Flow	77.5	71.4

TABLE VII
SVM CLASSIFICATION ACCURACIES OF THE VARIOUS TRAFFIC CLASSES BASED ON MULTIPLE CONTIGUOUS FRAMES COVERING $T = 5s, 10s, 15s, 20s, 25s, 30s$ OF THE CUMULATIVE ACOUSTIC SIGNAL. FEATURE: MFCC500 (PRIMARY WINDOW=$500ms$, SHIFT=$100ms$)

Traffic Class	**T=5s**	**T=10s**	**T=15s**	**T=20s**	**T=25s**	**T=30s**
Jammed	95.0	97.7	98.9	99.5	100.0	100.0
Medium-Flow	96.1	98.5	99.0	99.4	99.9	100.0
Free-Flow	93.5	97.8	99.0	99.6	99.8	100.0

of the two classes depending on the sign of $'f(x)'$,

$$f(x) = \sum_{k=1}^{N_{sv}} \alpha_k y_k \Phi(s_k).\Phi(x) + b$$
$$= \sum_{k=1}^{N_{sv}} \alpha_k y_k K(s_k, x) + b$$

where $K(s_k, x) = \Phi(s_k).\Phi(x)$ directly computes the dot product between the transformed features $\Phi(s_k)$ and $\Phi(x)$.

In our experiments, we have used a radial basis function (RBF) kernel based SVM to perform the above classification task. The training and the test sets are the same as in the GMM based Bayes classifier in the subsection IV-C. The RBF kernel function used is,

$$K(x_i, x_j) = \exp\left(-0.5(x_i - x_j)^2\right) \quad (9)$$

where, x_i, x_j are the feature vectors and $K(x_i, x_j)$ is their dot-product in some high dimensional space as determined by the choice of the RBF kernel function. In general, the RBF kernel functions provide good classification performance and hence we have used them in these experiments. The LibSVM library[25] was used for training the SVM parameters and for testing. Tab. VI, shows the classification accuracy of a single frame (single feature vector) based on the SVM classifier for the MFCC500 feature which was one of the best performing feature for the Bayes classifier. As can be noted from this table, the SVM classifier achieved better performance than the GMM-Bayes classifier. We then, used multiple contiguous feature vectors over the time span ($T = 5, 30s$). Unlike the GMM-Bayes classifier which provided a likelihood for each feature vector (time frame) as in (6), RBF-SVM classifier only provided the class label (one out of three classes) to which it assigned a feature vector x_t. Therefore, we used a simple voting technique to arrive at a single class label for all the contiguous feature vectors that spanned $T = (5, 30)s$ of the cumulative acoustic signal. The classification accuracies are in Tab. VII. We note two important trends from this table. First, the SVM classification accuracies are higher than GMM-Bayes accuracies with the same MFCC500 features as in Tab. V. Second, SVM provides particularly better accuracy for the *medium-flow* class. This could be due to the better inter-class discrimination provided by the SVM in a non-linear and higher-dimensional space as determined by the RBF kernel in (9).

In summary, both the GMM-Bayes classifier and the SVM provided high accuracies while looking at the acoustic evidence of the order $20 - 30s$, with the SVM providing additional gains over the GMM-Bayes classifier. In terms of computational complexity, GMM-Bayes classifier may be preferable over the SVM as in the latter case the test-phase computational expense is directly proportional to the number of support vectors. The number of support vectors in the above experiments were $N_{sv} = 25113$ out of a total of $N = 51822$ training vectors. In comparison the GMM-Bayes complexity is limited by the number of Gaussian components in the mixture, which in this case, was 11 components per class.

VI. CONCLUSION

We have presented a simple technique that uses the discriminant information present in the MFCC features of the cumulative roadside acoustic signal, to classify the traffic density state (*Jammed, Medium-Flow, Free-Flow*) that is most likely to have produced it. As this technique uses simple omni-directional microphones (installed on the outer side of a street), its cost of installation and operation is significantly lower as compared to other techniques such as the magnetic loop detectors[19]. And, unlike the computer vision and tracking

based traffic monitoring techniques[16], [18], [17], the proposed technique is independent of the lighting conditions and the visual occlusion problem. Therefore, it works reasonably well even in the chaotic and *non-lane* traffic conditions that are very common in the developing regions (South Asia/South east Asia). As has been noted in [14], [15], honks are very commonplace in the traffic of the developing regions. Their usage is subtly ingrained within the driving culture of the developing countries (especially India, South Asia) where the drivers use it partly as a "safety measure" to "announce" their presence. However, unlike the techniques in [14], [15], which require robust and accurate detection of the honk signals, followed by the Doppler frequency shift estimation to arrive at the average speeds, our proposed technique is not based on the honk detection. We instead, extract the distinctive acoustic signatures of the various traffic density classes, which are characterized by the multiple types of noises induced by the vehicles, and may or may not include the honks.

In this background, our technique provides several advantages in terms of the low cost and high accuracy operation for a three class traffic density classification; (*Jammed:* $(0 - 10Kph)$, *Medium-flow:* $(10 - 40Kph)$ and *Free-Flow:* $(40 - 100Kph)$). However, we note that unlike the techniques such as the radar guns and a single vehicle profile extraction[1], [12], [13], the proposed technique cannot provide a very fine-grained average speed estimation. Nevertheless, we believe that this may not be a limitation for the traffic congestion monitoring in the developing regions. In such conditions, we are not interested in the precise average speeds of the traffic, but rather in its broad range of the average speeds – a quantity well captured by our three broad traffic density classes.

REFERENCES

[1] V. Cevher, R. Chellappa and J. H. McCllelan, " Vehicle Speed Estimation Using Acoustic Wave Patterns ", IEEE Trans. on Signal Processing, Vol. 57, No. 1, Jan 2009.

[2] S. A. Amman and M. Das, "An Efficient Technique for Modeling and Synthesis of Automotive Engine Sounds, " IEEE Trans. Ind. Electron., Vol. 48, pp. 225-234, Feb. 2001.

[3] U. Sandberg, "Tyre/Road Noise – Myths and Realities, " In Proc. 2001 Int. Congress and Exhibition on Noise Control Eng., Hague Netherlands, Aug 27-30, 2001.

[4] Road Directorate-Ministry of Transport, "Noise Reducing Pavement, " Road Directorate, Danish Road Institute Tech Report 141, Apr 2005.

[5] U. Sandberg and A. J. Ejsmont, "Tyre/Road Noise Reference Book, " Kisa, Sweden: Infomex, 2002, SE-59040

[6] R. A. G. Graf, C. Y. Kuo, A. P. Dowling, and W. R. Graham, " On the Horn Effect of a Tyre/Road Interface -Part I: Experiment and Computations, " J. Sound. Vibrations., Vol. 256, pp. 433-445, 2002.

[7] R. E. Eskridge, and J. C. R. Hunt, "Highway Modeling. Part I: Prediction of Velocity and Turbulence Fields in the Wake of Vehicles, " Amer. Meteorolog. Soc., Vol. 79, pp/. 387-400, 1979.

[8] N. Sarigul-Klijn, D. Dietz, D. Karnopp, and J. Dummer, "A computational Aeroacoustic Model for Near and Far Field Vehicle Noise predictions, " New York: The Amer. Inst. Aeronaut. Astronaut., 2001.

[9] J. G. Lilly, "Engine Exhaust Noise Control [Online], " Available: http:www.ashraeregion7.org

[10] S. Chen, Z. Sun and B. Bridge, "Traffic Monitoring Using Digital Sound Field Mapping ", IEEE Trans on Vehicular Technology, Vol. 50, No. 6, Nov 2001.

[11] J. Lee and A. Rakotonirainy, "Acoustic Hazard Detection for Pedestrians With Obscured Hearing, " To appear in IEEE Trans. on Intelligent Transportation Systems.

[12] B. G. Quinn, " Doppler Speed and Range Estimation Using Frequency and Amplitude Estimates, " J. Acoust. Soc. Amer., Vol. 98, pp. 2560-2566, Nov. 1996

[13] C. Couvreur and Y. Bresler, "Doppler-based Motion Estimation For Wide-band Sources From Single Passive Sensor Measurements, " In the Proc .of IEEE ICASSP, pp. 21-24, Apr. 1997.

[14] R. Sen, B. Raman and P. Sharma, " Horn-Ok-Please ", In the Proc of ACM MobiSys 2010, San Francisco, USA.

[15] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich Monitoring of Road and Traffic Conditions Using Mobile Smartphone ", In Sensys, Nov 2008.

[16] R. Cucchiara, M. Piccardi, and P. Mello, " Image Analysis and Rule Based Reasoning for a Traffic Monitoring System, " IEEE Trans on Intelligent Transportation Systems, vol. 1, no. 2, June 2000.

[17] S. Kamijo, Y. Matsushita, K. Ikeuchi, "Traffic Monitoring and Accident Detection at Intersections, " IEEE Trans. on Intelligent Transportation Systems, vol. 1, no. 2, June 2000.

[18] B. Coifman, D. Beymer, P. McLaughlin, J. Malik, "A Real-time Computer Vision System for Vehicle Tracking and Traffic Surveillance, " Transportation Research Part C 6, pp 271-288, 1998

[19] D. I. Robertson *et. al.* "Optimizing Networks of Traffic Signals in Real Time – The SCOOT Method, " IEEE Trans. Veh. Technol., Vol. 40, No.1, Feb. 1991.

[20] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition, " Data Mining and Knowledge Discovery, 2, pg. 121-167, Kluwer Academic Publishers, Boston, 1998.

[21] V. Vapnik, "The Nature of Statistical Learning Theory, " Springer-Verlag, New York, 1995.

[22] C. Cortes and V. Vapnik, "Support Vector Networks, " Machine Learning, 20:273-297, 1995.

[23] V. Banlz et. al., "Comparison of View Based Object Recognition Algorithms Using Realistic 3D Models, " In C. von der Malsburg et. al. eds., Artificial Neural Networks-ICANN'96, pg.251-256, Berlin 1996. Spring Lecture Notes in Computer Science, Vol. 112.

[24] M. Schmidt, "Identifying Speaker with Support Vector Networks, " In Proc. of Interface'96, Sydney 1996.

[25] C. Chang and Chih-Jen Lin, "LIBSVM: A Library for Support Vector Machines, 2001." Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm

[26] J. Q. Li and A. Barron, "Mixture Density Estimation, " Advances in Neural Information Processing Systems 12, S. A. Solla, T. K. Leen, and K. R. Muller, eds., 1999.

[27] A. Rakhlin, D. Panchenko, and S. Mukherjee, "Risk Bounds for Mixture Density Estimation, " ESAIM: Probability and Statistics, Vol. 9, pp. 220-229, June 2005.

[28] D. A. Reynolds, "Speaker Identification and Verification Using Gaussian Mixture speaker Models, " Speech Communication 17(1-2): 91-108 (1995).

[29] A. Michael Noll, "Short-time Spectrum and Cepstrum Techniques for Vocal-Pitch Detection, " Journal of Acoustical Society of America, Vol.36, Issue 2, pp. 296-302.

[30] S. Davis, P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, " IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 28, Issue: 4, Aug. 1980.

[31] V. Tyagi, I. McCown, H. Misra and H. Bourlard, "Mel cepstrum modulation spectrum (MCMS) features for robust ASR, " In the Proc. of IEEE Worskhop on Automatic Speech Recognition and Understanding, pp.399-404, Dec 2003.

[32] L. Rabiner, "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition ," Proc. of IEEE, Vol. 77, No. 2, February 1989.

[33] F. Jelinek, "Continuous Speech Recognition by Statistical Methods, " Proc. of IEEE, Vol. 64, pp.532-536, Apr. 1976.

[34] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm, " J. Royal Statist. Soc. Ser. B., 39, 1977.

[35] J. A. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, " ICSI and Computer Science Division, Dept. EECS, U.C. Berkeley, TR-97-021, April 1998.

Vivek Tyagi Vivek Tyagi is a Research Scientist in the Next Gen Systems & Smarter Planet Solutions Department at IBM India Research Labs. Prior to this, he was a Senior Research Scientist in the speech recognition R&D dept at the Nuance Communications Inc. Montreal, Canada. He received a B.Tech

degree in Electrical Engineering from the Indian Institute of Technology, Kanpur, India in July 2001 followed by a Ph.D. degree in 2006, at the School of Computer and Communications Sciences Dept, Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland. His research interests are in large scale speech recognition, machine learning with applications in speech recognition, signal processing, and recently in acoustic sensing for Smart Traffic and other environment sensing related applications. Along with his co-authors, he received the 2009 International Speech Communication Association (ISCA) Best Journal Paper award among all the papers published in the Elsevier Speech Communication journal during the period 2006-2008. He is a Member of IEEE and ISCA.

Shivkumar Kalyanaraman Shivkumar Kalyanaraman is a Senior Manager of the Next Gen Systems & Smarter Planet Solutions Department, and Senior Researcher at IBM India Research Labs Bangalore. He was a Manager of the Next Generation Telecom Research group and a Research Staff Member since 2008. Previously he was a Professor at the Department of Electrical, Computer and Systems Engineering at Rensselaer Polytechnic Institute inTroy, NY. He received a B.Tech degree in Computer Science from the Indian Institute of Technology, Madras, India in July 1993 followed by M.S. and Ph.D. degrees at the Ohio State University in 1994 and 1997 respectively. He also holds an Executive M.B.A. (EMBA) degree from Rensselaer Polytechnic Institute (2005). His current research in IBM is at the intersection of emerging wireless technologies and IBM middleware and systems technologies with applications to large-scale smarter planet problems (grids, traffic, finance etc). He was selected by MIT's Technology Review Magazine in 1999 as one of the top 100 young innovators for the new millenium. He served as the TPC Co-chair of IEEE INFOCOM 2008, and as the General co-chair of ACM SIGCOMM 2010 in New Delhi. He is on the editorial board of IEEE/ACM Transactions of Networking. He is a Fellow of IEEE and an ACM Distinguished Scientist.

Raghuram Krishnapuram Raghu Krishnapuram received his B. Tech degree from IIT-Bombay and his Ph.D. degree in electrical and computer engineering from Carnegie Mellon University. From 1987 to 2000, he served as a faculty member at the University of Missouri, Columbia, and Colorado School of Mines,Golden, Colorado. Dr. Krishnapuram joined IBM Research - India in 2000. His research encompasses many aspects of analytics, optimization, computational intelligence, and data mining. He has published over 160 papers in journals and conferences in these areas. He is currently Associate Director, Solutions, IBM Research - India, where he is leading research projects in the areas of information management, mobile solutions, IT/Wireless convergence, and analytics & optimization, with a particular focus on emerging markets. Dr. Krishnapuram is an IEEE Fellow, and a member of IBM's Technology Council of the IBM Academy of Technology.