

# 도달시간차 기반의 음원 위치 추정법의 정확도 향상을 위한 딥러닝 적용 연구

## Deep learning-based approach to improve the accuracy of time difference of arrival - based sound source localization

정일주,<sup>1</sup> 허현석,<sup>1</sup> 정인지,<sup>2†</sup> 이승철<sup>3†</sup>

(Iljoo Jeong,<sup>1</sup> Hyunsuk Huh,<sup>1</sup> In-Jee Jung,<sup>2†</sup> and Seungchul Lee<sup>3†</sup>)

<sup>1</sup>포항공과대학교 기계공학과, <sup>2</sup>한국표준과학연구원 물리표준본부, <sup>3</sup>한국과학기술원 기계공학과  
(Received December 4, 2023; revised January 10, 2024; accepted January 11, 2024)

**초 록:** 본 연구는 데이터 기반의 딥러닝 접근 방식을 통해 도달 방향 추정의 정확성과 정밀성의 개선을 통해 보다 강건하고 정확한 음원 위치 추적 기술을 제안한다. 본 연구에서는 도달시간 차 기반의 음원 위치 추적법을 개선함을 목적으로 하며, 이를 위해 상호상관함수로부터 정확하고 정밀한 시간 지연을 추정한다. 실제 마이크로폰으로부터 계측된 값은 많은 잡음이 혼입된 형태이므로, 따라서 실제 도달시간 차이를 정확히 추정하는 것이 여전히 이 분야의 한계로 남아있다. 또한, 마이크로폰으로부터 실제 신호를 계측하는 과정에서 신호는 디지털화가 되며, 계측 시스템의 샘플링 주파수에 의해 측정 정밀도가 한정되는 양자화 오류를 수반한다. 본 연구에서는 딥러닝 기반 접근법을 통해, 기존의 방법이 가지는 한계를 극복한다. 또한 본 연구에서는 획득된 상호상관함수로부터 시간 지연을 추정하는 원리를 분석하기 위해, 두 개 및 세 개의 마이크로폰으로 구성된 배열에 대한 검증을 수행한다. 마지막으로, 실험을 통해 본 방법의 실제 활용성을 검증한다.

**핵심용어:** 음원 위치 추정, 딥러닝, 상호상관함수, 도달시간차

**ABSTRACT:** This study introduces an enhanced sound source localization technique, bolstered by a data-driven deep learning approach, to improve the precision and accuracy of direction of arrival estimation. Focused on refining Time Difference Of Arrival (TDOA) based sound source localization, the research hinges on accurately estimating TDOA from cross-correlation functions. Accurately estimating the TDOA still remains a limitation in this research field because the measured value from actual microphones are mixed with a lot of noise. Additionally, the digitization process of acoustic signals introduces quantization errors, associated with the sampling frequency of the measurement system, that limit the precision of TDOA estimation. A deep learning-based approach is designed to overcome these limitations in TDOA accuracy and precision. To validate the method, we conduct comprehensive evaluations using both two and three-microphone array configurations. Moreover, the feasibility and real-world applicability of the suggested method are further substantiated through experiments conducted in an anechoic chamber.

**Keywords:** Sound source localization, Deep-learning, Cross-correlation function, Time Difference Of Arrival (TDOA)

**PACS numbers:** 43.60.Jn, 43.60.Pt

†Corresponding author: In-Jee Jung (injee@kriss.re.kr)

Division of Physical Metrology, Korea Research Institute of Standards and Science, 267, Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea

(Tel: 82-42-868-5300, Fax: 82-42-868-5643)

†Corresponding author: Seungchul Lee (seunglee@kaist.ac.kr)

Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, 291, Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

(Tel: 82-42-350-3026, Fax: 82-42-350-3210)



Copyright©2024 The Acoustical Society of Korea. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## I. 서론

마이크로폰 배열 기반의 음원 위치추정법은 국방, 산업 안전, 인명구조 및 로봇 분야 등 다양한 분야에서 꾸준한 요구가 있어 온 기술 분야이다.<sup>[1]</sup> 기존의 접근 방법으로는 도달시간차(Time-Different Of Arrival, TDOA), 빔포밍, 음향 홀로그래피 등이 있으며,<sup>[2]</sup> 특히 도달시간차를 이용한 방법은 배열의 기하학적 구조를 기반으로 계산되는 마이크로폰 간 상호상관함수를 통해 시간 지연이 결정되므로 계산이 간단한 특징이 있다.<sup>[3]</sup> 한편, 많은 연구에서 위치추정의 정확도를 높이기 위해 일반화된 상호상관 방법이 광범위하게 사용되었다.<sup>[4-6]</sup>

실제 환경에서는 다양한 잡음이 동시에 측정되며, 따라서 마이크로폰에서 측정된 음향 신호의 왜곡은 결과적으로 이를 통한 음원 위치추정 성능 저하를 유발한다. 또한, 디지털 신호처리 과정에서 나타나는 시간 분해능 한계 및 마이크로폰 배열에 도달하는 음원의 방향에 따른 내재적인 오차의 원인들은 측정 방법의 강건성을 낮추는 문제를 유발한다.<sup>[7]</sup>

최근에는 딥러닝을 적용하여 데이터의 숨겨진 패턴을 찾아내는 방법으로 음원 위치 추정하는 문제에 적용되어, 기존의 방식이 갖는 성능 한계를 극복하고 있다.<sup>[8]</sup> 도달시간차 방법을 적용하기 위하여, 상호상관함수를 통한 음원 위치추정을 위해 딥러닝의 적용이 최근 활발히 시도된 바 있으며,<sup>[9-11]</sup> 이러한 연구들은 대부분 측정된 원신호로부터 곧바로 음원의 위치를 추정하는 방식으로 접근하였으나 시간 또는 공간분해능을 제한함으로써 음원 위치 추정 성능의 본질적인 한계가 남아있다.

본 연구에서는 딥러닝을 통해 정확한 도달시간차 추정 기술을 제안한다. 이를 통해, 측정 과정에서 발생하는 양자화 오류를 극복함으로써 도달시간차 방법을 기반으로 한 음원 위치 추정 정확도를 향상시키는 것을 목적으로 한다. 또한, 마이크로폰이 2개 및 3개로 이루어진 배열에 대한 시뮬레이션 및 실험을 통해 본 연구에서 제안하는 방법을 검증하고 결과에 대해 논의한다.

## II. 배경 이론 및 문제 정의

### 2.1 상호상관함수를 이용한 음원 위치추정

마이크로폰의 기하적 배치 및 음원이 도달하는 방향에 따라 각 음원에 대한 마이크로폰의 도달시간이 달라지게 된다. 따라서 마이크로폰 배열을 통해 이러한 시간차를 역으로 계산하면 음원의 위치를 추정할 수 있다. 도달시간차는 마이크로폰 배열에서 계산되는 상호상관함수를 통해 계산되며 다음과 같은 수식으로 표현된다.

$$R_{12}(\tau) = E[s_1(t)s_2(t+\tau)] = R_{ss}(\tau + \delta_{GT}), \quad (1)$$

여기서  $R_{12}(\tau)$ 은 두 마이크로폰에서 측정된 음압 사이의 상호상관함수이며,

$$s_1(t) = s(t) + n_1(t), \quad (2)$$

$$s_2(t) = s(t + \delta_{GT}) + n_2(t), \quad (3)$$

여기서  $s(t)$ 은 음원 신호  $n_1(t)$ ,  $n_2(t)$ 는 비상관 잡음,  $R_{ss}(\tau)$ 는 음원 신호에 대한 자기상관함수,  $\delta_{GT}$ 는  $R_{12}(\tau)$ 의 최대값에 해당하는 시점이며 실제 시간지연에 해당한다. 따라서 마이크로폰 배열에서 계산되는 복수의 상호상관함수 및 마이크로폰의 기하적 배치에 따라 음원의 위치를 추정할 수 있다.

### 2.2 문제 정의

도달시간차를 통한 음원위치 추정 방법에서 마이크로폰에서 측정된 상관함수에서  $\delta_{GT}$ 를 추론하는 것이 가장 중요한 과정이다. 이를 추론하는 과정에서 산란 및 굴절과 같은 음향 전파에 따른 현상 또는 반사 및 잔향과 같은 요소들은 시간지연의 왜곡을 초래할 수 있다. 또한, 실제 측정에 따라 계산되는 이산화된 상호상관함수는 공간분해능을 떨어지게 하는 주요한 요인 중 하나이다. 따라서 이러한 오차의 요인들은 결국 위치추정 정확도를 떨어뜨리고 측정 결과의 불확도를 커지게 한다.<sup>[7]</sup>

본 연구의 주요 목적은 음향 신호의 시간 지연을 정확하게 추정할 수 있는 딥러닝 모델을 개발하는 것이다. 아래와 같이 특정 추정 모델을 획득하는 것을 목표로 한다.

$$f(CC_{biased}) = \delta_{pred}, \quad (4)$$

여기서  $CC_{biased}$ 은 실제 측정 결과를 통해 계산되는 상호상관함수를 나타내고,  $\delta_{pred}$ 은 모델에서 추론된 도달시간 차이다.

본 모델의 핵심 목표는 편향된 상호상관함수  $CC_{biased}$ 로부터 실제  $\delta_{GT}$ 값을 보상하는 것이다. 이러한 문제를 해결함으로써 기존의 도달시간차를 이용한 음원 위치추정 방법보다 정확한 시간지연 값을 계산할 수 있으며, 따라서 본 제안된 방법은 음원 위치추정 정확도를 향상시키는 데 적용할 수 있다.

### III. 딥러닝 기반 도달시간차 추정

본 연구의 주요 목표는  $CC_{biased}$ 로 구성된 데이터 세트에 대해 훈련된 딥러닝 모델을 활용하여  $\delta_{GT}$ 를 직접 예측하는 것이다. 문제 해결을 위해서 대표적인 딥러닝 모델인 Multi-Layer Perceptron(MLP)를 이용하였으며, 과적합이 발생하지 않는 선에서 높은 유추 성능을 확보하기 위해서 MLP의 은닉층을 4개로 구성하였다. 이에 따라 모델의 하이퍼파라미터는 직접 탐색을 통해 선정되었다. Fig. 1에 본 연구에서 제안하는 딥러닝 모델에 대한 구성을 나타내었다.

모델 학습에 있어서, 손실 함수로는 평균 제곱 오차(Mean Squared Error, MSE)가 사용되었다. 학습 과정에서의 학습률은 0.01로 설정되었으며, 모델 최적화는 Adam Optimizer를 사용하여 수행되었다.<sup>[12]</sup>

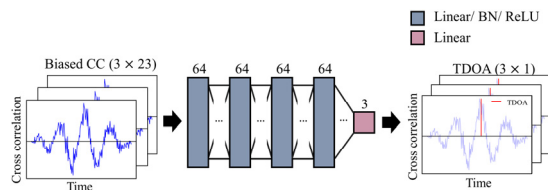


Fig. 1. (Color available online) Proposed TDOA estimation based on MLP architecture.

제안 모델의 도달시간차 추정 성능을 확인하기 위해 Fig. 2와 같이 마이크론 2개 및 3개로 구성된 배열을 사용하여 시뮬레이션을 진행하였다. 성능 비교를 위하여 실제 시간 지연과 예상 시간 지연 사이의 평균절대오차(Mean Absolute Error, MAE)를 사용하였다. 2개 마이크론으로 구성된 배열에 대해서는 1개의 상호상관함수, 그리고 3개 마이크론으로 구성된 배열에 대해서는 3개의 상호상관함수가 데이터 세트에 포함되며, 총 10,000개의 데이터 쌍을 포함한다. 각 데이터는 Matlab을 통한 시뮬레이션으로 생성되었다. 여기서 간격이 0.03 m인 마이크론 배열을 적용하였고, 데이터 형성에 적용된 음속 및 샘플링 주파수는 각각 343 m/s, 10 kHz이며, 대역 제한된 백색잡음에 대하여 상호상관함수를 계산하였다. 상호상관함수는 총 23개의 시간지연에 대한 계산 결과를 포함하며 그 범위는  $|\tau| \leq 1.1$  ms에 해당한다.

Fig. 3은 2개의 마이크론으로 구성된 배열에서,

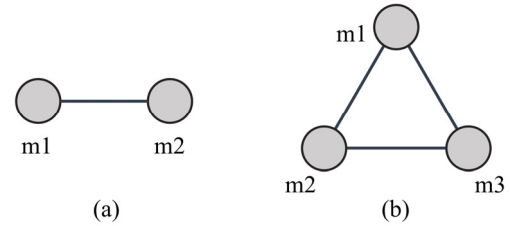


Fig. 2. Microphone array setups: (a) line array with two microphones; (b) equilateral triangle array with three microphones. Here,  $m_1$ ,  $m_2$ ,  $m_3$  indicate the position of microphones.

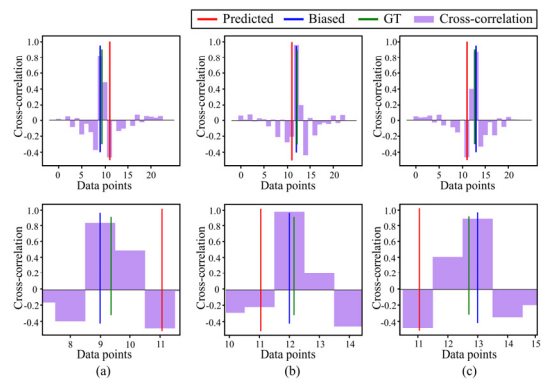


Fig. 3. (Color available online) Simulation results of TDOA estimation of two microphone for three different source DOA examples. DOA of (a)  $(60^\circ, -6^\circ)$ , (b)  $(-107^\circ, 24^\circ)$ , and (c)  $(-59^\circ, 10^\circ)$ . Here, the second row is the zoomed result of the first row.

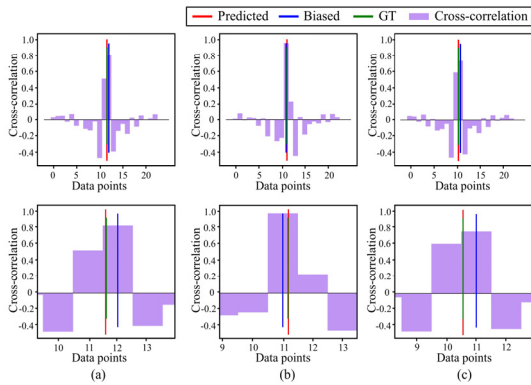


Fig. 4. (Color available online) Simulation results for three microphone for one DOA example with (azimuth, elevation) is  $(98^\circ, 42^\circ)$ . (a), (b) and (c) show the TDOA estimation result for  $CC_{12}$ ,  $CC_{13}$ , and  $CC_{23}$ . Here, the second row is the zoomed result of the first row.

서로 다른 3개 음원 방향에 대해 계산된 상호상관함수 및 도달시간차 추정 결과를 나타낸다. 여기서 x-축은 데이터 순서를 나타내고, 11번째 데이터가  $\tau=0$ 에 해당한다. Fig. 3(a)~(c) 결과를 비교해보면, 각기 다른 상호상관함수에 대해 모두  $\delta_{pred} = 0$ 로 예측하는 것을 알 수 있다.  $\delta_{GT}$ 를 중심으로 확대 도시된 결과를 통해 상세히 살펴보면, 도달시간차 예측 값인  $\delta_{pred}$ 가  $\delta_{biased}$ 에 비해서 오히려 정확도가 낮은 것을 볼 수 있다. 이를 통해, 제안된 모델은 단일 상호상관함수로부터 도달시간차를 예측하는 패턴을 제대로 학습하지 못하였음을 알 수 있고, 따라서 딥러닝을 통해 단일의 상호상관함수에서 획득할 수 있는 정보만으로는 도달시간차 예측을 수행하기 어려운 것을 확인할 수 있다.

Fig. 4는 3개의 마이크로폰으로 구성된 배열에서, 임의로 선정된 한 음원 방향에 대해 계산된 상호상관함수  $CC_{12}$ ,  $CC_{13}$ ,  $CC_{23}$  및 도달시간차 추정 결과를 나타낸다. Fig. 4(a)~(c) 결과를 통해 높은 정확도로  $\delta_{pred}$ 가  $\delta_{GT}$ 를 추정하는 것을 볼 수 있다. 한편,  $\delta_{biased}$ 는 시간분해능 제한에 따라  $\delta_{GT}$ 와 차이를 나타내는 것을 볼 수 있고, 이는 양자화 오차에 해당한다. 이를 통해, 제안된 딥러닝 모델을 이용하여 양자화 오차를 보상함으로써 도달시간차 기반의 음원 위치 추정법의 정확도를 향상시킬 수 있음을 알 수 있다.

Table 1은 마이크로폰 2개와 3개로 구성된 배열에

Table 1. Comparison of the mean absolute error for averaged TDOA estimation for 1000 test sets for random DOAs.

	$CC_{12}$	$CC_{13}$	$CC_{23}$	Average
Previous	1.98E-05	2.11E-05	1.81E-05	1.97E-05
2-microphone	1.30E-4	-	-	1.30E-04
3-microphone	7.52E-06	8.76E-06	7.64E-06	7.97E-06

대한 도달시간차 추정의 정량적 결과(MAE)를 나타낸다. 여기서는 1,000개의 데이터 세트가 계산에 활용되었다. 이를 토대로 분석한 결과, 마이크로폰을 3개 사용한 경우에는 도달시간차의 오차가 감소하지만, 마이크로폰이 2개인 경우에는 딥러닝 모델 적용 시 오히려 오차가 커지는 것을 볼 수 있다. 이로부터, 상호상관함수를 통한 도달시간차의 추정을 위한 핵심은 주어진 마이크로폰 배열에서 획득된 서로 다른 상호상관함수 간의 연관성에 있음을 유추할 수 있다.

이를 통해, 딥러닝 적용 결과의 유효성을 확립하기 위해서 최소 3개 이상의 마이크로폰이 필요한 것을 알 수 있다. 2개의 마이크로폰으로는 cone of confusion에 의하여 도달시간차를 통한 정확한 방위각 추정이 어려운 점을 고려하였을 때, 이는 충분히 보상 가능한 조건이라 생각할 수 있다.

본 연구에서는 음원 추정을 위한 가장 기본적인 마이크로폰 배열로부터 딥러닝 모델 적용 가능성을 검토하였으며, 향후에는 딥러닝 모델의 최적화를 위한 마이크로폰 배열 및 마이크로폰 개수에 대한 분석이 필요하다.

#### IV. 실험을 통한 검증

제안 모델의 도달시간차 추정 성능을 확인하기 위한 실험적 검증을 수행하였다. Fig. 2(b)와 같은 정삼각형 형태의 마이크로폰 배열을 사용하였으며, 무향실 내에서 임의의 252개 방향에서 전파되는 음원에 대한 측정을 수행하였다.

본 실험에서는 실제 마이크로폰을 사용함으로써, 실험 수행에 따라 발생할 수 있는 측정 오차를 포함한다. 또한, 마이크로폰의 간격을 0.14 m로 설정하여 딥러닝 모델이 다양한 크기를 갖는 마이크로폰 어레이에 적용될 수 있는지에 대한 여부를 확인하였다.

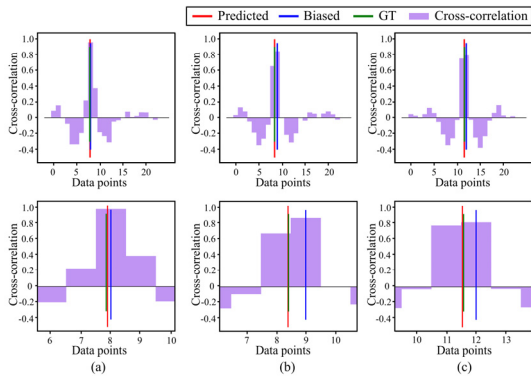


Fig. 5. (Color available online) Experiment results for TDOA estimation of three microphone at DOA is  $(-110^\circ, 50^\circ)$ . (a), (b) and (c) is according to  $CC_{12}$ ,  $CC_{13}$ , and  $CC_{23}$ . Here, the second row is the zoomed result of the first row.

Table 2. Experiment results of mean absolute error for averaged TDOA estimation using 3 microphones.

	$CC_{12}$	$CC_{13}$	$CC_{23}$	Average
Previous	2.30E-05	2.36E-05	1.50E-05	2.06E-05
Deep Learning based	1.25E-05	7.77E-06	8.18E-06	9.40E-06

따라서, 측정 데이터의 일부는 마이크로폰 간격조정 에 따른 정밀 보정을 위해 사용되었다.

실험 결과 중 하나를 선정하여 Fig. 5에 나타내었다. 그 결과,  $\delta_{biased}$ 에 비해서  $\delta_{pred}$ 가  $\delta_{GT}$ 를 더욱 정확하게 추정하는 것을 볼 수 있다. 딥러닝을 적용하지 않은 결과에서는 실험 조건에 따른 편차가 크게 나타나는 반면, 딥러닝을 적용한 실험 결과에서는 실험 조건이 바뀌어도 강건한 추정 특성을 나타내는 것을 보여준다. Table 2에 실험 결과에 대한 정량적인 분석 결과를 나타내었으며, 그 결과 기존 방법 대비 평균 오차가 54% 저감된 것을 볼 수 있다.

본 실험에서는 자유음장조건에 대한 딥러닝 모델의 특성을 관찰하였으며, 향후에는 배경소음 및 잡음을 포함하는 일반적인 공간에서 제안된 딥러닝 모델의 적용 가능성을 검토할 필요가 있다.

## V. 결론

본 연구에서는 마이크로폰 배열을 통해 계산된 상호상관함수에서 도달시간차를 추론하기 위한 딥러

닝 기반 접근 방법을 적용하였다. 이 방법은 실제 측정결과로부터 계산되는 상호상관함수에 내재하는 양자화에 따른 정보 손실로 인한 오차를 보상하여 시간지연 계산 정확도를 향상시킨다. 이를 검증하기 위해 간단한 구조의 배열 구성으로부터 시뮬레이션 및 실험을 수행하였고, 그 결과 본 연구에서 제시하는 딥러닝 기반의 접근법의 강건성을 검증하였다. 결론적으로 딥러닝 기반의 방법은 높은 해상도의 시간분해능을 가지는 도달시간차 계산을 수행할 수 있고, 이를 통해 높은 공간분해능을 갖는 음원 위치 추정자로써 음원의 위치를 정확하게 추정하는데 적용할 수 있다.

## 감사의 글

본 연구는 한국표준과학연구원의 연구 과제(KRISS-2023-GP2023-0002, KRISS-2023-GP2023-0004-05) 및 시장선도를위한 한국주도형 K-Sensor 기술개발 「센서소자 신뢰성 평가지원」(RS-2002-00154837)의 지원을 받아 수행되었습니다.

## References

1. Y. Bao, Q. Wang, K. Shen, and Y. Wang, "Microphone array-based sound source localization and tracking system," Proc. IEEE ITNEC, 515-519 (2023).
2. S. Gombots, J. J. Nowak, and M. Kaltenbacher, "Sound source localization—state of the art and new inverse scheme," E&I Elektrotech. Inf. Tech. **138**, 229-243 (2021).
3. M. R. Bai, S.-S. Lan, and J.-Y. Huang, "Time difference of arrival (TDOA)-based acoustic source localization and signal extraction for intelligent audio classification," Proc. IEEE SAM, 632-636 (2018).
4. C. Evers, H. W. Löllmann, H. Mellmann, A. Schmidt, H. Barfuss, P. A. Naylor, and W. Kellermann, "The LOCATA challenge: Acoustic source localization and tracking," IEEE/ACM TASLP, **28**, 1620-1643 (2020).
5. R. Lee, M. S. Kang, B. H. Kim, K. H. Park, S. Q. Lee, and H. M. Park, "Sound source localization based on GCC-PHAT with diffuseness mask in noisy and reverberant environments," IEEE Access, **8**, 7373-7382 (2020).
6. D. Salvati, C. Drioli, and G. L. Foresti, "Acoustic source localization using a geometrically sampled grid



SRP-PHAT algorithm with max-pooling operation,” IEEE Signal Process. Lett. **29**, 1828-1832 (2022).

7. I.-J. Jung and J.-G. Ih, “Comparison of the sound source localization methods appropriate for a compact microphone array” (in Korean), J. Acoust. Soc. Kr. **31**, 47-56 (2020).
8. P.-A. Grumiaux, S. Kitić, L. Girin, and A. Guérin, “A survey of sound source localization with deep learning methods,” J. Acoust. Soc. Am. **152**, 107-151 (2022).
9. X. Xiao, S. Zhao, X. Zhong, D. L. Jones, E. S. Chng, and H. Li, “A learning-based approach to direction of arrival estimation in noisy and reverberant environments,” Proc. IEEE ICASSP, 2814-2818 (2015).
10. F. Vesperini, P. Vecchiotti, E. Principi, S. Squartini, and F. Piazza, “A neural network-based algorithm for speaker localization in a multi-room environment,” Proc. IEEE MLSP, 1-6 (2016).
11. H. C. Maruri, P. L. Meyer, J. Huang, J. Ontiveros, and H. Lu, “GCC-PHAT cross-correlation audio features for simultaneous sound event localization and detection (SELD) in multiple rooms,” DCASE 2019 Tech. Rep. 2019.
12. D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv:1412.6980 (2014).

**저자 약력**

▶ 정 일 주 (Iljoo Jeong)



2016년 2월 : 아주대학교 기계공학과 학사  
2019년 9월 ~ 현재 : POSTECH 기계공학과 석박통합과정

▶ 허 현 석 (Hyunsuk Huh)



2018년 8월 : UNIST 기계공학과 학사  
2019년 2월 ~ 현재 : POSTECH 기계공학과 석박통합과정

▶ 정 인 지 (In-Jee Jung)



2014년 2월 : 영남대학교 기계공학과(공학사)  
2016년 2월 : KAIST 기계공학과(공학석사)  
2021년 2월 : KAIST 기계공학과(공학박사)  
2016년 1월 ~ 2017년 3월 : LIG nex1 해양연구소  
2017년 4월 ~ 2017년 8월 : KAIST 기계기술연구소 위촉연구원  
2021년 2월 ~ 2021년 7월 : KAIST 기계기술연구소 연수연구원  
2021년 7월 ~ 현재 : 한국표준과학연구원 선임연구원

▶ 이 승 철 (Seungchul Lee)



2001년 2월 : 서울대학교 기계항공공학부 학사  
2008년 4월 : 미시간대학교 기계공학과 석사  
2010년 8월 : 미시간대학교 기계공학과 박사  
2013년 7월 ~ 2017년 12월 : UNIST 기계공학과 조교수  
2018년 1월 ~ 2021년 2월 : POSTECH 기계공학과 조교수  
2021년 3월 ~ 2023년 8월 : POSTECH 기계공학과 부교수  
2023년 9월 ~ 현재 : KAIST 기계공학과 부교수