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

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
- I also served as Technical Research Personnel, which is a form of alternative military service in South Korea.

Publications

*: Equal contribution

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

- I contributed to the research and development of an on-device TTS system in eight different languages, which is included as a *Live Translation* feature and introduced as a *main AI feature in the Galaxy S24*.
- My contribution involved enhancing the model architecture and achieving a high-quality TTS system that supports various languages with a reduced model size.

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

- I contributed to the research and development of an on-device personalized TTS system, which was integrated into Samsung Galaxy Bixby's Custom Voice Creation and utilized within *Bixby Text-call* functionality.
- It can create a personalized TTS system by fine-tuning the TTS system directly on the user's device with just 10 utterances.

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

- I conducted research and crafted a multi-speaker TTS system capable of generating the voices of approximately *100 K-pop artists* within a single TTS system. This TTS system is utilized in the following two services integrated into UNIVERSE.
- **1. Fan Networking Service (FNS):** In this service, K-pop artists 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:** UNIVERSE offers a feature that allows 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

- I led the research and was the principal developer of an advanced TTS system capable of fine-grained prosody control. This allows users to generate speech with the specific prosodic characteristics they desire.
- It was launched as an *internal API*, and a website was also developed to enable users to generate speech with the required prosodic characteristics.
- As an example of its practical application, the TTS system was utilized to create speech for a video introducing the updated patch notes for the game "Trickster-M". (Youtube Link)

TTS System in Baseball Broadcast Scenario Mar 2019 - Mar 2021

- I played a pivotal role in the research and development of an expressive TTS system, specifically designed for diverse scenarios in baseball.
- The TTS system is capable of generating speech in four distinct emotional tones: highly expressive, expressive, neutral, and depressed.
- It can generate expressive speech responses based on input text symbols such as commas (,), tilde (∼), exclamation marks (!), and question marks (?).

- I published several demos on NCSOFT's official blog and news articles. (Demo Videos and Blog Post Links)

Stock Price Prediction Jan 2018 - Dec 2018

- I conducted research and developed a deep learning model capable of predicting the rise/fall of the KOSPI 200 index, employing a mixture-of-expert (MoE) method.
- I developed a model based on the RNN-VAE approach to predict future prices of 58 major factors using historical price data of various factors, sector indices, etc.

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

- I conducted research and developed a deep learning model capable of detecting faces in input images and predicting the gender and age of the detected individuals.
- I led the project and mainly developed the face detection and gender/age prediction model.

Academic Service

Conference Reviewer: AAAI 2025

Teaching

[EE635] Functional Brain Imaging (TA) | *KAIST* Sep 2018 - Dec 2018

[EE209] Programming Structure for Electrical Engineering (TA) | *KAIST* Sep 2017 - Dec 2017

Invited Talk

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