

## Interests

### Multimodal learning, Representation learning, Speech processing

REPRESENTATION LEARNING, MULTIMODAL LLM, AUDIO LLM, DIFFUSION MODEL, DUPLEX SPEECH-TO-SPEECH, SPEECH AI

- My recent research focuses on enabling simple, **low-cost integration** between **pretrained LLMs** and **pretrained multi-modal representation encoders** by **eliminating the modality gap** at the representation level. Specifically, my current work centers on modality-bridging translation across modalities, and going forward I aim to develop diffusion-based representation transport that can move representations into an **omni-space**, thereby achieving strong coupling without repeatedly building large paired multimodal datasets. This offers a practical path for enterprises and individuals to rapidly build domain-specific multimodal LLMs as diverse open-source and commercial LLMs become widespread.
- From an application perspective, I also highlight **duplex speech-to-speech** as a key application for audio LLMs, actively exploring end-to-end reasoning and conversational capabilities for natural voice interaction.
- I hope to expand this direction through close collaboration with various researchers interested in these topics.

## Education

### Korea Advanced Institute of Science and Technology (KAIST)

PH.D. IN SCHOOL OF ELECTRICAL ENGINEERING; ADVISOR: JOON SON CHUNG (MULTIMODAL AI LAB)

S.Korea

Sept. 2024 - Present

### Korea Advanced Institute of Science and Technology (KAIST)

M.S. IN SCHOOL OF ELECTRICAL ENGINEERING; ADVISOR: JOON SON CHUNG (MULTIMODAL AI LAB)

S.Korea

Aug. 2022 - Aug. 2024

### Hankuk University of Foreign Studies (HUFS)

B.S. IN COMPUTER SCIENCE; GPA: 4.1/4.5

S.Korea

Mar. 2015 - Aug. 2022

## Publication

### International Conferences

- **Nam KiHyun\***, Choi, J. M.\*, Lee, H. k., Heo, J. W., and Chung, J. S., "Diffusion-Link: Diffusion Probabilistic Model for Bridging the Audio-Text Modality Gap", preprint, submitted at *ICASSP 2026*.
- **Nam KiHyun**, Heo, J. W., Jung, J. W., Park, G., Jung, C., Yu, H. J., and Chung, J. S., "SEED: Speaker Embedding Enhancement Diffusion Model", *INTERSPEECH*, 2025.
- **Nam KiHyun**, Heo, H. S., Jung, J. W., and Chung, J. S., "Disentangled Representation Learning for Environment-agnostic Speaker Recognition", *INTERSPEECH*, 2024.
- Heo, H. S., **Nam KiHyun**, Lee, B. J., Kwon, Y., Lee, M., Kim, Y. J., and Chung, J. S., "Rethinking Session Variability: Leveraging Session Embeddings for Session Robustness in Speaker Verification", *ICASSP*, 2024.
- Jung, C.\*, Lee, S.\*, **Nam KiHyun**, Rho, K., Kim, Y. J., Jang, Y., and Chung, J. S., "TalkNCE: Improving Active Speaker Detection with Talk-Aware Contrastive Learning", *ICASSP*, 2024.
- Kwak, D.\*, Jung, J.\*, **Nam KiHyun**, Jang, Y., Jung, J. W., Watanabe, S., and Chung, J. S., "VoxMM: Rich Transcription of Conversations in the Wild", *ICASSP*, 2024.
- **Nam KiHyun\***, Kim, Y.\*, Huh, J., Heo, H.S., Jung, J.W., Chung, J.S., "Disentangled Representation Learning for Multilingual Speaker Recognition", *INTERSPEECH*, 2023.
- Ha, J.\*, **Nam KiHyun\***, Kang, J., Lee, S., Yang, S., Jung, H., Kim, H., Kim, E., Kim, S., Kim, H.A., Doh, K., Lee, C.K., Sung, N., Kim, S., "ClovaCall: Korean Goal-Oriented Dialog Speech Corpus for Automatic Speech Recognition of Contact Centers", *INTERSPEECH*, 2020.

### International Journals

- **Nam KiHyun\***, Song Y.J.\*, and Yun I.D., "SSS-AE: Anomaly detection using self-attention based sequence-to-sequence auto-encoder in smd assembly machine sound", *IEEE Access*, 2021.

## Work Experience

---

### NAVER Clova Speech

S.Korea

DEEP LEARNING RESEARCHER, INTERNSHIP

Sep. 2019 - Feb. 2020

- Implemented End-to-End Automatic Speech Recognition models.
- Implemented Active Learning system for Automatic Speech Recognition.

Mar. 2021 - Sep. 2021

- Implemented Speech Representation Model using Normalizing Flow.
- Implemented End-to-End Forced Aligner with Sentence-Alignment.

## Awards & Honors

---

### International Competitions

NIST 2024 SPEAKER RECOGNITION EVALUATION

Sep. 2024 - Oct. 2024

- **1st Place** (Audio Track) / **4th Place** (Audio-Visual Track) - Collaboration with Microsoft, KAIST MMAI Lab, PolyU, NUS and UEF

## Technical Skills

---

Programming: C++, Python, Pytorch

## Teaching

---

### Teaching Assitance at EE, KAIST

Daejeon, S.Korea

\* EE488: COMPUTER VISION (FALL, 2023)

\* EE738: SPEECH RECOGNITION SYSTEMS (FALL, 2022)