Min-Jae Hwang

Speech Scientist

(+1) 206-376-5502 | mjhwang@meta.com | minjaehwang1993 | Scholar

Education

Combined M.S. and Ph.D degree in Electrical and Electronics *Seoul, Korea*
Yonsei University *Sep. 2015 - Feb. 2020*
- Research topics: Speech synthesis, neural vocoder, and audio watermarking
- Thesis: <LP-WaveNet: Linear Prediction-based WaveNet Speech Synthesis>

B.S. degree in Electrical and Electronics *Seoul, Korea*
Yonsei University *Mar. 2011 - Aug. 2015*
- National Science & Technology Scholarship (2013 – 2015) from Department of Engineering

Work Experience

Postdoctoral Researcher *Seattle, WA, USA*
Seamless team at Meta AI *Oct. 2022 - Present*
- Presently researching expressive speech-to-speech translation system.

Research Scientist *Seongnam, Korea*
Voice & Avatar team at Naver Corporation *May. 2019 - Sep. 2022*
- Primarily researched the high-quality, fast neural vocoding system
- Developed and adopted various neural vocoders including LP-WaveNet and Multiband HN-PWG for various TTS services at Naver
- Developed PyTorch-based TTS toolkit to build high-quality, fast, and controllable GPU TTS system

Research Intern *Beijing, China*
Speech group at Microsoft Research Asia *Jan. 2018 - Nov. 2018*
- Researched the topic of WaveNet vocoders for high-quality TTS system
- Investigated the methodologies to adopt the traditional speech processing approach to the neural vocoding systems

Research Intern *Seongnam, Korea*
Voice team at Naver Corporation *Dec. 2017 - Dec. 2017*
- Researched the topic of glottal vocoder-based parametric TTS system

Honors and Awards

2023   Recognized SeamlessM4T as 100 best inventions of 2023, *TIME Magazine*, USA

2020   2nd place, *N Innovation* in Naver Corporation, Seongnam, Korea

2020   Best Paper Award, *APSIPA* Conference, Auckland, New Zealand

2019   1st place, *N Innovation* in Naver Corporation, Seongnam, Korea

2018   Award of Excellence, Microsoft Research Asia, Beijing, China

Program Committees

2021   Chairman, 2021 Interspeech, Session <Thu-M-V-3 source separation I> *Brno, Czech*

Presentations

- Voice Synthesis and Applications *Seongnam, Korea*
Invited talks at KAIST and SNU *Apr. - May. 2022*

- High-fidelity Parallel WaveGAN with Harmonic-plus-Noise Models *Seongnam, Korea*
2021 Engineering day at Naver Corporation *Jul. 2021*

- Low-cost and High-quality TTS based on TTS-driven Data Augmentation *Seongnam, Korea*
2020 N Innovation award at Naver Corporation *Jan. 2021*

- TTS-driven Data Augmentation for Fast and High-quality Speech Synthesis *Seongnam, Korea*
2020 Engineering day at Naver Corporation *Oct. 2020*

- High-quality DNN-TTS
2019 Engineering day at Naver Corporation

<div align="right">*Seongnam, Korea*
Oct. 2019</div>

- Toward WaveNet Speech Synthesis [Link]
Technical talk at Naver Corporation

<div align="right">*Seongnam, Korea*
Dec. 2018</div>

Publications

[PREPRINTS]
- Seamless: Multilingual Expressive and Streaming Speech Translation
Seamless Communication

<div align="right">*2023 Arxiv*</div>

- SeamlessM4T—Massively Multilingual & Multimodal Machine Translation
Seamless Communication

<div align="right">*2023 Arxiv*</div>

[JOURNAL]
- SVD-based Adaptive QIM Watermarking on Stereo Audio Signals
Min-Jae Hwang, JeeSok Lee, Misuk Lee, and Hong-Goo Kang

<div align="right">*IEEE Transactions on Multimedia*
3.977 impact factor at 2017</div>

[CONFERENCE]
- HierSpeech: Bridging the Gap between Text and Speech by Hierarchical Variational Inference using Self-supervised Representations for Speech Synthesis
Sang-Hoon Lee, Seung-Bin Kim, Ji-Hyun Lee, Eunwoo Song, **Min-Jae Hwang**, and Seong-Whan Lee

<div align="right">*2022 NeurIPS*</div>

- Language Model-Based Emotion Prediction Methods for Emotional Speech Synthesis Systems
Hyunwook Yoon, Ohsung Kwon, Hoyeon Lee, Ryuichi Yamaoto, Eunwo Song, Jae-Min Kim, and **Min-Jae Hwang**

<div align="right">*2022 Interspeech*</div>

- TTS-by-TTS 2: Data-selective Augmentation for Neural Speech Synthesis Using Ranking Support Vector Machine with Variational Autoencoder
Eunwoo Song, Ryuichi Yamamoto, Ohsung Kwon, Chan-Ho Song, **Min-Jae Hwang**, Suhyeon Oh, Hyun-Wook Yoon, Jin-Seob Kim, and Jae-Min Kim

<div align="right">*2022 Interspeech*</div>

- Linear Prediction-based Parallel WaveGAN Speech Synthesis
Min-Jae Hwang, Hyun-Wook Yoon, Chan-Ho Song, Jin-Seob Kim, Jae-Min Kim, and Eunwoo Song

<div align="right">*2022 ICEIC*</div>

- Effective Data Augmentation Methods for Neural Text-to-Speech Systems
Suhyeon Oh, Ohsung Kwon, **Min-Jae Hwang**, Jae-Min Kim, and Eunwoo Song

<div align="right">*2022 ICEIC*</div>

- High-Fidelity Parallel WaveGAN with Multi-Band Harmonic-Plus-Noise Model
Min-Jae Hwang*, Ryuichi Yamamoto*, Eunwoo Song, and Jae-Min Kim (*Equally contributed)

<div align="right">*2021 Interspeech*</div>

- LiteTTS: A Lightweight Mel-Spectrogram-Free Text-to-Speech Synthesizer Based on Generative Adversarial Networks
Huu-Kim Nhuyen, Kihyuk Jeong, Seyun Um, **Min-Jae Hwang**, Eunwoo Song, and Hong-Goo Kang

<div align="right">*2021 Interspeech*</div>

- TTS-by-TTS: TTS-driven Data Augmentation for Fast and High-quality Speech Synthesis
Min-Jae Hwang, Ryuichi Yamamoto, Eunwoo Song, and Jae-Min Kim

<div align="right">*2021 ICASSP*</div>

- Parallel Waveform Synthesis based on Generative Adversarial Networks with Voicing-aware Conditional Discriminators
Ryuichi Yamamoto, Eunwoo Song, **Min-Jae Hwang**, and Jae-Min Kim

<div align="right">*2021 ICASSP*</div>

- ExcitGlow: Improving a WaveGlow-based Neural Vocoder with Linear Prediction Analysis
Suhyeon Oh, Hyungseob Lim, Kyungguen Byun, **Min-Jae Hwang**, Eunwoo Song, and Hong-Goo Kang

<div align="right">*2020 APSIPA*</div>

- LP-WaveNet: Linear prediction-based WaveNet speech synthesis
Min-Jae Hwang, Frank Soong, Eunwoo Song, Xi Wang, Hyeonjoo Kang, and Hong-Goo Kang

<div align="right">*2020 APSIPA*</div>

- Neural Text-to-Speech with a Modeling-by-Generation Excitation Vocoder   *2020 Interspeech*
Eunwoo Song, **Min-Jae Hwang**, Ryuichi Yamamoto, Jin-Seob Kim, Ohsung Kwon, and Jae-Min Kim

- Improving LPCNet-based Text-to-Speech with Linear Prediction-structured Mixture Density Network   *2020 ICASSP*
Min-Jae Hwang, Eunwoo Song, Ryuichi Yamamoto, Frank Soong, and Hong-Goo Kang

- Parameter Enhancement for MELP Speech Codec in Noisy Communication Environment   *2019 Interspeech*
Min-Jae Hwang and Hong-Goo Kang

- A Unified Framework for the Generation of Glottal Signals in Deep Learning-based Parametric Speech Synthesis Systems   *2018 Interspeech*
Min-Jae Hwang, Eunwoo Song, Jinseob Kim, and Hong-Goo Kang

- Modeling-by-Generation-structured Noise Compensation Algorithm for Glottal Vocoding Speech Synthesis System   *2018 ICASSP*
Min-Jae Hwang, Eunwoo Song, Kyunggeun Byung, and Hong-Goo Kang

[WORKSHOP]
- Improved Parallel WaveGAN Vocoder with Perceptually Weighted Spectrogram Loss   *2021 IEEE SLT workshop*
Eunwoo Song, Ryuichi Yamamoto, **Min-Jae Hwang**, Jin-Seob Kim, Ohsung Kwon, and Jae-Min Kim

Patents

- Method and System for Synthesizing Emotional Speech based on Emotion Prediction   *KR 10-2022-0047188*
Hyunwook Yoon, **Min-Jae Hwang**, Ohsung Kwon, Hoyeon Lee, Ryuichi Yamaoto, and Eunwo Song   *Applied*

- Neural Network for Speech Synthesis Based on Selective Self-augmentation Algorithm   *KR 10-2022-0012736*
Ohsung Kwon, Suhyuon Oh, **Min-Jae Hwang**, and Eunwoo Song   *Applied*

- Method and System for Non-autoregressive Speech Synthesis   *KR 10-2021-0115859*
Min-Jae Hwang, Ryuichi Yamamoto, and Eunwoo Song   *Applied*

Additional Information

– **Language** : Korean, English

– **Programming** : Python, Bash, LaTex, Matlab

– **Deep Learning Framework** : PyTorch, Fairseq

– **Coorporation** : Git