

# Heeseung Kim

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

<b>EXPERIENCE</b>	<b>University of Seoul</b> , Seoul, Korea ▪ Assistant Professor	Mar 2026 – Present
	<b>Qualcomm AI Research Korea</b> , Seoul, Korea ▪ Senior Research Engineer	Sep 2025 – Feb 2026
<b>EDUCATION</b>	<b>Seoul National University</b> , Seoul, Korea ▪ Ph.D. candidate in Electrical & Computer Engineering • Focus: Deep Learning, Generative Models, Speech LLM, Spoken Dialog Model, Speech Synthesis • Cumulative GPA: 3.93 / 4.3	Mar 2019 – Aug 2025
	▪ B.S. in Electrical & Computer Engineering • Focus: Signal Processing, Machine Learning, Deep Learning • Cumulative GPA: 3.85 / 4.3 (Cum Laude)	Mar 2015 – Feb 2019
<b>RESEARCH INTERESTS</b>	Speech Large Language Model, Voice Assistant, Omni-modal Agents	
<b>SELECTED PUBLICATIONS</b>	[1] <a href="#">H. Kim</a> , S. Seo, K. Jeong, O. Kwon, S. Kim, J. Kim, J. Lee, E. Song, M. Oh, J. Ha, S. Yoon, and K. Yoo, “ <b>Paralinguistics-Aware Speech-Empowered Large Language Models for Natural Conversation</b> ,” in <i>NeurIPS 2024</i> , Vancouver, Canada, Dec 2024. [2] <a href="#">H. Kim*</a> , C. Lee*, S. Park, J. Yeom, N. Park, S. Yu, and S. Yoon, “ <b>Does Your Voice Assistant Remember? Analyzing Conversational Context Recall and Utilization in Voice Interaction Models</b> ,” <i>ACL 2025 Findings</i> , Jul 2025. [3] <a href="#">H. Kim*</a> , S. Kim*, and S. Yoon, “ <b>Guided-TTS: A Diffusion Model for Text-to-Speech via Classifier Guidance</b> ,” in <i>ICML 2022</i> , Baltimore, Maryland USA, Jul 2022. [4] <a href="#">H. Kim</a> , S. Kim, J. Yeom, and S. Yoon, “ <b>UnitSpeech: Speaker-adaptive Speech Synthesis with Untranscribed Data</b> ,” in <i>INTERSPEECH 2023, Oral Presentation</i> , Dublin, Ireland, Aug 2023. [5] <a href="#">H. Kim</a> , S. Lee, J. Yeom, C. Lee, S. Kim, and S. Yoon, “ <b>VoiceTailor: Lightweight Plug-In Adapter for Diffusion-Based Personalized Text-to-Speech</b> ,” in <i>INTERSPEECH 2024</i> , Kos Island, Greece, Sep 2024.	
<b>RESEARCH EXPERIENCE</b>	<b>Qualcomm AI Research Korea</b> ▪ Role: Senior Research Engineer	Sep 2025 – Feb 2026
	▪ <b>Project Overview:</b> • 2025.09 – 2026.02: Developed <b>Full-Duplex Voice Assistant</b> enabling natural and simultaneous speech interaction. Improved Full-Duplex Speech LLM performance by leveraging synthetic full-duplex text data.	
	<b>Naver &amp; Seoul National University Collaboration</b>	Mar 2022 – Jan 2025
	▪ Role: Industry-academia collaboration, Electrical & Computer Engineering (Ph.D. Candidate) ▪ <b>Project Overview:</b> • 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 <i>English spoken dialog system</i> with an end-to-end pipeline and paralinguistic awareness. This research led to a publication at <i>NeurIPS 2024</i> , 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.	
<b>REPOSITORIES</b>	<b>UnitSpeech</b> ★130+ Official Implementation of INTERSPEECH 2023 paper “UnitSpeech: Speaker-adaptive Speech Synthesis with Untranscribed Data”. (Kim et al., 2023)	
<b>OPEN-SOURCE CONTRIBUTION</b>	<b>NAVER USDM</b> ★90+ Official Implementation of our NeurIPS 2024 paper “Paralinguistics-Aware Speech-Empowered Large Language Models for Natural Conversation”. (Kim et al., 2024)	
<b>INVITED TALKS</b>	“Speech Synthesis to Voice Assistant”, Supertone, 2025	

	<p>“Latest Trends in Spoken Dialog Models and Voice Agents”, Qualcomm, 2024</p> <p>“Integrating Paralinguistics in Speech-Empowered Large Language Models for Natural Conversation”, HMG Tech. Summit, 2024</p> <p>“Speech and Spoken Dialog Modeling”, Neosapience, 2024</p> <p>“A case study of research and development at Seoul National University using Amazon Mechanical Turk”, AWS Summit Seoul, 2024</p> <p>“Guided-TTS: A Diffusion Model for Text-to-Speech via Classifier Guidance”, Kakao Enterprise, 2022</p>
<b>HONORS</b>	<p>Best Paper Award, ACML 2023, 2023</p> <p>Best Poster Award, 2022 AIIS Fall Retreat, 2022</p> <p>Outstanding Paper Award, Hyundai AI Consortium, 2022</p> <p>Cum Laude, Seoul National University, 2019</p> <p>Academic Performance Scholarship, Seoul National University: 2016-1, 2018-1,2</p>
<b>SERVICES</b>	<p><b>Reviewer:</b> ACM Multimedia, NeurIPS, ARR, ICML, IEEE Transactions On Multimedia 2025, CVPR, ICLR, AAAI, <b>Top Reviewer:</b> NeurIPS 2024</p>
<b>LANGUAGES</b>	<ul style="list-style-type: none"> <li>▪ Korean: Native language.</li> <li>▪ English: Intermediate (speaking, reading, writing).</li> </ul>
<b>OTHER PUBLICATIONS</b>	<p><b>CONFERENCES (OTHERS)</b></p> <p>[1] J. Choi, C. Shin, Y. Oh, <u>H. Kim</u>, J. Lee, S. Yoon, “Style-Friendly SNR Sampler for Style-Driven Generation,” <b>WACV 2026</b>, Mar 2026.</p> <p>[2] C. Lee, <u>H. Kim</u>, J. Yeom, and S. Yoon, “EdiText: Controllable Coarse-to-Fine Text Editing with Diffusion Language Models,” <b>ACL 2025</b>, Jul 2025.</p> <p>[3] N. Park, <u>H. Kim</u>, C. Lee, J. Choi, J. Yeom, S. Yoon, “NanoVoice: Efficient Speaker-Adaptive Text-to-Speech for Multiple Speakers,” <b>ICASSP 2025, Oral Presentation</b>, Hyderabad, India, Apr 2025.</p> <p>[4] J. Yeom, <u>H. Kim</u>, J. Choi, C. Lee, N. Park, S. Yoon, “VoiceGuider: Enhancing Out-of-Domain Performance in Parameter-Efficient Speaker-Adaptive Text-to-Speech via Autoguidance,” <b>ICASSP 2025</b>, Hyderabad, India, Apr 2025.</p> <p>[5] C. Shin, J. Choi, <u>H. Kim</u>, S. Yoon, “Large-Scale Text-to-Image Model with Inpainting is a Zero-Shot Subject-Driven Image Generator,” <b>CVPR 2025</b>, Nashville, Tennessee USA, Jun 2025.</p> <p>[6] C. Shin*, <u>H. Kim</u>*, C. Lee, S. Lee, and S. Yoon, “Edit-A-Video: Single Video Editing with Object-Aware Consistency,” <b>ACML 2023, Oral Presentation, Best Paper Award</b>, Istanbul, Turkey, Nov 2023.</p> <p>[7] S. Lee, <u>H. Kim</u>, 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,” <b>ICLR 2022 (Virtual)</b>, Apr 2022.</p> <p>[8] U. Hwang, <u>H. Kim</u>, D. Jung, H. Jang, H. Lee, and S. Yoon, “Stein Latent Optimization for Generative Adversarial Networks,” <b>ICLR 2022 (Virtual)</b>, Apr 2022.</p> <p>[9] S. Yu, J. Song, <u>H. Kim</u>, S. Lee, W. Ryu, and S. Yoon, “Rare Tokens Degenerate All Tokens: Improving Neural Text Generation via Adaptive Gradient Gating for Rare Token Embeddings,” <b>ACL 2022</b>, Dublin, Ireland, May 2022.</p> <p><b>JOURNALS</b></p> <p>[1] H. Yoo*, E. Kim*, J. Chung*, H. Cho, S. Jeong, <u>H. Kim</u>, 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,” <b>ACS Appl. Mater. Interfaces 2022</b>, Nov 2022. (Impact Factor: 9.229)</p> <p><b>ARXIV (OTHERS)</b></p> <p>[1] S. Kim*, <u>H. Kim</u>*, and S. Yoon, “Guided-TTS 2: A Diffusion Model for High-quality Adaptive Text-to-Speech with Untranscribed Data,” <i>arXiv:2205.15370</i>, May 2022.</p> <p>[2] HyperCLOVA X Team, “HyperCLOVA X Technical Report,” <i>arXiv:2404.01954</i>, Apr 2024.</p> <p><b>PATENTS</b></p> <p>[1] “Speech recognition using facial skin strain data”, S. Yoon, E. Kim, <u>H. Kim</u>. US Patent US11810549B2 (2021) &amp; KR Patent KR20220118583A (2021)</p> <p>[2] “Method and apparatus for training an unsupervised conditional generative model”, S. Yoon, U. Hwang, <u>H. Kim</u>. US Patent US20230394319A1 (2023) &amp; KR Patent KR20230168128A (2023)</p>

(\*: Equal contribution)