DAFTAR PUSTAKA

Ahn, E., & Chodroff, E. (2022). *VoxCommunis Corpus*. https://osf.io/t957v *2011 3rd Computer Science and Electronic Engineering Conference (CEEC) : 13th-14th July 2011, University of Essex, UK*. (2011). IEEE.

Geneva. (2023). *R 128 LOUDNESS NORMALISATION AND PERMITTED MAXIMUM LEVEL OF AUDIO SIGNALS Status: EBU Recommendation*.

Gorodetskii, A., & Ozhiganov, I. (2022). *Zero-Shot Long-Form Voice Cloning with Dynamic Convolution Attention*. http://arxiv.org/abs/2201.10375

Huang, Y. C., Wu, C. H., Chen, Y. Y., Shie, M. G., & Wang, J. F. (2017). Personalized Spontaneous Speech Synthesis Using a Small-Sized Unsegmented Semispontaneous Speech. *IEEE/ACM Transactions on Audio Speech and Language Processing*, *25*(5), 1048–1060. https://doi.org/10.1109/TASLP.2017.2679603

Hochreiter, S., & Urgen Schmidhuber, J. ¨. (1997). *Long Short-Term Memory*.

Huybrechts, G., Merritt, T., Comini, G., Perz, B., Shah, R., & Lorenzo-Trueba, J. (2021). Low-resource expressive text-to-speech using data augmentation. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2021-June*, 6593–6597. https://doi.org/10.1109/ICASSP39728.2021.9413466

Jia, Y., Zhang, Y., Weiss, R. J., Wang, Q., Shen, J., Ren, F., Chen, Z., Nguyen, P., Pang, R., Moreno, I. L., & Wu, Y. (2018). *Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis*. http://arxiv.org/abs/1806.04558

Kalchbrenner, N., Elsen, E., Simonyan, K., Noury, S., Casagrande, N., Lockhart, E., Stimberg, F., Oord, A. van den, Dieleman, S., & Kavukcuoglu, K. (2018). *Efficient Neural Audio Synthesis*. http://arxiv.org/abs/1802.08435

Kaoru Ishikawa. (1968). *Guide to Quality Control*.

Kumar, K., Kumar, R., de Boissiere, T., Gestin, L., Teoh, W. Z., Sotelo, J., de Brebisson, A., Bengio, Y., & Courville, A. (2019). *MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis*. http://arxiv.org/abs/1910.06711

Lestari, D., Koji, I., & Furui, S. (2014). *A large vocabulary continuous speech recognition system for Indonesian language*. https://www.researchgate.net/publication/228829709

Li, J., Meng, Y., Li, C., Wu, Z., Meng, H., Weng, C., & Su, D. (2021). *Enhancing Speaking Styles in Conversational Text-to-Speech Synthesis with Graph-based Multi-modal Context Modeling*. http://arxiv.org/abs/2106.06233

Mahmoodi, D., Soleimani, A., Marvi, H., Razzazi, F., Taghizadeh, M., & Mahmoodi, M. (2011). *2011 3rd Computer Science and Electronic Engineering Conference (CEEC):13th-14th July 2011, University of Essex, UK*. IEEE.

Mustafa, A., Pia, N., & Fuchs, G. (2021). Stylemelgan: An efficient high-fidelity adversarial vocoder with temporal adaptive normalization. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2021-June*, 6034–6038. https://doi.org/10.1109/ICASSP39728.2021.9413605

Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). *WaveNet: A Generative Model for Raw Audio*. http://arxiv.org/abs/1609.03499

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. Dalam *IEEE Transactions on Knowledge and Data Engineering* (Vol. 22, Nomor 10, hlm. 1345–1359). https://doi.org/10.1109/TKDE.2009.191

Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). *Librispeech: An ASR corpus based on public domain audio books*. http://www.gutenberg.org

Wan, L., Wang, Q., Papir, A., & Moreno, I. L. (2017). *Generalized End-to-End Loss for Speaker Verification*. http://arxiv.org/abs/1710.10467

Wang, Y., Skerry-Ryan, R., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Z., Bengio, S., Le, Q., Agiomyrgiannakis, Y., Clark, R., & Saurous, R. A. (2017). *Tacotron: Towards End-to-End Speech Synthesis*. http://arxiv.org/abs/1703.10135

Yamamoto, R., Song, E., & Kim, J.-M. (2019). *Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram*. http://arxiv.org/abs/1910.11480

Yasuda, Y., Wang, X., Takaki, S., & Yamagishi, J. (2018). *Investigation of enhanced Tacotron text-to-speech synthesis systems with self-attention for pitch accent language*. http://arxiv.org/abs/1810.11960

Zhang, J.-X., Ling, Z.-H., & Dai, L.-R. (2018). *Forward Attention in Sequence-to-sequence Acoustic Modelling for Speech Synthesis*. https://doi.org/10.1109/ICASSP.2018.8462020