Eunwoo Song

Last updated: Aug 2024

CONTACT INFORMATION

Email gregorio.song@gmail.com

Homepage https://sewplay.github.io/

Address 203, Buil-ro, Bucheon, 14598, South Korea

Phone +82-10-3191-9108

EXPERIENCE

Seongnam, Korea

Jan 2023 – present

Naver Cloud

Senior research scientist, Voice Synthesis team lead

Development of High-quality TTS api for cloud services

- Role: Management, research and development of DNN TTS models
- Related services
  - Clova Voice Pro, Nov 2022 – present

Development of automatic TTS modeling with smartphone recordings

- Role: Management, research and development of DNN TTS models
- Related services
  - Clova VoiceMaker, May 2022 – Oct 2023

Seongnam, Korea

Mar 2017 – Dec 2022

Naver Corp.

Senior research scientist, Voice Model team lead

Hybrid TTS combining deep learning and unit-selection models

- Role: Management, research and development of DNN TTS models
- Related services
  - Showhost voice for Naver Shopping, May 2021 – Dec 2022
  - Representative voice for Clova Contact Center, May 2019 - Dec 2022
  - Navigation voice for Naver Map, Jan 2019 - Sep 2020
  - Sangjin Oh's (Korean newscaster) voice for Naver News, Oct 2019 - May 2020
  - In-Na Yoo's (Korean actress) voice for Clova AI speaker, Apr 2018 - Dec 2018

TTS-based audiobook generation

- Role: Development of audiobook generation tool
- Related services
  - Clova Dubbing x Inmun 360, Sep 2021 - May 2022

Seoul, Korea

Aug 2022 – present

Seoul National Univ.

Adjunct professor, Artificial Intelligence Institute

San Diego, CA

Aug 2016 – Nov 2016

Qualcomm Technologies Inc.

Research intern, Multimedia Research and Development Laboratory

- Spatial audio: Fixed-point implementation of MPEG-H 3D Audio Decoder
- Mentor: Dr. Deep Sen

Beijing, China

Sep 2015 – Feb 2016

Apr 2016 – Jun 2016

Microsoft Research Asia

Research intern, Speech Group

- Speech synthesis: Deep learning-based TTS system using ITFTE vocoder
- Mentor: Dr. Frank Soong

EDUCATION

Seoul, Korea

Sep 2010 – Feb 2019

Yonsei University

Combined M.S. and Ph.D., Electrical and Electronic Engineering

- Dissertation: Improved time-frequency trajectory excitation vocoder for deep learning-based statistical parametric speech synthesis system
- Advisor: Prof. Hong-Goo Kang

Seoul, Korea

Mar 2006 – Aug 2010

Yonsei University

B.S., Electrical and Electronic Engineering

TALKS

1. "Speech synthesis and applications", SNU (2023)
2. "Parallel waveform synthesis", Samsung Research (2022)
3. "Data-selective TTS augmentation", Naver Engineering Day (2022)
4. "Voice synthesis and applications", KAIST and SNU (2022)
5. "Introduction to text-to-speech", Naver Engineering Day (2021)
6. "Deep learning-based text-to-speech", Yonsei Univ. and Korea Univ. (2021)
7. "Clova AI: Text-to-speech technology", Yonsei Univ. (2020)
8. "Parallel WaveGAN", Naver Engineering Day (2020)
9. "Speaker-adaptive WaveNet". Naver Engineering Day (2018)
10. "Clova voice: From unit-selection to deep learning-based TTS", ASK Conference (2018)
11. "Speaker-adaptive text-to-speech", Naver AI Colloquium (2018)
12. "Speech synthesis: Improved time-frequency trajectory excitation", Beijing Univ. (2015)

PUBLICATIONS

1. H. Yoon, J.-S. Kim, R. Yamamoto, R. Terashima, C.-H. Song, J.-M. Kim, **E. Song**, "Enhancing multilingual TTS with voice conversion based data augmentation and posterior embedding," in Proc. ICASSP, 2024, pp. 12186-12190.
2. H. Yoon, C. Kim, **E. Song**, H. Yoon , H.-G. Kang, "Pruning self-attention for zero-shot multi-speaker text-to-speech," in Proc. Interspeech, 2023, pp. 4299-4303.
3. Y. Shirahata, R. Yamamoto, **E. Song**, R. Terashima, J.-M. Kim, K. Tachibana, "Period VITS: Variational inference with explicit pitch modeling for end-to-end emotional speech synthesis," in Proc. ICASSP, 2023, pp.1-5.

PUBLICATIONS

4. S.-H. Lee, S.-B. Kim, J.-H. Lee, **E. Song**, M.-J. Hwang, S.-W. Lee, "HierSpeech: Bridging the gap between text and speech by hierarchical variational inference using self-supervised representations for speech synthesis," in Proc. NeurIPS, 2022, pp. 16624-16636.
5. **E. Song**, R. Yamamoto, O. Kwon, C.-H. Song, M.-J. Hwang, S. Oh, H.-W. Yoon, J.-S. Kim, J.-M. Kim, "TTS-by-TTS 2: Data-selective augmentation for neural speech synthesis using ranking support vector machine with variational autoencoder," in Proc. INTERSPEECH, 2022, pp. 1941-1945.
6. H. Yoon, O. Kwon, H. Lee, R. Yamamoto, **E. Song**, J.-M. Kim, M.-J. Hwang, "Language model-based emotion prediction methods for emotional speech synthesis systems," in Proc. INTERSPEECH, 2022, pp. 4596-4600.
7. R. Terashima, R. Yamamoto, **E. Song**, Y. Shirahata, H.-W. Yoon, J.-M. Kim, K. Tachibana, "Cross-speaker emotion transfer for low-resource text-to-speech using non-parallel voice conversion with pitch-shift data augmentation," in Proc. INTERSPEECH, 2022, pp. 3018-3022.
8. M.-J. Hwang, H.-W. Yoon, C.-H. Song, J.-S. Kim, J-.M. Kim, **E. Song**, "Linear prediction-based Parallel WaveGAN speech synthesis," in Proc. ICEIC, 2022, pp. 1-4.
9. S. Oh, O. Kwon, M.-J. Hwang, J.-M. Kim, **E. Song**, "Effective data augmentation methods for neural text-to-speech systems," in Proc. ICEIC, 2022, pp. 1-4.
10. M.-J, Hwang, R. Yamamoto, **E. Song**, J.-M. Kim, "High-fidelity Parallel WaveGAN with multi-band harmonic-plus-noise model," in Proc. INTERSPEECH, 2021, pp. 2227-2231.
11. H.-K. Nguyen, K. Jeong, S. Um, M.-J. Hwang, **E. Song**, H.-G. Kang, "LiteTTS: A decoder-free lightweight text-to-wave synthesis based on generative adversarial networks," in Proc. INTERSPEECH, 2021. pp. 3595-3599.
12. R. Yamamoto, **E. Song**, M.-J. Hwang, J.-M. Kim, "Parallel waveform synthesis based on generative adversarial networks with voicing-aware conditional discriminators," in Proc. ICASSP, 2021, pp. 6039-6043.
13. M.-J. Hwang, R. Yamamoto, **E. Song**, J.-M. Kim," TTS-by-TTS: TTS-driven data augmentation for fast and high-quality speech synthesis," in Proc. ICASSP, 2021, pp. 6598-6602.
14. **E. Song**, R. Yamamoto, M.-J. Hwang, J.-S. Kim, O. Kwon, J.-M. Kim, "Improved Parallel WaveGAN with perceptually weighted spectrogram loss," in Proc. SLT, 2021, pp. 470-476.
15. M.-J. Hwang, F. K. Soong, **E. Song**, X. Wang, H. Kang, H.-G. Kang, "LP-WaveNet: Linear prediction-based WaveNet speech synthesis," in Proc. APSIPA, 2020, pp. 810-814.

PUBLICATIONS

16. S. Oh, H. Lim, K. Byun, M.-J. Hwang, **E. Song**, H.-G. Kang, "ExcitGlow: Improving a WaveGlow-based neural vocoder with linear prediction analysis," in Proc. APSIPA, 2020, pp. 831-836.
17. **E. Song**, M.-J. Hwang, R. Yamamoto, J.-S. Kim, O. Kwon, J.-M. Kim, "Neural text-to-speech with a modeling-by-generation excitation vocoder," in Proc. INTERSPEECH, 2020, pp. 3570-3574.
18. **E. Song**, J.-S. Kim, K. Byun, H.-G. Kang, "Speaker-adaptive neural vocoders for parametric speech synthesis systems," in Proc. MMSP, 2020, pp. 1-5.
19. R. Yamamoto, **E. Song**, J.-M. Kim, "Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram," in Proc. ICASSP, 2020, pp. 6194-6198.
20. M.-J. Hwang, **E. Song**, R. Yamamoto, F. K. Soong, H.-G. Kang, "Improving LPCNet-based text-to-speech with linear predictions-structured mixture density network," in Proc. ICASSP, 2020, pp. 7214-7218.
21. R. Yamamoto, **E. Song**, J.-M. Kim, "Probability density distillation with generative adversarial networks for high-quality parallel waveform generation," in Proc. INTERSPEECH, 2019, pp. 699-703.
22. **E. Song**, K. Byun, H.-G. Kang, "ExcitNet vocoder: A neural excitation model for parametric speech synthesis systems," in Proc. EUSIPCO, 2019, pp. 1179-1183.
23. K. Byun, **E. Song**, J. Kim, J.-M. Kim, H.-G. Kang, "Excitation-by-SampleRNN model for text-to-speech," in Proc. ITC-CSCC, 2019, pp. 356-359.
24. J. Y. Lee, S. J. Cheon, B. J. Choi, N. S. Kim, **E. Song**, "Acoustic modeling using adversarially trained variational recurrent neural network for speech synthesis," in Proc. INTERSPEECH, 2018, pp. 917-921.
25. M.-J. Hwang, **E. Song**, J.-S. Kim, H.-G. Kang, "A unified framework for the generation of glottal signals in deep learning-based parametric speech synthesis systems," in Proc. INTERSPEECH, 2018, pp. 912-916.
26. M.-J. Hwang, **E. Song**, H.-G. Kang, "Modeling-by-generation-structured noise compensation algorithm for glottal vocoding speech synthesis system," in Proc. ICASSP, 2018, pp. 5669-5673.
27. **E. Song**, F. K. Soong, H.-G. Kang, "Perceptual quality and modeling accuracy of excitation parameters in DLSTM-based speech synthesis systems," in Proc. ASRU, 2017, pp. 671–676.
28. **E. Song**, F. K. Soong, H.-G. Kang, "Effective spectral and excitation modeling techniques for LSTM-RNN-based speech synthesis systems," IEEE/ACM Trans. Audio, Speech, and Lang. Process., vol. 25, no. 11, pp. 2152–2161, 2017.
29. **E. Song**, F. K. Soong, H.-G. Kang, "Improved time-frequency trajectory excitation vocoder for DNN-based speech synthesis," in Proc. INTERSPEECH, 2016, pp. 874–878.

PUBLICATIONS

30. **E. Song**, H.-G. Kang, "Multi-class learning algorithm for deep neural network-based statistical parametric speech synthesis," in Proc. EUSIPCO, 2016, pp. 1951–1955.
31. **E. Song**, H.-G. Kang, "Deep neural network-based statistical parametric speech synthesis system using improved time-frequency trajectory excitation model," in Proc. INTERSPEECH, 2015, pp. 874–878.
32. K. Byun, **E. Song**, H. Sim, H. Lim, H.-G. Kang, "A constrained two-layer compression technique for ECG waves," in Proc. EMBC, 2015, pp. 6130–6133.
33. **E. Song**, Y. S. Joo, H.-G. Kang, "Improved time-frequency trajectory excitation modeling for a statistical parametric speech synthesis system," in Proc. ICASSP, 2015, pp. 4949–4953.
34. **E. Song**, H.-G. Kang, J. Lee, "Fixed-point implementation of MPEG-D unified speech and audio coding decoder," in Proc. DSP, 2014, pp. 110–113.
35. **E. Song**, J. Ryu, H.-G. Kang, "Speech enhancement for pathological voice using time-frequency trajectory excitation modeling," in Proc. APSIPA, 2013, pp. 1–4.

PREPRINT

1. H. Kim, S. Seo, K. Jeong, O. Kwon, J. Kim, J. Lee, **E. Song**, M. Oh, S. Yoon, K. Yoo, "Unified speech-text pretraining for spoken dialog modeling," arXiv preprint arXiv: 2402.05706, 2024
2. O. Kwon, **E. Song**, J.-M. Kim, H.-G. Kang, "Effective parameter estimation methods for an ExcitNet model in generative text-to-speech systems," arXiv preprint arXiv:1905. 08486, 2019.

PATENTS

1. KR10-2661751, "Method and system for generating speech synthesis model based on selective data augmentation," Apr 2024 (registered).
2. KR10-2626618, "Method and system for synthesizing emotional speech based on emotion prediction," Jan 2024 (registered).
3. KR10-2621842, "Method and system for non-autoregressive speech synthesis," Jan 2024 (registered).
4. KR10-2198598, "Method for generating synthesized speech signal, neural vocoder, and training method thereof," Dec 2020 (registered).
5. KR10-2198597/JP7274184, "Neural vocoder and training method of neural vocoder for constructing speaker-adaptive model," Dec 2020 (registered).
6. KR10-2023-0016624, "Method, computer device, and computer program for pruning self-attention for speaker-adaptive text-to-speech system," Feb 2023 (applied).
7. KR10-2022-0117469, "Method and system for synthesizing speech," Sep 2022 (applied).

8.

HONORS & AWARDS

1. Innovators Under 35 Korea, MIT Technology Review — Dec 2022
2. Ranked No. 2 in N Innovation Award 2020, Naver Corp. — Dec 2020

HONORS & AWARDS

3.	The Best Paper Award, APSIPA ASC 2020	Dec 2020
4.	Ranked No. 1 in N Innovation Award 2019, Naver Corp.	Dec 2019
5.	Ranked No. 1 in N Innovation Award 2018, Naver Corp.	Nov 2018
6.	Excellent intern award, Microsoft Research Asia	Jun 2016
7.	Excellent intern award, Microsoft Research Asia	Feb 2016