Vol.2019-SLP-126 No.5
2019/2/27

# End-to-End Neural TTS and Parallel Wave Generation

Wei Ping[1]

概要：There are two phenomenal trends in speech synthesis research: 1) directly generating waveform samples through state-of-the-art generative models, and 2) building end-to-end TTS systems without too much expert knowledge. In this talk, I will first present our recent work on parallel wave generation, which removes the autoregressive inference bottleneck in WaveNet. I will also compare various non-autoregressive generative models for waveform synthesis. In addition, “end-to-end”speech synthesis actually refers to the text-to-spectrogram models with a separate vocoder in previous studies. I will introduce the first text-to-wave neural architecture for TTS, which is fully convolutional and enables truly end-to-end training from scratch.

## 経歴

Wei Ping obtained his Ph.D. in Computer Science from University of California, Irvine (UCI) in 2016. Before that, he received his B.E. and M.E. degrees from Harbin Institute of Technology and Tsinghua University in 2008 and 2011, respectively. He is currently a Senior Research Scientist at Baidu Research (USA), leading their team on speech synthesis. His research area is machine learning and speech synthesis, with interests spread over deep learning, generative models, graphical models and variational inference.

He has published a series of advanced deep learning papers on text-to-speech synthesis, including Deep Voice 2 (NIPS’17), Deep Voice 3 (ICLR’18), Neural Voice Cloning (NIPS’18) and ClariNet (Submitted to ICLR 2019), which are well-known among speech synthesis researchers.

[1] Baidu Research, USA