Heeseung Kim

gmltmdsnlp@gmail.com • https://gmltmd789.github.io/ • https://www.linkedin.com/in/gmltmd789/ • https://github.com/gmltmd789

EDUCATION

Seoul National University, Seoul, Korea

- B.S. in Electrical & Computer Engineering — Mar 2015 – Feb 2019
  - Focus: Signal Processing, Machine Learning, Deep Learning
  - Cumulative GPA: 3.85 / 4.3 (Cum Laude)
- Ph.D. candidate in Electrical & Computer Engineering — Mar 2019 – Aug 2025 (Exp)
  - Focus: Deep Learning, Generative Models, Speech Synthesis, Speech LLM, Spoken Dialog Model
  - Cumulative GPA: 3.93 / 4.3

RESEARCH INTERESTS

Deep Generative Model, Text-to-Speech, Speech LLM, Voice Interaction Models

SELECTED PUBLICATIONS

[1] H. Kim, S. Seo, K. Jeong, O. Kwon, S. Kim, J. Kim, J. Lee, E. Song, M. Oh, J. Ha, S. Yoon, and K. Yoo, "**Paralinguistics-Aware Speech-Empowered Large Language Models for Natural Conversation**," in ***NeurIPS 2024***, Vancouver, Canada, Dec 2024.

[2] H. Kim*, S. Kim*, and S. Yoon, "**Guided-TTS: A Diffusion Model for Text-to-Speech via Classifier Guidance**," in ***ICML 2022***, Baltimore, Maryland USA, Jul 2022.

[3] H. Kim, S. Kim, J. Yeom, and S. Yoon, "**UnitSpeech: Speaker-adaptive Speech Synthesis with Untranscribed Data**," in ***INTERSPEECH 2023, Oral Presentation***, Dublin, Ireland, Aug 2023.

[4] H. Kim, S. Lee, J. Yeom, C. Lee, S. Kim, and S. Yoon, "**VoiceTailor: Lightweight Plug-In Adapter for Diffusion-Based Personalized Text-to-Speech**," in ***INTERSPEECH 2024***, Kos Island, Greece, Sep 2024.

[5] C. Shin*, H. Kim*, C. Lee, S. Lee, and S. Yoon, "**Edit-A-Video: Single Video Editing with Object-Aware Consistency**," in ***ACML 2023, Oral Presentation, Best Paper Award***, Istanbul, Turkey, Nov 2023.

RESEARCH EXPERIENCE

Naver & Seoul National University Collaboration — Mar 2022 – May 2025 (Exp)

- **Role**: Graduate Research Student, Electrical & Computer Engineering (Ph.D. Candidate)
- **Project Overview**:
  - **2022.03 – 2023.07**: Built a speaker-adaptive TTS model for more natural voice generation. Incorporated speech input into a GPT-2–scale model by attaching an encoder for ASR (Automatic Speech Recognition) and SER (Speech Emotion Recognition), aiming to enrich speech understanding capabilities.
  - **2023.08 – 2024.05**: Developed an *English spoken dialog system* with an end-to-end pipeline and paralinguistic awareness. This research led to a publication at **NeurIPS 2024**, while simultaneously laying the groundwork for Naver's proprietary Speech LLM.
  - **2024.06 – 2025.01**: Focused on developing a text-to-speech model for synthetic spoken dialog generation.
  - **2025.02 – 2025.05 (Ongoing)**: Beyond supervised fine-tuning, exploring *RLHF (e.g., DPO, GRPO)* methods to capture more natural and affective nuances in spoken dialog. Currently investigating strategies for reward/preference data collection, either from human annotators or synthetic approaches, to enhance voice interaction quality.

REPOSITORIES

UnitSpeech ★130+
Official Implementation of INTERSPEECH 2023 paper "UnitSpeech: Speaker-adaptive Speech Synthesis with Untranscribed Data". (Kim et al., 2023)

OPEN-SOURCE CONTRIBUTION

NAVER USDM ★80
Official Implementation of our NeurIPS 2024 paper "Paralinguistics-Aware Speech-Empowered Large Language Models for Natural Conversation". (Kim et al., 2024)

INVITED TALKS

"Speech Synthesis to Voice Assistant", Supertone, 2025
"Latest Trends in Spoken Dialog Models and Voice Agents", Qualcomm, 2024
"Integrating Paralinguistics in Speech-Empowered Large Language Models for Natural Conversation", HMG Tech. Summit, 2024
"Speech and Spoken Dialog Modeling", Neosapience, 2024
"A case study of research and development at Seoul National University using Amazon Mechanical Turk", AWS Summit Seoul, 2024
"Guided-TTS: A Diffusion Model for Text-to-Speech via Classifier Guidance", Kakao Enterprise, 2022

HONORS

Best Paper Award, ACML 2023, 2023
Best Poster Award, 2022 AIIS Fall Retreat, 2022
Outstanding Paper Award, Hyundai AI Consortium, 2022
Cum Laude, Seoul National University, 2019
Academic Performance Scholarship, Seoul National University: 2016-1, 2018-1,2

SERVICES

Reviewer, ICML 2025, IEEE Transactions On Multimedia 2025, CVPR 2025, ICLR 2025
Top Reviewer, NeurIPS 2024

LANGUAGES

- Korean: Native language.
- English: Intermediate (speaking, reading, writing).

OTHER PUBLICATIONS

CONFERENCES (OTHERS)

[1] J. Yeom, H. Kim, J. Choi, C. Lee, N. Park, S. Yoon, "VoiceGuider: Enhancing Out-of-Domain Performance in Parameter-Efficient Speaker-Adaptive Text-to-Speech via Autoguidance," ***ICASSP 2025***, Hyderabad, India, Apr 2025.

[2] C. Shin, J. Choi, H. Kim, S. Yoon, "Large-Scale Text-to-Image Model with Inpainting is a Zero-Shot Subject-Driven Image Generator," ***CVPR 2025***, Nashville, Tennessee USA, Jun 2025.

[3] N. Park, H. Kim, C. Lee, J. Choi, J. Yeom, S. Yoon, "NanoVoice: Efficient Speaker-Adaptive Text-to-Speech for Multiple Speakers," ***ICASSP 2025***, Hyderabad, India, Apr 2025.

[4] S. Lee, H. Kim, C. Shin, X. Tan, C. Liu, Q. Meng, T. Qin, W. Chen, S. Yoon, and T. Liu, "PriorGrad: Improving Conditional Denoising Diffusion Models with Data-Dependent Adaptive Prior," ***ICLR 2022*** (Virtual), Apr 2022.

[5] U. Hwang, H. Kim, D. Jung, H. Jang, H. Lee, and S. Yoon, "Stein Latent Optimization for Generative Adversarial Networks," ***ICLR 2022*** (Virtual), Apr 2022.

[6] S. Yu, J. Song, H. Kim, S. Lee, W. Ryu, and S. Yoon, "Rare Tokens Degenerate All Tokens: Improving Neural Text Generation via Adaptive Gradient Gating for Rare Token Embeddings," ***ACL 2022***, Dublin, Ireland, May 2022.

JOURNALS

[1] H. Yoo*, E. Kim*, J. Chung*, H. Cho, S. Jeong, H. Kim, D. Jang, H. Kim, J. Yoon, G. Lee, H. Kang, J. Kim, Y. Yun, S. Yoon, Y. Hong, "Silent Speech Recognition with Strain Sensors and Deep Learning Analysis of Directional Facial Muscle Movement," ***ACS Appl. Mater. Interfaces 2022***, Nov 2022. (Impact Factor: 9.229)

ARXIV (OTHERS)

[1] H. Kim*, C. Lee*, S. Park, J. Yeom, N. Park, S. Yu, and S. Yoon, "Does Your Voice Assistant Remember? Analyzing Conversational Context Recall and Utilization in Voice Interaction Models," *arXiv:2502.19759*, Feb 2025.

[2] S. Kim*, H. Kim*, and S. Yoon, "Guided-TTS 2: A Diffusion Model for High-quality Adaptive Text-to-Speech with Untranscribed Data," *arXiv:2205.15370*, May 2022.

[3] HyperCLOVA X Team, "HyperCLOVA X Technical Report," *arXiv:2404.01954*, Apr 2024.

[4] C. Lee, H. Kim, J. Yeom, and S. Yoon, "EdiText: Controllable Coarse-to-Fine Text Editing with Diffusion Language Models," *arXiv:2502.19765*, Feb 2025.

[5] J. Choi, C. Shin, Y. Oh, H. Kim, S. Yoon, "Style-Friendly SNR Sampler for Style-Driven Generation," *arXiv:2411.14793*, Nov 2024.

PATENTS

[1] "Speech recognition using facial skin strain data", S. Yoon, E. Kim, H. Kim.
US Patent US11810549B2 (2021) & KR Patent KR20220118583A (2021)

[2] "Method and apparatus for training an unsupervised conditional generative model", S. Yoon, U. Hwang, H. Kim.
US Patent US20230394319A1 (2023) & KR Patent KR20230168128A (2023)

(: Equal contribution)*

REFERENCES

- **Professor Sungroh Yoon**
  Professor of Electrical & Computer Engineering, Seoul National University
  1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
  sryoon@snu.ac.kr • +82-2 880-1401