

Hounsung Kim

PhD Student, KAIST | Speech and Audio Research

hanshounsung@kaist.ac.kr | hanshounsung.github.io | [GitHub](#)

I study structured speech representations for controllable generation, with a focus on neural speech codecs, zero-shot text-to-speech, and discrete diffusion. My earlier work spans music transcription and expressive instrument synthesis.

Education

KAIST | PhD in Culture Technology Sep 2023 – Present

Music and Audio Computing Lab · Advisor: Juhan Nam

KAIST | MSc in Culture Technology Sep 2021 – Aug 2023

Music and Audio Computing Lab · Advisor: Juhan Nam

KAIST | BSc in Mechanical Engineering Mar 2015 – Aug 2021

Publications and Preprints

Beyond Token Revision: Investigating Mask-and-Replace Diffusion for Zero-Shot Text-to-Speech

Hounsung Kim, Joonyong Park, Yuki Saito, Satoru Fukayama, Juhan Nam

Preprint · Under review, 2026 | [Paper](#)

Mask-and-replace training improves zero-shot TTS even without token revision; disentangles noisy-context augmentation from replacement supervision.

VocalCoachBench: Benchmarking Audio-Language Models on Expert Feedback for Singing

Hayeon Bang, Hounsung Kim, Wonil Kim, Juhan Nam

NeurIPS 2026 · Evaluations & Datasets Track | [Paper](#) | [Venue](#)

SDP-Codec: A Speaker-Decoupled Speech Codec with Pitch Injection for Low-Bitrate Coding and Zero-Shot Voice Conversion

Hounsung Kim, Juhan Nam

Interspeech 2026 | [Paper](#) | [Code](#) | [Demo](#)

Separates global speaker information from local content and pitch for low-bitrate reconstruction and zero-shot voice conversion, with reduced speaker leakage.

D3PIA: A Discrete Denoising Diffusion Model for Piano Accompaniment Generation From Lead Sheet

Eunjin Choi, Hounsung Kim, Hayeon Bang, Taegyun Kwon, Juhan Nam

ICASSP 2026 | [Paper](#) | [Code](#) | [Demo](#)

Publications and Preprints *continued*

D3RM: A Discrete Denoising Diffusion Refinement Model for Piano Transcription

Hounsung Kim, Taegyun Kwon, Juhan Nam

ICASSP 2025 | [Paper](#) | [Code](#)

CoNMod: Controllable Neural Frame-based Modulation Effects

Gyubin Lee, Hounsung Kim, Junwon Lee, Juhan Nam

DAFx 2024 • Best Student Paper Award | [Paper](#)

Expressive Acoustic Guitar Sound Synthesis with an Instrument-Specific Input Representation and Diffusion Outpainting

Hounsung Kim, Soonbeom Choi, Juhan Nam

ICASSP 2024 | [Paper](#) | [Demo](#)

Research Experience

AIST, Japan | Technical Trainee

May 2026 – Aug 2026

Research on speech and audio representation learning and neural codec-based speech modeling.

Neutune Co. Ltd. | Research Intern

Jun 2021 – Sep 2021

Honors and Awards

VoiceMOS Challenge 2026, Track 2 – 1st place

Speech naturalness and emotion-similarity prediction for emotional TTS systems.

DAFx 2024 — Best Student Paper Award for CoNMod (co-author)

Research Mentoring

Donghoon Kang, Computer Science, KAIST

Dec 2024 – Apr 2025

Gyubin Lee, Electrical Engineering, KAIST

Sep 2023 – Mar 2024

Teaching

Teaching Assistant, KAIST

Musical Applications of Machine Learning

Fall 2022, Spring 2024, Spring 2025

Fundamentals of Computer Music

Fall 2023, Fall 2024

Introduction to Culture Technology

Spring 2022

Technical Skills

Programming and tools: Python, PyTorch, Linux, Docker, Slurm, Git

Research methods: Neural audio codecs, audio tokenization, discrete diffusion, speech and music signal processing, model training and evaluation