





















- [4] Fisher Yu, Vladlen Koltun, “Multi-Scale Context Aggregation by Dilated Convolutions,” 於 *arXiv:1511.07122v3*, 2015.
- [5] Daniel Griffin and Jae Lim, “Signal estimation from modified short-time fourier transform,” 於 *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 2, pp. 236 – 243, 1984.
- [6] Masanori MORISE, Fumiya YOKOMORI, Kenji OZAWA, “WORLD: A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications,” 於 *IEICE Trans. on Information and Systems*, vol. 99, no. 7, pp. 1877 – 1884, 2016.
- [7] Ryan Prenger, Rafael Valle, Bryan Catanzaro, “WaveGlow: A Flow-based Generative Network for Speech Synthesis,” 於 *arXiv:1811.00002v1*, 2018.