

# 다중 소리 감지 및 모션 인식 기반의 청각 장애인을 위한 반려로봇: 히어베어 v2

홍현지, 이현서, 곽민정, 김유진, 오찬영\*

국립공주대학교

hhj@smail.kongju.ac.kr, hs381717@smail.kongju.ac.kr, 202202186@smail.kongju.ac.kr,  
kimyj4852@smail.kongju.ac.kr, \*cyoh@kongju.ac.kr

## HearBear v2: A Companion Robot for the Hearing Impaired Based on Multi-sound Detection and Motion Recognition

Hyeon-Ji Hong, Hyeon-Seo Lee, Min-Jung Kwak, Yu-Jin Kim, Chanyoung Oh\*

Kongju National Univ.

### 요약

청각 장애인의 안전 확보와 정보 접근성 향상을 위한 반려로봇 ‘히어베어’는 실내 소리를 실시간으로 감지 및 분석하여 청각 장애인 사용자에게 애플리케이션 알림을 제공한다. 본 논문은 ‘히어베어’의 실용성을 높이기 위한 고도화 연구를 제안한다. 기존 시스템의 한계를 보완하기 위해 다중 소리 감지 모델과 가변 임계값 후처리 기법을 도입하여 위급 상황 인지 능력을 개선하였다. 또한, 카메라를 활용한 음원 발생 위치의 현장 사진 전송 기능과 모션 인식 제어, 그리고 도트 매트릭스(Dot Matrix) 기반의 시각적 피드백 기능을 통합하여 사용자 상호작용을 강화하였다. 이를 통해 청각 장애인의 안전 확보와 정서적 교감에 실질적인 도움을 줄 수 있을 것으로 기대된다.

### I. 서론

청각 장애인은 일상 속 청각 정보의 부재로 인해 안전사고 위험에 노출되어 있다. 이를 해결하고자 선행 연구[1]에서는 단일 소리 감지 (Single-label) AI 모델 및 텍스트 기반 인터페이스 기능을 갖춘 반려로봇 ‘히어베어’를 제안한 바 있다. 하지만 실제 사용 환경에서는 여러 소리가 복합적으로 발생하는 상황이 빈번하며, 기존의 텍스트 기반 인터페이스는 즉각적인 상황 파악을 위한 시각적 정보가 부족하다는 한계가 있었다.

본 논문에서는 이러한 문제를 보완하기 위해서 다중 레이블 분류와 가변 임계값 후처리 기법을 통해 동시 감지 성능을 최적화하고, 카메라를 및 모션 인식 인터페이스를 통해 직관성과 편의성을 높인 고도화된 ‘히어베어 v2’를 구현하였다.

### II. 본론

#### 1. ‘히어베어’ 개요

선행 연구에서 제안된 초기 ‘히어베어’ 시스템은 Robotis사의 Turtlebot3[2]을 기반으로 제작되었다. 추가로 부착한 ReSpeaker[3]로 음원 방향(DOA, Direction of Arrival)을 추정해 이동하고, AST(Audio Spectrogram Transformer)[4] 모델 학습을 통해 단일 소리 이벤트를 분류하여 사용자에게 전달하는 시스템이다. 사용자와의 상호작용은 애플리케이션 내에서 사용할 수 있는 텍스트 기반 감정분석 챗봇과 예약어 입력을 통한 동작 제어 방식으로 구성했다.

그러나 이 시스템은 단일 소리만 감지할 수 있어 복합 소음 환경에 취약하며, 사용자에게 시각 정보가 부족하고, 텍스트 기반의 인터페이스는 사용 편의성이 낮다는 한계가 있었다.

### 2. 시스템 구성

본 연구에서 구현한 ‘히어베어 v2’는 선행 연구의 하드웨어 구성을 기반으로 하되, 기능 확장을 위해 카메라와 도트 매트릭스를 추가로 장착하였다. 구현된 로봇의 본체와 사용자 애플리케이션 화면은 그림 1과 같다. 새롭게 추가된 카메라는 사용자 모션 인식과 애플리케이션에 전송할 현장 사진 촬영을 수행하며, 도트 매트릭스는 정서적 교감을 위한 시각적 피드백 기능을 수행한다. 전체 시스템의 아키텍처는 그림 2와 같으며, 기존과 동일하게 ROS 2 환경에서 통합 제어된다.



(a) (b)

그림 1. (a) 로봇 본체 및 (b) 애플리케이션 메인화면

Fig. 1. (a) Robot body and (b) application main

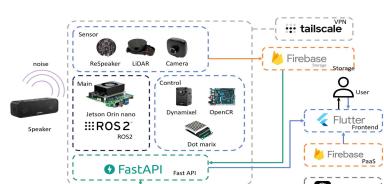


그림 2. 전체 시스템 설계도

Fig. 2. Overall system architecture

### 3. 주요 기능 및 개발 세부 내용

#### 3.1 다중 상황 인지를 위한 소리 감지 알고리즘 개선

해당 시스템의 기본 동작 방식은 선행 연구에서 구현된 음원 방향 추정 방식을 계승한다. 다만, 도착 후 사용자에게 현장 상황을 실시간으로 공유할 수 있도록 카메라 기능을 통합하여 시각적 피드백 체계를 보완하였다. 소리 분류 모델은 다중 상황 인지를 위해 고도화되었다. 기존의 단일 레이블(Single-label) 분류 방식에서 여러 소리가 혼재된 환경에서도 개별 이벤트를 탐지할 수 있도록 다중 레이블(Multi-label) 분류 방식을 도입하였다. 이를 위한 AST 모델을 다중 레이블 데이터셋인 FSD50K[5]로 미세 조정(fine-tuning)하였다.

또한, 탐지된 이벤트의 종류에 따라 가변 임계값 후처리 기법을 적용하였다. 화재 경보나 비명 등 생명 및 안전과 직결된 주요 사운드 카테고리에는 낮은 임계값을 설정함으로써 탐지 누락을 최소화하였으며, 이를 위해 위험 상황을 사용자에게 우선적으로 신속하게 알릴 수 있도록 설계하였다. 데이터셋에서 실내 환경에서 발생 빈도가 높은 이벤트를 중심으로 5가지 카테고리로 재구성하였으며, 각 카테고리의 중요도와 소리 특성을 고려하여 임계값을 차등 설정하였다. 상세한 카테고리별 임계값 설정 수치는 표 1과 같다.

표 1 카테고리별 소리 임계값 설정

Table 1. Threshold Settings by Sound Category

카테고리	임계값
안전(Safety)	0.40
일상(Household)	0.65
사람(Human)	0.70
동물(Animal)	0.70
기타(Etc)	0.50

#### 3.2 제스처 인식 기반의 상호작용 기능 고도화

‘히어베어’는 사용자와의 유대감을 강화하기 위해 특정 명령에 따라 반응하는 상호작용 동작 기능을 제공한다. 선행 연구에서는 사용자가 애플리케이션에 예약어를 직접 입력하는 방식을 사용하였으나, 본 연구에서는 사용자 편의성과 직관성을 높이기 위해 제스처 인식 기반의 인터페이스로 고도화하였다.

이를 위해 Google MediaPipe Hand Landmarker[6]를 활용하여 손의 특징점 데이터를 수집하였으며, 정적인 이미지가 아닌 시간적 흐름이 포함된 동적 제스처를 정확히 인식하고자 LSTM(Long Short-Term Memory)[7] 신경망을 도입하였다. 이러한 시계열 데이터 학습을 통해 사용자의 연속적인 동작 변화를 효과적으로 파악하여 인식된 제스처에 정의된 로봇의 상호작용 동작이 즉각적으로 실행되도록 구현하였다.

#### 3.3 도트 매트릭스를 활용한 로봇 감정 피드백 구현

선행 연구에서는 애플리케이션 내 챗봇 기능과 감정 분석을 통해 사용자와의 정서적 상호작용을 할 수 있도록 구축하였다. 그러나 이러한 기능은 로봇 본체와의 직접적인 상호작용이 결여되어 물리적 실재감이 부족하다는 한계가 있었다.

이를 보완하기 위해 로봇 전면에 8x8 도트 매트릭스 디스플레이를 부착하였다. 사용자의 대화에서 분석된 감정 결과에 따라, 그림 3과 같이 도트 매트릭스에 다양한 눈 모양의 그래픽을 출력함으로써, 로봇이 사용자의 감정에 공감하고 반응하는 모습을 시각적으로 형상화하였다. 이러한 시각적 피드백을 통해 사용자와 로봇 간의 유대감을 강화하고 더욱 직관적인 상호작용이 가능하도록 구현하였다.

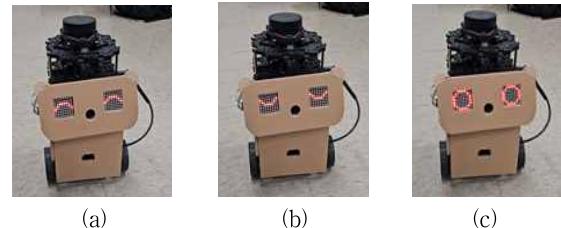


그림 3. 감정별 도트 매트릭스 출력 형태 예시 (a) 행복 (b) 슬픔 (c) 놀람  
Fig. 3. Examples of emotional facial expressions on the Dot Matrix

## III. 결론

본 논문에서는 선행 연구의 ‘히어베어’ 시스템을 개선하여 청각 장애인을 위한 보조 기능을 고도화하였다.

다중 소리 감지 및 가변 임계값 후처리 기법을 통해 여러 소리가 혼재된 상황에서도 위급 상황을 우선적으로 탐지할 수 있도록 하였으며, 제스처 인식 인터페이스로 사용자의 조작 편의성을 극대화하였다. 또한, 카메라와 Dot Matrix를 활용한 시각적 피드백을 사용자에게 제공함으로써 로봇의 실용성과 정서적 상호작용 능력을 향상시켰다.

## ACKNOWLEDGMENT

본 연구는 2026년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음 (2024-0-00073)

## 참 고 문 헌

- [1] 홍현지, 이현서, 광민정, 김유진, 오찬영, “청각 장애인을 위한 반려로봇 : 히어베어,” 2025 대한임베디드공학회 추계학술대회 논문집, pp. 335-336, 2025.
- [2] ROBOTIS, “TurtleBot3 e-Manual,” Accessed on Sep. 30, 2025.
- [3] Seeed Studio, “ReSpeaker Mic Array v2.0,” Accessed on Sep. 30, 2025.
- [4] Y. Gong, Y. A. Chung, and J. Glass, “AST: Audio Spectrogram Transformer,” in Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech), pp. 571-575, 2021.
- [5] E. Fonseca, X. Favory, J. Pons, F. Font, and X. Serra, “FSD50K: an open dataset of human-labeled sound events,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 2229-2244, 2021.
- [6] F. Zhang et al., “MediaPipe Hands: On-device Real-time Hand Tracking,” arXiv preprint arXiv:2006.10214, 2020.
- [7] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.