일반논문 (Regular Paper)
방송공학회논문지 제27권 제1호, 2022년 1월 (JBE Vol.27, No.1, January 2022)
https://doi.org/10.5909/JBE.2022.27.1.91
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

사각영역이 없는 전방향 음원인식을 위한 QRAS 기반의 알고리즘

김 영 언[a], 박 구 만[b)‡]

QRAS-based Algorithm for Omnidirectional Sound Source Determination Without Blind Spots

Youngeon Kim[a] and Gooman Park[b)‡]

요 약

음원의 음량, 방향 및 음원까지의 거리와 같은 음원의 특성을 인식하는 것은 자율주행차, 로봇 시스템, AI 스피커 등 무인 시스템에서 중요한 기술 중의 하나이다. 음원의 방향이나 거리를 인식하는 방법은 레이다, 라이더, 초음파 및 고주파와 소리를 이용하는 방법이 있다. 그러나 이러한 방법은 신호를 발신하여야 하며, 장애물에 의한 비가시 영역에서 발생하는 음원은 정확하게 인식할 수 없다. 본 논문에서는 비가시 영역을 포함한 주변에서 발생하는 음원의 음량, 방향 및 음원까지의 거리를 인식하는 방법으로 가청 주파수 대역의 소리를 검출하여 인식하는 방법을 구현하고 평가하였다. 음원을 인식하기 위하여 주로 사용하는 교차형 기반의 음원인식 알고리즘은 음원의 음량과 방향을 인식할 수 있으나 사각영역이 발생하는 문제가 있다. 뿐만아니라 이 알고리즘은 음원까지의 거리를 인식할 수 없다는 제약이 있다. 이러한 기존 방법의 한계를 탈피하기 위하여, 본 논문에서는 교차형 기반의 알고리즘보다 더 발전된 직사각형 기법을 사용한 QRAS 기반의 알고리즘으로 음원의 음량, 방향 및 음원까지의 거리를 인식하여 음원의 특성을 파악할 수 있는 음원인식 알고리즘을 제안한다. 전방향 음원인식을 위한 QRAS 기반의 알고리즘은 직사각형으로 배치된 4개의 음향센서에 의하여 도출되는 6쌍의 음향 도착 시간차를 사용한다. QRAS 기반의 알고리즘은 기존 교차형 기반의 알고리즘으로 음원을 인식할 때 발생하는 사각영역과 같은 문제점을 해결할 수 있으며, 음원까지의 거리도 인식할 수 있다. 실험을 통하여 제안된 전방향 음원 인식을 위한 QRAS 기반의 알고리즘은 사각영역없이 음원의 음량, 방향 및 음원까지의 거리를 인식할 수 있음을 확인하였다.

Abstract

Determination of sound source characteristics such as: sound volume, direction and distance to the source is one of the important techniques for unmanned systems like autonomous vehicles, robot systems and AI speakers. There are multiple methods of determining the direction and distance to the sound source, e.g., using a radar, a rider, an ultrasonic wave and a RF signal with a sound. These methods require the transmission of signals and cannot accurately identify sound sources generated in the obstructed region due to obstacles. In this paper, we have implemented and evaluated a method of detecting and identifying the sound in the audible frequency band by a method of recognizing the volume, direction, and distance to the sound source that is generated in the periphery including the invisible region. A cross-shaped based sound source recognition algorithm, which is mainly used for identifying a sound source, can measure the volume and locate the direction of the sound source, but the method has a problem with “blind spots”. In addition, a serious limitation for this type of algorithm is lack of capability to determine the distance to the sound source. In order to overcome the limitations of this existing method, we propose a QRAS-based algorithm that uses rectangular-shaped technology. This method can determine the volume, direction, and distance to the sound source, which is an improvement over the cross-shaped based algorithm. The QRAS-based algorithm for the OSSD uses 6 AITDs derived from four microphones which are deployed in a rectangular-shaped configuration. The QRAS-based algorithm can solve existing problems of the cross-shaped based algorithms like blind spots, and it can determine the distance to the sound source. Experiments have demonstrated that the proposed QRAS-based algorithm for OSSD can reliably determine sound volume along with direction and distance to the sound source, which avoiding blind spots.

Keywords : Rectangular-shaped, blind spot, unmanned system, direction, distance

Ⅰ. Introduction

As key technology trends have been evolving towards unmanned systems with AI, recognizing sound source characteristics such as the sound volume, direction and distance to the sound source is also becoming an important technology for all unmanned systems like autonomous vehicles, robot systems and AI speakers.

As methods of recognizing the direction and distance of the sound source, there are multiple methods using a radar, a rider, an ultrasonic wave and a RF signal with a sound[1]. These methods require the transmission of a signal and cannot accurately determine the sound source generated in the invisible region due to obstacles. However, there is a method of recognizing the characteristics of the sound source by using the sound of an audible frequency without transmitting a signal.

Most traditional methods which use the sound of an audible frequency without transmitting a signal for the sound source recognition use two microphones[2,3]. These methods measure the time difference and the level difference of arrived-sounds and estimate the direction in which the sound source originates[4,5]. But these methods have limitations in recognizing the direction of the sound source that occurs in the range of 360 degrees.

To determine the direction of the sound source within the range of 360 degrees, there are several methods like the three-microphone array type by Hung-Yan Gu, et al[6], the triangle microphone array type by Catur Hilman Adritya, et al[7], the RRAS type by Ben Rudzyn, et al[8], the cross-shaped based algorithm using four microphones by Bahaa Al-Sheikh, et al[9] and etc. One of methods is the cross-shaped based algorithm that uses four microphones[9]. The cross-shaped based algorithm measures the time difference between the sound source and the microphones, and determines the direction of the sound source within the range of 360 degrees. But this method cannot determine the distance between the sound source and the microphones. Therefore, we have to study an algorithm that can determine not only the volume and direction of the sound source without blind spots, but also the distance between the sound source and the microphones.

In this paper, we propose the Quadruple Right-angle Acoustic Sensing (QRAS)-based algorithm by the rectangular-shaped technique which also uses four microphones. This algorithm consists of three functions: an input function, a processing function, and an output function, as described in detail in Chapter III. An experimental system for this method consists of 3 blocks: a computing block, a location block and a recognition block, as described in detail in Chapter IV.

Most importantly, the QRAS-based algorithm for Omnidirectional Sound Source Determination (OSSD) can eliminate the existing problems such as blind spots and calculating errors which are generated by the conventional cross-shaped based algorithm as described in Chapter V.

Ⅱ. Theory of sound source recognition methods

Since Lord Rayleigh announced the Spherical Head Model (SHM) in 1907[10], a lot of studies have been done to estimate the direction and to determine the position of the sound source. It has been reported that humans listen to sounds with their ears and auricles, estimate the location of the sound, and recognize the characteristics of the

a) 동오정밀(주)부설연구소 (R&D Center, DongOh Technology Co., Ltd.)
b) 서울과학기술대학교 나노IT디자인융합대학원 (The Graduate School, Seoul National University of Science and Technology)
‡ Corresponding Author : 박구만(Gooman Park)
E-mail: gmpark@seoultech.ac.kr
Tel: +82-2-970-6430
ORCID:https://orcid.org/0000-0002-7055-5568
※ This research was supported by Brain Pool program funded by the Ministry of Science and ICT through the National Research Foundation of Korea(2020H1D3A2A02048280).
· Manuscript received September 30, 2021; Revised December 14, 2021; Accepted January 3, 2022.

sound[11]. For example, humans can detect the arrival time and level difference of the sound coming into the ears, and determines the direction of the sound source. The two key elements for recognizing the direction are an Interaural Time Difference (ITD) and an Interaural Level Difference (ILD)[12,13]. The factor of recognizing altitude is determined from the spectral shape properties of the acoustic input that arise from the complex geometry and associated direction dependent filtering of the pinna[14].

There are also multiple methods that use the microphones instead of human ears for unmanned systems like autonomous vehicles, robot systems and AI speakers. Most traditional methods use an ITD and an angle Θ by two microphones to calculate the direction of the sound source within the range of 180 degrees as shown in Fig. 1. But these methods cannot determine D, the distance between the sound source and the center of two microphones because we don't know the time that occurs the sound in Fig. 1. S_0 and S_3 are the microphones in left side and right side. X is the distance between the microphone S_0 and S_3. $ITDs_{01}$ is the interaural time difference between S_0 and S_3. θ_1 and θ_1' are the angles between the vertical line and the D or D' line. In other word, θ_1 is an angle that represents the arrival of the sound source. Sound Source and θ_1 are examples when the sound source is located at left side, and Sound Source' and θ_1' are examples when the sound source is located at right side as shown in Fig. 1.

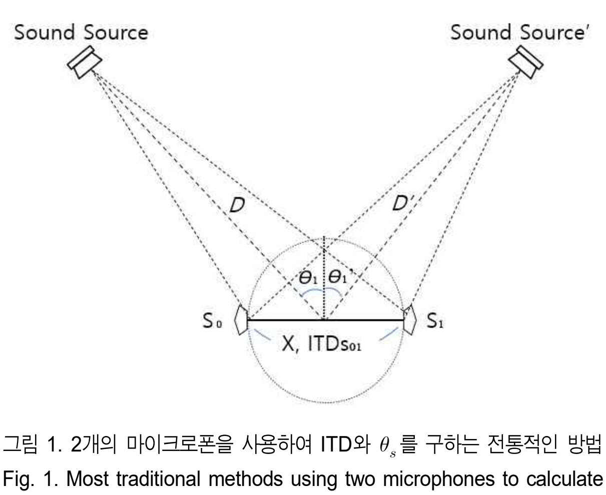

그림 1. 2개의 마이크로폰을 사용하여 ITD와 θ_s를 구하는 전통적인 방법
Fig. 1. Most traditional methods using two microphones to calculate an ITD and θ_s

$ITDs_{01}$ is the interaural time difference between S_0 and S_1 that is obtained as Equation (1)[15]. $R_{S_0S_1}(n)$ is the Generalized Cross Correlation with Phase Transform (GCC-PHAT) function that is obtained as Equation (2)[15].

$$ITDs_{01} = \underset{n}{\arg max}\ Rs_0s_1(n) \quad (1)$$

where

$$Rs_0s_1(n) = \sum_{k=0}^{n-1} \Phi(k) Ss_0s_1(k) e^{i\left(\frac{2\pi nk}{N}\right)} \quad (2)$$

The GCC_PHAT is mainly used to create the ITD between two microphones because it has a higher value at the sound source direction than the others like the Generalized Cross Correlation (GCC). The Phase Transform (PHAT) is the most commonly used weight function for the GCC because this can effectively reduce the magnitude of the cross-correlation value of the part requiring suppression in a simple manner, and provides the same weight for each phase in each frequency band[15].

The $ITDs_0s_1$ between a microphone (S_0) and another microphone (S_1) is calculated using the GCC_PHAT function. After that the $\Theta s_0 s_1$ is finally calculated according to the $ITDs_0s_1$ as Equation (3) and Equation (4). The microphone (S_0) and microphone (S_1) are input signals that were received to microphones from the sound source. GP(f) is the function of the GCC-PHAT. F is the sampling frequency of input signals. X is the distance between S_0 and S_1. C is the speed of the sound.

$$ITDs_0s_1 = GP(f)[S_0, S_1, F, X/c] \quad (3)$$

$$\theta s_0 s_1 = \sin^{-1}[Ds_{01}/(X/c)]*(180/\pi) \quad (4)$$

We have to use four microphones to determine the direc-

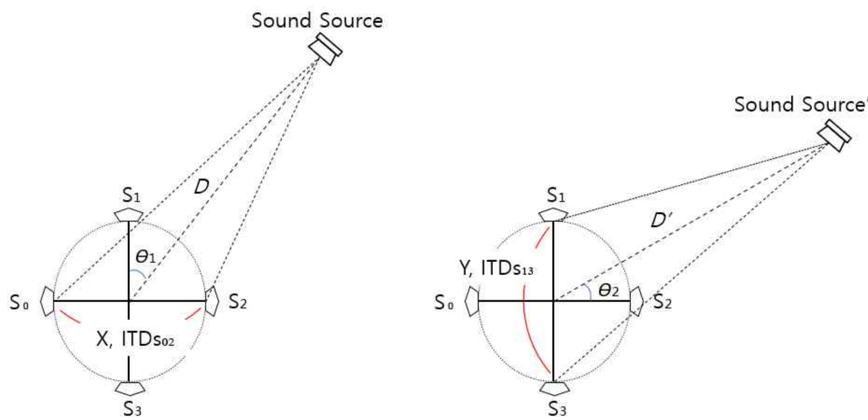

그림 2. 4개의 마이크로폰을 사용하여 ITDs와 θs를 구하는 교차형 기반의 방법
Fig. 2. Cross-shaped based method using four microphones to calculate ITDs and θs

tion of the sound source within the range of 360 degrees. The cross-shaped-based algorithm using four microphones is one of the multiple methods. This method uses the cross-shaped technique by four microphones as shown in Fig. 2.

This method can create two ITDs but only uses an ITD depend on the direction of the sound source. If the sound source is close to the vertical line by S_1 and S_3, this algorithm chooses the ITDs_{02}. Otherwise, the sound source is close to the horizontal line by S_0 and S_2, this algorithm chooses the ITDs_{13} instead of ITDs_{02} as shown in right side of Fig. 2. The microphone S_0, microphone S_1, microphone S_2 and microphone S_3 are four input signals that are arrived to four microphones from the sound source. Sound Source is close to the vertical line by S_1 and S_3, and Sound Source' is close to the horizontal line by S_0 and S_2 as shown in Fig. 2. X is the distances between the microphone S_0 and S_2, and Y is the distances between S_1 and S_3.

This method calculates the direction of the sound source as below. This method pairs S_0 and S_2, and S_1 and S_3 based on the cross-shaped based algorithm. ITDs_{02} and ITDs_{13} are interaural time differences between S_0 and S_2, and S_1 and S_3 according to above pairing. D is the distance between the sound source and the center of four microphones. θ_1 and θ_2 are angles that represent the arrival of the sound source as shown in Fig. 2.

First of all, if the sound source is close to the vertical line by S_1 and S_3, this method uses the θ_1 based on ITDs_{02} to determine the direction of the sound source within the range of each 90 degrees for the forward and backward direction. Otherwise, this method uses the θ_2 based on ITDs_{13} to determine the direction of the sound source within the range of each 90 degrees for the left and right direction. And so, the determined direction is uncleared if the sound source locates in the middle direction between the two microphones, or at the center of four microphones. Therefore, this method has five blind spots at least. Also, this method cannot recognize D, the distance between the sound source and the center of four microphones because we don't know the time that occurs the sound in Fig. 2. And so, this method cannot recognize the distance between the sound source and the microphone.

Ⅲ. Proposal of the QRAS-based algorithm for the OSSD

In order to determine the sound volume, direction and distance to the sound source for the OSSD without the blind spots, we propose a new QRAS-based algorithm. The QRAS-based algorithm consists of three functions: an input function, a processing function and an output function as shown in Fig. 3.

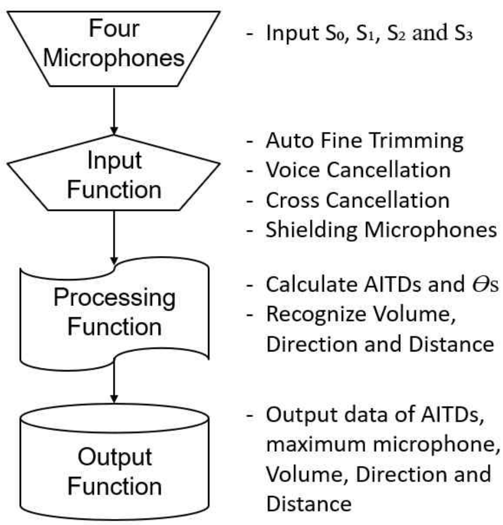

그림 3. QRAS에 기반한 3 역할의 구성
Fig. 3 Architecture of three functions based on QRAS

The input function of this algorithm performs the Automatic Fine Tuning (AFT), voice cancellation, cross-cancellation and shielding microphones. First of all, the input function detects the surrounding sounds by four microphones installed rectangularly on the same plane. As_0, Bs_1, Cs_2 and Ds_3 signals are the output signals of the four microphones. Due to the different sensitivity of the microphones, there are many differences in the signal levels of the four signals extracted from the four microphones. Therefore, the AFT minimizes the deviation between each channel. The extracted signals are automatically fine-tuned by the AFT installed in the system based on the minimum output channel level. Signals processed to the same level as the minimum level are output to the cross-cancellation. The cross-cancellation improves the recognition ability by deleting the voice signal using the built-in voice canceller. Then, the noises such as random noises generated uniformly from each microphone near the unmanned system is suppressed by smoothing the waveforms using the moving average method. After that, the cross-cancellation outputs the minimized signal as minimum_As_0 to minimum_Ds_3 , as shown in Fig. 4.

The input function attenuates the external noises by shielded-microphones as shown in Fig. 5 because the unmanned system has many external noises which vary de-

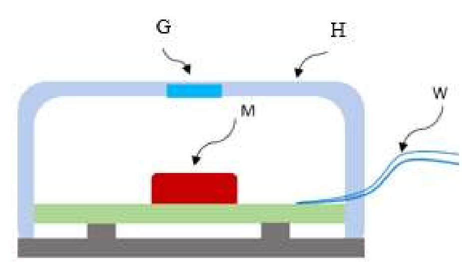

그림 5. 바람과 노이즈에 대한 마이크로폰의 실드
Fig. 5. Shielding the microphone from wind and noises

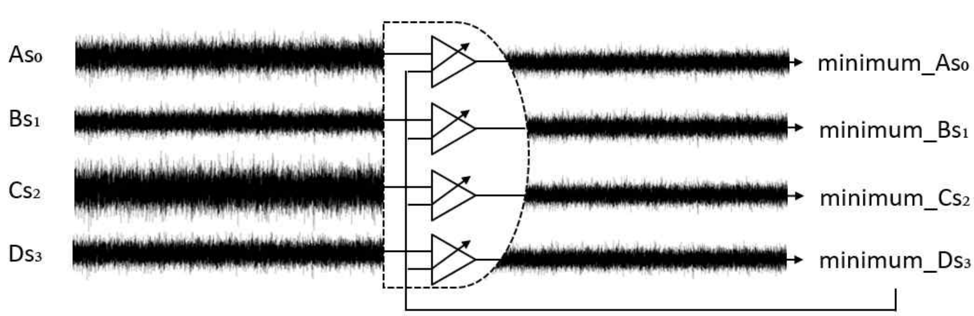

그림 4. 자동 미세 트리밍과 상쇄로 처리된 4개의 신호
Fig. 4. Four signals after that processed by auto fine trimming and countervail

pending on positions. G is the gate to pass the sound signals. M is the microphone and W is the wire connection to connect to the system. H is the housing to shield the wind for the sound signals.

Finally, the input function outputs only the high-weighted signals that have the more increased signal to noise ratio than the detected signals to the processing function. The volume level measures the size of the y signal to which the As_0, Bs_1, Cs_2, and Ds_3 signals are added. The size of the y signal will be measured by the experimental system using the software.

The processing function determines the volume level of the sound source as the sum signal y, regardless of the direction of the sound source, and determines the approximate direction of the sound source. The direction and distance of the sound source are calculated by two sets of 6 pairs of Advanced Interaural Time Difference (AITD) generated from As_0, Bs_1, Cs_2, and Ds_3 signals. To eliminate existing issues such as blind spots, the processing function must share one of the two microphones needed when creating two AITDs.

The ITD is calculated by the different arrival times from the sound source to microphones S_0 and S_1 on the same plane. However, AITDs by the process function are calculated not only by two different arrival times that arrives at S_0S_1 on the same plane, but also by two different arrival times that arrives at S_0S_3 on the same plane. In order to create two AITDs, two pairs of microphones must share a microphone which is located at the crossing of right angle as shown in Fig. 6.

In the processing function, a microphone should be the common microphone as above S_1 when the sound source is located at left side and S_2 when the sound source is located at right side as Fig. 6. When two microphones share a microphone which is located at the crossing of right angle, the processing function choose a microphone which outputs the maximum level among four microphones. So, three microphones provide the output signals and create two AITDs. The common microphone is variable. It is chosen one among four microphones depend on the output level of each microphone. In this method, if the sound source is located at forward-center between the microphone S_1 and S_2, the algorithm selects $AITDs_{13}$ diagonally which is one of 6 AITDs. And so, this method solves the problem of blind spots and computing errors that are occurring in the conventional algorithms. It's because a pair of

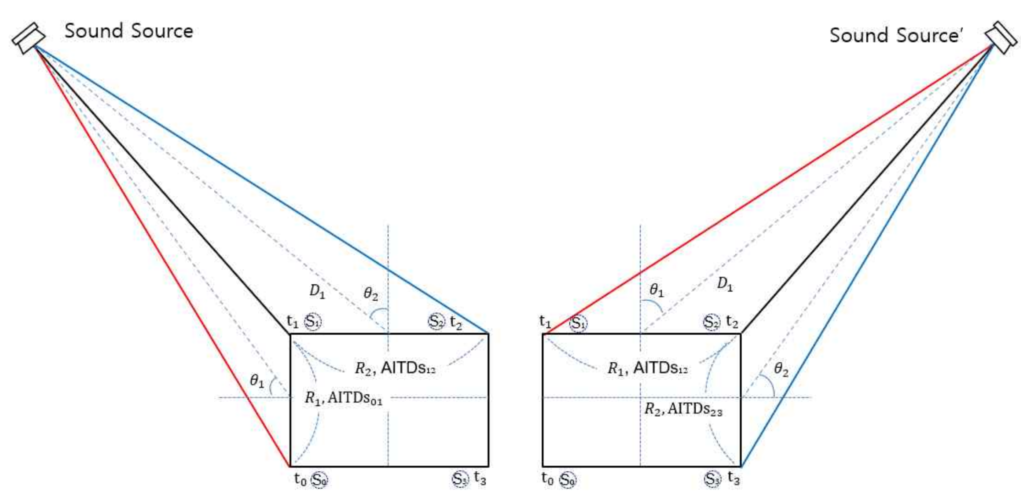

그림 6. 음원과 직사각형 형의 마이크로폰들 사이의 AITDs 구조
Fig. 6. AITDs structure between the sound source and microphones by the rectangular-shaped

relevant AITDs can find the sound source wherever it is located including blind spots in the conventional methods. AITDs_{01} is obtained as Equation (5) and Equation (6) as follows[15].

$$AITDs_{01} = \underset{n}{\arg max} Rs_0s_1(n) \tag{5}$$

where

$$Rs_0s_1 = \sum_{k=0}^{n-1}\Phi(k)Ss_0s_1(k)e^{i\left(\frac{2\pi nk}{N}\right)} \tag{6}$$

For example, θ_1 or θ_2 is computed by AITDs$_{01}$ or AITDs$_{03}$ extracted on the QRAS-based algorithm. θ_1 or θ_2 is the angle of the direction which points out the location of the sound source on the same plane.

D_1 is the distance between the sound source and the microphone S_3 that consists of d_{11} and d_{12} as shown in Fig. 7.

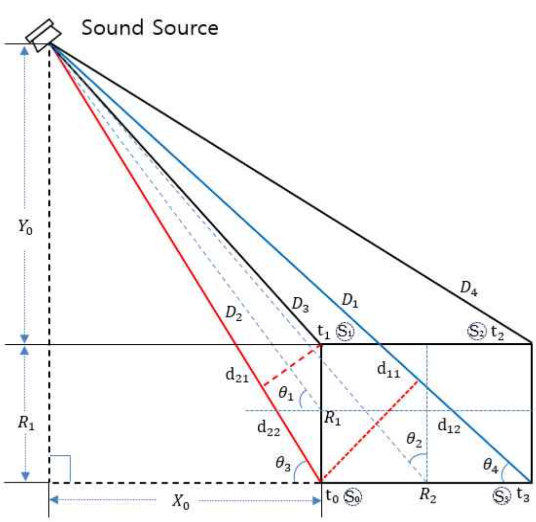

그림 7. 음원과 직사각형 형의 마이크로폰들 사이의 거리 구조
Fig. 7. Distance structure between the sound source and microphones by the rectangular-shaped

d_{11} is the distance between the sound source and the microphone S_0 and d_{12} is the distance that the sound signal travels from the position reaching the microphone S_0 to the microphone S_3. The distance of D_1 can be obtained by Equation (7). t_0 is the time that the sound signal reaches the microphone S_0 from the sound source, and t_3 is the time that the sound signal reaches the microphone S_3. θ_1, θ_2, θ_3 and θ_4 can be calculated by the AITDs with t_0 ~ t_3.

$$D_1 = d_{11} + d_{12} = \frac{R_2 \sin(\theta_4)\sin(90+\theta_4-\theta_3)}{\cos(90+\theta_4-\theta_3)} + (t_3-t_0)*c \tag{7}$$

The distance D_1 is given by Equation (7) but an error is contained. Because, the theory model assumes that the sound source is located at the infinite position, it sets the tangent of D_1 and the red dot line between S_0 and d_{11} • d_{12} at right angle as shown in Fig. 7. Therefore, the correction of the estimated distance needs to be done as given by Equation (8) and Equation (9).

In the processing function, the calculation error occurs because the actual angle of the tangent line is not a right angle at the short distance rather than the infinite position. So, the simulation error, E_S occurs between the actual distance and the computing distance. The processing function requires to add the correction step to minimize the error. CEs^0 and CEs^3 in Equation (8) and Equation (9) are the error corrections, which are created by the computer simulation in the processing function. The error correction is performed in the processing function.

$$\theta_3 = 90 - \left\{\cos^{-1}\frac{(t_0-t_1)*c + R_1 * CEs^0}{R_1}\right\} \tag{8}$$

$$\theta_4 = \cos^{-1}\frac{(t_3-t_0)*c + R_2 * CEs^3}{R_2} \tag{9}$$

As a processing result, the output function shows the volume level by dB, direction of the sound source and distance between the sound source and the microphones on the display. It also shows the maximum input microphone, and the AITDs.

Ⅳ. Implementation of an experimental system which is built in the QRAS-based algorithm

The experimental system consists of 3 blocks: a computing block, a location block and a recognition block as shown in Fig. 8. Rs is a reference signal that controls the input gate: Process or Stop in the input function. Rs level is set by the system software.

The computing block directly performs arithmetic processing to make more faster processing time. In the location block, the detected signals are analyzed firstly to determine surrounding environment. Subsequently it generates 6 pairs of AITDs by the QRAS-based algorithm, and forms the functions needed for the recognition of the location. The computing block and location block calculate the sound volume, direction and distance between the sound source and microphones with the computed result and generated AITDs. In the recognition block, it determines the sound sources as a result. Also, the entity of the sound source is finally recognized by combining the analyzed data at the location block and the computing block.

When the sound signal, y is detected, the QRAS-based algorithm defines the signal, y which is normal signal or abnormal signal by comparing it to the reference signal. If the QRAS-based algorithm detects an abnormal signal, the algorithm manages the detected signals. The QRAS-based algorithm outputs the high weighted signals which minimized the noises contained in the signals.

They are computed to estimate the volume level and approximate the direction of the sound source by the largest signal channel. The y signal is outputted to the computing block to perceive an entity of the sound source. If the situation continues, the direction of the sound source is perceived by As_0, Bs_1, Cs_2 and Ds_3 signals in the computing

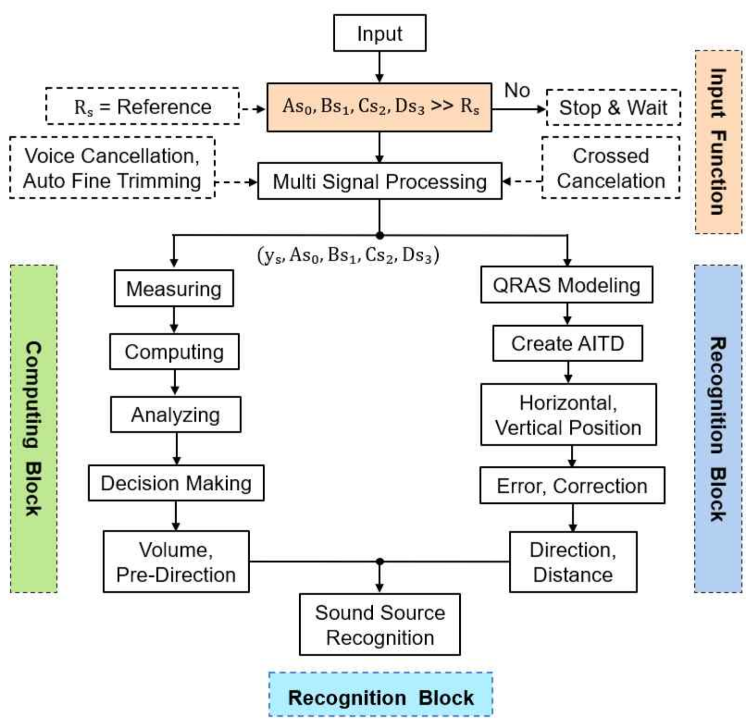

그림 8. QRAS에 기반한 알고리즘의 구조
Fig. 8. Structure of the QRAS-based algorithm

block. The location block generates the AITDs to calculate the position and distance between the sound source and microphones in real time. The recognition block finally extracts the sound volume level, direction and distance to the sound source from 4 signals and the summed y signal, and perceives the sound source.

When a vehicle like an ambulance is approaching the microphone S_3 of the unmanned system, the time Ts_3 for the sound source to reach the microphone S_3 can be expressed as Equation (10). Ts_3 is used for a simulation in the experimental system.

$$Ts_3 = t_3 = \frac{\sqrt{(R_2 + x_0)^2 + (R_1 + y_0)^2}}{c} \quad (10)$$

The distance D_1 between the sound source and the microphone S_3 is given by Equation (11), Equation (12) and Equation (13).

$$D_1 = \frac{\overline{S_0S_3}\sin(\theta_4)\sin(90+\theta_4-\theta_3)}{\cos(90+\theta_4-\theta_3)} + (t_3 - t_0)*c \quad (11)$$

$$\theta_3 = 90 - \left\{\cos^{-1}\frac{(t_0-t_1)*c + \overline{S_0S_1}*CEs^0}{\overline{S_0S_1}}\right\} \quad (12)$$

$$\theta_4 = \cos^{-1}\frac{(t_3-t_0)*c + \overline{S_0S_3}*CEs_3}{\overline{S_0S_3}} \quad (13)$$

$\overline{S_0S_3}$ and $\overline{S_0S_1}$ are given by Fig. 7. $t_0 \sim t_3$ are measured and 6 AITDs are computed by the QRAS-based algorithm.

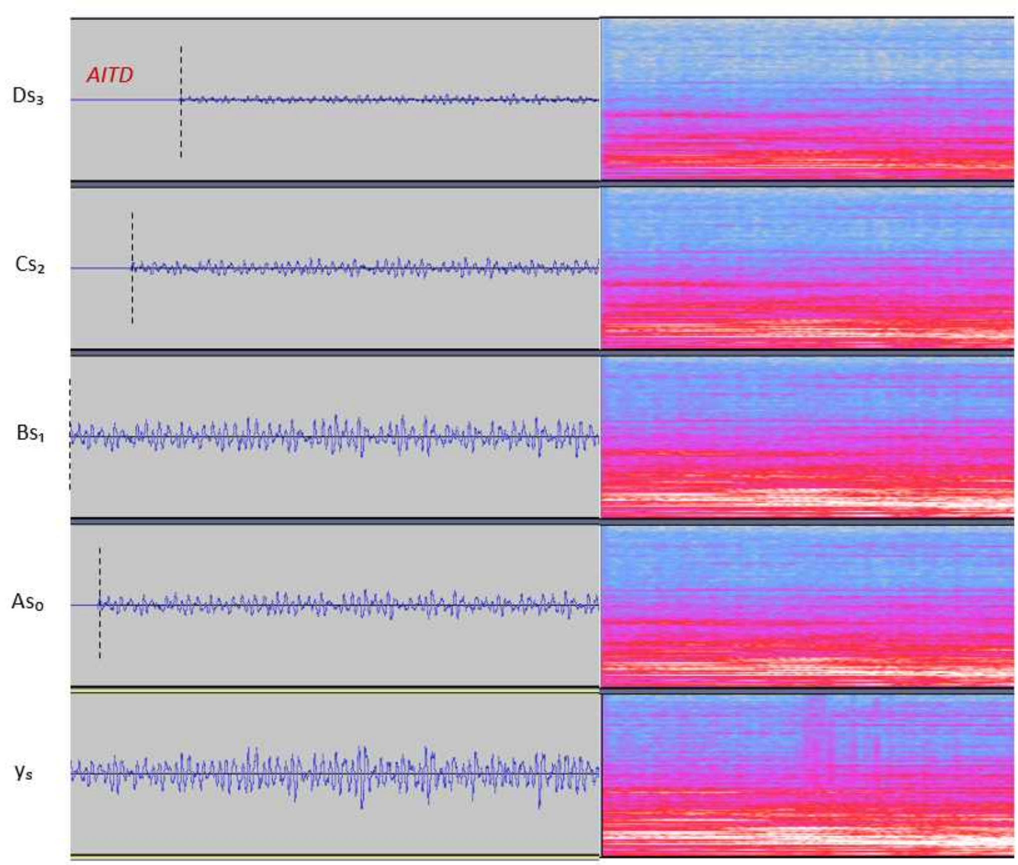

그림 9. AITD를 갖는 As_0, Bs_1, , Cs_2 , Ds_3 신호와 y_s 신호의 파형
Fig. 9. Waveforms of As_0, Bs_1, , Cs_2 , Ds_3 and y_s signals with AITDs

The y_s signal is used with As_0, Bs_1, Cs_2 and Ds_3 signals including the AITDs for these experiments. The frequency characteristics and spectra of these signals are presented as shown in Fig. 9.

The experimental system designed by the QRAS-based algorithm for the OSSD outputs the test results as shown in Fig. 10. It displays the volume by “Max input channel”, the direction with LEDs and the distance by “F Range” as right-side. It also captures the AITD of S_0S_1 by “Vertical tau” and the AITD of S_0S_3 by “Horizontal tau” as the top of right-side screen. The top of left-side screen shows the real time data of test results.

Ⅴ. Experimental results and discussion

A series of test results shows that the rectangular-shaped QRAS-based algorithm is capable of determining the direction of the sound source without blind spots. Table 1 compares the success rate of directional determination for the rectangular-shaped QRAS-based algorithm versus and the cross-shaped algorithms. An ambulance sound at 75dB was used as the test source in 12 directions separated by 30 degrees. Tests were performed with 50 trials at each directional position. The data in the tables 1 shows that the rectangular-shaped QRAS-based algorithm has a better direc-

표 1. 방위각 30도로 분리된 12 방향의 방향 인식률
Table 1. Direction determination rates at 12 directions-each separated 30 degrees in azimuthal angle

Test Methods \ Directions		12	1	2	3	4	5	6	7	8	9	10	11
Rectangular- Shaped (50 Trials)	Correctly determined (Trials)	50	50	50	50	50	50	50	50	50	50	50	50
	Rate (%)	100	100	100	100	100	100	100	100	100	100	100	100
Cross-Shaped (50 Trials)	Correctly determined (Trials)	50	49	48	50	48	47	50	47	48	50	48	49
	Rate (%)	100	98	96	100	96	94	100	94	96	100	96	98

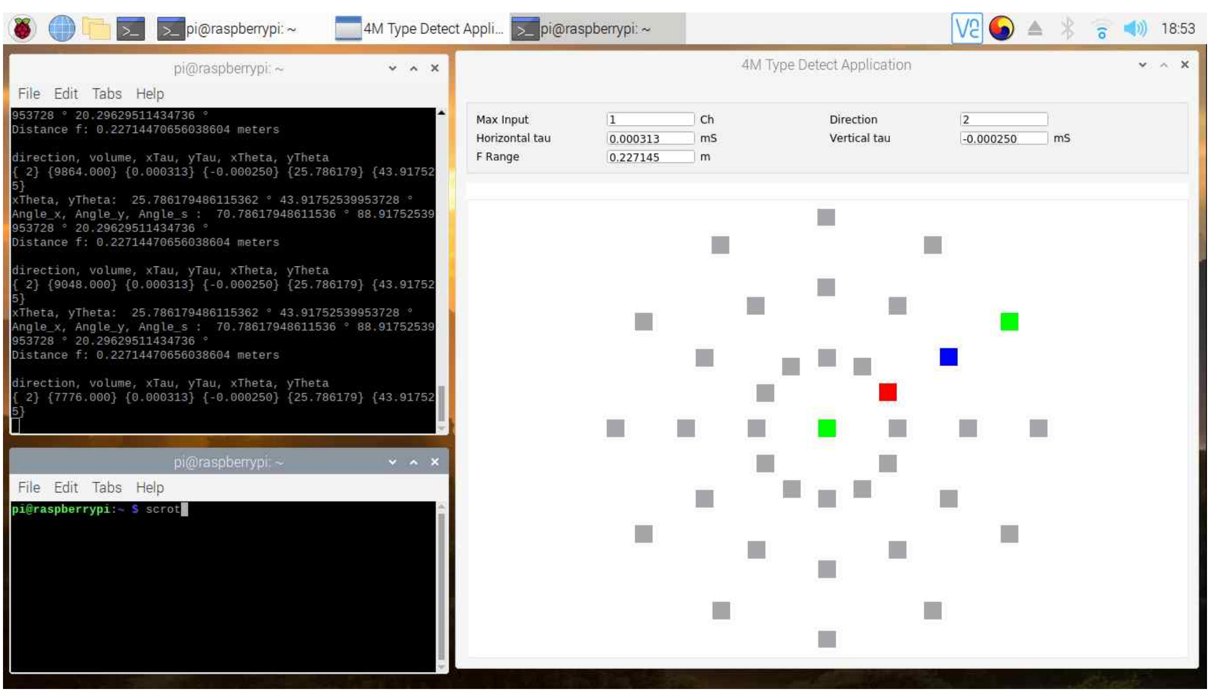

그림 10. 음량, 방향과 거리를 보여주는 화면
Fig. 10. A display screen showing the volume, direction and distance

tional detection success rate than the cross-shaped based algorithm at 8 of the positions corresponding to the cardinal compass points. The lower success rate of the cross-shaped based algorithm means that it had some blind spots in those tests. It is reasonable to expect that the cross-shaped based algorithm would have problems in the middle direction between the two microphones.

The measured volume level is used for the rectangular-shaped QRAS-based algorithm, and it only shows the position of the microphone with maximum volume on the display, as shown in Fig. 10.

Trials were performed to test the capability of the QRAS-based algorithm determine distances to sound sources. This type of test can only work with the QRAS-based algorithm, because the cross-shaped based algorithm with four microphones, gives no useful information to determine the distance between the source and the microphones.

The distance between the sound source and the microphone S_3 was computed with QRAS-based algorithm. In the two-dimensional case, the calculated distances (after CE_S compensation) produced favorable results with errors in the range of 7.2% ~ 1.6% as shown in Table 2. The trials were performed at 3 distances: 14.26m (minimum); 70.83m (intermediate); and 141.54m (maximum). The errors shown in Table 2 are the differences between the real distance and calculated distance.

표 2. CE_S로 보상한 후의 실제 거리와 연산 거리
Table 2. Actual distances and calculated distances after CE_S compensation

X_0,Y_0 Distance in Fig. 7 (m)	Actual Distance: (m)	Calculated Distance: $D_{SA}w/CE_S$ (m)	Error (%)
10	14.26	13.23	7.2
50	70.83	69.2	2.3
100	141.54	139.34	1.6

The sound source was identified as ‘an ambulance’ using the extracted features of y signals. Results are shown Table 3 for various values of Signal to Noise Ratio (SNR). The table displays shows ‘Detections’ and ‘Errors’ for 100 trials at each of the indicated values of SNR: 40dB, 30dB, 20dB and 10dB, which were associated with using the y_s signal. Tests confirmed that the correct identification rate is more than 97% for ‘an ambulance’ above 20 dB SNR . However, the identification rate of the signal source with a large amount of noise (for example 10dB SNR) is drastically decreased. This is related to the performance of signal activity detection, and will be elevated using an advanced signal activity detection.

An ambulance sound at 75dB signal level was used for these tests. The test sound was merged with an ambulance sound and a noise signal that was provided by the noise signal generator. Even through the cross-shaped based algorithm using four microphones shows no information about sound source identification with comparable SNR, we think its capability might be similar to the QRAS-based algorithm of the rectangular-shaped.

표 3. 잡음을 포함한 y_s 신호로 구급차를 인식하는 시험 결과
Table 3. Test results for an ambulance recognition by y_s signal with noises

Count \ SNR dB	40	30	20	10	0
Detection (Trials)	100	100	97	23	0
Error (Trials)	0	0	3	77	100
Total (Trials)	100	100	100	100	100
Detection Rate	100 %	100 %	97 %	23 %	0 %

Ⅵ. Conclusion

In this paper, we propose the QRAS-based algorithm using four microphones for the OSSD without blind spots. This algorithm can determine volume, direction and distance between the sound source and microphones in real time; and it shows the recognized sound source on the display. It performs well, even when the experimental system is located at the long distance from the sound source. The QRAS-based algorithm gives better recognition performance even though the experimental system requires only four microphones - same as conventional methods. Therefore, it is confirmed that the QRAS-based algorithm for the OSSD can be applied to implement an auditory function into the unmanned system.

As future work, the distance estimation error can be examined at various distances between the sound source and the microphones, and elevated using the proposed QRAS-based algorithm.

참 고 문 헌 (References)

[1] Y. T. Cho, C. M. Park, Y. G. Kim, Y. Kim, and I. B. Jung, "Distance measurement system between sensor in the wireless sensor network", Korea information processing society, Vol. 13 No. 1, pp. 1263-1266, 2006.
[2] D. Lee, B. Jang, S. Im, and J. Song, "A new sound source localization approach using stereo directional microphones", International conference on information communication and signal processing", pp.504-508, 2019.
[3] H. Liu, J. Zhang, and Z. Fu, "A new hierarchical binaural sound source localization method based on interaural matching filter", International conference on robotics & automation, pp. 1598-1605, 2014.
[4] C. M. Zannini, R. Parisi, and A. Uncini, "Binaural sound source localization in the presence of reverberation", International conference on digital signal processing 2011, pp. 1-6, 2011.
[5] P. Parhizkari, "Binaural hearing human ability of sound source localization", Blekinge institute of technology, pp. 01-57, 2008.
[6] H. Gu, and S. Yang, "A sound-source localization system using three-microphone array and cross power spectrum phase", International conference on machine learning and computing, pp. 1-5, 2012.
[7] C. H. Adritya, and H. M. Saputra, "Azimuth estimation based on generalized cross correlation phase transform using equilateral triangle microphone array", International conference on radar, antenna, microwave, electronics and telecommunications, pp. 89-93, 2019.
[8] B. Rudzyn, W. Kadous, and C. Sammut, "Real time robot audition system incorporating both 3D sound source localization and voice characterization", International conference on robotics and automation, pp. 4733-4738, 2007.
[9] B. Al-Sheikh, A. Elshebli, A. Al-Assaf, S. Almashaqbeh, W. Alrawabdeh, and A. Al-Tahat, "Sound source direction estimation in horizontal plane using microphone array", International conference on applied electrical engineering and computing technologies, pp. 1-6, 2013.
[10] L. Rayleigh, "On our perception of sound direction", Philosophical magazine, pp. 214-232, 1907.
[11] B. Grothe, M. Pecka, and D. McAlpine, "Mechanisms of sound localization in mammals", The American physiological society, Physiol rev. 90, pp. 983-1012, 2010.
[12] M. Zohourian, and R. Martin, "Binaural speaker localization and separation based on a joint ITD/ILD model and head movement tracking", International conference on acoustics, speech, & signal processing 2016, pp. 430-434, 2016.
[13] M. Raspaud, H. Viste, and G. Evangelista, "Binaural source localization by joint estimation of ILD and ITD", Transactions on audio, speech, and language processing, vol. 18, pp. 68-77, 2010.
[14] E. Visser, M. P. Zwiers, C. C. Kan, L. Hoekstra, A. J. Opstal, and Jan K. Buitelaar, "A typical vertical sound localization and sound-onset sensitivity in people with autism spectrum disorders", Canadian medical association, pp. 398-406, 2013.
[15] J. D. Chen, J. Benesty, and Y. Huang, "Performance of GCC- and AMDF-based time-delay estimation in practical reverberant environments", European association for signal processing journal on applied signal processing, pp. 25-36, 2005.

저 자 소 개

김 영 언

- 1970년 ~ 2000년 : 삼성전자(주) 시스템 LSI 연구소 수석연구원
- 2000년 ~ 2016년 : LSI, ST-Micro(Genesis), Synaptics, Sigma Desin Director
- 2019년 : 서울과학기술대학교 정보통신미디어공학 박사
- 2020년 ~ 현재 : 동오정밀(주) 부설연구소 수석연구위원
- ORCID : https://orcid.org/0000-0002-4626-1674
- 주관심분야 : Autonomous Driving, AI, 비언어인식

저 자 소 개

박 구 만

- 1984년 : 한국항공대학교 전자공학과 공학사
- 1986년 : 연세대학교 대학원 전자공학과 석사
- 1991년 : 연세대학교 대학원 전자공학과 박사
- 1991년 ~ 1996년 : 삼성전자 신호처리연구소 선임연구원
- 1999년 ~ 현재 : 서울과학기술대학교 전자IT미디어공학과 교수
- 2006년 ~ 2007년 : Georgia Institute of Technology, Dept.of ECE. Visiting Scholar
- 2016년 ~ 2017년 : 서울과학기술대학교 나노IT디자인융합대학원 원장
- ORCID : https://orcid.org/0000-0002-7055-5568
- 주관심분야 : 컴퓨터비전, 실감미디어, 딥뉴럴네트워크