Jaesung Bae

jb82@illinois.edu | https://jaesungbae.github.io | Google Scholar | LinkedIn | +1 (447) 410-9728

Education

University of Illinois Urbana-Champaign — Aug 2024 - May 2028 (Expected)
PhD in Computer Science — *Illinois, USA*
- Advisor: Prof. Minje Kim and Prof. Paris Smaragdis

Korea Advanced Institute of Science and Technology (KAIST) — Feb 2017 - Feb 2019
MS in School of Electrical Engineering — *Daejeon, South Korea*
- GPA: 3.95/4.30 (3.84/4.00)
- Advisor: Prof. Dae-Shik-Kim (Brain Reverse Engineering and Imaging Lab)
- Thesis: Speech Command Recognition using Capsule Network

University of Applied Sciences Upper Austria — Sep 2015 - Jan 2016
Exchange student — *Upper Austria, Austria*

Yonsei University — Mar 2013 - Feb 2017
BS in Electrical and Electronics Engineering — *Seoul, South Korea*
- GPA: 3.69/4.30 (3.61/4.00)
- Honors - 2nd Semester, 2016

Work Experience

Meta — May 2025 - Aug 2025
CS Research Scientist Intern — *Menlo Park, CA, USA*
- Voice Modeling Team, Meta Superintelligence Lab
- Research topics: Improving generation and understanding capabilities in the pretraining of speech large language models.

Samsung Research, Samsung Electronics — May 2022 - Jun 2024
Full Time, Speech AI Researcher — *Seoul, South Korea*
- Language & Voice Team, Global AI Center
- Research topics: Zero-shot Text-to-Speech (TTS), personalized TTS, on-device TTS, and expressive TTS

NCSOFT — Mar 2019 - Apr 2022
Full Time, Speech AI Researcher — *Seongnam, South Korea*
- Speech AI Lab, AI Center
- Research topics: Expressive TTS, fine-grained prosody control of TTS, and multi-speaker TTS
- Served as Technical Research Personnel, fulfilling alternative military service in South Korea.

Publications

*: Equal contribution

2026 [17] **Jaesung Bae**, Minje Kim, "Semantics-Aware Generative Latent Data Augmentation for Learning in Low-Resource Domains," *arXiv preprint arXiv:2602.02841*, 2026.

2025 [16] **Jae-Sung Bae**, Anastasia Kuznetsova, Dinesh Manocha, John Hershey, Trausti Kristjansson, and Minje Kim, "Generative Data Augmentation Challenge: Zero-Shot Speech Synthesis for Personalized Speech Enhancement," *in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW): Generative Data Augmentation for Real-World Signal Processing Applications (GenDA 2025)*, 2025.

2024 [15] **Jae-Sung Bae**, Joun Yeop Lee, Ji-Hyun Lee, Seongkyu Mun, Taehwa Kang, Hoon-Young Cho, Chanwoo Kim, "Latent Filling: Latent space data augmentation for zero-shot speech synthesis," *in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2024.

[14] Heejin Choi, **Jae-Sung Bae**, Joun Yeop Lee, Seongkyu Mun, Jihwan Lee, Hoon-Young Cho, Chanwoo Kim, "MELS-TTS : Multi-emotion multi-lingual multi-speaker text-to-speech system via disentangled style tokens," *in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2024.

2023 [13] Joun Yeop Lee, **Jae-Sung Bae**, Seongkyu Mun, Jihwan Lee, Ji-Hyun Lee, Hoon-Young Cho, Chanwoo Kim, "Hierarchical timbre-cadence speaker encoder for zero-shot speech synthesis," in *Proc. Interspeech*, 2023.

[12] Taejun Bak, Junmo Lee, Hanbin Bae, Jinhyeok Yang, **Jae-Sung Bae**, Young-Sun Joo, "Avocodo: Generative adversarial network for artifact-free vocoder," in *Proc. AAAI*, 2023.

2022 [11] **Jae-Sung Bae**, Jinhyeok Yang, Tae-Jun Bak, Young-Sun Joo, "Hierarchical and multi-scale variational autoencoder for diverse and natural non-autoregressive text-to-speech," in *Proc. Interspeech*, 2022.

[10] Jihwan Lee, **Jae-Sung Bae**, Seongkyu Mun, Heejin Choi, Joun Yeop Lee, Hoon-Young Cho, Chanwoo Kim, "An Empirical Study on L2 Accents of Cross-lingual Text-to-Speech Systems via Vowel Space," *arXiv preprint arXiv:2211.03078,* 2022.

[9] Jihwan Lee, Joun Yeop Lee, Heejin Choi, Seongkyu Mun, Sangjun Park, **Jae-Sung Bae**, Chanwoo Kim, "Into-TTS: Intonation template based prosody control system," *arXiv preprint arXiv:2204.01271,* 2022.

2021 [8] **Jae-Sung Bae**, Tae-Jun Bak, Young-Sun Joo, and Hoon-Young Cho, "Hierarchical context-aware transformers for non-autoregressive text to speech," in *Proc. Interspeech*, 2021.

[7] Jinhyeok Yang*, **Jae-Sung Bae***, Taejun Bak, Youngik Kim, and Hoon-Young Cho, "GANSpeech: Adversarial training for high-fidelity multi-speaker speech synthesis," in *Proc. Interspeech*, 2021.

[6] Taejun Bak, **Jae-Sung Bae**, Hanbin Bae, Young-Ik Kim, and Hoon-Young Cho, "FastPitchFormant: Source-filter based decomposed modeling for speech synthesis," in *Proc. Interspeech*, 2021.

[5] Hanbin Bae, **Jae-Sung Bae**, Young-Sun Joo, Young-Ik Kim, and Hoon-Young Cho, "A neural text-to-speech model utilizing broadcast data mixed with background music," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2021.

[4] **Jae-Sung Bae**, Hanbin Bae, Young-Sun Joo, Junmo Lee, Gyeong-Hoon Lee, Hoon-Young Cho, "Speaking speed control of end-to-end speech synthesis using sentence-level conditioning," in *Proc. Interspeech*, 2020.

2019 [3] Juntae Kim, **Jae-Sung Bae**, "Phase-aware speech enhancement with a recurrent two stage network," *arXiv preprint arXiv:2001.09772*, 2019.

[2] Juntae Kim*, **Jae-Sung Bae***, Minsoo Hahn, "End-point detection with state transition model based on chunk-wise classification," *arXiv preprint arXiv:1912.10442*, 2019.

2018 [1] **Jae-Sung Bae**, Dae-Shik Kim, "End-to-end speech command recognition with capsule network," in *Proc. Interspeech*, 2018.

Projects

On-device TTS System in various languages for Galaxy S24's Live Translation Mar 2023 - Jun 2024

- Researched and developed an on-device multilingual TTS system supporting eight languages, integrated as the *Live Translation* feature and introduced as a *main AI capability in the Galaxy S24*.
- Enhanced model architecture to achieve a high-quality TTS system across multiple languages while reducing model size for efficient on-device deployment.

On-device Personalized TTS System for Bixby Custom Voice Creation May 2022 - Jun 2024

- Researched and developed an on-device personalized TTS system integrated into Samsung Galaxy Bixby's *Custom Voice Creation* and utilized in *Bixby Text-Call* functionality.
- Enabled personalized TTS generation by fine-tuning the base model directly on the user's device with just ten recorded utterances.

TTS System of K-pop Fandom Platform, "UNIVERSE" Mar 2019 - Apr 2022

- Conducted research and developed a multi-speaker TTS system capable of generating voices for approximately *100 K-pop artists* within a single TTS system, powering two features in the *UNIVERSE* platform.
- **1. Fan Networking Service (FNS):** In this feature, K-pop artists were able to create posts with photos and short comments, similar to Instagram. The TTS system reads these comments aloud in the respective artists' voices, enhancing the fan experience.
- **2. Private Call:** This enabled fans to receive simulated phone calls with the voices of their favorite artists. The TTS system was used to generate the voices for these artist phone calls, providing fans with a unique and exciting interaction with their beloved artists.

Fine-grained Prosody Control of TTS System Mar 2021 - Apr 2022

- Led the research and development of an advanced TTS system enabling fine-grained prosody control, allowing users to synthesize speech with customized prosodic characteristics.

- Released as an *internal API* with a companion website for interactive prosody generation and testing.
- Applied in a promotional video for the game named *Trickster-M*, generating expressive narration for updated patch notes.

TTS System in Baseball Broadcast Scenario Mar 2019 - Mar 2021

- Developed an expressive TTS system optimized for baseball broadcast environments, capable of adapting to multiple emotional contexts.
- Enabled speech synthesis across *four emotional tones* (highly expressive, expressive, neutral, and depressed) using symbolic cues such as commas (,), tilde (∼), exclamation marks (!), and question marks (?).
- Published demos and blog articles on NCSOFT's official channels. (Demo Videos and Blog Post Links)

Stock Price Prediction Jan 2018 - Dec 2018

- Built a deep learning model to forecast the rise and fall of the KOSPI 200 index using a mixture-of-experts (MoE) framework.
- Designed and trained an RNN-VAE-based model incorporating 58 major factors, sector indices, and historical price data for multi-factor time series prediction.

Human Facial Expression, Behavior Recognition, and Tracking Mar 2017 - Dec 2017

- Developed a deep learning-based system for facial expression and behavior recognition, capable of detecting faces in input images and predicting detected individuals' gender and age.
- Led the model design and implementation of face detection and gender/age classification components.

Academic Service

Conference Reviewer: AAAI, NeurIPS, MLSP 2025, ICASSP 2026

Teaching

[CS448] Audio Computing Laboratory (TA) | *University of Illinois Urbana-Champaign* Jan 2026 - May 2026
[EE635] Functional Brain Imaging (TA) | *KAIST* Sep 2018 - Dec 2018
[EE209] Programming Structure for Electrical Engineering (TA) | *KAIST* Sep 2017 - Dec 2017

Invited Talk

Data Augmentation for Speech Dec 2025
Korea University *Seoul, South Korea*
End-to-End Speech Command Recognition with Capsule Network Sep 2018
Naver Corp. *Seongnam, South Korea*

- Youtube Link

Media

Latent Filling: Latent Space Data Augmentation for Zero-shot Speech Synthesis Apr 2024
Samsung Research's official blog (Link)
MELS-TTS: Multi-Emotion Multi-Lingual Multi-Speaker Text-To-Speech System via Disentangled Style Tokens Apr 2024
Samsung Research's official blog (Link)
Hierarchical Timbre-Cadence Speaker Encoder for Zero-shot Speech Synthesis Sep 2023
Samsung Research's official blog (Link)
Introducing Four Papers Accepted at Interspeech 2021 Sep 2021
NCSOFT's official blog (Link)
NCSOFT's Speech AI Lab: Creating Achievements Together and Growing Together - Four Papers Accepted at Interspeech 2021 Sep 2021
NCSOFT's official blog (Link)
Preserving the Realism of Baseball Game with "Broad-Casting Style" TTS System that Mimics Sports Commentators Dec 2020
Yonhap News (Link), NCSOFT's official blog (Link)
Speed Control of AI TTS Systems Enhancing Naturalness of Synthesized Speech Nov 2020
NCSOFT's official blog (Link)

Skills

- Python, pytorch, C++, git, docker, and tensorflow