

# 협소 공간 내에 생존자 탐색을 위한 인명 탐지 모듈과 최소 제곱법 기반 음원 위치 추정 알고리즘에 대한 연구

## A Study on the Survivor Detection Module and Least-Squares Sound Source Localization Algorithm for Victim Search in Narrow Spaces

석윤정<sup>1</sup>·김성재<sup>2</sup>·박서연<sup>3</sup>·서진호<sup>†</sup>

Yun-Jeong Seok<sup>1</sup>, Sung-Jae Kim<sup>2</sup>, Seo-Yeon Park<sup>3</sup>, Jin-Ho Suh<sup>†</sup>

**Abstract:** In this study, we designed a survivor detection module for victim search in narrow spaces and proposed Least Squares-Time Delay of Arrival (LS-TDOA) based sound source localization and visualization algorithms to be integrated into the module. The module was designed as a compact structure capable of navigating through confined spaces, with an expandable umbrella-like mechanism for size adjustment when needed. Additionally, the module is equipped with a gas sensor for monitoring the surrounding environment and a thermal camera for detection in low-light conditions, alongside a microphone for sound source detection. To improve the sound source localization algorithm, we utilize voice activity detection (VAD) to distinguish human voices, thereby improving robustness to environment noise. The least squares method was employed to calculate the coordinates of the sound source, minimizing localization errors. Finally, the performance of the proposed algorithms was verified through simulations across various environments, and field experiments confirmed the effectiveness of the sound source visualization and localization.

**Keywords:** Sound Source Localization, Time Difference of Arrival, Narrow Space, Survivor Detection Module, Least Squares

### 1. 서 론

국내외를 불문하고 여전히 많은 인명 피해를 동반한 붕괴 사고가 지속적으로 발생하고 있다. 일본 고베 대지진 생존율 통계에 따르면, 붕괴 사고 발생 후 3일 이내에 구조가 이루어질 때 생존율이 가장 높다<sup>[1]</sup>. 그러나 탐색 작업이 선행되지 않을 경우,

증장비 투입이 오히려 2차 붕괴를 초래해 매몰자의 생명을 더욱 위협할 수 있다. 따라서, 생존자 구조를 위해서는 매몰자의 위치를 파악하는 점이 선행되어야 한다. 기존의 구조 방법은 구조조거나 구조 인력이 직접 투입되는 수동적인 방식을 사용하였으나, 최근에는 탐지 로봇<sup>[2,3]</sup>을 사용하여 매몰자의 위치를 탐색하는 방법이 도입되고 있다. 그러나 이러한 로봇 활용에는 여전히 많은 한계가 존재한다. 특히, 2017년 멕시코 대지진 현장에 백형 로봇을 투입하여 수색을 진행한 연구팀에 따르면 추가적인 탐지 수단의 부족<sup>[4]</sup>과 더불어, 매몰 공간 내 비산 먼지 등으로 인해 카메라의 시야 확보가 어려워<sup>[5]</sup> 탐색 작업에 어려움을 겪었다. 이를 개선한 최신 선행 연구<sup>[6]</sup>에서는 협소한 공간에 적합한 센서 배치와 함께 다양한 센서를 탑재한 로봇 시스템을 제안하고 있으며, 영상 정보에만 의존하지 않고 음원 기반의 위치 추정 방식을 도입하였다. 해당 연구에서는 TDOA (Time Difference of Arrival)<sup>[7]</sup>를 활용하여 생존자의 위치를 파악하고,

Received : Oct. 24. 2024; Revised : Nov. 14. 2024; Accepted : Dec. 6. 2024

※ This work was supported by the Technology Innovation Program (No. 20018110, "Development of a wireless teleoperable relief robot for detecting searching and responding in narrow space") funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

1. Master Student, Intelligent Robot Engineering, Pukyong University, Seoul, Korea (amcnjeong00@gmail.com)

2. Senior Researcher, Industrial Science Technology Research center, Pukyong University, Seoul, Korea (bbman7020@gmail.com)

3. Assistant Researcher, Korea Institute of Robotics & Technology Convergence, Pohang, Korea (sypark@kro.re.kr)

† Professor, Corresponding author: Mechanical System Engineering, Pukyong University, Seoul, Korea (suhgang@pknu.ac.kr)

이를 보정하기 위해 Calibrated GCC-PHAT (Generalized Cross Correlation-Phase Transform) 알고리즘을 적용하여 강건한 추정을 유지하고 있다.

그러나 이러한 방식은 잡음에 취약하고, 필터 알고리즘의 느린 반응으로 인해 탐지 속도가 느리다는 단점이 지적되고 있다. 특히, 사람의 목소리만을 정확히 탐지하지 못해 유사한 주파수 대역의 소리가 감지될 때 오류가 발생할 수 있다는 문제가 확인되었다. 또한, 두 개의 마이크로폰만을 사용하여 계산하는 방식은 고유의 변환 방식으로 인해 0~30도 사이의 각도를 정확하게 추정하는 데 한계가 있다. 더불어, 탐지 결과를 직접적으로 가시화하지 못하여 구조자가 직관적으로 결과를 이해하기 어려운 점도 문제로 지적된다.

현재 사용되고 있는 TDOA 방식은 적은 수의 마이크로폰으로도 구현이 가능하고, 계산 방식이 단순하여 저비용으로 탐지가 가능하다는 점에서 널리 활용되고 있다<sup>[8,9]</sup>. 또한 원형<sup>[10]</sup>이나 사각형<sup>[11]</sup>과 같은 다양한 마이크로폰 배열을 사용하여 실시간으로 높은 정확도의 3D 음원 추정을 가능하게 한다는 점에서 많은 연구가 진행되고 있다<sup>[12]</sup>. 그러나 복잡한 환경에서의 성능에는 한계가 있으며, 기존 AoA (Angle of Arrival) 방식은 삼각함수의 특성상 음원의 위치가 배열의 양 끝단으로 갈수록 정확도가 크게 저하된다는 단점이 있다.

이러한 부분을 해소하기 위해 다양한 개선 방법들이 제안되었다. 최근에는 데이터 클러스터링을 통해 음원의 위치를 추정하는 방식<sup>[13]</sup>이나, 딥러닝을 사용하여 CNN (Convolutional Neural Network)을 사용하여 음원을 처리하는 방식<sup>[14]</sup> 등 고도화된 방법들이 등장하였다. 그러나 이러한 방식들은 높은 정확도를 달성하기 위해 고차원의 데이터를 처리해야 하므로 계산 비용이 크게 증가하는 단점이 있다. 따라서 최신 연구와 같이 네트워크를 사용하는 추정 방식의 경우 협소한 공간에서는 사용하기 어렵다. 이에 따라, 적절한 계산 비용과 나은 위치 추정 성능을 제공하는 방법으로, 선형적 통계 방법을 적용한 연구<sup>[15-17]</sup>를 사용하고자 한다.

따라서 본 연구에서는 협소한 100 mm × 100 mm의 공간에서도 운용할 수 있도록 다양한 센서를 통합하고, 고도각과 방위각을 탐지할 수 있는 마이크로폰 배열을 활용하여 좁은 공간에서도 효율적으로 사람을 탐지할 수 있는 유닛을 설계하고자 한다. 유닛에는 사람의 목소리를 구분하기 위한 VAD (Voice Activity Detection) 기술을 도입하였다<sup>[18-20]</sup>. 또한, 적은 계산량으로 음원의 위치를 정밀하게 추정할 수 있도록 최소 제곱법 (Least Squares, LS) 기반의 향상된 LS-TDOA 음원 추정 알고리즘을 적용하였다. 이와 더불어, 시스템을 효율적으로 운영하고 시각화할 수 있는 음원 가시화 알고리즘을 구현하였다. 최종적으로, 다양한 상황에 적용할 수 있도록 시뮬레이션을 통해 성능을 검증하였다.

이 논문은 다음과 같은 기여를 갖는다.

- 협소한 100 mm × 100 mm의 공간에서도 운용할 수 있도록 다양한 센서와 마이크로폰 배열을 통해 좁은 공간에서도 효율적으로 탐지할 수 있는 유닛을 설계한다. 이 유닛에는 인명 탐지에 효과적으로 사용할 수 있는 열화상 카메라, 마이크로폰, 가스 센서, 이산화탄소 센서를 포함한다.
- 이 논문에서는 MCU에 구현 가능한 GCC-PHAT TDOA 기반 음원 위치 추정 알고리즘을 사용한다. 견고한 음원 위치 추정 알고리즘을 구현하기 위해 VAD 알고리즘을 통해 사람 음성에 대해 알고리즘이 활성화될 수 있도록 구성하였다. 추가적으로 최소 제곱법을 도입하여 음원 위치 추정 알고리즘의 해상도를 향상했다.
- 고안한 알고리즘을 효과적으로 가시화하기 위해 열화상 카메라에 매핑하는 알고리즘을 고안하여 GUI에 표시했다.
- 고안한 인명 탐지 모듈을 제작하여 실험을 통해 성능을 검증했다.

본 논문은 다음과 같이 구성된다. 2장에서는 인명 탐지 모듈의 하드웨어 설계 및 구조를 다루며, 3장에서는 모듈에 탑재된 알고리즘을 설명한다. 4장에서는 시뮬레이션과 실증 실험 결과를 제시하며, 5장에서는 본 연구의 결론을 논의한다.

## 2. 인명 탐지 모듈 개발

### 2.1 공간적 제약 사항

본 연구에서는 협소한 공간에서의 효과적인 음원 위치 추정을 목표로 하였으며, 이를 위해 플랫폼의 크기를 최소화하는 설계가 필요하다. TDOA 방식은 마이크로폰 간의 간격에 따라 위치 추정의 해상도가 달라지므로, 이론적으로는 마이크로폰 간의 거리를 넓히는 것이 유리하다. 그러나 제한된 공간에서 운용을 고려할 때, 마이크로폰 간의 거리를 늘리는 것에는 한계가 있다. 따라서 소형 플랫폼 내에서 적절한 마이크로폰 배열을 통해 정확한 위치 추정을 가능하게 하는 설계가 필수적이다. 그러므로 본 연구에서는 최종적으로 플랫폼의 크기를 70 mm 이하로 설계하되, 마이크로폰을 포함한 다양한 센서를 배치하고 음원 위치 추정의 정확도를 향상하는 방향으로 고도화한다.

### 2.2 인명 탐지 모듈의 센서 선정

앞서 언급한 공간적 제약을 고려하여, 5 cm 이하의 소형 센서를 탑재하였으며, 탐지 범위 내에서 비교적 높은 해상도를 제공할 수 있는 센서를 선정하였다. 가시화 센서로는 저조도 환경에서도 인명 탐지가 가능하며 3 m 이상의 거리에서도 사람의 형상을 선명하게 구분할 수 있는 해상도를 고려하여 Boson 640

열화상 카메라를 사용하였다. 음향 신호 수음을 위해 회로 설계가 용이하며 크기가 작은 4개의 MEMS 마이크론을 전방에 배치하였다. 다음으로, 생존자 주변의 기체 환경에서 유독 가스 유무를 확인하기 위해 일산화탄소 등을 감지할 수 있는 MQ2 가스 센서와 생존 징후인 호흡을 확인하기 위한 CO2 센서를 장착하였다. 가스 센서의 경우, 빠른 반응 시간과 회복 속도를 고려하여 선정하였다. 각 센서로부터 데이터를 안정적으로 수집하고 DSP 알고리즘을 효율적으로 처리하기 위해 STM32F4 기반의 통합 제어 보드를 설계하였다. 이 제어 보드는 센서 데이터를 처리한 후, 데이터를 메인 PC(이하 조종기 PC)로 전송하는 역할을 수행한다. 이후 조종기 PC에서는 그래픽 사용자 인터페이스(GUI)를 통해 탐지된 정보를 시각적으로 확인하고, 구조자가 신속하게 대응할 수 있도록 직관적인 위치 가시화 기능을 제공한다.

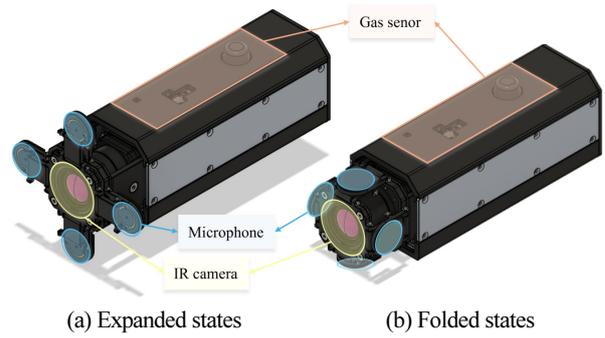
### 2.3 인명 탐지 모듈 하드웨어의 구조

하드웨어 구조는 소형화를 필두로 각 센서에서 효과적으로 데이터를 수집할 수 있도록 고안하였다. TDOA 기반 음원 추정 알고리즘의 특성상, 마이크론 간의 간격이 넓을수록 위치 추정의 정확도가 높아지므로, 공간적 제약을 극복하기 위한 접이식 메커니즘을 설계하였다. [Fig. 1]에 제시된 우산형 접이식 구조를 통해, 이동 시에는 장치가 55 mm로 작게 접혀 소형화되고, 내부 수색 시에는 100 mm로 넓게 펼쳐져 마이크론 간의 적절한 간격을 확보할 수 있다. 이는 마이크론의 해상도를 높여 협소한 공간에서도 효율적인 탐색을 가능하게 한다. 최종 설계에서는 [Fig. 2]의 (a)와 같이 열화상 카메라를 중심으로, 방위 각과 고도각 추정을 위해 가로축과 세로축에 각각 2개의 마이크론을 배치하였다. 가스 센서는 탐지 효율성을 높이기 위해 전방 상단에 위치시켰다.

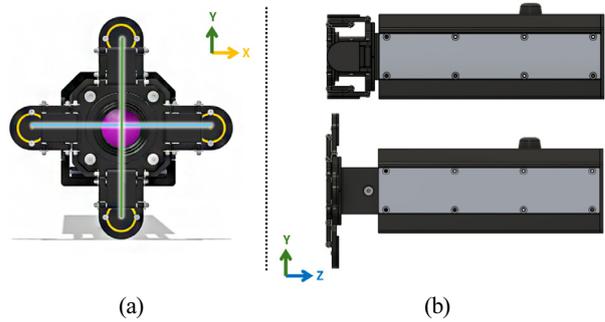
## 3. 생존자 탐색 알고리즘

### 3.1 제안하는 음원 위치 추정 알고리즘

음원 위치 추정 알고리즘으로는 MUSIC과 TDOA가 대표적으로 사용된다. 그러나 협소한 공간에서 플랫폼 크기를 최소화해야 한다는 제약으로 인해, 다수의 마이크론을 요구하는 MUSIC 알고리즘의 적용에는 한계가 있다. 또한, 해상도 증가를 위해 MUSIC 알고리즘은 많은 연산량을 요구하며, 이는 연산 자원이 제한된 MCU 환경에서 구현하기 어렵다. 이에 따라, 본 연구에서는 TDOA 알고리즘을 기반 음원 위치 추정 알고리즘을 제안한다.



[Fig. 1] Umbrella mechanism in expanded (a) and folded states (b)



[Fig. 2] Front (a) and side (b) views of the compact survivor detection module. The horizontal axis (blue) and vertical axis (green) are centered on the thermal camera. A gas sensor is mounted on top. Hardware dimensions: Maximum Length ( $x \times y \times z$ ): 100 mm  $\times$  100 mm  $\times$  178 mm, Minimum Length ( $x \times y \times z$ ): 55 mm  $\times$  55 mm  $\times$  178 mm

#### 3.1.1 GCC-PHAT 기반 TDOA

생존자 탐색을 위한 기존 방법은 GCC-PHAT를 사용한 TDOA 방식이다<sup>[6]</sup>. 이 방식은 음원에서 발생한 소리가 각 마이크론에 도달하는 시간을 기반으로 음원의 위치를 추정한다. 소리가 마이크론에 도달할 때 발생하는 거리 지연  $d$  를 소리의 속도인 342 m/s로 나누어 시간 지연  $\tau$  를 계산하며, 이를 통해 음원의 위치를 추정한다. 이때 시간 지연은 다음과 같이 정의된다.

$$\tau = \frac{d}{c} \quad (1)$$

이후, 마이크론에 도달하는 신호의 시간 차이( $\tau$ )를 추정하기 위해 GCC-PHAT가 사용된다. 실제 환경에서는 잡음, 반향 등의 영향으로 인해 정확한 시간 차이를 추정하기 어렵는데, GCC-PHAT는 신호의 위상 정보를 강조하고 진폭 정보를 제거하여 이러한 문제를 완화하고, 정확한 TDOA 추정을 가능하게 한다.

GCC-PHAT는 두 신호  $x(t)$ 와  $y(t)$  간의 교차 상관함수식의 교차 상관함수를 사용하여 신호 간의 시간 차이를 추정한다.

두 신호 간의 교차 상관함수는 시간 영역에서 다음과 같이 정의된다.

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)y(t-\tau)dt \quad (2)$$

이를 주파수 영역으로 변환하기 위해 푸리에 변환을 적용하면,  $x(t)$ 와  $y(t)$ 는 각각  $X(w)$ 와  $Y(w)$ 로 나타낼 수 있다. 교차 상관함수는 주파수 영역에서 다음과 같이 계산된다.

$$R_{xy}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(w)Y^*(w)e^{jw\tau}dw \quad (3)$$

여기서  $Y^*(w)$ 는  $Y(w)$ 의 켈레 복소수를 의미하며, 두 신호의 위상 차이를 추출하기 위해 사용된다. 이후 가중치를 통해 교차 스펙트럼의 진폭 정보를 제거하고 위상 정보만 남긴다. 교차 스펙트럼의 크기(Amplitude)를 사용하여 가중치를 계산하며, 이 가중치는 식 (4)와 같이 표현한다.

$$W(w) = \left| \frac{1}{X(w)Y(w)} \right| \quad (4)$$

이를 기반으로, PHAT 가중치를 적용한 교차 상관함수는 주파수 영역에서 다음과 같이 표현된다.

$$R_{xy}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(w)X(w)Y^*(w)e^{jw\tau}dw \quad (5)$$

여기서  $W(w)$ 는 PHAT 가중치이며 식 (5)는 GCC-PHAT 를 나타낸다. 마지막으로, 시간 지연은  $\tau$ 는 교차 상관 함수  $R_{xy}(\tau)$ 에서 최댓값을 갖는  $\hat{\tau}$ 를 찾는 방식으로 결정된다.

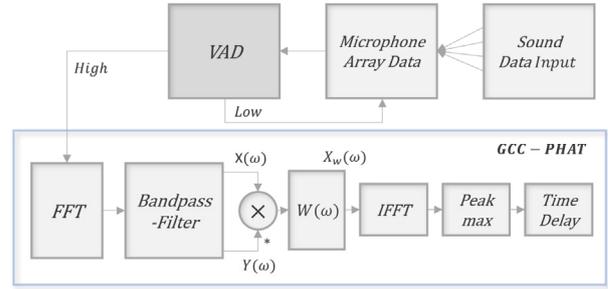
$$\hat{\tau} = \operatorname{argmax}_{\tau} R_{xy}(\tau) \quad (6)$$

이렇게 추정된  $\hat{\tau}$  값을 TDOA 방정식에 대입하여 음원의 위치를 추정할 수 있다.

### 3.1.2 VAD를 적용한 TDOA

사람의 목소리 대역 주파수에만 필터를 적용할 경우, 동일한 주파수 대역의 잡음이 함께 유입되면 음원 위치 추정 성능이 저하될 수 있다. 이를 극복하기 위해, 본 연구에서는 사람의 목소리만을 효과적으로 추정하기 위해 VAD (Voice Activity Detection)를 적용하여 성능을 향상시켰다<sup>[20]</sup>. VAD는 음성 신호에서 말하는 구간과 말하지 않는 구간을 구분하는 기술로, 이를 통해 음성 신호만을 추출하여 잡음의 영향을 최소화할 수 있다.

먼저, 적절한 오디오 프레임에서 사람의 말소리 주파수 이외의 저주파와 고주파 잡음을 제거한다. 그 후, 음성 신호에서 높



[Fig. 3] Flowchart of the TDOA algorithm using VAD

은 에너지를 감지하고, 일정한 패턴을 가진 신호를 구분한다. 이렇게 필터링된 신호에서 배경 소음과 잡음을 제거하면, 소음이 많은 환경에서도 사람의 목소리를 정확하게 감지할 수 있다. VAD가 적용된 알고리즘의 흐름은 [Fig. 3]과 같다.

VAD를 통해 사람 음성을 정확하게 추출한 후, TDOA 알고리즘을 적용함으로써 음성 기반의 위치 추정 성능을 개선할 수 있다.

### 3.1.3 최소제곱법을 적용한 TDOA

앞서 구현한 음원 추정 기술을 좌표로 표현하기 위해 최소제곱법(Least Squares Method)을 사용한다. 최소제곱법은 주어진 데이터에 가장 적합한 해를 구하기 위한 통계적 방법으로, 잔차(residual)의 제곱합을 최소화하는 방식으로 모델의 매개변수를 추정한다. 이 방법은 관찰된 데이터와 모델이 예측하는 값 사이의 차이를 최소화하는 데 사용된다. 이를 통해 음원의 좌표를 정확하게 추정하고, 다양한 환경에서 발생하는 오차를 보정할 수 있다.

마이크로폰의 위치를  $(x_i, y_i)$ 라고 할 때, 각각의 위치는  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 으로 정의할 수 있다. 음원의 위치  $(x_s, y_s)$ 라고 할 때 마이크로폰과 음원 간의 거리 차이는 다음과 같이 나타낼 수 있다.

$$d_i = \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2} \quad (7)$$

여기서 음속은  $c$ 로 나타낸다. 각각의 거리 차이는 TDOA (Time Difference of Arrival) 방식을 통해 각 마이크로폰에 도달하는 신호의 시간 지연을 기반으로 계산되며, 이를 수식화하면 다음과 같다.

$$\tau_i = \frac{d_i - d_1}{c} \quad (8)$$

이때, 각 마이크로폰과 음원의 거리 차이를 행렬 방정식으로 변환할 수 있다. 위의 관계를 일반적인 행렬 방정식 형태로 나타내면 다음과 같다.

$$\tau \cdot c = AX_s \quad (9)$$

$$A = \begin{bmatrix} (x_2 - x_1) & (y_2 - y_1) \\ (x_3 - x_1) & (y_3 - y_1) \\ (x_4 - x_1) & (y_4 - y_1) \end{bmatrix}, X_s = \begin{bmatrix} x_s - x_1 \\ y_s - y_1 \end{bmatrix} \quad (10)$$

여기서  $\tau$ 는 마이크로폰과 음원 간의 시간 지연 값을 나타내는 벡터,  $A$ 는 각 마이크로폰 좌표와 관련된 행렬,  $X_s$ 는 음원의 좌표를 나타내는 벡터이다. 이 방정식을 통해 음원의 위치를 추정할 수 있다.

일반적으로 최소 제곱법의 원리는 주어진 관찰된 데이터  $y$ 가 모델  $Ax$ 에 의해 설명된다고 가정할 때, 실제로는 모델이 데이터와 정확하게 일치하지 않으므로 잔차  $\gamma = y - Ax$ 가 발생한다. 최소 제곱법은 이러한 잔차  $\gamma$ 의 제곱합을 최소화하는  $x$ 를 찾는 방법이다. 이를 수식으로 나타내면 다음과 같다.

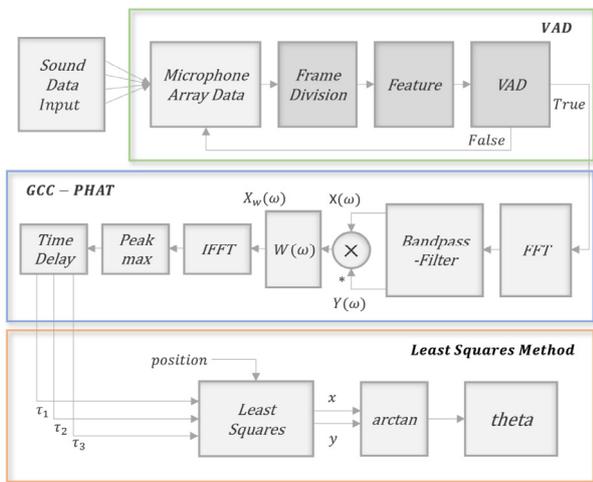
$$\min_x (y - Ax)^T (y - Ax) \quad (11)$$

이를 TDOA 기반 음원 위치 추정에 적용할 경우,  $y$ 는 시간 지연 값을 나타내고,  $A$ 는 마이크로폰의 위치 정보를 담은 행렬이다. 최소 제곱법에 따라 음원의 위치  $x_s$ 는 다음과 같이 추정할 수 있다.

$$X_s = (A^T A)^{-1} (A^T (c \cdot \tau)) \quad (12)$$

위와 같이 음원 위치 추정 방식에 최소제곱법을 적용함으로써, 여러 마이크로폰으로부터 수집된 시간 지연 데이터의 기반으로 위치를 가장 적합하게 추정할 수 있다.

본 논문에서는 위와 같은 최소 제곱법을 사용한 TDOA를 LS-TDOA로 명하며 최종 알고리즘의 선도는 다음 [Fig. 4]와



[Fig. 4] Flowchart of the LS-TDOA Algorithm

같이 나타낼 수 있다. 각 마이크로폰에서 수집된 4개의 신호를 10 ms 프레임 단위로 나누어, 말소리 패턴에 따라 음성 신호와 비음성 신호로 분류한다. 음성 신호로 분류된 데이터는 GCC-PHAT 알고리즘을 통해 각각의 마이크로폰 배열에 대한 시간 지연(Time-Delay) 값을 계산한다. 이후, 각각의 시간 지연에 대해 식 (12) 최소 제곱법을 적용하여 음원의 위치를 좌표로 추정하고, 추정된 좌표를 기반으로 arctan 함수를 사용하여 방향각 ( $\theta$ )를 도출한다.

### 3.2 음원 가시화 알고리즘

음원 가시화는 Boson 640 열화상 카메라를 기준으로 수행된다. 카메라의 비율을 유지하면서, 추정된 음원 좌표에 스케일 벡터를 적용하여 GUI에 반영한다. 마이크로폰은 전방 180도 내의 음원을 측정할 수 있으나, 열화상 카메라는 화각(Field of View, FOV)에 따라 가장자리의 시야가 제한되므로, 열화상 카메라의 FOV에 맞추어 측정 범위를 조정해야 한다. 이를 위해 FOV의 최솟값(Low)과 최댓값(High)을 계산하여, 카메라가 감지할 수 있는 유효 범위 내에서 음원을 시각화한다. 이후, 앞서 언급한 LS-TDOA 알고리즘을 통해 추정된 음원 좌표를 열화상 카메라의 시야 내로 매핑하여, 실시간으로 음원의 위치를 화면에 표시한다. 이에 대한 의사코드(Pseudocode)는 [Fig. 5]에 제시되어 있다.

#### Algorithm 1 Display Scaled SSL Coordinates on GUI

```

Require: camera_size, ssl_coordinates
Ensure: Adjust and display SSL coordinates on GUI
1: Define GUI size and initialize:
2: gui_size ← (width, height)
3: Scale camera size to match GUI:
4: scale_vector ← (gui_size / camera_size)
5: scaled_camera ← camera_size × scale_vector
6: Scale SSL coordinates to fit GUI:
7: for each ssl_coordinate in ssl_coordinates do
8:   scaled_ssl_x ← ssl_coordinate.x × scale_vector.x
9:   scaled_ssl_y ← ssl_coordinate.y × scale_vector.y
10:  Append (scaled_ssl_x, scaled_ssl_y) to scaled_ssl
11: end for
12: Determine FOV boundaries based on camera FOV:
13: low_value ← FOV.min
14: high_value ← FOV.max
15: Display SSL coordinates on GUI:
16: for each scaled_ssl in scaled_ssl do
17:   if scaled_ssl.x = low_value then
18:     Display at left edge of GUI
19:   else if scaled_ssl.x = high_value then
20:     Display at right edge of GUI
21:   else
22:     Display at corresponding (x, y) position on GUI
23:   end if
24: end for
25: Repeat step 5 for Y-axis for all SSL coordinates
26: Draw coordinates on GUI:
27: for each coordinate in scaled_ssl do
28:   Set position of point: (coordinate.x, coordinate.y)
29:   Draw point on GUI at (coordinate.x, coordinate.y)
30: end for
    
```

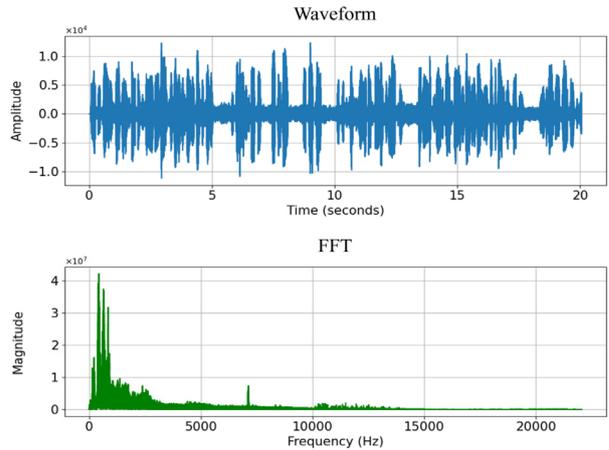
[Fig. 5] Pseudocode for the graphical user interface (GUI)

## 4. 실험 및 결과

### 4.1 실험 환경 및 평가 방법

제안하는 알고리즘을 검증하기 위해 시뮬레이션을 진행한다. 시뮬레이션의 환경은 다음과 같다. 시뮬레이션 공간은  $20 \times 20 \times 5$ , SNR 10 dB, T60 0.3 s인 공간으로 설정하고 마이크로폰 배열은 실제 설계와 동일하게 배치한다. 마이크로폰의 위치는 [Fig. 6]에 나타낸 바와 같이 방의 중심점인 (10, 10)으로부터 0.04 m 간격으로 배치되며 각 마이크로폰의 위치 좌표는 M1 (10.04, 10, 0), M2 (10, 10.04, 0), M3 (9.96, 10, 0), M4 (10, 9.96, 0)으로 설정된다. 음원은 사람 목소리 주파수 대역인 1 kHz와 실제 목소리가 포함된 음원을 연속 재생한다. [Fig. 7]은 연속 음원의 주파수 스펙트럼이다. 실험 동안 음원의 위치는 지속적으로 변화한다고 가정한다.

실증 실험에 대한 평가는 평균 절대 오차(Mean Absolute Error, MAE)와 정확도 (Accuracy, Acc)로 평가한다. 각 음향 조건에 대해 실제 DOA[°]와 추정된 DOA[°] 간의 평균 절대 오차는



[Fig. 7] Sound source frequency spectrum

아래와 같이 계산한다.

$$MAE [^\circ] = \frac{1}{L} \sum_{i=1}^L |\theta_s - \hat{\theta}_i| \quad (13)$$

여기서 L은 데이터의 수이며 실제 DOA와 추정된 DOA는  $\theta_s$ ,  $\theta_i$ 로 표시한다.

위치 추정 정확도는 다음과 같이 계산한다.

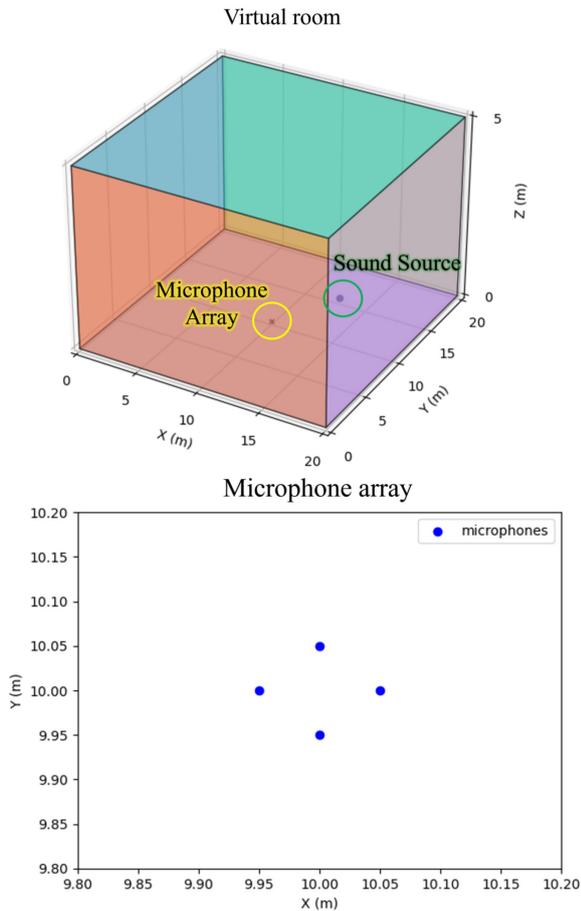
$$Acc. [\%] = \frac{\hat{C}_{acc}}{C} \times 100 \quad (14)$$

전체 데이터의 개수를 C, 실제 값과 예측값의 차이가 1도 이하일 때 위치 추정이 정확한 것으로 간주하고 이를  $\hat{C}_{acc}$ 라고 표시한다.

### 4.2 알고리즘 시뮬레이션 결과

#### 4.2.1 출력 해상도 비교

TDOA 알고리즘은 샘플링 주파수와 마이크로폰 간의 간격을 기반으로 각도를 계산할 수 있다. 그러나 하드웨어의 한계로 인해 샘플링 주파수를 높이는 데 제약이 있으며, 이를 보완하기 위해 마이크로폰 간의 거리를 조정하여 해상도를 높이는 방법을 사용한다. 하지만 협소한 공간에서는 마이크로폰 간의 거리를 충분히 늘리기 어려운 한계가 있다. 따라서, 마이크로폰의



[Fig. 6] Positions of microphones and sound sources in the room

[Table 1] Number of angles by microphone spacing

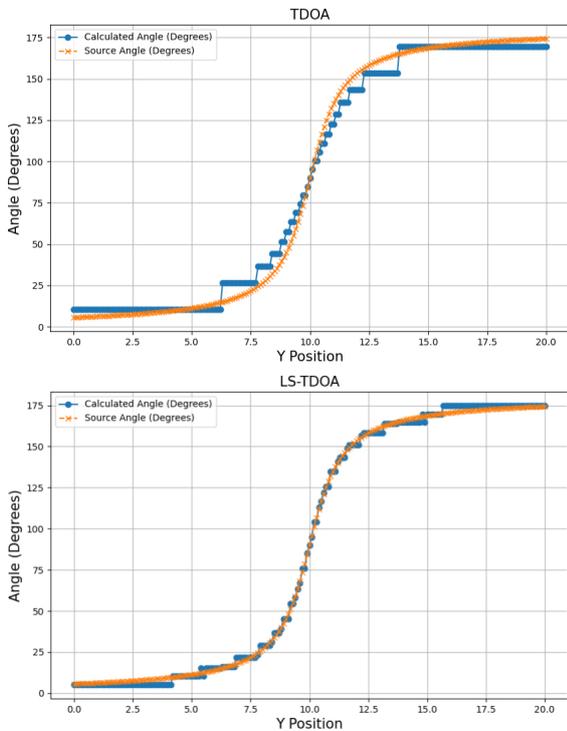
|         |   | L [m] |      |     |
|---------|---|-------|------|-----|
|         |   | 0.05  | 0.08 | 0.1 |
| TDOA    | n | 15    | 23   | 29  |
| LS-TDOA | n | 24    | 34   | 42  |

간격을 0.1 m 이하로 가정할 때, 기존 알고리즘과 제안하는 알고리즘으로 계산할 수 있는 각도의 개수는 [Table 1]에 나타나 있다. 제안된 알고리즘은 기존 TDOA 알고리즘에 비해 약 1.5 배 더 많은 각도를 추정할 수 있는 것으로 확인되었다.

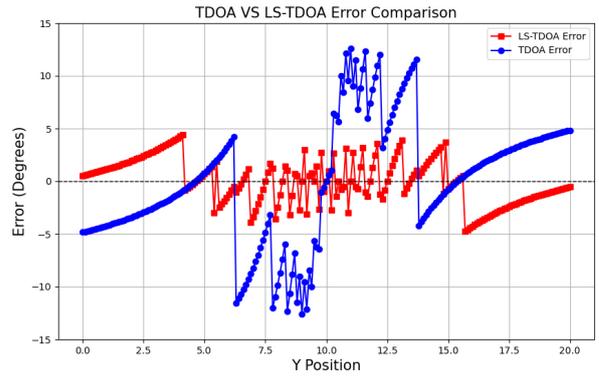
**4.2.2 참조 신호에 따른 추적 결과**

개발된 알고리즘의 시나리오는 붕괴 현장에서 고정된 사람의 위치를 탐지하고, 이를 화면에 시각적으로 표현하는 상황을 가정하였다. 따라서 구조 작업 중 로봇이 이동하면서 구조자의 위치는 변할 수 있지만, 로봇의 직진 주행 상황을 고려하여 알고리즘 검증 시 x축 좌표는 고정하고, y축만 변경하여 성능을 평가하였다. 기존 TDOA 방식과의 성능 비교를 위해 동일한 환경에서 각각 실험을 진행하였으며, 실험에 사용된 음원은 1000 Hz를 사용한다.

[Fig. 8]에서 보듯이, 기존 TDOA 방식은 y 값이 4, 5, 6일 때 각각 10.56°이고 7일 때부터 26.6°로 표현하는 각도가 급격하게 변하는 것을 볼 수 있다. 이는 0도에 가까워질수록 시간 지연의 변화 폭이 매우 작아져, 각도 계산 시 오류가 발생하기 때문이다. 반면, 제안된 LS-TDOA 방식은 최소 제곱법을 적용하여 오차를 최소화하였기 때문에, y 값이 4일 때 5.2°, 5일 때 10.6°, 6일 때 15.2°, 7일 때 21.8°로 각도 변화 폭이 상대적으로 크지 않음을 확인할 수 있다.



[Fig. 8] Tracking based on reference signal: Reference signal (orange), TDOA (top), LS-TDOA (bottom)

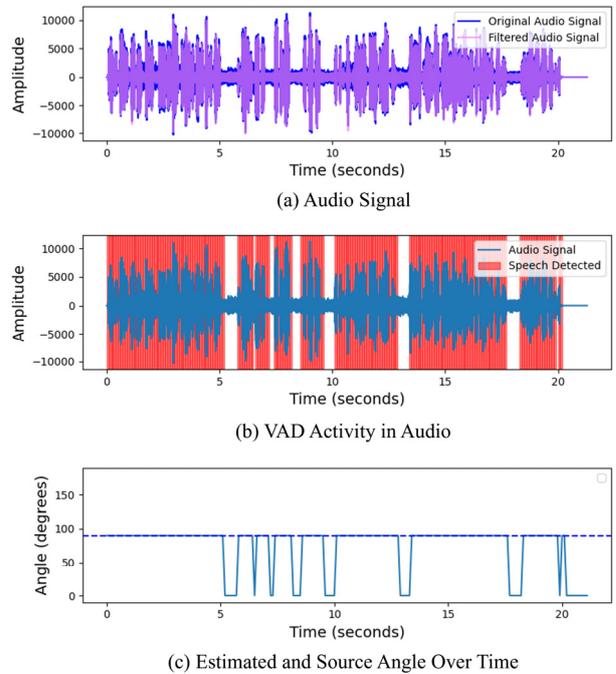


[Fig. 9] Angle error for reference signal, TDOA (Blue), and LS-TDOA (Red)

전체적으로, 참조 신호와 비교했을 때 제안된 알고리즘은 [Fig. 9]에서와 같이 평균적으로 5도 이내로 오차를 줄이며, 기존 방식 보다 음원 위치를 더 정확하게 추정하는 것을 확인할 수 있다.

**4.2.3 VAD 적용된 LS-TDOA 결과**

기존의 300 - 3000 Hz 대역 주파수를 구분하는 필터 방식은 [Fig. 10(a)]와 같이 5초 구간의 비음성 영역을 구분하지 못한다. 이는 잡음이 심한 환경에서 음원 추정의 정확성을 저해하는 요인으로 작용한다. 이에 따라, 본 연구에서는 사람의 목소리를 효과적으로 구분하기 위해 VAD를 적용하여 음성 구간과 비



[Fig. 10] (a) Band-pass filtered source (Violet), (a)(b) Original source (blue), (b) Band-pass filtered with VAD-applied source (Red), and (c) voice segment angle

음성 구간을 나눈 후, LS-TDOA를 적용하였다. 음원을 10 ms 단위로 나누어 VAD를 적용한 결과는 [Fig. 10(b)]에 나타나 있다.

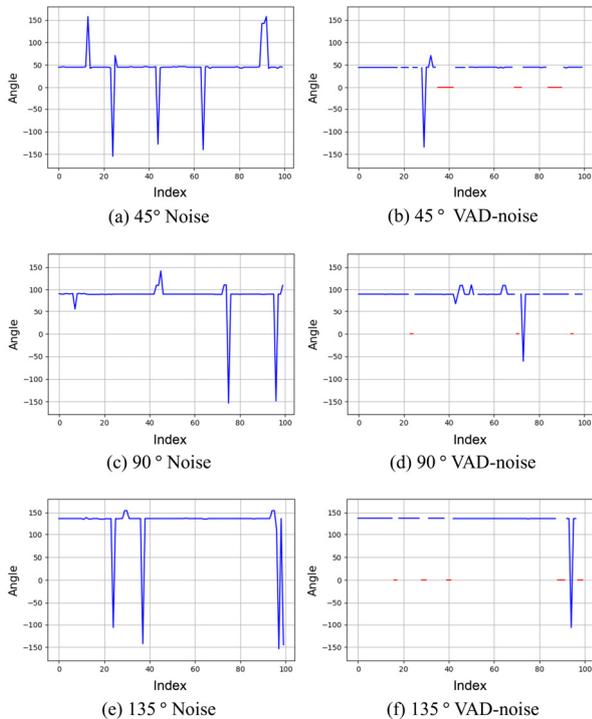
5초, 10초, 13초, 17초 부근의 무음 구간에서는 비음성 구간으로 분류되어 위치 추정을 하지 않는 모습을 [Fig. 10(c)] 확인할 수 있었다. 또한 배경 소음이 강하게 유입된 부분에서도 각도를 감지하지 않는 것을 확인할 수 있었다. 이를 통해 기존 주파수 필터 방식이 처리하지 못한 잡음을 효과적으로 구분할 수 있음을 알 수 있다.

### 4.3 실증 실험 결과

45°, 90°, 135°에 음원을 위치시키고, 각 위치에서 총 100개의 데이터를 수집하여 실험을 진행하였다. 실험은 백색 소음

[Table 2] Results at 45°, 90°, and 135° for each evaluation metric

|           | $\theta$ [°] |         |         |         |         |         |
|-----------|--------------|---------|---------|---------|---------|---------|
|           | 45           |         | 90      |         | 135     |         |
|           | MAE [°]      | Acc [%] | MAE [°] | Acc [%] | MAE [°] | Acc [%] |
| Clean     | 0.79         | 95      | 0.58    | 100     | 0.49    | 97      |
| Noise     | 10.4         | 86.7    | 6.86    | 85      | 12.05   | 83      |
| VAD-noise | 2.13         | 93      | 2.89    | 92      | 2.42    | 99      |



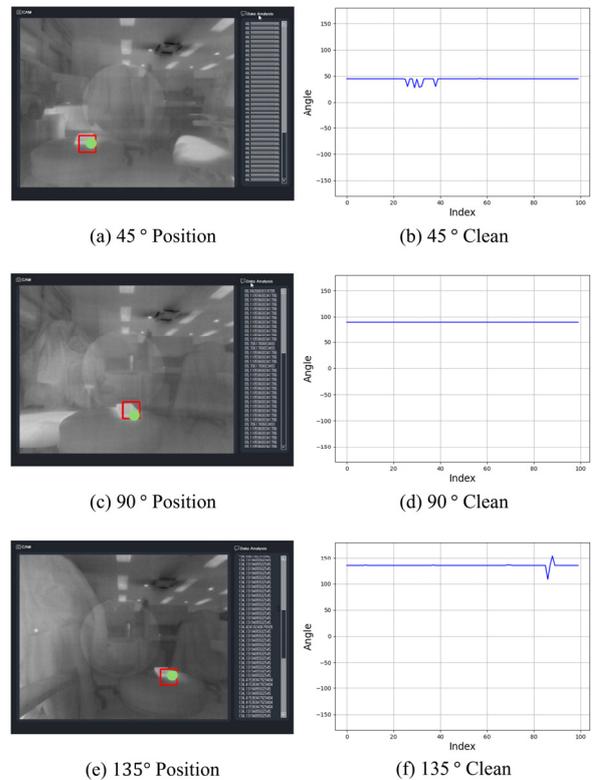
[Fig. 11] Sound source localization results at 45°, 90°, and 135° for each environment (a), (c), (e) LS-TDOA without VAD in an environment with snap sounds, (b), (d), (f) LS-TDOA with VAD applied in an environment with snap sounds : calculated angle values (blue), VAD-filtered angle values (red)

(White Noise) 환경에서 스냅 소리(Snap Noise)가 없는 환경과 있는 환경에서 각각 수행되었으며, 스냅 소리가 있는 환경에서는 VAD가 적용된 LS-TDOA와 적용되지 않은 LS-TDOA 알고리즘으로 나누어 실험을 진행하였다. 이후, 스냅 소리는 노이즈로 정의하여 서술한다. 이를 통해, 노이즈 환경에서 VAD의 적용 여부에 따른 음원 위치 추정 성능을 비교하였다.

[Table 2]에서 확인할 수 있듯이, 노이즈가 없는 환경에서는 LS-TDOA 알고리즘이 음원의 위치를 95% 이상 정확하게 추정하는 것을 확인할 수 있었다. 반면, 노이즈가 있는 환경에서는 [Fig. 11(a)], [Fig. 11(c)], [Fig. 11(e)] 그래프에서 확인할 수 있듯이, 불안정한 값이 발생하여 정확도가 약 80%로 감소하는 것을 확인할 수 있었다. 이는 주파수 필터만으로는 제대로 제거되지 않은 유효하지 않은 값들이다. 이러한 문제를 해결하기 위해 VAD를 적용한 결과, [Fig. 11(b)], [Fig. 11(d)], [Fig. 11(f)] 그래프에서 확인할 수 있듯이, 약 8% 이상의 정확도 향상을 확인할 수 있었다. VAD는 비음성 구간을 효과적으로 구분하여 노이즈의 영향을 감소시킴으로써, 음원 위치 추정의 정확도를 개선하였다.

### 4.4 음원 가시화 알고리즘 결과

구현한 LS-TDOA 알고리즘을 모듈에 탑재하여 가시화 작업



[Fig. 12] Source visualization algorithm applied using LS-TDOA (a), (c), (e) Angle positions displayed on the camera, (b), (d), (f) Angle values calculated in a noise-free environment

을 진행하였다. 이를 위해 카메라의 좌표값, GUI의 좌표값, 그리고 음원의 좌표값을 일치시킨다. 이후 음원의 위치를 카메라 화면 위에 [Fig. 12]와 같이 바운딩 박스 형태로 표현하였다. 알고리즘이 음원을 화면에 얼마나 정확하게 표현하는지를 확인하기 위해 실험을 진행하였다.

각각 45°, 90°, 135°에 음원을 위치시키고 [Fig. 7]의 음원을 재생하였다. 총 100개의 좌표 데이터를 수신하여, 각 위치에 대한 그래프를 그려 데이터를 분석하였다. [Fig. 12(a)], [Fig. 12(c)], [Fig. 12(e)]에서 볼 수 있듯이, 녹색 원은 추정해야 할 음원의 위치를 나타내고, 적색 박스는 추정된 각도를 중심으로 가로세로 25 pixel 크기의 바운딩 박스를 그린 모습이다. [Fig. 12(b)], [Fig. 12(d)], [Fig. 12(f)]는 노이즈가 없는 환경에서 GUI로 수신된 좌표를 기반으로 계산된 각도 값을 보여준다. 계산된 각도 값은 비교적 일정하게 유지되며, 추정된 좌표가 음원 신호에 맞게 정확하게 계산되어 화면에 표시되고 있음을 확인할 수 있다.

## 5. 결 론

본 연구에서는 협소한 공간 내에서 생존자를 탐지하기 위한 소형 인명 탐지 모듈을 개발하고, 이를 통한 음원 위치 추정 및 가시화 알고리즘을 구현하여 검증하였다. 음원 가시화를 위해 카메라 정면 시야각 내에서 음원의 위치를 2차원 좌표계로 표현할 수 있도록, 우선 구조의 하드웨어를 설계하고 마이크로폰과 열화상 카메라를 전방에 배치하였다. 또한, 생존자의 주변 환경과 생존 신호를 확인하기 위해 가스 센서를 상단에 배치하였다. 개선된 LS-TDOA를 적용함으로써 이전 연구에 비해 위치 추정의 정확도를 향상하였다. 또한, 음성과 비음성을 구분하는 VAD를 적용하여 잡음에 대한 강건성을 높였다. 마지막으로, 구조자가 효율적으로 탐지 결과를 활용할 수 있도록 가시화 기능을 구현하고 적용하였다.

로봇의 이동을 고려한 시뮬레이션 결과, 기존 TDOA 방식에 비해 제안된 알고리즘이 참조 신호를 더 정확하게 추종하며, 약 5도 이하의 낮은 오차를 기록하였다. 특히, 0도에서 30도까지의 각도를 기존 방식에 비해 약 2배 이상 더 세밀하게 표현할 수 있음을 확인하였다. 또한, 제안된 알고리즘은 노이즈가 있는 환경에서도 사람의 목소리에만 반응하여, 기존 주파수 대역 필터링 방식에 비해 각도 오차가 135°에서 12.05°에서 2.42°로 약 5배로 감소하였으며, 정확도는 최소 7%에서 최대 19%까지 더 향상된 결과를 보였다. 마지막으로, 음원 가시화 기능을 통해 추정된 음원의 위치를 카메라 화면에 근접하게 표현하는 것을 확인할 수 있었다. 이를 통해 실제 구조 상황에서도 음원의 위치를 직관적으로 파악할 수 있으며, 시스템의 실용성을 입증하였다.

향후에는 협소 공간 탐색 로봇을 제작하여 실물 협소 공간 내에서 실증 실험을 진행할 계획이다.

## References

- [1] *Chapter 1: Reducing the Number of Fatalities*, [Online], <https://www.kkr.mlit.go.jp/plan/daishinsai/1.html>, Accessed: Oct. 24, 2024.
- [2] P. Liljebäck, K. Y. Pettersen, Ø. Stavdahl, and J. T. Gravdahl, "A review on modelling, implementation, and control of snake robots," *Robotics and Autonomous Systems*, vol. 60, no. 1, pp. 29-40, Jan., 2012, DOI: 10.1016/j.robot.2011.08.010.
- [3] S. Han, S. Chon, J. Y. Kim, J. Seo, D. G. Shin, S. Park, J. T. Kim, J. Kim, M. Jin, and J. Cho, "Snake Robot Gripper Module for Search and Rescue in Narrow Spaces," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1667-1673, Apr., 2022, DOI: 10.1109/LRA.2022.3140812.
- [4] E. Ackerman, *What CMU's Snake Robot Team Learned While Searching for Mexican Earthquake Survivors*, [Online], <https://spectrum.ieee.org/cmu-snake-robot-mexico-earthquake>, Accessed: Oct. 24, 2024.
- [5] E. Ackerman, *Why Robots Can't Help Find Survivors in the Florida Building Collapse*, [Online], <https://spectrum.ieee.org/why-robots-cant-help-find-survivors-in-the-florida-building-collapse>, Accessed: Oct. 24, 2024.
- [6] S. W. Han, S. J. Kim, D. G. Shin, J. H. Pyo, M. S. Lee, M. Kim, and J. H. Suh, "A Study on Sound Source Localization of Survivors for the Robot Searching Victims in a Narrow Space," *Journal of the Korean Society for Precision Engineering*, vol. 39, no. 7, pp. 509-516, Jul., 2022, DOI: 10.7736/JKSPE.022.019.
- [7] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320-327, Aug., 1976, DOI: 10.1109/TASSP.1976.1162830.
- [8] W. Oh, "Wide Coverage Microphone System for Lecture Using Ceiling-Mounted Array Structure," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 22, no. 4, pp. 624-633, Apr., 2018, DOI: 10.6109/jkiice.2018.22.4.624.
- [9] M. A. Chung, H. C. Chou, and C. W. Lin, "Sound localization based on acoustic source using multiple microphone array in an indoor environment," *Electronics*, vol. 11, no. 6, 2022, DOI: 10.3390/electronics11060890.
- [10] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-Time Multiple Sound Source Localization and Counting Using a Circular Microphone Array," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2193-2206, Oct., 2013, DOI: 10.1109/TASL.2013.2272524.
- [11] X. Alameda-Pineda and R. Horaud, "A Geometric Approach to Sound Source Localization from Time-Delay Estimates," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 6, pp. 1082-1095, Jun., 2014, DOI: 10.1109/TASLP.2014.2317989.
- [12] A. Parsayan and S. M. Ahadi, "Real Time High Accuracy 3-D PHAT-Based Sound Source Localization Using a Simple 4-Microphone Arrangement," *IEEE Systems Journal*, vol. 6, no. 3, pp. 455-468, Sept., 2012, DOI: 10.1109/JSYST.2011.2176766.

- [13] H. Liu, B. Yang, and C. Pang, "Multiple sound source localization based on TDOA clustering and multi-path matching pursuit," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, pp. 3241-3245, 2017, DOI: 10.1109/ICASSP.2017.7952755.
- [14] T. H. Tan, Y. T. Lin, Y. L. Chang, and M. Alkhaleefah, "Sound source localization using a convolutional neural network and regression model," *Sensors*, vol. 21, no. 23, Dec., 2021, DOI: 10.3390/s21238031.
- [15] Z. Xu, N. Liu, and B. M. Sadler, "A Simple Closed-Form Linear Source Localization Algorithm," *MILCOM2007 - IEEE Military Communications Conference*, Orlando, FL, USA, pp. 1-7, 2007, DOI: 10.1109/MILCOM.2007.4454975.
- [16] X. Dang, W. Ma, E. A. P. Habets, and H. Zhu, "TDOA-Based Robust Sound Source Localization With Sparse Regularization in Wireless Acoustic Sensor Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 1108-1123, 2022, DOI: 10.1109/TASLP.2022.3153251.
- [17] R. Jyothi and P. Babu, "SOLVIT: A reference-free source localization technique using majorization minimization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2661-2673, 2020, DOI: 10.1109/TASLP.2020.3021500.
- [18] J. H. Chang, N. S. Kim, and S. K. Mitra, "Voice activity detection based on multiple statistical models," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 1965-1976, Jun., 2006, DOI: 10.1109/TSP.2006.874403.
- [19] S. Gazor and W. Zhang, "A soft voice activity detector based on a Laplacian-Gaussian model," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 498-505, Sept., 2003, DOI: 10.1109/TSA.2003.815518.
- [20] *WebRTC documentation*, [Online], <https://webrtc.googlesource.com/src/+refs/heads/main/docs/>, Accessed: Oct. 24, 2024.



**석 윤 정**

2023 국립부경대학교 기계시스템공학전공 (학사)  
2023~현재 국립부경대학교 지능로봇공학과 (석사과정)

관심분야: 회로 설계, 음원 추정, 자연어 처리, 필드 로봇



**김 성 재**

2024 국립부경대학교 지능로봇공학과(박사)  
2024~현재 국립부경대학교 산업과학기술 연구소 선임연구원

관심분야: 제어이론, 지능 제어, 강인 제어, 로봇 제어



**박 서 연**

2023 국립부경대학교 지능로봇공학과(석사)  
2023~현재 한국로봇융합연구원 주임연구원

관심분야: 로봇 설계, 동역학 해석, 모바일 로봇, 기구 설계



**서 진 호**

2002 도쿄 공업대 제어공학과(박사)  
2018 한국로봇융합연구원 단장  
2018~현재 국립부경대학교 기계시스템공학 전공 교수

관심분야: 필드로봇, 재난대응로봇, 로봇시스템 통합