The Development of the Vehicle Sound Source Localization System

Yoonho Jang[*], Jaekwang Kim[*] and Jinhak Kim[*]
[*]Research and Development Division
Hyundai Motor Company
Uiwang-si, Gyeonggi-do, Korea, Republic of
E-mail: {yoonho, jaekwang, jinhak}@hyundai.com

Abstract— **Recently many different kinds of sensors such as RADAR, LIDAR and camera have been widely developed by automotive OEMs and component suppliers. Most OEMs are using SRR for BSD(Blind Spot Detection) applications, however, its high cost prevents BSD from wider vehicle application. In this paper, a new alternative sensor system is proposed for BSD system. We proposed the method to detect the presence of moving vehicles and find their locations using Time Difference of Arrival (TDOA) between microphones. The vehicle sound source localization systems were implemented using the array of microphones. We improved the system accuracy based on the proposed localization algorithm. The system results were discussed that has equivalent performance with SRR systems**

I. INTRODUCTION

Sound source localization technique using an array of sensors has generated wide interest in the signal processing community for more than 30 years [1].

Sound source localization using multiple microphones has been researched in various ways and being utilized in the intelligent robots, video conference system, speech recognition system and so on. However in the automotive development field, the technique was not considered applicable.

In automotive part, the consumer interests have been increased about the car safety and the regulations for them of governments have strengthened in recent years. Various new sensors were considered to construct new safety systems for the cognitive ability improvement of drivers. BSD (Blind Spot Detection) is an alarm system to the driver by detecting a vehicle in the driver's blind spot that cannot be seen through the side mirror. The BSD system mostly uses two radars that are installed in the rear lateral of the vehicle. Several developers have researched the safety systems using low cost sensors. For this reason, the new sensor systems were constructed using inexpensive microphones than automotive radars. This paper proposes the way to detect the presence of moving vehicles and find their locations using multiple microphones.

The rest of this paper is organized as follows: Section 2 provides a brief review of previous vehicle sound detection and sound localization techniques. Section 3 explains the proposed vehicle sound source recognition methods. In Section4, we describe a vehicle sound source localization algorithm. Section 5 deals with experimental results. Finally, conclusions and future works are presented in Section 6.

II. A REVIEW OF VEHICLE DETECTION AND LOCALIZATION

Acoustic vehicle detection system includes signal pre-processing, feature extraction, decision making. Most of the sound source tracking system was developed for use in a room or in a state of being fixedly microphones attached to the external infrastructure to track the location of a certain sound source. When the microphone array was installed on vehicle, the sound localization system has almost found sound the car horn[2] or siren[3] rather than the position of moving vehicle.

A. Acoustic vehicle detection and recognition

In the literature, a particular interests in development of moving vehicle detection and recognition systems using acoustic signals started to appear in the 1970s[4].

Depending on the category of target vehicles, these can be divided into two groups: military vehicle recognition and road vehicle recognition. Early publications regarding automated acoustic vehicle recognition algorithms were focused mainly on military vehicle signals, in order to develop a system that improves surveillance for security. However the research has recently been conducted to recognize the vehicle on the road for car safety.

Vehicle detection and recognition consist of signal pre-processing, feature extraction, machine learning, and decision making. Sound data is obtained by acoustic sensors. The collected input signals contain unwanted signals, including wind noise and sampling noise. Some kind of noise reduction algorithm is applied to improve recognition performance. Then the pre-processed signals are treated further to extract selected features of signals that represent some characteristic source properties. There are some decent research results about vehicle sound feature extraction. In [5], wavelet based feature extraction of moving vehicles was suggested to detect the approaching vehicles when other noises are present. Munich [6] compared conventional methods used for speaker recognition, namely, systems based on Mel-frequency cepstral coefficients (MFCC) and either Gaussian mixture models (GMM) or hidden Markov models (HMM), with Bayesian subspace method based on the short term Fourier transform (STFT) of the vehicles' acoustic feature. In this paper we propose an algorithm for vehicle detection from the characteristic feature of MFCC and Sub-band power.

B. Sound localization

Sound localization is a fundamental human ability. By applying this function in vehicles, it can be warnings against the risk of a blind spot. Sound localization systems process the sound signals acquired from multiple omnidirectional microphones.

The IID(Interural Intensity Difference) is entirely based on the relative intensity difference between the signal input to each microphone. However IID has a disadvantage that has poor performance for low frequency sound signal localization.

The most common sound source localization technique is indeed the ITD (Interaural Time Difference). The systems measure the time difference of each signal. Among all the approaches proposed in the literature, numerous ones are based on Time Difference Of Arrival (TDOA) [7] at different microphone pairs. Many of them are based on the Generalized Cross-Correlation PHAse Transform (GCC-PHAT).

Beam-forming is a widely applied technique. It is based on the output power of the beam-former from the signal phase difference. However beamforming techniques need a lot of microphones

III. Vehicle Sound Recognition and Detection

This section describes the process to extract the vehicle characteristics from the acoustic data and to classify the various sorts of sounds. We acquired much data in driving conditions and analyzed them in detail. Then we extracted two kinds of features from the data in various conditions including road type, velocity, and models. Afterward, we use machine learning techniques for classification of vehicle sounds.

A. Data acquisition

Three microphones are mounted horizontally in the rear bumper of test vehicle at intervals of thirty centimeters in order to collect the sound signals in various driving conditions. Fig1 shows the mounting positions of microphones. We recorded the sound data that contain the five classes. Table 1 presents the classes of the obtained data from data acquisition device. In our database, there are sounds including two kinds of classes, for example, vehicle sounds can be recorded when the wind is blowing toward the microphones. The sounds that contain the vehicle sounds with other sounds are categorized as Vehicle Sound Class, because our purpose is to judge the vehicle existence. For other cases, we do not use sounds with uncertainty to build a database. The length of audio data is 10 seconds per test. And sampling rate is 51200 samples per second. The vehicles were traveling at constant speeds that varied from 50km/h to 100km/h. The obtained samples of vehicle sounds are segmented by the certainty. We selected the vehicle signals by playing back and listening to the collected. Human auditory sense is involved in the vehicle sound selection. We simply extracted moving vehicle sounds from the test records.

Fig. 1 the equipped multiple microphones in the rear bumper of test vehicle.

TABLE I
THE OBTAINED DATA CATEGORIZED BY 5 CLASSES

Class	Test vehicle movement	Number of test	Number of features
Own test car noise	Move	50	1604
Vehicle sound	Move/Stop	100/20	606/458
Wind noise	Stop	10	480
Train	Stop	10	60
Car horn (single horn, dual horn)	Move/Stop	20/20	60/60
Other noises (construction noise)	Stop	10	821

B. Feature extraction and Classification

The recognition system works directly with the raw acoustic spectrum. We found the regularity of the sub-band power in frequency domain. So we extracted the most prominent segments on spectrogram of vehicle sounds. The sub-band power has been applied at 400 to 1200Hz. The first sub-band power feature contains 100 data points during 0.1 seconds.

We also utilized MFCC(Mel-frequency cepstral coefficients) features that have been originally proposed for speech recognition and speaker recognition[8]. The frame size is set 0.1 seconds and the number of mel-scale parameter is 50. The each sound samples contain 84 MFCC characteristics. Therefore, the sequential 5120 signals can be represented by 184 elements feature vectors for vehicle sound recognition.

We applied the conventional neural network that could classify by the multiple classes. We constructed the training data sets from the data samples for learning and generalization. We applied 10 hidden layers of the neural network.

IV. Vehicle Sound Source Localization Algorithm

After a vehicle is classified, we focus on a way to estimate where the vehicle is. By the sound source localization techniques, we can determine which direction the target vehicle is located. The most common sound source localization technique is Time-Difference Of Arrival, TDOA. The TDOA is based on the estimation scheme of time delay between two signals from a microphone pair located different positions. Generalized Cross-Correlation, GCC and peak detection are generally used to find the phase of two signals.

However, TDOA has problems when there are two or more sound sources simultaneously TDOA cannot determine where the sound source is exactly, since the output of GCC and peak detection is the maximum scalar value that represents a phase of two signals[9]. To solve this problem, we estimate each probability that the sound source is located at each degree of the direction.

The overview of our localization algorithm is illustrated in Fig. 2. First of all, we calculate the time delays with the assumption that the sound source emerges at each degree with pre-defined microphone positions. The time delays are calculated and stored in off-line process and applied in on-line process. After then, we estimate the degree of the sound source by the modified convolution scheme in real-time. The rest of the section is devoted to a detailed discussion of our proposed method.

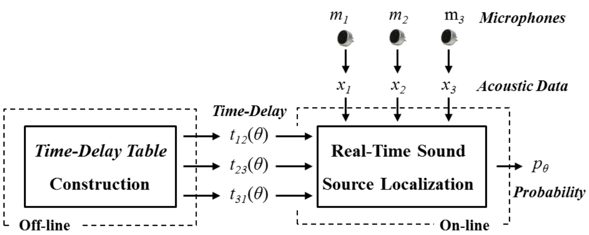


Fig. 2 The overview of sound localization algorithm

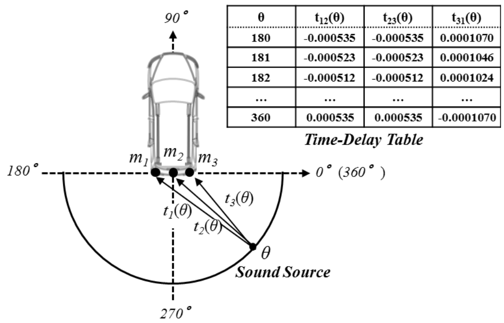
A. *Time-Delay Table*

The time delays between microphones are depended on the directions of the sound sources with fixed positions of microphones. We can calculate the time delays when a sound source emerges at some degree from 180 to 360 degree in four-quadrant coordinate since positions of microphones are not changed after set. As shown in Fig. 3, each transfer time, $t_1(\theta)$, $t_2(\theta)$, or $t_3(\theta)$, is different for each microphone, m_1, m_2, or m_3, because of the consistent velocity of the sound wave. By this concept, *Time-Delay Table* is stored in off-line process. The time delay is defined by (1).

$$t_{12}(\theta) = t_1(\theta) - t_2(\theta) \qquad (1)$$

B. *Real-Time Sound Source Localization*

The direction of the object is decided by the modified convolution scheme in real-time. To achieve our goal, a similarity between delayed signals that are affected by pre-defined time delays at each degree is estimated. These similarities are probabilities that the sound source is located at each degree of direction. The probability vector, $\mathbf{P} = \{p_{180}, p_{181}, \ldots, p_{360}\}$ is estimated by the similarity calculation method. p_θ means the probability that the sound source is located at θ degree. Fig 4 shows the example of the **P**. We can know the distribution of the sound sources. There are two approached vehicles, one is at around 235 degree direction and another is at around 300 degree direction, in this situation. In this paper, we target to detect the most dangerous vehicle and give warning to the drivers. Therefore, the position of the



θ	$t_{12}(\theta)$	$t_{23}(\theta)$	$t_{31}(\theta)$
180	-0.000535	-0.000535	0.0001070
181	-0.000523	-0.000523	0.0001046
182	-0.000512	-0.000512	0.0001024
...	...	...	...
360	0.000535	0.000535	-0.0001070

Time-Delay Table

Fig. 3 Estimated time delay (sec) for various angles

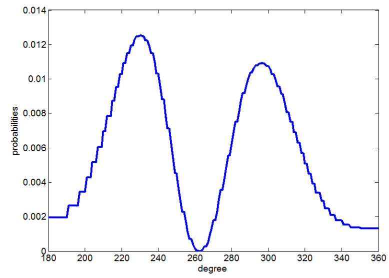


Fig. 4 The probability from similarity calculation algorithm

object that has the most probability is detected. The average filter is applied to the **P** to remove the impulse noise and the degree that has maximum probability is selected for the position of the targeted vehicle.

$$s_\theta = \sum_{\text{all } t} \{x_1'(t) \cdot x_2'(t - t_{12}(\theta)) + x_1'(t) \cdot x_3'(t + t_{31}(\theta)) + x_2'(t) \cdot x_3'(t - t_{23}(\theta))\} \qquad (2)$$

To estimate the probabilities, we need to define the similarity calculation method. A similarity of a particular degree is calculated by (2). x_1, x_2, and x_3 are the time-domain signals from microphones. The three terms in (2) can be calculated by general convolution. Therefore, the convolutions of signal pairs are pre-calculated to get the similarity, s_θ. The probability, p_θ, is estimated by (3).

$$p_\theta = \frac{s_\theta}{\sum_{\theta=180}^{360} s_\theta} \qquad (3)$$

We use the differential value, *x'*, instead of original value, *x*, to calculate the similarity value. There is significant reason to make the algorithm robust. When there is much noise in the signals, the similarities are very affected if we use the original signals. However, similarities based on the differential value are hard to be defected by the noise since the variance of the signals efficiently is lower than before the differential calculation. As shown in Fig. 5, differential-based calculation well worked in noisy situation. When two vehicles are passing from left-rear to left-front at less noisy situation, two methods well estimate the location of passing vehicles (in Fig. 5-(a), (b)). At much noisy situation, while original value-based

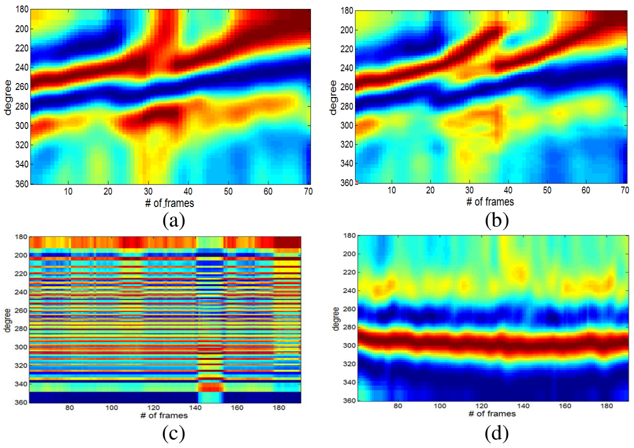


Fig. 5 the sound localized degrees based on differential calculation method, (a) original value-based localization at less noisy situation, (b) differential value-based localization at less noisy situation, (c) original value-based localization at much noisy situation, (d) differential value-based localization at much noisy situation.

localization cannot estimate the target vehicle, differential value-based localization can estimate the target vehicle well.

V. Experimental Results

In this section, we present the results of our acoustical vehicle localization scheme for a given set of measured data. In order to investigate the performance of our method, we equipped the vehicle with triple microphones in the same environmental conditions when the data acquisition was conducted. We utilized the surface microphones Type 40PS with fairing. The array of three surface microphones was mounted on the external surface of our test vehicle. Our system has been implemented on real time linux based embedded board. The board includes Freescale i.MX6Q, 1G RAM, 4CH audio codec, and 4CH microphone input. The real signal experiments are carried out using real-time linux C language in real time and the signals are detected using microphones array.

The Real-time embedded system, as shown in Fig. 6 was implemented in the consecutive order. The system is executed every 0.1 seconds repeatedly, and recognition and tracking system use the cumulative data during 0.5 seconds. And memory release function is performed simultaneously.

We carried out the real-time performance evaluation test at the proving ground in Hyundai R&D center as according to our BSD test scenario. The test scenario is made up of 2 parts. First part is to detect the overtaking and passed vehicles. Last one is to localize the following vehicles with uniform motion of straight line. The speeds of test vehicle are set 50, 75, 100 km/h. And tests progressed 5 times at each different speed. We assumed the success cases that the system detects the approaching vehicles exactly within range of 3 to 15 meters. The detection accuracy of vehicle sound localization system is 93.3 percentages of whole tests. Since our database includes vehicle sounds in various situation, our system can robustly recognize the vehicle when other sounds occur simultaneously.

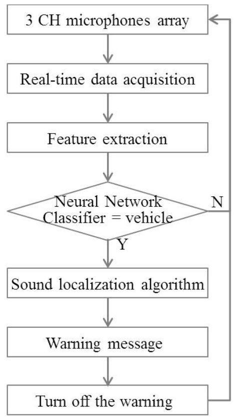


Fig. 6 Flowchart of the real time vehicle sound localization system

VI. Conclusions and Future Works

In this paper, we developed the real-time sound source localization system using the microphone array, by proposing the recognition and tracking algorithm. The system was tested in proving ground environment, with various velocities of vehicles and noise conditions. In particular, we demonstrated the feasibility of developing the acoustic sensor systems mounted on the exterior of vehicles. In order to improve the performance, it seems to construct the mass database in various environmental driving conditions that includes weather, relative velocity, and the segments of object vehicles.

References

[1] H. Krim and M. Viberg, “Two decades of array signal processing research - the parametric approach,” IEEE Sig. Proc. Mag., pp. 67-94, July 1996.
[2] J. Chen, J. Benesty, and Y. Huang, “Time delay estimation in room acoustic environments: An overview,” EURASIP Journal on Appl. Sig.Proc., vol. 2006, pp. 1–19.
[3] S. Shon, E. Kim, J. Yoon, H. Ko, “Sudden noise source localization system for intelligent automobile application with acoustic sensors,” IEEE ICCE, pp 233-234, Jan. 2012.
[4] K. Kodera, , A. Itai , and H. Yasukawa, “Sound Localization of Approaching Vehicles Using Uniform Microphone Array,” IEEE ITSC , pp1054-1058, Oct. 2007.
[5] D.W. Thomas and B. R.Wilkins. “Determination of engine firing rate from acoustic waveform,” Electronics Letters, 6(7):193–194, 1970.
[6] A. Averbuch, V. A. Zheludev, N. Rabin, and A. Schclar “Wavelet-based acoustic detection of moving vehicles,” Multidimensional Systems and Signal Processing, pp 55-80, Mar.2009.
[7] Mario E. Munich, ‘Bayesian Subspace Methods For Acoustic Signature Recognition Of Vehicles,” Proc. of the 12th European Signal Processing Conf. (EUSIPCO 2004), Sep, 2004,
[8] L. Rabiner and B. Juang. Fundamentals of Speech Recognition. Prentice Hall, Inc., 1993.
[9] X. Anguera, C. Wooters, B. Peskin and M. Aguilo, “Robust Speaker Segmentation for Meetings: The ICSI-SRI Spring 2005 Diarization System,” in *MLMI 2005 Meeting Recognition Workshop*, Edinburgh, UK, July 2005.