Seungheon Doh

seungheon.doh@gmail.com

blog • google scholar • github

Research Interest

Multimodal Representation Learning, Music Information Retrieval, Language Models

My research focuses on developing music intelligence for understanding, retrieval, generation, and post-production tasks. This work is structured around three interconnected research directions. First, I investigate **multi-modal representation learning** methods that establish semantic correspondences between music and complementary modalities including natural language, visual content, and other media formats. These representations serve as foundational components for transfer learning in music-related applications. Second, I explore **multi-modal large language models** (MLLMs) for music applications, with a focus on three key subtopics: reasoning, chain-of-thought processes, and tool calling, leveraging the intermediate generation capabilities of language models. Third, I design and evaluate **conversational interfaces** through real-world applications, focusing on user experience, usability, and satisfaction in diverse music-related scenarios, ensuring that the developed technologies meet practical needs and deliver measurable value to end users.

Tutorial

Connecting Music Audio and Natural Language
Seungheon Doh, Ilaria Manco, Zachary Novack, Jong Wook Kim and Ke Chen
The 25nd International Society for Music Information Retrieval Conference (ISMIR), 2024

Selected Publications

TALKPLAY: Multimodal Music Recommendation with Large Language Models
Seungheon Doh, Keunwoo Choi, and Juhan Nam
Proceedings of the conference on Neural Information Processing Systems (NeurIPS-Submitted), 2025

Can Large Language Models Predict Audio Effects Parameters from Natural Language?
Seungheon Doh, Junghyun Koo, Marco A. Martínez-Ramírez, Wei-Hsiang Liao, Juhan Nam, Yuki Mitsufuji
The IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA-Submitted), 2025

Music Discovery Dialogue Generation Using Human Intent Analysis and Large Language Model
Seungheon Doh, Keunwoo Choi, Daeyong Kwon, Taesu Kim, and Juhan Nam
Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR), 2024

LP-MusicCaps: LLM-based Pseudo Music Captioning
Seungheon Doh, Keunwoo Choi, Jongpil Lee, Juhan Nam
Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR), 2023
(Best Paper Award Nomination - Long Talk)

Enriching Music Descriptions with a Finetuned-LLM and Metadata for Text-to-Music Retrieval
Seungheon Doh, Minhee Lee, Dasaem Jeong, Juhan Nam
Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024

More Publication @ `https://scholar.google.com/citations?user=MCkggcgAAAAJ` (Total Citation: 403)

Education

Korea Advanced Institute of Science and Technology (KAIST) 2021 - 2025
Ph.D in Graduate School of Culture Technology
Music and Audio Computing Lab (Advisor: Juhan Nam)

Korea Advanced Institute of Science and Technology (KAIST) 2019 - 2021
MSc. in Graduate School of Culture Technology
Music and Audio Computing Lab (Advisor: Juhan Nam)

Ulsan National Institute of Science and Technology (UNIST) 2014 - 2019
B.S. in School of Business administration & Industrial Design
Academic Performance Scholarship Recipient for every semester

Experience

SonyAI Apr 2025 - Nov 2025
Tokyo, Japan
Research Intern (Advisor: Junghyun Koo, Marco A. Martínez-Ramírez, Wei-Hsiang Liao)
Topic: Large Language Model for Music Production (Multimodal LLM)

Adobe Research Jun 2024 - Sep 2024
San Francisco, United States
Research Intern (Advisor: Nick Bryan, Ge Zhu)
Topic: Large Language Model for Music Generation (Multimodal LLM)

Chart Metric Dec 2023 - Mar 2024
Remote
Research Intern (Advisor: Keunwoo Choi)
Topic: Artist Similarity Retrieval System (Artist-to-Artist)

Naver Corp. Dec 2022 - Mar 2023
Seongnam, South Korea
Research Intern in Now AI Team (Advisor: Jeong Choi)
Topic: Multimodal Video Recommendation System (User-to-Item)

NYU, Computational Intelligence, Learning, Vision, and Robotics (CILVR) Jun 2022 - Aug 2022
New York, United States
Visiting Student (Advisor: Kyunghyun Cho)
Topic: Self-supervised Learning Model for Music

ByteDance/Tiktok Jul 2021 – Jan 2022
Remote
Research Intern in Speech, Audio & Music Intelligence Team (Advisor: Keunwoo Choi, Minz Won)
Topic: Multimodal Music Retrieval System (Text-to-Music, Speech-to-Music)

NYU, Music and Audio Research Laboratory (MARL) Dec 2019 - Feb 2020
New York, United States
Visiting Student (Advisor: TaeHong Park)
Topic: Word Embedding for Music Annotation and Retrieval

Open-Source Software

- **LP-MusicCaps:** Music Captioning System (Released 2023, 300+ stars):
- **TTMR++:** Text-to-Music Retrieval System (Released 2025, 36+ stars):
- **TTMR:** Text-to-Music Retrieval System (Released 2022, 130+ stars):

Advising

Have advised research internship students. Academic papers have been presented in top-tier conferences and workshops (ISMIR, HCMIR, NLP4MusA).

- Daeyong Kwon, Retrieval Augmented Generation for Music Question Answering (2025)
- Hayeon Bang, Audio-Text-Midi Joint Embedding Space (2024)
- Minhee Lee, Audio-Text Joint Embedding Space (2023)
- Yoonjin Chung, Music-to-Image Generation (2022)
- Youngjune Choi, Hierarchical Audio Few-shot Learning (2022)
- Haven Kim, Playlist Title Generation (2022)
- Junwon Lee, Playlist Title Generation (2021)
- Hounsu Kim, Playlist Continuation (2020)
- Nabin Kim, Music Emotion Recognition (2020)

Teaching Activity

- TA, GCT731 Topics in Music Technology: Generative AI for Music (Mar 2023)
- TA, GCT634 Musical Applications of Machine Learning (Sep 2022)

- TA, GCT634 Musical Applications of Machine Learning (Sep 2021)
- TA, GCT731 Topics in Music Technology: Cognitive Science of Music (Sep-2020)
- TA, GCT576 Social Computing (Sep 2019)

Talk

- Invited Talk, LLM-Powered Music Annotation and Retrieval, KAIST
  Jun 2024, Host: Juhan Nam
- Invited Talk, Text-based Music Annotation and Retrieval, UC San-Diego
  Mar 2024, Host: Hao-Wen Dong
- Invited Talk, Multimodal Music Retrieval for Listener and Contents Creator, Seoul National Univ.
  Feb 2023, Host: Junghyun Koo
- Invited Talk, Music Informational Retrieval with Natural Language Processing, YONSEI Univ.
  Dec 2022, Host: Saebyul Park

Service

- Organizing Committee, Newcomer Initiative Chair, ISMIR 2025
- Organizing Committee, 1st Workshop on Large Language Models for Music & Audio, ISMIR 2025
- Organizing Committee, 3rd Workshop on Natural Language Processing for Music Audio, ISMIR 2024
- Reviewer, ISMIR, ICASSP, ICML, NeurIPS, IEEE TASLP

Languages & Skill & Interests

- English(fluent), Korean(native)
- Python (Pytorch, Tensorflow), HTML/CSS, React.js, C++
- Media Art, Graphic Design, Prototyping & Making, Piano, Movie, Crossfit, Lifting

References

- Juhan Nam, Professor, KAIST — juhan.nam@gmail.com
- Keunwoo Choi, Senior Research Director, Gaudio Lab — gnuchoi@gmail.com
- Minz Won, Research Scientist, Suno — minz.s.won@gmail.com
- Jongpil Lee, CEO, Mix.Audio — jongpillee@neutune.com