Jae-Sung Bae

Phone: (+82) 10-8763-3279
E-mail: bjsd3@alumni.kaist.ac.kr
Web page: https://jaesungbae.github.io
Google Scholar: https://scholar.google.co.kr/citations?user=ay1zanAAAAAJ&hl=ko
LinkedIn: http://www.linkedin.com/in/jaesung-bae-955410157

Interests & Skills

Over the past few years, I have been focused on research related to expressive text-to-speech (TTS) systems and zero-shot TTS systems. Through my research experiences, I realized the importance of properly representing information inherent in speech such as emotion, speaking style, and speaker identity. Looking ahead, I would like to explore how we can effectively represent and interpret speech characteristics. Additionally, I am also interested in multimodal models, which model the relationship between various information, such as vision, language, and speech, just as humans naturally do.

- Interests:
  - Speech representation learning
    - e.g., Interpreting and controlling the representation of speech
  - Multi-modal modeling utilizing speech
    - Interpreting and learning the relationship between visual and language representations and speech representation.
    - e.g., spontaneous speech-to-speech translation, video dubbing
  - Control of generative models
    - e.g., Control of LLM model and LLM-like generative models (e.g., VALLE)
- Skills: python, pytorch, (docker, git, and tensorflow, c++)

Work Experiences

Samsung Research, Samsung Electronics - *Speech AI Researcher, full-time*
Seoul, Republic of Korea
May 2022 - Present

- Language & Voice team, Global AI center
- Research topics:
  - Zero-shot and personalized TTS system
  - On-device TTS system
  - Expressive TTS system

NCSOFT - *Speech AI Researcher, full-time (as a technical research personnel[1])*
Seong-Nam, Republic of Korea
Mar 2019 - Apr 2022

- Speech AI Lab, AI Center
- Research topics:
  - Expressive TTS systems, speaking style (prosody) modeling/control
  - Fine-grained prosody control of TTS system
  - Prosody modeling/control via text symbols

[1] The technical research personnel system is a form of alternative military service that enables companies (research institutes) to utilize research personnel for the advancement of science and technology in Korea. Master's degree and 36 months of service in science/engineering research is required. (Source: https://natsci.kaist.ac.kr/eng/sub040507)

    - Improving speech intelligibility of TTS system
    - Multi-speaker TTS system (about 100 celebrities' voices)
    - Korean TTS system
- Experiences in various generative models such as VAE, GAN, and Flow

Education

Korea Advanced Institute of Science and Technology (KAIST)
Master's degree in School of Electrical Engineering
Mar 2017 - Feb 2019, *Daejeon, Republic of Korea*

- GPA: 3.95/4.30
- Advisor: Prof. Daeshik-Kim (Brain Reverse Engineering and Imaging Lab)
- Research Area: deep learning, speech (command) recognition
- Research grant project experiences:
    - Face recognition, with Kolon Industries
    - Time-series (stock price) prediction, with Eastspring Investments

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

Yonsei University
Bachelor's degree in School of Electrical and Electronics Engineering
Mar 2013 - Feb 2017, *Seoul, Republic of Korea*

- Overall GPA: 3.69/4.30
- Major GPA: 3.79/4.30

Publications

*: Equal Contribution

2023 [14] **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," *arXiv preprint, arXiv: 2310.03538,* 2023. (submitted to ICASSP 2024)

[13] 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,". (submitted to ICASSP 2024)

[12] 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.

[11] 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 [10] **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.

- Presentation link: https://youtu.be/3U5cEu0gFYY

[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.

2020 [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.

- Presentation link: https://youtu.be/WyDfc53Ez_A

2019 [3] Jun,tae Kim, **Jae-Sung Bae,** "Phase-Aware Speech Enhancement with a Recurrent Two Stage Network," *arXiv preprint arXiv:2001.09772,* 2019.

[2] Juntae Kim*, **Jaesung 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[2]

On-device Personalized TTS System (Live Service)

May 2022 - Present

- I contributed to the research and development of the on-device personalized TTS system which is integrated into Samsung Galaxy devices 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.

TTS System of K-pop Fandom Platform, "UNIVERSE" (Live Service)

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 two services integrated into UNIVERSE:

[2] Please visit the personal web page (https://jaesungbae.github.io) for demo videos.

- **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 idols' 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 idol phone calls, providing fans with a unique and exciting interaction with their beloved idols.

Fine-grained Prosody Control of TTS System (Prototype Web Service)

Mar 2021 - Apr 2022

- I conducted research and developed the *TTS system that is able to control the prosody of speech in a fine-grained level*. This allows users to generate speech with the specific prosodic characteristics they desire.
- This *fine-grained controllable TTS system* was released as a prototype web service, which was opened as an in-company service.
- As an example of its practical application, the TTS system was utilized to create speech for a video introducing the latest patch notes for the game "Trickster-M."
    - Youtube link: https://youtu.be/_Ssb9y73XtI?si=gZha2nn1NiKJAZh9&t=110

TTS System in Baseball Broadcast Scenario

Mar 2019 - Mar 2021

- I played a significant role in the research and development of *an expressive Text-to-Speech (TTS) system tailored for various baseball scenarios.* This system can produce speech with dynamic emotional nuances, which are well-suited for different situations in the baseball game.
    - 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 (?).
- Published several demos on NCSOFT's official blog and news articles. (Demo and blog post link: https://about.ncsoft.com/news/article/prosody-control-ai-20201210)

Speaking Speed Control of TTS System

Mar 2019 - Mar 2021

- I conducted research on the TTS system that can control the speaking speed of speech.
- Published as NCSOFT's official blog and news articles. (Demo and blog post link: https://about.ncsoft.com/news/article/gst-tts-ai-20201106)

Teaching

[EE635] Functional Brain Imaging (TA)

Sep 2018 - Dec 2018, *KAIST*

[EE209] Programming Structure for Electrical Engineering (Extra TA)

Sep 2017 - Dec 2017, *KAIST*

Honors

Honors - 2nd Semester, 2016 *(at Yonsei University)*

Invited Talks

End-to-End Speech Command Recognition with Capsule Network
NAVER Corp., *Seong-Nam, Republic of Korea*
Sep 2018

- Youtube link: https://youtu.be/iFtZqjedoWE

Media

Hierarchical Timbre-Cadence Speaker Encoder for Zero-shot Speech Synthesis
Samsung Research's official blog, Sep 2023

- Link: https://research.samsung.com/blog/Hierarchical-Timbre-Cadence-Speaker-Encoder-for-Zero-shot-Speech-Synthesis (Go to Link)

Introducing Four Papers Accepted at Interspeech 2021
NCSOFT's official blog, Sep 2021

- Link: https://about.ncsoft.com/news/article/interspeech-02-20210913

NCSOFT's Speech AI Lab: Creating Achievements Together and Growing Together - Four Papers Accepted at Interspeech 2021
NCSOFT's official blog, Aug 2021

- Link: https://about.ncsoft.com/news/article/interspeech-01-20210830

Preserving the Realism of Baseball Game with "Broad-Casting Style" TTS System that Mimics Sports Commentators
NCSOFT's official blog, Dec 2020

- Link: https://about.ncsoft.com/news/article/prosody-control-ai-20201210

Speed Control of AI TTS Systems Enhancing Naturalness of Synthesized Speech
NCSOFT's official blog, Nov 2020

- Link: https://about.ncsoft.com/news/article/gst-tts-ai-20201106