Eunwoo Song

CONTACT INFORMATION	Naver Corp. 6, Buljeong-ro, Bundang-gu, Gyeonggi-do, 13561, Korea	gregorio.song@gmail.com +82-10-3191-9108

RESEARCH INTERESTS

Speech Signal Processing
Speech Synthesis
Deep Learning

RESEARCH EXPERIENCE

Naver Corp., Seongnam, Korea   Mar 2017 – present

- Senior Research Scientist
- HDTS Team Lead, Clova AI
- Topic: Speech synthesis
  - Research and development of hybrid speech synthesis system, combining deep learning and unit-selection TTS models. Implementing cloud-based real-time TTS products for
    - Naver AI news anchor (Korean Celeb voice, May 2020)
      https://blog.naver.com/clova_ai/221981676372
    - Gatebox (Japanese Character voice, Oct 2019).
      https://gatebox.ai/home
    - Line Car Navi (Japanese Navigation, Sep 2019),
      https://carnavi.line.me
    - Naver Maps (Korean Navigation, Sep 2019),
    - Naver Clova AI speaker (Korean Celeb voice, Nov 2018),
      https://clova.ai/ko/events/celeb_voice

  - Research and development of ExcitNet vocoder, incorporating linear prediction filter to WaveNet architecture for quality improvement.
    https://sewplay.github.io/demos/excitnet

  - Research and development of Parallel WaveGAN vocoder, incorporating generative adversarial network to non-auto-regressive WaveNet generator.
    https://sewplay.github.io/demos/wavegan-pwsl

  - Research and development of end-to-end expressive speech synthesis system, leveraging global style token-based emotion embedding methods.
    https://sewplay.github.io/demos/gst_tacotron2_excitnet

  - Implementing and evaluating state-of-the-art speech synthesis models, such as Tacotron, Tacotron 2, Transformer, WaveNet, WaveRNN, WaveGlow. Experimenting on these models by architectural and feature-level modifications.

  - Implementing and evaluating parametric vocoders for speech synthesis back-end, such as ITFTE, WORLD, STRAIGHT, Glottal Vocoder, HNM, MBE, MELP. Experimenting on these vocoders by architectural modifications for TTS.

Qualcomm Technologies Inc., San Diego, CA   Aug 2016 – Nov 2016

- Intern for Multimedia Group (Mentor: Dr. Deep Sen)
- Topic: Spatial audio
  - Fixed-point implementation of MPEG-H 3D Audio Decoder

Microsoft Research Asia, Beijing, China

Apr 2016 – Jun 2016
Sep 2015 – Feb 2016

- Student Consultant for Speech Group (Mentor: Dr. Frank Soong)
- Topic: Speech synthesis
    - Deep learning-based TTS system using ITFTE vocoder

Yonsei University, Seoul, Korea

Sep 2010 – Feb 2019

- Research Assistant for DSP Lab.

EDUCATION **Yonsei University**, Seoul, Korea

- Combined M.S. and Ph.D., Electrical and Electronic Engineering, Feb 2019
    - Dissertation: Improved time-frequency trajectory excitation vocoder for deep learning-based statistical parametric speech synthesis system
    - Advisor: Prof. Hong-Goo Kang
- B.S., Electrical and Electronic Engineering, Aug 2010

PUBLICATIONS
1. **E. Song**, R. Yamamoto, M.-J. Hwang, J. Kim, O. Kwon, J.-M. Kim, "Improved Parallel WaveGAN with perceptually weighted spectrogram loss," *in Proc. SLT*, 2021 (in press).
2. M.-J. Hwang, F. Soong, **E. Song**, X. Wang, H. Kang, and H.-G. Kang, "LP-WaveNet: Linear prediction-based WaveNet speech synthesis," *in Proc. APSIPA,* 2020 (in press).
3. 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 (in press).
4. **E. Song**, M.-J. Hwang, R. Yamamoto, J. Kim, O. Kwon, J.-M. Kim, "Neural text-to-speech with a modeling-by-generation excitation vocoder," *in Proc. INTERSPEECH*, 2020, pp. 3570-3574.
5. **E. Song**, J. Kim, K. Byun, H.-G. Kang, "Speaker-adapted neural vocoders for parametric speech synthesis systems," *in Proc. MMSP,* 2020, pp.
6. 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.
7. 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.
8. 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.
9. **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.
10. 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.
11. 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.
12. 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.
13. 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.
14. **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.
15. **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.
16. **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.

17. **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.
18. **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.
19. 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.
20. **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.
21. **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.
22. **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. R. Yamamoto, **E. Song**, M.-J. Hwang, J.-M. Kim, "Parallel waveform synthesis based on generative adversarial networks with voicing-aware conditional discriminators," *arXiv preprint arXiv:2010.14151*, 2020.
2. M.-J. Hwang, R. Yamamoto, **E. Song**, J.-M. Kim, "TTS-by-TTS: TTS-driven data augmentation for fast and high-quality speech synthesis," *arXiv preprint arXiv:2010.13421*, 2020.
3. 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.

HONORS & AWARDS

Ranked No. 1 in N Innovation Award 2019, Naver Corp.	Dec 2019
Ranked No. 1 in N Innovation Award 2018, Naver Corp.	Nov 2018
Excellent intern award, Microsoft Research Asia	Jun 2016
Excellent intern award, Microsoft Research Asia	Feb 2016
Full scholarship, Yonsei University	Mar 2006 – Aug 2010

Last updated: 9 Nov 2020