一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
SeMI2019-116(2020-01)

Initial Evaluation of a Compressive Measurement-Based Acoustic Vehicle Detection and Identification System

Billy DAWTON[†], Shigemi ISHIDA[†], Yuki HORI[†], Masato UCHINO[†], Yutaka ARAKAWA[†], and Akira FUKUDA[†]

[†] Faculty/Graduate School of Information Science and Electrical Engineering, Kyushu University

Abstract As society becomes increasingly interconnected, the need for sophisticated signal processing and data analysis techniques becomes increasingly apparent, particularly in the field of Intelligent Transportation Systems (ITS) where various sensing applications generate data at an exponential rate. In this paper, we put a forward a compressive sensing-based system to extract information from passing vehicle sounds sampled at sub-Nyquist rates for Acoustic Vehicle Detection and Identification (AVDI) applications. The obtained compressive measurements are used to detect and identify passing vehicles. Initial evaluation is performed using data obtained from roads on a university campus and with a back-end ADC sample rate of 3 kHz.

Key words Intelligent Transportation Systems (ITS), Acoustic Vehicle Detection, Compressive Sensing (CS), Feature Extraction.

1. Introduction

The past years have seen a marked increase in the development of Intelligent Transportation System (ITS) technologies. A growing number of novel applications such as smart navigation, traffic monitoring, and road safety have been accompanied by a corresponding improvement in overall system performance and efficiency.

However, with this increase in performance comes an increase in computational cost and complexity, requiring more data and processing power than even before. This is particularly apparent in traffic monitoring applications, where the methods used in the detection and identification of vehicles often come with high computational and installation costs. In an effort to mitigate this, low-cost, low-complexity vehicle detection systems based on acoustic sensors have been proposed.

Most recently, the authors have presented a stereo microphone-based vehicle detection and identification in [1]. Despite the low installation costs associated with acoustic sensing, the subsequent analysis and leveraging of the acquired data is often costly in terms of computational complexity, reducing the overall efficiency of the sensing system.

Our aim is to find a way to reduce the amount of data involved at every stage of ITS sensing systems. We are seeking to lower the overall computational cost, complexity, and power consumption when compared to existing setups whilst maintaining high classification accuracy.

To achieve this goal, we make use of a technique called compressive sensing (CS). First presented in [2], CS is a technique that enables the reconstruction of sparse or compressible signals from a reduced set of linear, non-adaptive measurements.

In this paper, we propose a system that takes advantage of the dimensionality reduction properties of CS to acquire vehicle signals at sub-Nyquist sample rates and uses them in conjunction with a range of machine learning techniques for particular use in Acoustic Vehicle Detection and Identification (AVDI) applications.

2. Related Work

Vehicle detection and identification using features extracted from vehicle audio in tandem with supervised learning has been widely explored. Methods using Support Vector Machine (SVM) classifiers [3], k-Nearest Neighbor (KNN) classifiers [4], Gaussian Mixture Models, and Hidden Markov Models [5] used with the frequency domain information of vehicle signals have been proposed. Whilst these systems share a similar goal and basic approach, they differ in their applications, performance and features.

A method for identifying passing vehicles based on the shape of the frequency-domain representation of their sound signature has been proposed in [6]. Instead of using the signal's individual frequency components as features, the proposed system uses information obtained from the frequency domain envelope itself. The unique shape of each passing vehicle's frequency envelope enables the system to accurately distinguish individual vehicles from one another. However, this same uniqueness makes it impossible for the system to identify the type (i.e. the class label) of a passing vehicle.

In [7], a system capable of analyzing the acoustic signature of vehicles independently of any changes in engine speed is presented. By using wavelet packet analysis instead of more traditional time or frequency domain-based techniques and a Multilayer Perceptron (MLP) classifier, the system is able to extract engine speed-independent features from sounds emitted by passing vehicles. This improves system accuracy performance in a range of sensing environments, however the the computational and hardware requirements entailed by the use of a neural network makes it difficult to deploy the system in low-power low-cost situations.

The authors have, in previous works, proposed several acoustic vehicle detection systems. SAVeD, the sequential acoustic vehicle detector put forward in [8] works by fitting S-curve models to points on a sound map using a random sample consensus (RANSAC) estimation method. Once a vehicle is successfully detected, the sound map is purged of the corresponding points, and the detection process is repeated to detect subsequent vehicles. The system F-measure is 83 %. In [1], the authors designed a stereo microphone-based detection system, which identifies passing vehicles based on frequency-domain features extracted from their sound signature. By time-shifting and combining the two signals to produce an emphasized sound signal, vehicle type estimation accuracy is improved, particularly when faced with simultaneously and successively passing vehicles. The system accuracy is 95 %. Whilst both the above systems perform well when compared to existing microphone-based detection methods, the computational cost associated with the two methods is high and makes low-power, embedded applications of these systems difficult.

The authors also propose in [9] an ultra low-power vehicle detector (ULP-VD) capable of detecting passing vehicles with minimal computation cost. This system however is only able to detect the presence of a vehicle and must be used in conjunction with other techniques to identify them.

Traditionally, digital signal processing techniques are performed on a full set of samples acquired by sampling an analog signal at the Nyquist rate. In [10] the concept of using compressive sensing as a tool for signal processing on samples acquired at sub-Nyquist sample rates is explored. The authors find that it is possible to succesfully perform a variety of processes including filtering, detection and classification directly on a reduced set of linear samples, without reconstructing the signal beforehand.

In [11], it is shown that it is viable to use the linear measurements as features in machine learning with only minimal pre-processing; by carefully selecting the sensing matrix parameters, the authors demonstrate that it is possible to obtain enough relevant signal information to detect faulty solenoids.

The authors in [12] put forward a license plate recognition system which uses an SVM to identify the numbers on the plate by sub-sampling the sparse, flattened 1-D representations of images obtained from traffic monitoring cameras.

The above methods serve to illustrate the viability of using a CS measurement-based sensing system for audio signals.

3. Proposal

The aim of the proposed research is to design a supervised learning-based sensing system utilizing CS-based techniques. We are seeking to improve upon existing AVDI methods by exploiting the compressible nature of the signals under consideration to sample them at sub-Nyquist rates, thus reducing the amount of data and computational complexity involved at each successive stage of the detection and identification process.

Current AVDI systems contain, more often than not, a stage presenting relatively high computational complexity. This occurs either prior to the initial detection or classification stage like in [1] or [13] where the use of successive DWTs or DFTs are used to analyze and process the data, or during the classification stage itself where complex supervised learning methods such as deep neural networks (DNN) [14] or MLPs [7] are employed. In either case, this computational cost associated with these stages somewhat mitigates the savings made using acoustic sensing methods. In this paper, we propose a CS-based AVDI system with a pre-processing stage consisting only of successive filtering and mixing, and that performs classification on easy-to-extract features using a simple machine learning classifier.

To the best of our knowledge, there are no currently existing compressive-measurement based acoustic vehicle detection and identification systems.

4. Proposed System

4.1 System Overview

Figure 1 shows the average sound signals obtained from passing cars and scooters, and from periods without a passing vehicle: we can see that the overwhelming majority of the frequency content is contained below 6 kHz, and that we can distinguish the different signal classes by the power contained in their respective frequency components from 3 kHz onwards. Rather than sample the signals at the Nyquist rate, our proposed system uses a CS-based approach to ac-

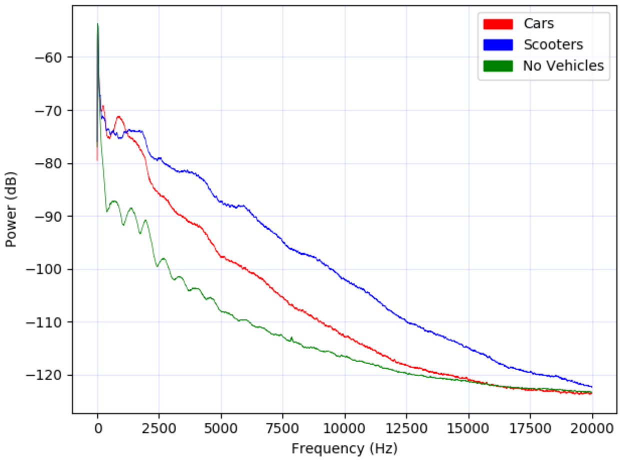

Figure 1 Average audio signals for three vehicle classes

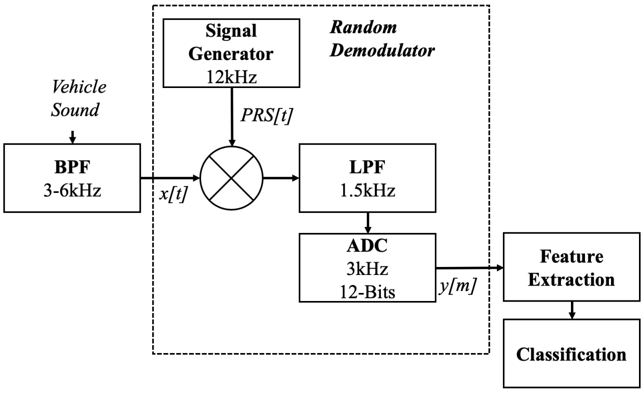

Figure 2 Proposed system overview

quire the information located in the 3–6 kHz frequency band directly at sub-Nyquist rates. Passing vehicles are then detected and identified using features extracted from the samples acquired in this manner. The system can be seen in Figure 2 and is made up of three stages: filtering, random demodulation, and classification. The filtering stage consists of a single band-pass filter (BPF) operating over the 3–6 kHz band and removes unwanted frequency content. The input signal is then combined with a random chipping sequence before being sampled at a sub-Nyquist rate in the random demodulation stage. Finally, in the classification stage, features are extracted from the samples obtained in the previous stage and are used as inputs to a classifier for vehicle type detection and identification. The workings of these stages are explained further in sections 4. 2 and 5. 2. 2.

4. 2 Random Demodulator

The sub-Nyquist sampling performed in the random demodulation stage is done using a CS-based approach. CS as a means for efficiently sampling sparse or compressible signals (a signal can be called compressible if only a small amount of its non-zero components have significant magnitude) was first put forward in [2] and [15]. The procedure can described as follows: $x \in R^N$ is an input signal which can be represented as a combination of a unitary sparsity basis $\mathbf{\Psi} \in C^{N\times N}$ and a K-sparse coefficient vector $\alpha \in C^N$ such as $x = \mathbf{\Psi}\alpha$. We define $y \in R^M$ as the set of linear measurements obtained by performing a sequence of sampling operations represented by $\mathbf{\Phi} \in R^{M\times N}$, such that $y = \mathbf{\Phi}x$ and crucially, $N > M$ (N and M are positive integers). We define $\mathbf{\Theta} \in C^{M\times N}$ as $\mathbf{\Theta} = \mathbf{\Phi}\mathbf{\Psi}$, and $y = \mathbf{\Theta}\alpha$. CS establishes that if $\mathbf{\Theta}$ satisfies the incoherence and RIP (restricted isometry property) conditions outlined in [16], it is possible to recover α, and thus x, from y with much fewer samples than would be required in traditional Nyquist sampling. The recovery process is typically performed using l_1 minimization.

The initial theoretical work on CS only considers discrete signals, however our proposed system looks to obtain continuous-time audio signals which have a sparse or compressible representation in the frequency domain. To that end, our system takes inspiration from an architecture developed by Tropp et al. in [17] called the Random Demodulator (RD), which allows for analog signals to be used in CS applications. The intuition behind the system is as follows: instead of sampling an analog signal at the Nyquist rate, the RD modulates the signal with a random chipping sequence, spreading the K-sparse input signal across the entirety of the frequency spectrum. This smeared signal is then low-passed before being sampled at a sub-Nyquist rate, and the original signal is obtained from these samples via a recovery algorithm.

More formally, our analog input signal of length T_s can be written as combination of discrete coefficients $\alpha \in C^N$ and continuous basis elements ψ_n(which correspond here to the columns of the IDFT matrix) for a given time window:

$$x(t) = \sum_{n=1}^{N} \alpha_n \psi_n(t)\ ,\ t \in [0, T_s) \tag{1}$$

The chipping sequence can be expressed as:

$$PRS(t) = \sum_{n=0}^{W-1} \epsilon_n(t)\ ,\ t \in \left[\frac{n}{W}, \frac{n}{W} + 1\right) \tag{2}$$

ϵ_n is a random sequence which switches between ±1 with equal probability (Rademacher sequence) at or above $f(t)$'s Nyquist rate.

The combined signal $x(t).PRS(t)$ is passed through an LPF $h(t)$ and sampled at a rate R below the Nyquist rate W with $R < W$ to obtain linear compressive samples $y[m]$. In the time domain, this corresponds to a multiplication followed by a convolution:

$$y[m] = \int_{-\infty}^{\infty} x(\tau) PRS(\tau) h(t-\tau) d\tau \Big|_{t=mR} \tag{3}$$

$$= \sum_{n=1}^{N} \alpha_n \int_{-\infty}^{\infty} \psi_n(\tau) PRS(\tau) h(mR-\tau) d\tau \tag{4}$$

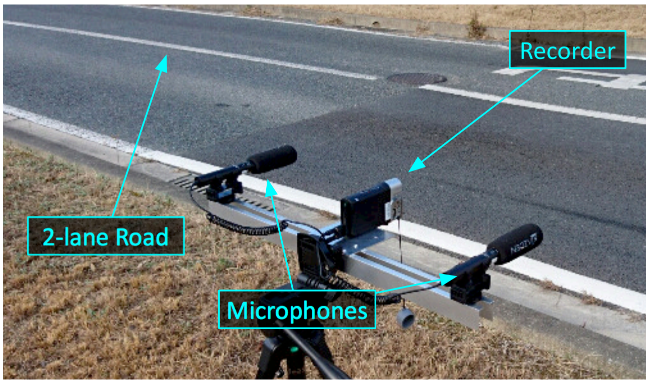

Figure 3 Experimental setup

From which we can obtain the expression for the sensing matrix $\mathbf{\Theta}$, where each entry is defined as $\theta_{m,n}$ for row m and column n.

$$\theta_{m,n} = \int_{-\infty}^{\infty} \psi_n(\tau) PRS(\tau) h(mR - \tau) d\tau \qquad (5)$$

The RD in our proposed system presents two major modifications. First, the presence of the BPF operating over the 3–6 kHz band at the RD's input, whose role is to remove redundant sub-3 kHz information from the signals, improving classification performance, and low-pass the signal to 6 kHz lowering the required chipping sequence frequency to 12 kHz. Second, the absence of a reconstruction stage: we are not looking to reconstruct x and will instead extract features directly from y for use in classification, reducing the computational load of the system by bypassing the computationally expensive reconstruction phase.

It is important to note that in our proposed system the front-end band-pass filtering (3–6 kHz) and mixing (input signal with 12 kHz chipping sequence) are performed in the analog domain, and that the sampling operation only occurs at the end of the signal acquisition process. Thus all pre-processing has been completed by the time the signal is sampled by the back-end ADC.

5. Evaluation

An initial software-based evaluation of our proposed system is performed using audio data collected from the roads on a university campus.

5. 1 Data Acquisition

The data acquisition setup can be seen in Figure 3. Two microphones are installed at the side of a two-lane two-way road at a height of 1m from the ground, parallel to the road and connected to a video camera. The microphones record the sound of passing vehicles for a duration of 20 minutes and the video camera records the ground truth video footage. The microphones used are a pair of AZDEN SGM-990s, recording at a sample rate of 48 kHz and bit depth of 16 bits, the video camera is a SONY HDR-MV1. The intra-microphone distance is 50 cm, the distance between the microphones and the center of the front lane is 3 m, and the distance between the microphones and the center of the back lane is 6 m. The signals received by the two microphones are averaged in order to obtain a single-channel mono signal for use in subsequent analysis. The setup was used to record vehicle sounds on two separate occasions, with the one set of data being used as a training set, and the other as a testing set.

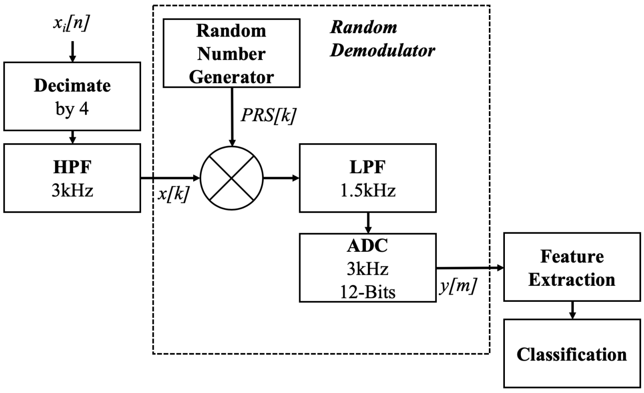

Figure 4 System software implementation overview

The first set contains 178 vehicle sounds: 57 cars, 94 scooters/motorbikes, 25 buses, and 2 trucks. The second set contains 162 vehicle sounds: 63 cars, 76 scooters/motorbikes, 21 buses, and 2 trucks. Classification was performed for 3 classes: cars, scooters/motorbikes, and no passing vehicle; referred to as: “Car”, “Scooter”, and “NoVeh” respectively.

During this initial evaluation, we are only looking to perform classification on individual, non-overlapping vehicle sounds. We call the time at which a given vehicle passes in front of the mid-point between the two microphones as t_p, and define the range $T_r = \left[\frac{t_p - T_s}{2}; \frac{t_p + T_s}{2}\right]$ where $T_s = 2s$. By using information about the t_p of each passing vehicle obtained from the ground truth data, we are able to extract the “Car” and “Scooter” signals whose T_r do not overlap with that of preceding or following vehicles. In this manner we are able to obtain 40 “Car” and 57 “Scooter” signals from the first set, and 52 “Car” and 50 “Scooter” signals for the second set. We obtain “NoVeh” signals by splitting the parts of the signal who do not correspond to the T_r of any vehicle into sections of length T_s. We obtain 115 “NoVeh” signals for the first set and 112 for the second for a total of 212 and 214 signals across the three classes under consideration for the first and second sets respectively.

5. 2 System Simulation

5. 2. 1 Overview

Figure 4 shows the software implementation of the system proposed in Figure 2.

We described in section 4. 1 the real-world analog im-

TABLE 1　System parameters.

Nyquist Rate	R	B	T_s	N	K	M
48 kHz	3 kHz	12 bits	$2s$	96000	24000	6000

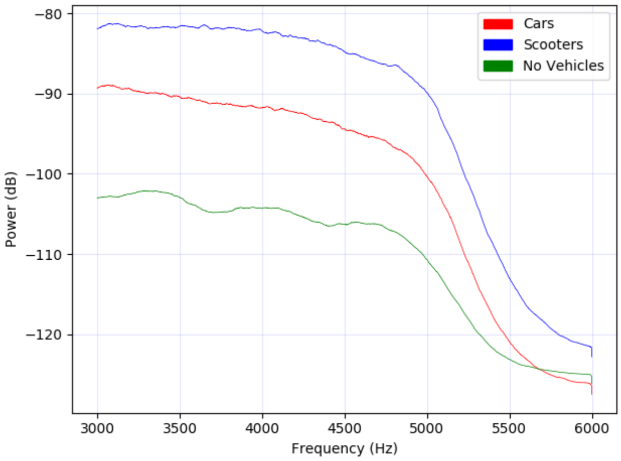

Figure 5　Filtered average audio signals for three vehicle classes

plementation of the system. In this section, we perform an initial evaluation of the system by simulating its operation using a software-based digital-domain representation.

$x_i[n] \in R^N$ denotes a discrete version of the analog input signal $f(t)$ sampled at 48 kHz. The input signal is decimated (low-pass filtered and downsampled) by a factor of 4, bringing the Nyquist rate down to 12 kHz. This downsampled version of the signal is high-pass filtered at 3 kHz through a Type II Chebyshev filter, and the resulting signal is referred to as $x[k] \in R^K$. The chipping sequence $PRS[k] \in R^K$ is a vector containing an equiprobable random distribution of values from the set $\{-1.1\}$. The LPF preceding the ADC is set as a 2nd order Butterworth filter and the linear measurements $y[m] \in R^M$ are obtained by uniformly sampling and quantizing every $\frac{N}{R}$th entry from the combined $x[k] \odot PRS[k]$ signal at rate $R = \frac{12kHz}{4}$ and bit-depth B. When compared to the initial Nyquist rate of 48 kHz, the reduction in sampling rate, and thus in the amount of samples from which we extract the features necessary for classification is $\frac{(\frac{48kHz}{4})}{3kHz} = \frac{N}{M} = 16$.

The action of the successive decimation and high-pass filtering on the frequency content of $x[k]$ can be seen in Figure 5. The sub-3 kHz content is strongly attenuated with a short 500 MHz passband, whereas the roll off towards 6 kHz is much less pronounced. This simultaneous filtering operation has the effect of further sparsifying the input signal by suppressing the unwanted information contained in the signals' lower frequency range, whilst also performing anti-aliasing by attenuating the frequencies above the signal's Nyquist frequency of 6 kHz. As a result, in the unattenuated 3-6 kHz band, the frequency information of each different signal class is clearly distinct.

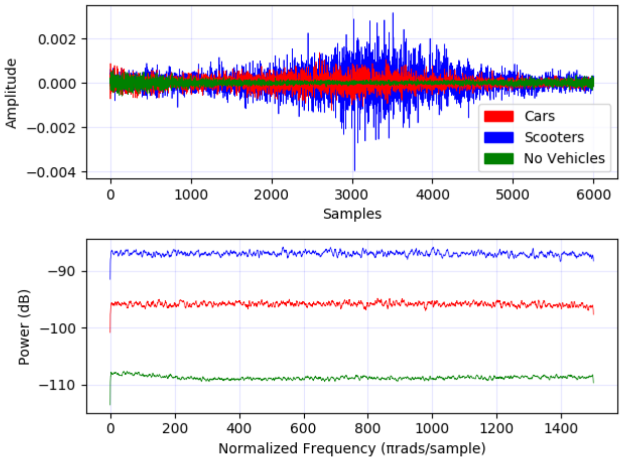

Figure 6　Linear measurements of average audio signals for three vehicle classes

Finally, we can see in Figure 6 the linear samples $y[m]$ and their frequency domain representations. The action of the proposed system causes a distinct separation between the signal classes, observable in the frequency-domain representation of the linear measurements, as well as the linear measurements $y[m]$ themselves. The different $y[m]$ plots, present various statistical features which are extracted for use in the subsequent classification stage.

5. 2. 2　Feature Extraction

We perform classification on a set of features extracted from the $y[m]$ measurements obtained during the sampling process. The advantages of reducing the amount of data used in the classification process by selecting relevant features are threefold: improved system performance due to the removal of redundant information, mitigating the effects of overfitting due to an excessive amount of features, and reducing the system's computational cost and complexity.

The 9 following features are selected:

- *mean*
- *standard deviation*
- *median*
- *absolute max value*
- *peak-to-peak range*
- *interquartile range*
- *data percent in 1st standard deviation*
- *data percent between 2nd & 1st standard deviation*
- *data percent between 3rd & 2nd standard deviation*

Prior to classification, the extracted feature data is randomly undersampled to obtain classes with equal amounts of entries.

5. 3　Classification Results

A random forest classifier is trained on the first dataset and tested on the second, and the results are averaged over 100 runs to obtain an average system accuracy of 86.2 %. In

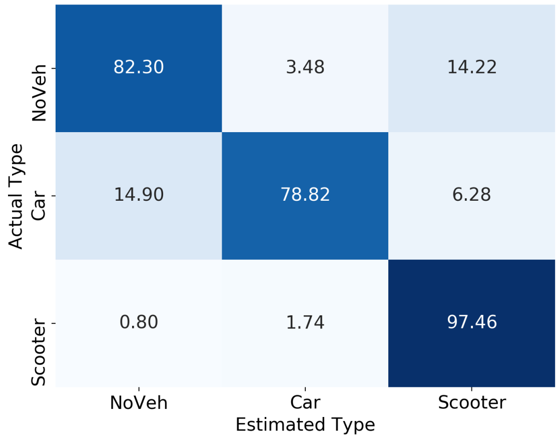

Figure 7 System confusion matrix

this initial evaluation, the detection and identification processes are performed simultaneously: given an unknown signal of length T_s, our system will determine whether or not a vehicle is passing, and the type of passing vehicle.

From the confusion matrix in Figure 7 we can see that system was able to identify “Scooter” signals with high accuracy, but was less effective at identifying “Car” and “NoVeh” signals. Looking at Figure 1 we can see that over the 3–6kHz band the signal power of “Scooter” signals is higher than that of the two other signals, which translates to higher amplitude and crucially, higher variance in the corresponding linear samples. The nature of the differences between the $y[m]$ of each signal combined with the selected feature set lead to a model that is biased towards the “Scooter” class.

6. Conclusion

This paper serves as an initial evaluation of a compressive measurement-based method for the detection and identification of vehicles based on their sound signature. We designed and created a software implementation of a modified RD architecture capable of classifying different passing vehicle sounds with an accuracy of 86.2 % and a back-end ADC sample rate 16 times smaller than the conventional Nyquist rate. Future work includes the fine-tuning of the existing system, in particular removing the bias towards the “Scooter” class, adding additional functionality to the system, such as a separate vehicle presence detection or a steady-state noise reduction stage, and finally working towards a hardware implementation of the system for use in ITS applications.

Acknowledgment The work presented in this paper is partially supported by JSPS KAKENHI Grant Number JP15H05708 and the Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University.

References

[1] S. Ishida, M. Uchino, D. Koike, S. Tagashira, and A. Fukuda, “Initial evaluation of vehicle type estimation using sidewalk stereo microphones,” IPSJ Multimedia, Distributed, Cooperative and Mobile Symposium (DICOMO), pp.1682–1687, 2019.

[2] E.J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” IEEE Transactions on Information Theory, vol.52, no.2, p.489509, Feb. 2006.

[3] A. Aljaafreh and L. Dong, “An evaluation of feature extraction methods for vehicle classification based on acoustic signals,” Proc. IEEE Int. Conf. on Networking, Sensing, and Control, pp.570–575, April 2010.

[4] Z. Changjun and C. Yuzong, “The research of vehicle classification using SVM and KNN in a ramp,” Proc. Int. Forum on Computer Science-Technology and Applications, pp.391–394, Dec. 2009.

[5] M.E. Munich, “Bayesian subspace methods for acoustic signature recognition of vehicles,” Proc. European Signal Processing Conf. (EUSIPCO), pp.2107–2110, Sept. 2004.

[6] S.S. Yang, Y.G. Kim, and H. Choi, “Vehicle identification using wireless sensor networks,” Proc. IEEE SoutheastCon, pp.41–46, March 2007.

[7] H. Gksu, “Engine speed-independent acoustic signature for vehicles,” Measurement and Control, pp.94–103, SAGE, April 2018.

[8] S. Ishida, J. Kajimura, M. Uchino, S. Tagashira, and A. Fukuda, “SAVeD: Acoustic vehicle detector with speed estimation capable of sequential vehicle detection,” Proc. IEEE Conf. Intelligent Transportation Systems (ITSC), pp.906–912, Nov. 2018.

[9] K. Kubo, C. Li, S. Ishida, S. Tagashira, and A. Fukuda, “Design of ultra low power vehicle detector utilizing discrete wavelet transform,” Proc. ITS AP Forum, pp.1052–1063, May 2018.

[10] M.A. Davenport, P.T. Boufounos, M.B. Wakin, and R.G. Baraniuk, “Signal processing with compressive measurements,” IEEE Journal of Selected Topics in Signal Processing, vol.4, no.2, pp.445–460, April 2010.

[11] C. Knoebel, H. Wenzl, J. Reuter, and G. Clemens, “A compressed sensing feature extraction approach for diagnostics and prognostics in electromagnetic solenoids,” Proc. Annual Conf. Prognostics and Health Management Society (PHM), pp.16:1–16:6, Oct. 2017.

[12] A. Jokić and N. Vuković, “License plate recognition with compressive sensing based feature extraction,” arXiv preprint, pp.1–4, Feb. 2019. arXiv:1902.05386 [cs.CV]. `https://arxiv.org/abs/1902.05386`

[13] H. Gksu, “Vehicle speed measurement by on-board acoustic signal processing,” Measurement and Control, vol.51, no.5-6, pp.138–149, July 2018.

[14] X. Liu, E. Gnlta, and C. Studer, “Analog-to-feature (a2f) conversion for audio-event classification,” 2018 26th European Signal Processing Conference (EUSIPCO), pp.2275–2279, Sep. 2018.

[15] D.L. Donoho, “Compressed sensing,” IEEE Transactions on Information Theory, vol.52, no.4, pp.1289–1306, April 2006.

[16] E.J. Candes and T. Tao, “Decoding by linear programming,” IEEE Transactions on Information Theory, vol.51, no.12, pp.4203–4215, Dec. 2005.

[17] J.A. Tropp, J.N. Laska, M.F. Duarte, J.K. Romberg, and R.G. Baraniuk, “Beyond nyquist: Efficient sampling of sparse bandlimited signals,” IEEE Transactions on Information Theory, vol.56, no.1, p.520544, Jan. 2010.