DOPPLER BASED SPEED ESTIMATION OF VEHICLES USING PASSIVE SENSOR

Shubhranshu Barnwal[1], Rohit Barnwal[1], Rajesh Hegde[1], Rita Singh[2], Bhiksha Raj[2]
1. Indian Institute of Technology Kanpur, India; 2. Carnegie Mellon University, USA
{barnwal, rbarnwal, rhegde}@iitk.ac.in; {rsingh, bhiksha}@cs.cmu.edu

ABSTRACT

This paper aims to develop a system for estimating a vehicle's speed by analyzing its drive by acoustics with a passive audio microphone. Analysis of the vehicle's acoustics would primarily use the phenomenon of Doppler shift, and the instant at which vehicle is at closest-point-of approach. This approach uses a technique called Seam carving to track harmonics formed by vehicle particularly its engine noise. The method proposed is computationally inexpensive and can very easily be developed into mobile application.

Index Terms— Speed Estimation, Doppler Effect, Engine Acoustics, Seam Carving

1. INTRODUCTION

The estimation of the motion of a point source from the signal received at fixed passive sensors is a classical problem in statistical signal processing and array processing. In a classical Doppler system, like radar or active sonar, the signal emitted by the source is in fact a reflection of a signal generated by the active sensor. In this paper we formulate a solution for Doppler-based motion estimation from measurements obtained from a single passive sensor. Our formulation is motivated by the following practical application: how to estimate the speed and position (possibly time varying) of a vehicle on a known path (like a car on a road) from its acoustic signature at a microphone located next to its path.

This effect of Doppler shift on automotive sounds has previously been studied by Couvreur and Bresler [1]. The speed-related Doppler shifts in the sound produced by the vehicle manifests in the form of ``seams'' in spectrographic characterizations of the audio. We propose that the speed and position of the vehicle can be estimated by analysis of these seams through seam tracking techniques. Analysis of seams in image-like representations is by itself not a new concept. In fact, seams of information across images have been effectively used in image processing for resizing images, and for changing image content non-linearly without affecting the overall image quality [2]. Spectrographic Seams have also been used to characterize feature vectors of spoken words for keyword spotting [3]. However, to the best of our knowledge, they have not been previously used for audio analysis tasks such as determining velocities from Doppler signatures.

The following paper is arranged as follows: in section 2 we present the basic theory involved behind the structure of harmonics formed in a spectrogram of vehicle's drive by acoustics. In section 3 we show how seams can be used to track harmonics and then be used to estimate speed. In section 4 we describe the Doppler Effect followed by its application in speed estimation. Finally in section 6 we present our experimental setup and results followed by conclusions and future work.

2. VEHICLE ACOUSTICS

The sources of sound in a vehicle are many like vehicle induced air turbulence, friction of vehicle tires with ground, etc., but the primary contributor is its engine. Thus the spectral content of the vehicle's acoustics include both wideband processes and harmonic components.

2.1. Spectral Characteristics of vehicle acoustics

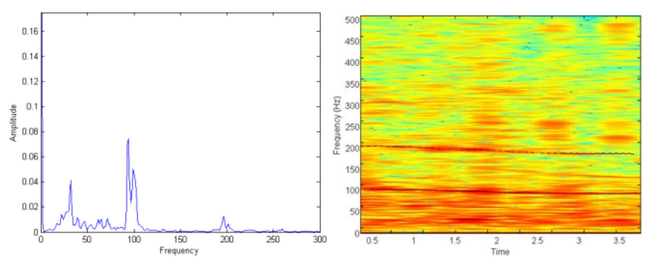

Figure 1. The harmonic components (f_0 and F_0) can be seen as a clear peak in the spectrum of vehicle acoustics (left); Spectral peak on spectrogram of vehicle's pass-by acoustics (right).

The noise generated from an internal combustion engine contains a deterministic harmonic train and a stochastic component so it can be modeled by the same methods used for human speech. The stochastic component of the engine noise is largely due to the turbulent air flow in the air intake (or intercooler), the engine cooling systems, and the alternator fans. This stochastic component is wideband in nature. The deterministic component is caused by the fuel combustion in the engine cylinders and has more power than the stochastic component. The lowest deterministic tone is called the cylinder fire rate f_0, defined as the firing rate of any one cylinder in the engine. Since each cylinder fires once every two engine revolutions in a four-stroke engine, there is a simple relationship between f_0 and the *RPM* of a vehicle:

$$f_0 = RPM / (60 \times 2) \text{ Hz} \tag{1}$$

The strongest tone in the engine noise is called the engine fire rate F_0, and it is related to f_0 in a simple manner:

$$F_0 = f_0 \times p \tag{2}$$

where p is the no. of cylinders in the engine [4].

The spectrum of acoustics of a moving vehicle received at a stationary observer is time varying and there are mainly two factors for this change:

1) Change in engine *RPM* (a result of varying vehicle speed or a gear shift), which ultimately results in change in the emitted f_0 and F_0
2) The Doppler effect, which either positively or negatively shifts the entire spectrum depending on the speed of source and observer relative to each other

In this paper we would only be considering the case of vehicles with constant speed, meaning a change in engine RPM does not occur. The shift in the spectrum due to the Doppler Effect can then be used for estimating the speed of the vehicle.

2.2. Spectrographic representation of vehicle acoustics

In order to analyze time-varying spectral variations from Doppler shifts we will first need a spectral characterization of the signal in which these variations are clearly apparent. For this we use a spectrographic representation of the signal.

A spectrogram is a visual representation of the short-time Fourier transform of a signal. A digital signal $x[n]$ is converted to a matrix of time-frequency components $X[\tau,f]$, where each time-frequency component is computed as:

$$X[\tau, f] = \sum_{n=\tau K}^{\tau K+N-1} x[n]w[n-\tau K]\ e^{\frac{-j2\pi nf}{N}} \tag{3}$$

where $w[n]$ is a window function, N is the size of the window and K is the shift between adjacent analysis windows. In practice we operate on $log\,|X[\tau,f]|$, the *log magnitude* of the spectrogram.

In order to obtain the optimal representation for tracking velocities, it is essential to utilize the appropriate parameter settings for the spectrogram. We employ a Hamming window for the analysis window. Frame shifts are set to be a quarter of the frame width. The critical component is the analysis frame width N. Long analysis windows, which result in so-called "narrow-band" spectrograms, enable fine characterization of spectral variations and reveal individual harmonics, which is essential for detecting distinct spectral peaks. On the other hand, this comes at the cost of time resolution. Shorter analysis window – which result in "wide-band" spectrograms – provide greater time resolution, but at the cost of frequency resolution. For our work we have found narrowband analysis to be most effective, as the ability to distinguish spectral peaks is important for our task, as we explain in our experiments section.

3. SEAM CARVING

As Figure 1 shows, Doppler frequency shifts in the recorded sound from an automobile manifest as "seams" – smooth curves that represent the variation of spectral peaks with time – in the spectrogram. In order to analyze the velocity pattern of the vehicle, it is sufficient to locate and characterize these seams. The seams have a distinct characteristic: if the spectrogram is viewed as an image, within the seam the value of the pixels does not vary dramatically from on instant to the next. On the other hand, the pixel values change drastically as we move from the seam, away from it. Thus the seam may be viewed as a *local* minimum-variation trajectory that traverses the breadth of the spectrogram. We will employ a technique called *seam carving* [2] to trace it.

3.1. Computing seams

We view the spectrogram as a two dimensional matrix of energy terms, where $E[i,j]$, the energy in the i^{th} row and j^{th} column of the matrix, is given by $E[i,j] = |X[\tau,f]|^2$. We compute seams through this matrix (which may also be viewed as an image with pixel elements equal to $E[i,j]$ using a simple dynamic programming technique. We use just three parameters while computing the seams – the element energy $E[i,j]$, the cumulative energy of the best seam leading up to that element, denoted by $C[i,j]$, and path matrix, denoted by $P[i,j]$, which stores the index to the preceding element in the best seam to (i,j).

$$C[i,j] = E[i,j] + max\{C[h,j-1]\},\ where\ h \in [i-1,i+1] \tag{4}$$

At the end of this process, the index of the maximum value of the last column of matrix C indicates the end of the best overall seam. In the final step, we backtrack from this maximum cumulative energy cell to find the required seam. Note that the process results in seams that only include one pixel from every column of the spectrogram, *i.e.* one frequency at each instant.

3.2. Seam smoothing

Since a spectrogram is much coarser in the information it represents per pixel and in pixel continuity than a standard image of the same size, the seams obtained on a spectrogram are usually jagged. To smooth these out, we use a very basic smoothing filter where a linear penalty is now imposed while computing the seam's path as follows:

$$C[i,j] = E[i,j] + max\{Pen[k,j].d.C[k,j]\} \tag{5}$$

where d is deviation distance, and $k \in [i-1,i+1]$.

The penalty imposed not only smoothen the seams, but also forces the seams to not change tracks to harmonics formed above or below the currently tracked harmonic. Seams are thus rendered more robust in spatial location on the spectrogram.

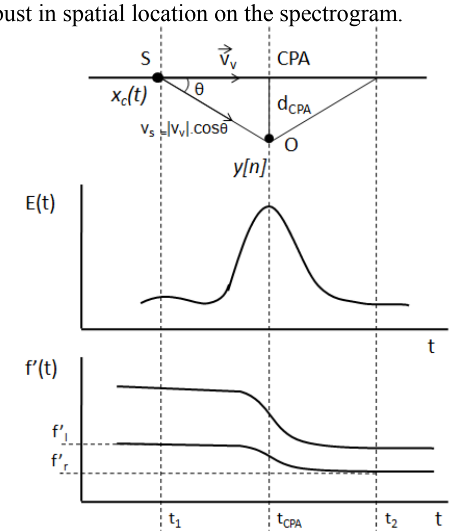

Figure 2. Geometry of our problem; source moving at a constant speed on a straight path (top). Plots of energy of vehicle acoustics received at sensor (middle) and Doppler shift of vehicle's acoustic harmonics (below) with time for constant velocity constraint

4. DOPPLER EFFECT

A well known phenomenon, the Doppler Effect, is observed when signals (in our case, sound) are emitted or received by objects moving relative to each other. The Doppler Effect law states that if an object emits a signal and moves relative to an observer, the frequency of the observed signal will be Doppler shifted and the magnitude of this shift depends on the frequency of the signal and the velocity of the source and observer relative to

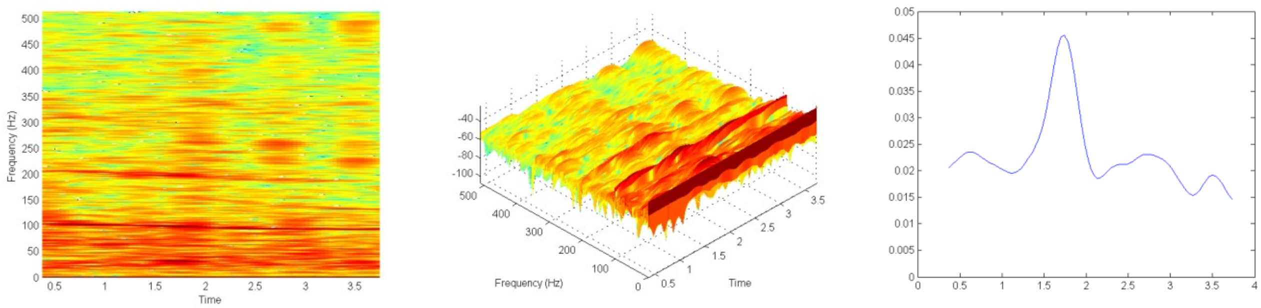

Figure 3. Seams tracking harmonics on spectrogram of engine acoustics with constant speed (left), isometric view of the spectrogram of engine acoustics with constant speed (middle), plot energy of vehicle acoustics at constant speed with time

Table 1. Comparison of estimated speed to actual speed for different variations of vehicles, speed and dCPA from 1 seam

	BMW (Car)	BMW (Car)	Peug (Car)	Panda (Car)	Bus	Yamaha (Bike)	Yamaha (Bike)
Actual speed (approximately) (in kmph)	20	80	40	50	40	40	50
Estimated Speed (Mic.1) (in kmph)	27	83	45	50	38	39	47
Estimated Speed (Mic. 2) (in kmph)	15	83	26	50	41	NA	NA

each other. The equation for this Doppler shift is given as

$$f' = \frac{c + v_o}{c + v_s} \times f^0 \quad (6)$$

where, v_o and v_s are component of velocities of observer and source relative to each other, f^0 is the real frequency generated from source, while f' is the apparent frequency observed at observer. c is the speed of the signal in the medium, in our case signal being sound. For this work we have assumed $c=343m/s$, which is the speed of sound in dry air at 20℃ (68°F) and at sea level. If v_o, $v_s << c$, then relation (6) can be reduced to

$$f' - f^0 = \frac{v_o - v_s}{c} \times f^0 \quad (7)$$

Our current case deals with a vehicle (like car or a bus) moving with constant speed v_v on a straight road, and a stationary observer with $v_o=0$ as illustrated in the figure 2. Let $x_c(t)$ be the continuous signal emitted by our source, and $y[n]$ be the sampled signal recorded by the fixed microphone situated d_{CPA} *mts* away from the *closest point of approach* (CPA) on the road. The observed frequency in $y[n]$ would then be Doppler shifted by relation (7) and the observed energy, $E(t)$ of $y[n]$ would have an attenuation governed by the following relation (8), where E^0 is the emitted energy, and t_{CPA} is the time at the instant where vehicle passes through CPA

$$E(t) = E^0 \times \frac{1}{\sqrt{{v_v}^2 (t_{CPA} - t)^2 + d_{CPA}^2}} \quad (8)$$

5. SPEED ESTIMATION

As described earlier, speed estimation will have two parts to it – due to change in engine's rpm and due to a Doppler shift of the observed signal. For this paper, we will be considering cases of vehicles with constant speed only, thus restricting our case only to Doppler shift.

5.1. Finding the instance at which vehicle crosses CPA

Before using Doppler Shift for speed estimation, we would require to know when the vehicle passes through the CPA. This is necessary because Doppler shift is not significantly present when the vehicle is on either side of the CPA. From equation 8 we can infer that when energy of vehicle acoustic received at observer is maximum, it would be expected that vehicle would be on CPA at that instant. Hence t_{CPA} could be estimated from the following method

$$t_{CPA} = \arg\max_t \left(\sum_f \|X(t,f)\|^2 \right) \quad (8)$$

5.2. Obtaining estimated velocity from seams

For a constant speed v_v, if we are able to determine f' before (f_l') and after (f_r') vehicle passing through the CPA, we can calculate its speed by the following relation (9).

$$|v_v|.\cos\theta = c \times \frac{f_l' - f_r'}{f_l' + f_r'} \quad (9)$$

Seams are expressed as a sequence of time-frequency pixel locations (t,f). From the previous subsection we have already estimated the value of t_{CPA}. Equidistant points $t_1=t_{CPA}-t$ and $t_2=t_{CPA}+t$ are then used for estimating f_l' and f_r'. f_l' corresponds to the frequency at t_1 (t_1,f_l') and similarly f_r' corresponds to the frequency at t_2 (t_2,f_r'). This can be clearly understood by referring to figure 2. If a sufficiently large value of t is chosen, cosθ can be approximated to 1, as distance travelled by car would be sufficiently large compared to d_{CPA}. Empirically, on our experimental data set, we found 't' as 2 seconds (and more) gave us the best results. For real deployment this however would have to be optimized experimentally. If d_{CPA} in itself is large, one would require knowing d_{CPA} before hand and estimating d_{CPA} can altogether be a separate problem. To shift 't' seconds in time domain, on spectrogram we will have to shift our index by a factor of $t.fs/K$, where K is the shift between adjacent analysis windows and fs is the sampling frequency of the signal used to construct the spectrogram. Details about choice for K and fs are discussed in next section.

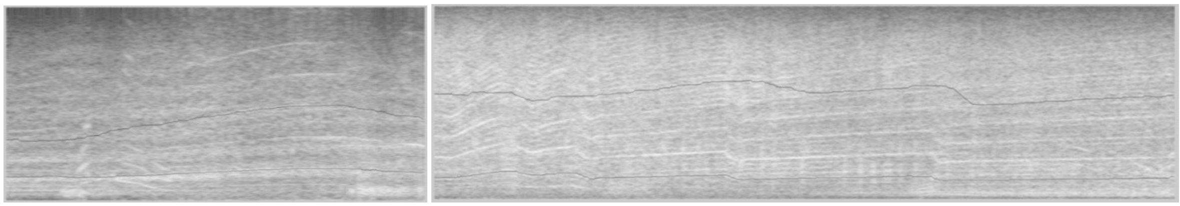

Figure 4. Harmonic tracking seams on spectrogram of engine acoustics with increasing speed without gear shift (left) and increasing speed with gear shift (right)

Other assumptions made while using relation (9) include
1) The vehicle is a point source
2) The seam represents a time-varying pure tone emitted by the vehicle.

6. EXPERIMENTS AND RESULTS

This speed estimation approach has been tested on data used by Valcarce et al [5][6]. They had used two omni-directional microphones mounted on a 6.4 mts tall pole, with a d_{CPA} of 14.5 mts and 17.3 mts to the two lanes of road. Their data was recorded with a sampling frequency of 14.6 kHz. This speed estimation algorithm has also been tested on data collected by author. It was collected using a single omni-directional microphone mounted on a 2.3 mts tall pole, with a d_{CPA} of 5 mts to the center of road. This data was recorded with a sampling frequency of 16 kHz.

Our approach was used on spectrograms obtained using different combinations of sampling frequency and FFT-size. Sampling frequency ranged from 1000 to 8000 Hz, while FFT-size was ranged from 512 to 2048. Also different window sizes were considered to compare seams on narrowband vs. wideband spectrogram. Too small a window, and we won't find the spectral peaks as clear seams; too long, and the changes in the velocity will result in smearing of the peaks. The optimal setting must hence be determined empirically.

The best results were obtained on a combination of using a FFT-size of 1024, and a window size of ¾ times of FFT-size on a signal down sampled to 1024 Hz. The use of 1024 Hz sampling frequency resulted in best narrowband spectrograms, for everyday use cars, as most of the harmonics for such cars are in the lower frequencies. For estimating speed of F1 and other exotic cars, sampling frequency may be down sampled to 8000Hz only as these cars have higher speeds and higher engine RPM (as observed on data from YouTube). The use of 1024 FFT-size gave the best frequency resolution. Comparisons of various parameters along with other interesting results are supplied with a supplementary presentation.

Generally, the estimated speed of vehicles varied by 0-10kmph from the actual speed. Error was higher typically for low speed (like 10-30kmph) since the low frequency of harmonics results in lower Doppler shit, and/or presence of noise in lower frequencies masks such harmonics. The Doppler shift is proportional both to vehicle velocity and signal frequency (primarily the engine RPM). In turn, vehicle velocity and engine RPM are directly related to each other by the relation

$$v_v = \frac{c_{t \times RPM}}{gr} \qquad (9)$$

where, v_v is the vehicle speed, c_t is the circumference of vehicle's tire, and gr is the current combined transmission and differential gear ratio of the engine. For higher speeds a larger number of harmonics is dominant on the spectrogram, and thus the average of speed estimation from multiple seams can result in more accurate estimation. We also compared results of lower gear vs higher gear for the same speed. Harmonics of a lower gear are louder and higher in frequency and thus display a larger shift. This gives more accurate results, as evident.

7. CONCLUSION AND FUTURE WORK

Evaluation on data has been done for simplistic cases, and this shows promising results. Estimates are within an acceptable range given that a passive sensor is being used and the approach is computationally inexpensive.

In a few cases the harmonics were not dominantly visible and preprocessing of audio to enhance spectral peaks is a viable option. For the controlled experiments reported in this paper, though, such preprocessing was not required. One can also notice clearly in figure 4, when a gear shift takes place (both the up shift, and down shift), and state if the speed is increasing or decreasing. We will be exploring approaches to combine Doppler Effect and the engine RPM–vehicle speed relation to track variable vehicle speeds. A state space model for such calculation can be used as we hypothesize that a combination of dv/dt and df/dt would sufficiently characterize given seam trajectories.

8. ACKNOWLEDGEMENT

The authors would like to express our gratitude to Valcarce et al. for allowing us to use their data set.

9. REFERENCES

[1] Christophe Couvreur and YoramBresler, "Doppler-based motion estimation for wide-band sources from single passive sensor measurements," ICASSP 1997
[2] ShaiAvidan and Ariel Shamir, "Seam Carving for content aware resizing," SIGGRAPH 2007
[3] ShubhranshuBarnwal, Kamal Sahni, Rita Singh and Bhiksha Raj, "Spectrographic Seam Patterns for Discriminatve Keyword Spotting," ICASSP 2012
[4] VolkanCevher, Rama Chellappa, James H. McClellan, "Vehicle Speed Estimation using Acoustic Wave Patterns," IEEE Transactions on signal processing 2009
[5] F. Perez-Gonzalez, R. Lopez-Valcarce and C. Mosquera, "Road Vehicle Speed Estimation from a two microphone array," ICASSP 2002
[6] R. Lopez-Valcarce, C. Mosquera, and F. Perez-Gonzalez, "Estimation of road vehicle speed using two omnidirectional microphones: A maximum likelihood approach," EURASIP Journal on Applied Signal Processing 2004