Eunwoo Song

Last updated: Oct 2022

CONTACT INFORMATION

Email gregorio.song@gmail.com

Homepage https://sewplay.github.io/

Address 95, Jeongjail-ro, Seongnam, 13561, South Korea

Phone +82-10-3191-9108

EXPERIENCE

Seongnam, Korea

Mar 2017 - present

Naver Corp.

Senior research scientist, DNN TTS team lead, Clova Voice

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 - present
  - Representative voice for Clova Contact Center, May 2019 - present
  - 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

Development of automatic TTS modeling with smartphone recordings

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

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

PUBLICATIONS

1. 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," Proc. NeurIPS, 2022 (in press).
2. **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," Proc. INTERSPEECH, 2022, pp. 1941-1945.
3. 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," Proc. INTERSPEECH, 2022, pp. 4596-4600.
4. 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," Proc. INTERSPEECH, 2022, pp. 3018-3022.
5. M.-J. Hwang, H.-W. Yoon, C.-H. Song, J.-S. Kim, J-.M. Kim, **E. Song**, "Linear prediction-based Parallel WaveGAN speech synthesis," Proc. ICEIC, 2022, pp. 1-4.
6. S. Oh, O. Kwon, M.-J. Hwang, J.-M. Kim, **E. Song**, "Effective data augmentation methods for neural text-to-speech systems," Proc. ICEIC, 2022, pp. 1-4.

7. M.-J, Hwang, R. Yamamoto, **E. Song**, J.-M. Kim, "High-fidelity Parallel WaveGAN with multi-band harmonic-plus-noise model," Proc. INTERSPEECH, 2021, pp. 2227-2231.
8. 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," Proc. INTERSPEECH, 2021. pp. 3595-3599.
9. R. Yamamoto, **E. Song**, M.-J. Hwang, J.-M. Kim, "Parallel waveform synthesis based on generative adversarial networks with voicing-aware conditional discriminators," Proc. ICASSP, 2021, pp. 6039-6043.
10. M.-J. Hwang, R. Yamamoto, **E. Song**, J.-M. Kim," TTS-by-TTS: TTS-driven data augmentation for fast and high-quality speech synthesis," Proc. ICASSP, 2021, pp. 6598-6602.
11. **E. Song**, R. Yamamoto, M.-J. Hwang, J.-S. Kim, O. Kwon, J.-M. Kim, "Improved Parallel WaveGAN with perceptually weighted spectrogram loss," Proc. SLT, 2021, pp. 470-476.
12. M.-J. Hwang, F. K. Soong, **E. Song**, X. Wang, H. Kang, H.-G. Kang, "LP-WaveNet: Linear prediction-based WaveNet speech synthesis," Proc. APSIPA, 2020, pp. 810-814.
13. 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," Proc. APSIPA, 2020, pp. 831-836.
14. **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," Proc. INTERSPEECH, 2020, pp. 3570-3574.
15. **E. Song**, J.-S. Kim, K. Byun, H.-G. Kang, "Speaker-adaptive neural vocoders for parametric speech synthesis systems," Proc. MMSP, 2020, pp. 1-5.
16. R. Yamamoto, **E. Song**, J.-M. Kim, "Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram," Proc. ICASSP, 2020, pp. 6194-6198.
17. 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," Proc. ICASSP, 2020, pp. 7214-7218.
18. R. Yamamoto, **E. Song**, J.-M. Kim, "Probability density distillation with generative adversarial networks for high-quality parallel waveform generation," Proc. INTERSPEECH, 2019, pp. 699-703.
19. **E. Song**, K. Byun, H.-G. Kang, "ExcitNet vocoder: A neural excitation model for parametric speech synthesis systems," Proc. EUSIPCO, 2019, pp. 1179-1183.
20. K. Byun, **E. Song**, J. Kim, J.-M. Kim, H.-G. Kang, "Excitation-by-SampleRNN model for text-to-speech," Proc. ITC-CSCC, 2019, pp. 356-359.

21. 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," Proc. INTERSPEECH, 2018, pp. 917-921.
22. 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," Proc. INTERSPEECH, 2018, pp. 912-916.
23. M.-J. Hwang, **E. Song**, H.-G. Kang, "Modeling-by-generation-structured noise compensation algorithm for glottal vocoding speech synthesis system," Proc. ICASSP, 2018, pp. 5669-5673.
24. **E. Song**, F. K. Soong, H.-G. Kang, "Perceptual quality and modeling accuracy of excitation parameters in DLSTM-based speech synthesis systems," Proc. ASRU, 2017, pp. 671–676.
25. **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.
26. **E. Song**, F. K. Soong, H.-G. Kang, "Improved time-frequency trajectory excitation vocoder for DNN-based speech synthesis," Proc. INTERSPEECH, 2016, pp. 874–878.
27. **E. Song**, H.-G. Kang, "Multi-class learning algorithm for deep neural network-based statistical parametric speech synthesis," Proc. EUSIPCO, 2016, pp. 1951–1955.
28. **E. Song**, H.-G. Kang, "Deep neural network-based statistical parametric speech synthesis system using improved time-frequency trajectory excitation model," Proc. INTERSPEECH, 2015, pp. 874–878.
29. K. Byun, **E. Song**, H. Sim, H. Lim, H.-G. Kang, "A constrained two-layer compression technique for ECG waves," Proc. EMBC, 2015, pp. 6130–6133.
30. **E. Song**, Y. S. Joo, H.-G. Kang, "Improved time-frequency trajectory excitation modeling for a statistical parametric speech synthesis system," Proc. ICASSP, 2015, pp. 4949–4953.
31. **E. Song**, H.-G. Kang, J. Lee, "Fixed-point implementation of MPEG-D unified speech and audio coding decoder," Proc. DSP, 2014, pp. 110–113.
32. **E. Song**, J. Ryu, H.-G. Kang, "Speech enhancement for pathological voice using time-frequency trajectory excitation modeling," Proc. APSIPA, 2013, pp. 1–4.

PREPRINT

1. 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. KR10-2198598, "Method for generating synthesized speech signal, neural vocoder, and training method thereof," Dec 2020 (registered).

2. KR10-2198597, "Neural vocoder and training method of neural vocoder for constructing speaker-adaptive model," Dec 2020 (registered).
3. KR10-2022-0117469, "Method and system for synthesizing speech," Sep 2022 (applied).
4. KR10-2022-0116409, "Method and system for generating speech synthesis model based on selective data augmentation," Sep 2022 (applied).
5. KR10-2022-0047188, "Method and system for synthesizing emotional speech based on emotion prediction," Apr 2022 (applied).
6. KR10-2021-0115859, "Method and system for non-autoregressive speech synthesis," Aug 2021 (applied).

HONORS & AWARDS

1.	Ranked No. 2 in N Innovation Award 2020, Naver Corp.	Dec 2020
2.	The Best Paper Award, APSIPA ASC 2020	Dec 2020
3.	Ranked No. 1 in N Innovation Award 2019, Naver Corp.	Dec 2019
4.	Ranked No. 1 in N Innovation Award 2018, Naver Corp.	Nov 2018
5.	Excellent intern award, Microsoft Research Asia	Jun 2016
6.	Excellent intern award, Microsoft Research Asia	Feb 2016