Traffic State Detection Using Smartphone Based Acoustic Sensors
Nitakshi Sood, Arshvir Kaur, Dinesh Vij, Naveen Aggarwal
UIET, Panjab University

Objectives
The traffic and number of vehicles on roads are increasing with an unstoppable pace, which in turn leads to the problem of traffic congestion. We propose the use of Acoustics to determine:
- How efficient and accurate do acoustic signals turn out to be in the detection of Traffic events.
- If acoustics could be used as a standalone parameter to detect Traffic events.
The figure 1 describes the problem statement in detail.

Figure 1: Traffic State Detection Using Acoustics

Basis
Acoustic signals can prove to be efficient in monitoring the traffic state of the road, as various types of sounds can be acquired from the road like engine noise, honking etc. Smartphones have good microphones, which can acquire acoustic signals without much disruption. Since large numbers of people have smartphones, the traffic state detected using other smartphones can be transferred to them. This can help in saving their time from being stuck in traffic jams.

Methodology
The acquired audio is passed through various steps of preprocessing followed by feature extraction and later the classification process using various classifiers like SVM, NN and GMM. Figure 2 describes the methodology followed in detail.

Figure 2: Overview of Methodology

Features used
Experiments have shown that feature set having features MFCC, SC and EE give the best results of all.

Frame Size Variation
The frame size of 8192 and Hamming window function with overlapping size of 50% give the best results.

Data Collection
Crowd sourced data collected from 7 different Smart Phones using ‘Smart Voice Recorder by SmartMolv1.7’ application
Phones used Nexus Samsung S3/S4/Grand Lenovo, Mi4
1000 Recordings of length 30 seconds approximately for each traffic scene, collected in almost 15 sessions.
Recorded at 16 KHz Sample Rate and 16-bit Bit Rate.
Additional data captured such as location of the phone, whether the phone was in hand/pocket/bag, associated AndroSensor file, number of microphones in the phone that was used for recording, air (still or turbulent), and environmental conditions.

Two- Level Classification
The acoustic data collected from commuter’s smartphone is segmented into fixed size frames. Various relevant factors affecting the classification accuracy are also tested like frame size, window functions, overlapping size and different combination of features. The frame size of 8192 and hamming window function proved to be more efficient than others. Experiments have shown that feature set having features MFCC, SC and EE, on doing two level classification, results in better classification accuracy of 91.8% with Neural Network and 93.2% with SVM.

Future Scope
- Early Detection: Minimum amount of recording needed by the classifier to correctly identify a scene.
- Triggering Mechanism: Audio would not be recorded continuously. Some event such as accelerometer would trigger the microphone of smartphone.

Features
- Zero-Crossing Rate (ZCR)
- Short Time Energy (STE)
- Root Mean Square (RMS)
- Energy Entropy (EE)
- Spectral Centroid (SC)
- Spectral Roll-off Point (SRP)
- Spectral Flux (SF)
- Mel Frequency Cepstral Coefficients (MFCC)

Busy vs Quiet
- The experiment of classification of two extreme traffic scenes i.e Busy Street and Quiet Street was done with the classifiers mentioned above.
- The SVM showed the best accuracy with RBF kernel and tuned valued of cost and gamma function.
- Results are summarized in table below.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Cong/Nocong</th>
<th>Med/Free</th>
<th>Total accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>92.7%</td>
<td>94.2%</td>
<td>93.2%</td>
</tr>
<tr>
<td>NN</td>
<td>90%</td>
<td>93.5%</td>
<td>91.8%</td>
</tr>
</tbody>
</table>

References