Initial Evaluation of Vehicle Type Identification using Roadside Stereo Microphones

Billy Dawton*, Shigemi Ishida*, Yuki Hori*, Masato Uchino*, Yutaka Arakawa* Shigeaki Tagashira† Akira Fukuda*
*Graduate School/Faculty of Information Science and Electrical Engineering, Kyushu University, Japan
{bdawton,hori,uchino}@f.ait.kyushu-u.ac.jp, {ishida,arakawa,fukuda}@ait.kyushu-u.ac.jp
†Faculty of Informatics, Kansai University, Japan
shige@res.kutc.kansai-u.ac.jp

***Abstract*—A key feature of Intelligent Transport Systems (ITS) is the ability to detect and identify vehicles. In this paper, we put forward a stereo microphone-based system capable of detecting and identifying the type of individually, sequentially, and simultaneously passing vehicles in multi-lane environments based on their sound. We find that our proposed system shows improved performance over single-microphone systems thanks to its improved sequential and successive vehicle detection performance. Initial evaluation results using sound data collected from roads on a university campus show a classification accuracy of 95.01 %.**

***Index Terms*—Acoustic Vehicle Detection, Vehicle Type Identification, Classification, RANSAC, FFT.**

I. Introduction

The increasing development of information and communication technology in recent years has led to similar advances in the field of Intelligent Transport Systems (ITS). A growing number of ITS applications such as navigation, traffic dependent guidance and auto-cruise systems have been proposed and realized with the aim of improving road traffic safety, efficiency, convenience, and reliability.

The detection and identification of vehicles passing on a road is of paramount importance in a wide variety of ITS applications, and several methods have already been put forward for the purpose of vehicle detection. The authors have themselves proposed a low-cost vehicle detection method using a stereo microphone pair [1]–[3].

These proposed technologies, however, only focus on vehicle detection and no consideration is given to the identification of vehicle type. Whilst conventional sensing systems aim for low-cost and high accuracy detection, there is a growing demand for these detected vehicles to be identified with similar accuracy. Current state-of-the-art systems offering such functionality work by using a video camera and image detection techniques to detect and identify passing vehicles [4], [5].

In this paper, we put forward a system capable of performing vehicle detection and classification using a stereo microphone pair. Vehicle detection is achieved using a method proposed by the authors in [3], and classification is performed by analyzing frequency domain information obtained from a detected vehicle's sound in conjunction with supervised machine learning techniques. Due to the multi-lane environment, it is necessary to use a stereo microphone pair to detect vehicles passing simultaneously in different lanes.

Initial evaluation using sound data collected from roads on Kyushu University's Ito campus show that a passing vehicle's type can be determined with an average accuracy of 95.01 %.

The paper is structured as follows: in Section II we explore related research on vehicle type classification, before describing our proposed method in Section III, and finally presenting our initial evaluation results in Section IV.

II. Related Work

A. Non-Acoustic Vehicle Type Classification

Examples of non-acoustic vehicle type classification methods include Electronic Toll Collection (ETC) and camera-based systems. In ETC-based methods, the vehicle type is identified by the registration information contained in the ETC onboard equipment. Whilst ETC systems enjoy widespread use on motorways around the world, the high installation and maintenance costs of the infrastructure make it difficult for them to be installed on standard roads for the sole purpose of vehicle detection and identification.

Two principal methods have been proposed for camera-based vehicle classification. In [6] Hongliang et al. propose a system capable of automatically detecting a vehicle's number plate, and thus its information from a single image using edge statistics. This method requires the use of a high-performance computer for analysis, and the installation of a camera in front of the vehicle passing point to achieve high accuracy. In [7] Avery et al. put forward a classification method using vehicle length: by taking the background difference information from images taken from roadside surveillance cameras, the authors are able to obtain the passing time, and thus the length of passing vehicles which is then used to determine vehicle type. Whilst this method is effective for detecting long vehicles such as trucks or buses, it is not suited for shorter ones like cars or motorbikes. In addition, accuracy performance suffers in rainy and foggy situations.

To the best of our knowledge, none of the above methods have shown any additional results or progress.

B. Acoustic Vehicle Type Classification

Low-cost acoustic vehicle type classification methods have been proposed by Aljaafreh et al. [8] and Changjun et al. [9]. Both of these methods make use of frequency domain features in supervised learning setups using Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN) classifiers respectively. Munich et al. also used supervised learning, namely a Gaussian Mixture Model (GMM) and a Hidden Markov Model (HMM), in conjunction with frequency domain features to identify vehicles; a comparison of the classification accuracy

doi: 10.1109/SAS48726.2020.9220076

of these techniques is shown in [10]. The results obtained help determine the optimal machine learning algorithm and amount of features when estimating a vehicle type from an emitted sound.

Yang et al. have developed a method for estimating a vehicle's type based on the shape of its sound in the frequency domain [11]. Rather than focusing on the individual frequency components that make up the signal, it uses the frequency domain envelope as the feature value: as each vehicle has a unique frequency spectrum shape, the system is able to accurately distinguish individual vehicles from one another. However, the inherent uniqueness of each frequency spectrum shape makes it impossible for the system to classify a passing vehicle's type (to determine its class label).

Göksu [12] has proposed a method of analyzing the acoustic signature of vehicles independently of any changes in engine speed. By making use of wavelet packet analysis in conjunction with a Multilayer Perceptron (MLP) classifier instead of more traditional time or frequency domain-based techniques, the author is able to extract features from a passing vehicle's sound signature, independently of its engine speed. Whilst this affords the system greater accuracy in a variety of situations, the use of a neural network makes it difficult to use in low-power low-cost situations due to the computational and hardware requirements.

Wieczorkowska et al. present a vehicle classification framework using a wide variety of features extracted from the time and frequency domain representations of vehicle sounds. The relevancy of the extracted features is determined by combining data obtained in live roadside recording situations with data obtained in controlled test environments. The selected features are then inputted to a wide selection of classifiers in order to determine the most appropriate one for a given situation. The results of this can be seen in [13].

A multimodal sensing framework using both video and acoustic sensors for vehicle detection and tracking has been presented by Chellappa et al. By using the data obtained from a microphone to determine a vehicle's initial direction of approach, the system is able to roughly estimate a target vehicle location. From this initial information, the vehicle's location is precisely determined and monitored using video data. Whilst not explicitly identifying the type of each passing vehicle, the improved performance of the system proposed in [14] lays the groundwork for a hybrid audio-video vehicle detection and identification sensing system.

None of the works mentioned above have considered vehicle classification in a situation with mixed vehicle sounds. In a real-world environment multiple vehicles could pass simultaneously in different lanes, or in quick succession in the same lane in front of a microphone, resulting in mixed vehicle sounds and reduced system accuracy.

III. Design

A. Intuition

The intuition behind our proposed vehicle type classification system is as follows: a stereo microphone pair is placed on the side of the road to track a vehicle's position relative to both microphones. As the vehicle passes in front of each microphone successively, its sound is recorded and the time difference between both microphones is calculated from which we can obtain the direction and speed of the passing vehicle. The passing vehicle's sound signature is then emphasized by aligning and superimposing the signals obtained by each microphone.

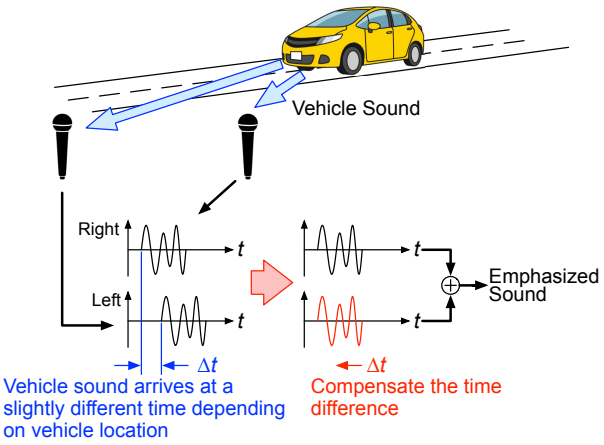


Fig. 1. Proposed stereo microphone vehicle type classification system overview. A passing vehicle's sound signature is emphasized by aligning and superimposing the signals obtained by the microphones.

Figure 1 shows an example of a vehicle moving from right to left: the left channel microphone is located farther from the vehicle than the right channel microphone. The arrival time between the sound emanating from the vehicle and the left microphone is larger than the arrival time between the sound emanating from the vehicle and the right microphone. The difference between the arrival times is Δt.

By shifting the left channel sound by $-\Delta t$ and adding it to the right channel sound we obtain our combined emphasized sound. We set $s_L(t)$, $s_R(t)$, as the left and right channel audio signals respectively and $s_{\text{emph}}(t)$ as the emphasized signal:

$$s_{\text{emph}}(t) = s_R(t) + s_L(t + \Delta t). \tag{1}$$

The vehicle type is estimated from this emphasized signal using supervised learning methods. Since the vehicle is assumed to travel continuously along the road, Δt changes with time and is thus a function of time t:

$$s_{\text{emph}}(t) = s_R(t) + s_L\left[t + \Delta t(t)\right]. \tag{2}$$

B. System Overview

Figure 2 shows our proposed stereo microphone-based vehicle type classification system consisting of the following components: a sound retrieval block, a vehicle detection block, an emphasis synthesizer block, and a vehicle type classification block.

The sound retrieval and vehicle detection blocks listen for sounds and analyze them to detect passing vehicles; if a vehicle is detected, then the blocks will also acquire the vehicle passing time and the reception time difference Δt. The detection block is designed using the SAVeD method established in previous research [3]. Using the acquired Δt, the sound signals acquired by the left and right microphones are superimposed in the emphasis synthesizer block to enhance the vehicle sound in the direction of travel. Frequency domain

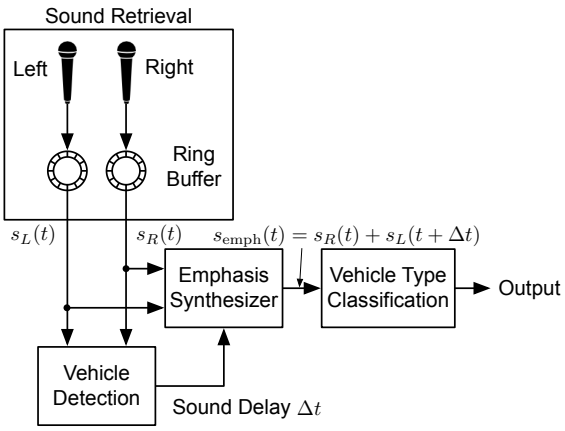


Fig. 2. System overview. The proposed stereo microphone vehicle type classification system consists of a sound retrieval, vehicle detection, emphasis synthesizer, and vehicle type classification block.

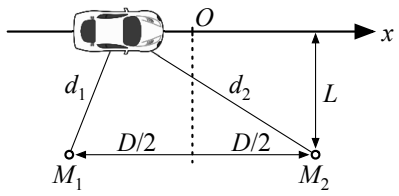


Fig. 3. Microphone setup.

feature values are extracted from this emphasized audio signal, and the vehicle is identified using supervised learning in the vehicle type classification block.

The workings of each block are explained in the following sections.

C. Sound Retrieval Block

The sound retrieval block is composed of a stereo microphone pair. Figure 3 shows the experimental microphone layout: the two microphones M_1 and M_2 are installed at a distance D from each other and a distance L from the road. Since the distances d_1 and d_2 from the vehicle to each microphone change over time, so does the time delay between a sound being emitted by a vehicle and it reaching both microphones. This time difference is used in both the detection and classification processes; for this purpose the audio signals acquired by both microphones are temporarily held in ring buffers.

D. Vehicle Detection Block

The vehicle detection block uses the audio signals stored in the ring buffers. Using these signals, a vehicle is detected by drawing a *soundmap*, which is a plot of the change in sound arrival time difference between the two microphones (estimated using the cross-correlation function), as a function of time. We write the audio signals received by the two microphones as $s_1(t)$ and $s_2(t)$, and the cross-correlation function $R(t)$ as:

$$R(t) = \int s_1(t)\, s_2(t+\tau)\, d\tau. \tag{3}$$

If the two microphones receive a signal with a time difference of Δt such as: $s_1(t) = s_2(t + \Delta t)$, then $R(t)$ reaches its maximum value at $t = \Delta t$. As a result of this, the time difference Δt can be estimated by looking for the peak of $R(t)$; the actual value of Δt is calculated using GCC-PHAT (Generalized Cross-Correlation Phase Transform) which calculates the time difference in the frequency domain.

Additionally, on Fig. 3 we can see that the difference in reception time (or sound delay) Δt between microphones M_1 and M_2 is proportional to the distance between the sound source and each microphone respectively. We set the initial passing time of a vehicle in front of the center of the microphones as $t = t_0$. We thus also derive $\Delta t(t)$, which is a function of time, in the following manner:

$$\begin{aligned}\Delta t(t) = & \frac{d_1 - d_2}{c} \\ = & \frac{1}{c}\left\{\sqrt{\left[v(t-t_0) + \frac{D}{2}\right]^2 + L^2}\right. \\ & \left. -\sqrt{\left[v(t-t_0) - \frac{D}{2}\right]^2 + L^2}\right\},\end{aligned} \tag{4}$$

where c is the speed of sound.

From Eq. (4), we can see that as a vehicle passes in front of a microphone with a constant speed v, an S-shaped curve is drawn on the soundmap. The vehicle detection block works by detecting this curve using a random sample consensus (RANSAC) robust estimation algorithm [15]. The unknown parameters in Eq. (4) are the speed v and the initial passing time t_0; these are estimated by fitting Eq. (4) to a "high likelihood" point cloud on the soundmap.

Figure 4 shows an example of vehicle detection using a soundmap and RANSAC, with the blue dots indicating the sound delay at each time t, and the orange line the result of the RANSAC fitting process. The red points were judged as being of "high likelihood" during the fitting process. It should be noted that, using RANSAC, it is possible to estimate the values v and t_0 even in conditions where the points themselves deviate significantly from the S-curve. For each detected vehicle, the vehicle detection block outputs the speed v and the passing time t_0 to the emphasis synthesizer block.

E. Emphasis Synthesizer Block

The emphasis synthesizer block begins by calculating the passing sound time difference Δt at each time t using the speed v and the initial passing time t_0 for each vehicle detected by the vehicle detection block. Using this information, the emphasis synthesizer block shifts the received sound signal at one of the microphone channels in time and adds the sounds of both channels together, creating an emphasized sound.

If the signal obtained by the first microphone enables us to broadly estimate the passing vehicle type, then the signal at the second microphone gives us information about any successively or simultaneously passing vehicles. For instance, the presence of frequency information corresponding to a high-amplitude signal at the second microphone would suggest a simultaneously passing vehicle in the opposite lane, whilst that of a lower-amplitude signal would suggest a successively

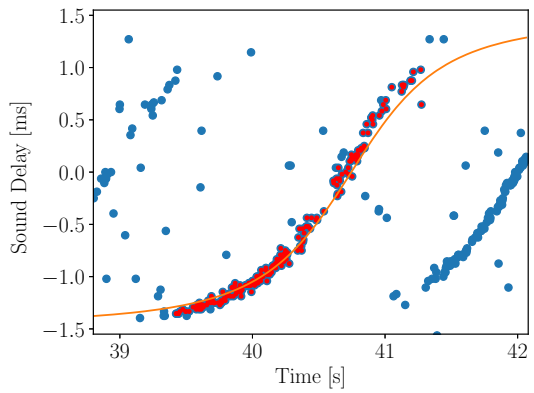


Fig. 4. Vehicle detection using RANSAC. The blue points indicate the sound delays at each time t, and the red points are points judged as being of "high likelihood" during the RANSAC fitting process. The orange line is the result of RANSAC fitting.

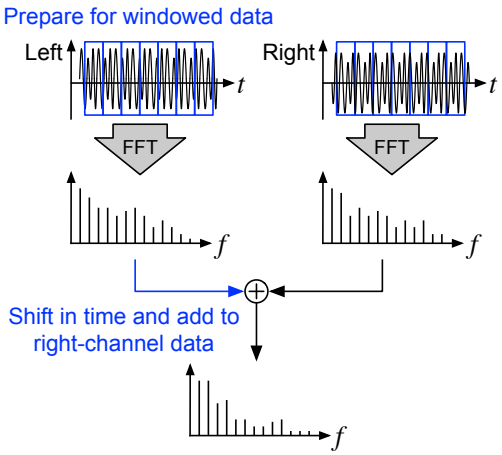


Fig. 5. Emphasis synthesis. For each fixed-width window, the frequency domain representation is derived using an FFT. The FFT'd signals on left and right channels are aligned and added together.

passing vehicle in the same lane, or no other vehicle at all. The emphasized signal obtained from the combination of one of these frequency signatures at the second microphone with the frequency signature at the first microphone gives us information about both the passing vehicle type and any successively or simultaneously passing vehicles.

Figure 5 shows an overview of the emphasis synthesis process: the audio signals of the left and right channels are subdivided into multiple fixed-width windows with the time shift being performed in the frequency domain in order to process each window sequentially. We obtain the time-frequency domain representation of each individual window by performing a Fast Fourier Transform (FFT) on each of them sequentially, before using the speed v and the passing time t_0 obtained beforehand to calculate the appropriate value of Δt as seen in (4). Finally, one of the signals is shifted by Δt to cancel out the time difference and the two signals are summed together.

Time shifting the signal in the frequency domain amounts to shifting the phase of each of its frequency components. Let $s[n]$ be the discrete time representation of the original signal and $S[k]$ its frequency domain representation obtained via DFT:

$$\begin{aligned} S[k] &= \mathrm{DFT}(s[n]) \\ &= \sum_{n=0}^{N-1} s[n]\ e^{-j2\pi k \frac{n}{N}}. \end{aligned} \tag{5}$$

Here, DFT( ) is the discrete Fourier transform, and N is the amount of points used in the DFT operation (= window size). The DFT of the signal $s[n-m]$, which is the signal obtained by delaying the time domain representation of the signal $s[n]$ by m points, can be represented as follows:

$$\begin{aligned} \mathrm{DFT}(s[n-m]) &= \sum_{n=0}^{N-1} s[n-m]\ e^{-j2\pi k \frac{n}{N}} \\ &= e^{-j2\pi k \frac{m}{N}} S[k]. \end{aligned} \tag{6}$$

From Eq. (6), we can see that shifting the time shifts the phase of each frequency component by $-2\pi k \frac{m}{N}$.

F. Vehicle Type Classification Block

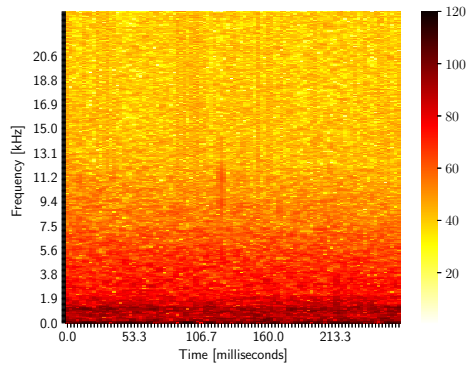
The vehicle type classification block extracts the features used for vehicle type classification from a frequency domain representation of the emphasized vehicle sound produced by the emphasis synthesizer block and determines the vehicle type using supervised learning. In this paper, we are looking to distinguish between multiple vehicle types, and so a supervised learning method capable of multi-class classification is necessary.

We use an SVM classifier due to the large number of features for each data point. The kernel used is the linear kernel as it offers good separability for our particular dataset whilst being less complex and less prone to overfitting than other kernels.

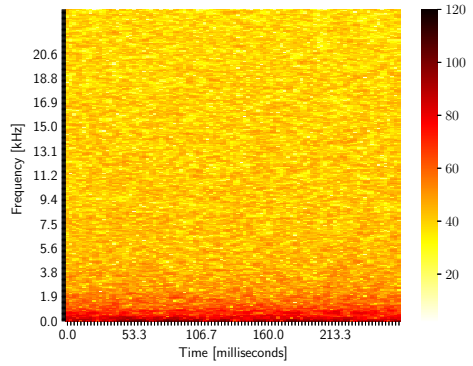
Our proposed method uses only the low-frequency components of the emphasized audio signal as features. Figure 6 shows the frequency spectrum of the audio signal acquired when a (a) vehicle was passing and (b) no vehicle was passing. We can see that the majority of the frequency content contained in a passing vehicle's signal is located in the sub-10 kHz band.

In order to reduce the influence of environmental noise, a low pass filter (LPF) is applied in the time domain to the individual frequency components prior to classification. Looking at the horizontal axes of Fig. 6, we can see that whilst the frequency spectrum of the actual audio signal does not change significantly in the short period of several hundred milliseconds, there are changes in the spectrum of the signal acquired by the microphone that are due to the influence of environmental noise. Given that the time required for a vehicle to pass in front of the microphone is relatively long (on the order of a few seconds) the effect of this small change can be reduced by applying a moving average filter over a shorter time span than the vehicle passing time. Based on our preliminary experimental results, the length of the moving average is set to 320 ms in our evaluations.

To improve system accuracy and efficiency, standardization is applied to all features before classification: $(xf[i]-\mu_f)/\sigma_f$, where $xf[i]$ is the $[i]$th entry in a feature vector, μ_f is the vector's average value, and σ_f its standard deviation.



(a)



(b)

Fig. 6. Sound spectrogram [dB] when (a) a vehicle is passing and (b) no vehicle is passing.

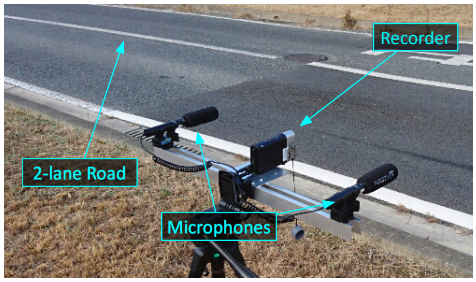


Fig. 7. Experimental setup. Two microphones installed on the roadside.

IV. Evaluation

In order to prove the viability of our proposed method, we performed an initial system evaluation using data collected on the roads of Kyushu University's Ito campus.

A. Evaluation Environment

The experimental setup is shown in Fig. 7. Sounds were acquired from two-way, two-lane roads and classification was performed on vehicles from both lanes.

Two microphones are installed on the roadside at approximately 1 m from the ground, parallel to the road and connected to a video camera. The sound of passing vehicles is then recorded for approximately 20 minutes. The video camera used is a SONY HDR-MV1 and the microphone an AZDEN SGM-990, recording at a sample rate of 48 kHz and bit depth of 16 bits. As in [3], the distance between both microphones is $D = 50\,\text{cm}$, the distance between the microphones and the center of the front lane is $L = 3\,\text{m}$ and the distance between the microphones and the back lane is $L = 6\,\text{m}$. The method outlined in Fig. 3 was applied to the acquired audio signal, and the vehicle type was determined through SVM multi-class classification.

The total number of vehicles that passed during the experiment was 178 (57 cars, 94 scooters/motorbikes, 25 buses, and 2 trucks). Classification was performed for 3 classes: cars, scooters/motorbikes, and buses. Because we only perform classification on vehicles that were actually detected by the detection block, we end up using 142 vehicles (46 cars, 78 scooters/motorbikes, 18 buses) in our evaluation.

The time taken by a vehicle to pass in front of the first microphone is defined as T_{pass}. We set the initial passing time of a vehicle in front of the microphone as $t = t_0$ and evaluate signals over the range $[t_{0,i} - T_{\text{pass}}/2;\, t_{0,i} + T_{\text{pass}}/2]$ where i corresponds to each successive passing vehicle, and $t_{0,i}$ is the initial passing time of that particular vehicle. We record each passing vehicle for $[t_{0,i} - T_{\text{pass}}/2\,;\, t_{0,i} + T_{\text{pass}}/2]$ before splitting the acquired audio signals into a sequence of windows which are sequentially FFT'd and LPF'd resulting in a spectrogram from which we extract frequency domain features (Fig. 6). The vehicle data is randomly undersampled to obtain classes with equal amount of entries, and the features are inputted to a 10-fold cross-validated classifier.

The evaluation compared the classification accuracy of the following two methods:

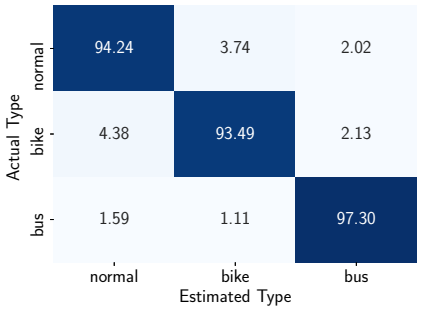
- ***Stereo classification method:*** Our proposed method illustrated in Fig. 3. By using the information obtained during the detection process, the sound obtained by both of the microphones is combined to emphasize the vehicle sound, and the vehicle type is determined by supervised learning using features obtained from the emphasized signal.
- ***Mono classification method:*** This method determines the vehicle type using only one microphone. As the evaluation environment in this paper uses two microphones, in this case the vehicle type was determined using features obtained from the left microphone's audio signal only.

B. System Performance

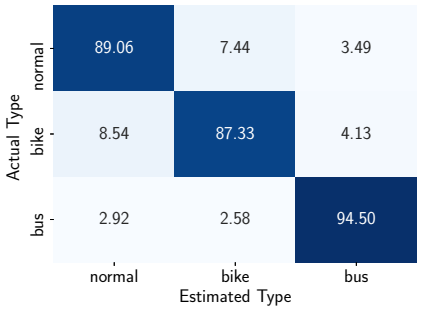
To determine the accuracy of our proposed method, we run our 10-fold cross validated SVM classifier 100 times and average the results, leaving us with our final system accuracy values and confusion matrices. The FFT window length was set to 4096 points, and the features used in classification were obtained by shifting the FFT window along the captured audio signals with a 25 % overlap. We set $T_{\text{pass}} = 2.0\,\text{s}$ based on the results of preliminary experiments.

Figure 8 shows the confusion matrices for (a) the stereo classification method, and (b) the mono classification method. The accuracy ratings are 95.01 % and 90.30 % respectively: vehicle identification accuracy is improved by 4.71 % when using the stereo classification method rather than the mono classification method.

Table I shows the proportion of simultaneously and sequentially passing vehicles compared to the overall amount of passing vehicles. We define a vehicle as "simultaneously passing" if it passes within a previous vehicle's T_{pass} period in



(a)



(b)

Fig. 8. Confusion matrices for: (a) stereo estimation method, (b) mono estimation method. Average accuracy is 95.01 % and 90.30 %, respectively.

TABLE I
NUMBER AND RATIO OF SUCCESSIVE AND SIMULTANEOUS PASSING VEHICLES.

	Normal	Bike	Bus	Total
Detected	46	78	18	142
Simultaneous	11 (23.91%)	33 (42.31%)	2 (11.11%)	46 (32.39%)
Successive	6 (13.04%)	10 (12.82%)	3 (16.67%)	19 (13.38%)
Total	17 (36.96%)	43 (55.13%)	5 (27.78%)	65 (45.77%)

the opposite direction, and "successively passing" if it passes within a previous vehicle's T_{pass} period in the same direction.

The authors believe the improvement in overall system accuracy to be mainly due to the improved detection of simultaneously and successively passing vehicles achieved thanks to the stereo classification method. By looking at Table I and Fig. 8 we can see that the simultaneously and successively passing vehicles make up only a relatively small proportion of the overall detected vehicles, which is why the overall system accuracy shows only a slight improvement.

The accuracy of our proposed system shows an 11-point improvement over that of SAVeD, an existing acoustic vehicle detection method designed to deal with the problem of simultaneously and successively passing vehicles.

If we were to test our setup on an environment with a larger proportion of simultaneously and successively passing vehicles, we would expect a correspondingly proportional increase in the accuracy of our proposed stereo microphone method compared to the mono microphone method.

V. CONCLUSION

In this paper, we put forward a system capable of both detecting and classifying passing vehicles using a stereo microphone setup. Vehicle detection is performed using a soundmap-based method based on previous works, and vehicle classification is performed using an emphasized signal obtained from the time-shifted sum of the microphone signals. An initial evaluation using data collected from roads on Kyushu University's Ito campus shows that our proposed method yields a vehicle type classification accuracy of 95.01 %.

ACKNOWLEDGMENTS

Part of the research in this paper was supported by JSPS KAKENHI Grant Numbers JP15H05708 and JP17H01741 as well as the Cooperative Research Project of RIEC, Tohoku University.

REFERENCES

[1] S. Ishida, K. Mimura, S. Liu *et al.*, "Design of simple vehicle counter using sidewalk microphones," in *Proc. ITS EU Congress*. EU-TP0042, Jun. 2016, pp. 1–10.

[2] S. Ishida, S. Liu, K. Mimura *et al.*, "Design of acoustic vehicle count system using DTW," in *Proc. ITS World Congress*. AP-TP0678, Oct. 2016, pp. 1–10.

[3] S. Ishida, J. Kajimura, M. Uchino *et al.*, "SAVeD: Acoustic vehicle detector with speed estimation capable of sequential vehicle detection," in *Proc. IEEE Conf. Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 906–912.

[4] N. Buch, M. Cracknell, J. Orwell *et al.*, "Vehicle localisation and classification in urban CCTV streams," in *Proc. ITS World Congress*, Sep. 2009, pp. 1–8.

[5] A. Nurhadiyatna, B. Hardjono, A. Wibisono *et al.*, "ITS information source: Vehicle speed measurement using camera as sensor," in *Proc. Int. Conf. on Advanced Computer Science and Information Systems (ICACSIS)*, Dec. 2012, pp. 179–184.

[6] B. Hongliang and L. Changping, "A hybrid license plate extraction method baed on edge statistics and morphology," in *Proc. Int. Conf. Pattern Recognition (ICPR)*, vol. 2, Aug. 2004, pp. 831–834.

[7] R. P. Avery, Y. Wang, and G. S. Rutherford, "Length-based vehicle classification using images from uncalibrated video cameras," in *Proc. IEEE Conf. Intelligent Transportation Systems (ITSC)*, Oct. 2004, pp. 1–6.

[8] A. Aljaafreh and L. Dong, "An evaluation of feature extraction methods for vehicle classification based on acoustic signals," in *Proc. IEEE Int. Conf. on Networking, Sensing, and Control*, Apr. 2010, pp. 570–575.

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

[10] M. E. Munich, "Bayesian subspace methods for acoustic signature recognition of vehicles," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Sep. 2004, pp. 2107–2110.

[11] S. S. Yang, Y. G. Kim, and H. Choi, "Vehicle identification using wireless sensor networks," in *Proc. IEEE SoutheastCon*, Mar. 2007, pp. 41–46.

[12] H. Göksu, "Engine speed-independent acoustic signature for vehicles," *Measurement and Control*, vol. 51, no. 3–4, pp. 94–103, Apr. 2018.

[13] A. Wieczorkowska, E. Kubera, T. Słowik *et al.*, "Spectral features for audio based vehicle and engine classification," *J. Intell. Inf. Syst.*, vol. 50, no. 2, pp. 265–290, Apr. 2018.

[14] R. Chellappa, Q. Gang, and Z. Qinfen, "Vehicle detection and tracking using acoustic and video sensors," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Mar.–Apr. 2004, pp. 265–290.

[15] M. A. Fischler and R. C. Bolles, "Random sample censensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.