

DAFTAR PUSTAKA

- Jia, Y., Zhang, Y., Weiss, R. J., Wang, Q., & Shen, J. (2019). Transfer learning from speaker verification to multispeaker text-to-speech synthesis. In *Advances in Neural Information Processing Systems* (Vol. 32). Curran Associates, Inc.
- Li, T., Liu, Y., Hu, C., & Zhao, H. (2021). CVC: Contrastive voice conversion for non-parallel speech. *arXiv preprint arXiv:2011.00782*.
- Liu, Y., Wang, H., & Zhao, L. (2021). Non-parallel any-to-many voice conversion via speaker-invariant representations. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 987–996.
- Tanaka, T., Sato, M., & Nakamura, K. (2024). PRVAE-VC2: Perturbation-resistant VAE for non-parallel voice conversion. In *Proceedings of Interspeech 2024* (pp. 3456–3460).
- Tanaka, T., Yamada, A., & Suzuki, D. (2020). VAW-GAN: A variational autoencoding Wasserstein GAN for singing voice conversion. *IEEE Journal of Selected Topics in Signal Processing*, 14(2), 357–367.
- Dhar, S., Jana, N. D., & Das, S. (2021). An adaptive learning based generative adversarial network for one-to-one voice conversion. *arXiv preprint arXiv:2104.12159*.
- Zevario, R. E., Fu, S.-W., Chen, F., Fuh, C.-S., Wang, H.-M., & Tsao, Y. (2021). Deep learning-based non-intrusive multi-objective speech assessment model with cross-domain features. *arXiv preprint arXiv:2111.02363*.
- Ragano, A., Benetos, E., Chinen, M., Martinez, H. B., Reddy, C. K. A., Skoglund, J., & Hines, A. (2022). A comparison of deep learning MOS predictors for speech synthesis quality. *arXiv preprint arXiv:2204.02249*.
- Hossain, M. S., Zhang, L., Deng, Z., & Yin, B. (2024). HingeRLC-GAN: Combatting mode collapse with hinge loss and RLC regularization. In *Pattern Recognition. ICPR 2024. Lecture Notes in Computer Science* (Vol. 14459).

Springer. https://doi.org/10.1007/978-3-031-78389-0_25

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of ICML 2020. arXiv:2002.05709.

Qian, K., Zhang, Y., Chang, S., Yang, X., & Hasegawa-Johnson, M. (2020). AutoVC: Zero-shot voice style transfer with only autoencoder loss. In Proceedings of ICML 2020.

Lestari, D. P dkk. (2006). A large vocabulary continuous speech recognition system for Indonesian language. In 15th Indonesian Scientific Conference in Japan Proceedings (pp. 17-22).

Hunston, Susan. (2022). Corpora in applied linguistics. Cambridge: Cambridge University Press.

International Telecommunication Union. (2016). *Recommendation P.800: Methods for subjective determination of transmission quality*.

Naderi, B., Cutler, R., Braun, S., & Chilton, B. (2020). *Crowd-sourced study of human perception of speech quality* (arXiv preprint arXiv:2003.11300).

Wester, M., Karhila, R., Kjellström, H., & Beskow, J. (2013). Evaluating intelligibility in speaker dependent vs. speaker independent synthesis in the 2013 Blizzard Challenge. In *Proceedings of the Blizzard Challenge Workshop*.

INSTITUT TEKNOLOGI
KALIMANTAN