# WEN-CHIN HUANG

wen.chinhuang@g.sp.m.is.nagoya-u.ac.jp ⋄ +81-52-789-3809
https://unilight.github.io/

## RESEARCH INTERESTS

Speech processing, mainly voice conversion, speech synthesis and speech quality assessment.

## EDUCATION

**Nagoya University** Apr. 2021 - Mar. 2024
Ph.D. (Informatics), Graduate School of Informatics
*Advisor*: Prof. Tomoki Toda

**Nagoya University** Apr. 2019 - Mar. 2021
M.S., Graduate School of Informatics
*Advisor*: Prof. Tomoki Toda

**National Taiwan University** Sep. 2014 - Jun. 2018
B.S. in Computer Science & Information Science

## PROFESSIONAL EXPERIENCES

**Nagoya University, Japan** Apr. 2024 - Now
Assistant Professor, Graduate School of Informatics.

**Institute of Information Science, Academia Sinica, Taiwan** Jul. 2017 - Mar. 2019, Aug. 2020 - Mar. 2021
*Research Assistant*. Supervisor: Dr. Hsin-Min Wang, Dr. Yu Tsao

## ACADEMIC ACTIVITIES

**Organizer/Committee Member**

1. Special session organizing committee, "The VoiceMOS Challenge 2024" at SLT 2024
2. Special session organizing committee, "The VoiceMOS Challenge 2023" at ASRU 2023
3. Special session organizing committee, "The Singing Voice Conversion Challenge 2023" at ASRU 2023
4. Special session organizing committee, "The VoiceMOS Challenge" at Interspeech 2022
5. Organizing Committee, Voice Conversion Challenge 2020

**Reviewer**

1. IEEE Transactions on Audio, Speech and Language Processing (2020-2024)
2. Speech Communication (2023)
3. IEEE Speech Processing Letters (2020-2022)
4. INTERSPEECH (2024)
5. IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) (2023)
6. IEEE Spoken Language Technology Workshop (SLT) (2022)

## HONORS

- **Outstanding Graduate Student Award**, from Nagoya University, Japan, June 2023
- **16th Student Journal Paper Award**, IEEE SPS Japan, Dec. 2022
- **Best Paper Award**, The 13th Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2021
- **Research Fellowship for Young Scientists (DC1)**, from Japan Society for the Promotion of Science (JSPS), April 2021 - March 2024

· **Scholarship for International Students**, JEES Docomo, April 2019 - March 2021

· **Best Student Paper Award**, The 11th International Symposium on Chinese Spoken Language Processing (ISCSLP), 2018

## PUBLICATIONS

### Journals

1. W.-C. Huang, S.-W. Yang, T. Hayashi, and T. Toda, "A Comparative Study of Self-Supervised Speech Representation Based Voice Conversion," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1308–1318, 2022
2. W.-C. Huang, T. Hayashi, Y. C. Wu, H. Kameoka, and T. Toda, "Pretraining techniques for sequence-to-sequence voice conversion," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 745–755, 2021
3. H. Kameoka, W.-C. Huang, K. Tanaka, T. Kaneko, N. Hojo, and T. Toda, "Many-to-many voice transformer network," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 656–670, 2021
4. W.-C. Huang, H. Luo, H.-T. Hwang, C.-C. Lo, Y.-H. Peng, Y. Tsao, and H.-M. Wang, "Unsupervised Representation Disentanglement Using Cross Domain Features and Adversarial Learning in Variational Autoencoder Based Voice Conversion," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 4, pp. 468–479, 2020
5. X. Wang, J. Yamagishi, M. Todisco, H. Delgado, A. Nautsch, N. Evans, M. Sahidullah, V. Vestman, T. Kinnunen, K. A. Lee, L. Juvela, P. Alku, Y.-H. Peng, H.-T. Hwang, Y. Tsao, H.-M. Wang, S. L. Maguer, M. Becker, F. Henderson, R. Clark, Y. Zhang, Q. Wang, Y. Jia, K. Onuma, K. Mushika, T. Kaneda, Y. Jiang, L.-J. Liu, Y.-C. Wu, W.-C. Huang, T. Toda, K. Tanaka, H. Kameoka, I. Steiner, D. Matrouf, J.-F. Bonastre, A. Govender, S. Ronanki, J.-X. Zhang, and Z.-H. Ling, "Asvspoof 2019: a large-scale public database of synthetized, converted and replayed speech," *Computer Speech & Language*, vol. 64, p. 101114, 2020

### Peer-reviewed Conferences and Workshops

1. W.-C. Huang, L. Violeta, S. Liu, J. Shi, and T. Toda, "Evaluating Methods for Ground-Truth-Free Foreign Accent Conversion," in *Proc. ASRU*, 2023
2. W.-C. Huang, E. Cooper, Y. Tsao, H.-M. Wang, T. Toda, and J. Yamagishi, "The VoiceMOS Challenge 2023: Zero-shot Subjective Speech Quality Prediction for Multiple Domains," in *Proc. ASRU*, 2023
3. R. Yamamoto, R. Yoneyama, L. Violeta, W.-C. Huang, and T. Toda, "A comparative study of voice conversion models with large-scale speech and singing data: the T13 systems for the Singing Voice Conversion Challenge 2023," in *Proc. ASRU*, 2023
4. B. Halpern, W.-C. Huang, L. Violeta, R. van Son, and T. Toda, "Improving Severity Preservation of Healthy-to-pathological Voice Conversion with Global Style Tokens," in *Proc. ASRU*, 2023
5. W.-C. Huang and T. Toda, "Evaluating Methods for Ground-Truth-Free Foreign Accent Conversion," in *Proc. APSIPA ASC*, 2023
6. W.-C. Huang, B. Peloquin, J. Kao, C. Wang, H. Gong, E. Salesky, Y. Adi, A. Lee, and P.-J. Chen, "A Holistic Cascade System, Benchmark, and Human Evaluation Protocol for Expressive Speech-to-Speech Translation," in *Proc. ICASSP*, 2023, pp. 1–5
7. L. P. Violeta, D. Ma, W.-C. Huang, and T. Toda, "Intermediate Fine-Tuning Using Imperfect Synthetic Speech for Improving Electrolaryngeal Speech Recognition," in *Proc. ICASSP*, 2023, pp. 1–5
8. W.-C. Huang, D. Markovic, A. Richard, I. D. Gebru, and A. Menon, "End-to-End Binaural Speech Synthesis," in *Proc. Interspeech*, 2022
9. W.-C. Huang, E. Cooper, Y. Tsao, H.-M. Wang, J. Yamagishi, and T. Toda, "The VoiceMOS Challenge 2022," in *Proc. Interspeech*, 2022
10. L. P. Violeta, W.-C. Huang, and T. Toda, "Investigating Self-supervised Pretraining Frameworks for Pathological Speech Recognition," in *Proc. Interspeech*, 2022
11. H.-S. Tsai, H.-J. Chang, W.-C. Huang, Z. Huang, K. Lakhotia, S.-w. Yang, S. Dong, A. Liu, C.-I. Lai, J. Shi *et al.*, "SUPERB-SG: Enhanced Speech processing Universal PERformance Benchmark for Semantic and Generative Capabilities," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 8479–8492
12. W.-C. Huang, S.-W. Yang, T. Hayashi, H.-Y. Lee, S. Watanabe, and T. Toda, "S3PRL-VC: Open-source Voice Conversion Framework with Self-supervised Speech Representations," in *Proc. ICASSP*, 2022

13. W.-C. Huang, E. Cooper, J. Yamagishi, and T. Toda, "LDNet: Unified Listener Dependent Modeling in MOS Prediction for Synthetic Speech," in *Proc. ICASSP*, 2022
14. W.-C. Huang, B. M. Halpern, L. P. Violeta, O. Scharenborg, and T. Toda, "Towards Identity Preserving Normal to Dysarthric Voice Conversion," in *Proc. ICASSP*, 2022
15. E. Cooper, W.-C. Huang, T. Toda, and J. Yamagishi, "Generalization ability of MOS prediction networks," in *Proc. ICASSP*, 2022
16. C. Xie, Y.-C. Wu, P. L. Tobing, W.-C. Huang, and T. Toda, "Direct Noisy Speech Modeling for Noisy-to-noisy Voice Conversion," in *Proc. ICASSP*, 2022
17. W.-C. Huang, T. Hayashi, X. Li, S. Watanabe, and T. Toda, "On Prosody Modeling for ASR+TTS based Voice Conversion," in *Proc. ASRU*, 2021
18. M.-C. Yen, W.-C. Huang, K. Kobayashi, Y.-H. Peng, S.-W. Tsai, Y. Tsao, T. Toda, J.-S. Jang, and H.-M. Wang, "Mandarin Electrolaryngeal Speech Voice Conversion with Sequence-to-Sequence Modeling," in *Proc. ASRU*, 2021, pp. 650–657
19. Y.-S. Liou, W.-C. Huang, M.-C. Yen, S.-W. Tsai, Y.-H. Peng, T. Toda, Y. Tsao, and H.-M. Wang, "Time Alignment using Lip Images for Frame-based Electrolaryngeal Voice Conversion," in *Proc. APSIPA ASC*, 2021
20. C. Xie, Y.-C. Wu, P. L. Tobing, W.-C. Huang, and T. Toda, "Noisy-to-noisy Voice Conversion Framework with Denoising Model," in *Proc. APSIPA ASC*, 2021
21. D. Ma, W.-C. Huang, and T. Toda, "Investigation of Text-to-speech-based Synthetic Parallel Data for Sequence-to-sequence Non-parallel Voice Conversion," in *Proc. APSIPA ASC*, 2021
22. W.-C. Huang, K. Kobayashi, Y.-H. Peng, C.-F. Liu, Y. Tsao, H.-M. Wang, and T. Toda, "A Preliminary Study of a Two-Stage Paradigm for Preserving Speaker Identity in Dysarthric Voice Conversion," in *Proc. Interspeech*, 2021, pp. 1329–1333
23. Y.-C. Wu, C.-H. Hu, H.-S. Lee, Y.-H. Peng, W.-C. Huang, Y. Tsao, H.-M. Wang, and T. Toda, "Relational Data Selection for Data Augmentation of Speaker-dependent Multi-band MelGAN Vocoder," in *Proc. Interspeech*, 2021, pp. 3630–3634
24. W.-C. Huang, C.-H. Wu, S.-B. Luo, K.-Y. Chen, H.-M. Wang, and T. Toda, "Speech recognition by simply fine-tuning bert," in *Proc. ICASSP*, 2021, pp. 7343–7347
25. W.-C. Huang, T. Hayashi, Y.-C. Wu, and T. Toda, "Any-to-One Sequence-to-Sequence Voice Conversion using Self-Supervised Discrete Speech Representations," in *Proc. ICASSP*, 2021, pp. 5944–5948
26. K. Kobayashi, W.-C. Huang, Y.-C. Wu, P. L. Tobing, T. Hayashi, and T. Toda, "CRANK: an Open-Source Software for Nonparallel Voice Conversion based on Vetor-Quantized Variational Autoencoder," in *Proc. ICASSP*, 2021, pp. 5934–5938
27. T. Hayashi, W.-C. Huang, K. Kobayashi, and T. Toda, "Non-autoregressive sequence-to-sequence voice conversion," in *Proc. ICASSP*, 2021, pp. 7068–7072
28. Y.-W. Chen, K.-H. Hung, S.-Y. Chuang, J. Sherman, W.-C. Huang, X. Lu, and Y. Tsao, "Ema2s: An end-to-end multimodal articulatory-to-speech system," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021, pp. 1–5
29. Z. Yi, W.-C. Huang, X. Tian, J. Yamagishi, R. K. Das, T. Kinnunen, Z.-H. Ling, and T. Toda, "Voice Conversion Challenge 2020 – Intra-lingual semi-parallel and cross-lingual voice conversion –," in *Proc. Joint Workshop for the Blizzard Challenge and Voice Conversion Challenge 2020*, 2020, pp. 80–98
30. R. K. Das, T. Kinnunen, W.-C. Huang, Z.-H. Ling, J. Yamagishi, Z. Yi, X. Tian, and T. Toda, "Predictions of Subjective Ratings and Spoofing Assessments of Voice Conversion Challenge 2020 Submissions," in *Proc. Joint Workshop for the Blizzard Challenge and Voice Conversion Challenge 2020*, 2020, pp. 99–120
31. W.-C. Huang, T. Hayashi, S. Watanabe, and T. Toda, "The Sequence-to-Sequence Baseline for the Voice Conversion Challenge 2020: Cascading ASR and TTS," in *Proc. Joint Workshop for the Blizzard Challenge and Voice Conversion Challenge 2020*, 2020, pp. 160–164
32. W.-C. Huang, P. L. Tobing, Y.-C. Wu, K. Kobayashi, and T. Toda, "The NU Voice Conversion System for the Voice Conversion Challenge 2020: On the Effectiveness of Sequence-to-sequence Models and Autoregressive Neural Vocoders," in *Proc. Joint Workshop for the Blizzard Challenge and Voice Conversion Challenge 2020*, 2020, pp. 165–169
33. W.-C. Huang, T. Hayashi, Y.-C. Wu, H. Kameoka, and T. Toda, "Voice Transformer Network: Sequence-to-Sequence Voice Conversion Using Transformer with Text-to-Speech Pretraining," in *Proc. Interspeech*, 2020, pp. 4676–4680
34. W.-C. Huang, Y.-C. Wu, K. Kobayashi, Y.-H. Peng, H.-T. Hwang, P. Lumban Tobing, T. Toda, Y. Tsao, and H.-M. Wang, "Generalization of Spectrum Differential based Direct Waveform Modification for Voice Conversion," in *Proc. 10th ISCA Speech Synthesis Workshop*, 2019, pp. 57–62

35. W.-C. Huang, Y.-C. Wu, C.-C. Lo, P. Lumban Tobing, T. Hayashi, K. Kobayashi, T. Toda, Y. Tsao, and H.-M. Wang, "Investigation of F0 Conditioning and Fully Convolutional Networks in Variational Autoencoder Based Voice Conversion," in *Proc. Interspeech*, 2019, pp. 709–713
36. W.-C. Huang, Y.-C. Wu, H.-T. Hwang, P. Lumban Tobing, T. Hayashi, K. Kobayashi, T. Toda, Y. Tsao, and H.-M. Wang, "Refined WaveNet Vocoder for Variational Autoencoder Based Voice Conversion," in *Proc. 27th European Signal Processing Conference (EUSIPCO)*, Sep 2019
37. W.-C. Huang, H.-T. Hwang, Y.-H. Peng, Y. Tsao, and H.-M. Wang, "Voice conversion based on cross-domain features using variational auto encoders," in *Proc. The 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Nov 2018
38. C.-C. Lo, S.-W. Fu, W.-C. Huang, X. Wang, J. Yamagishi, Y. Tsao, and H.-M. Wang, "MOSNet: Deep Learning based Objective Assessment for Voice Conversion," in *Proc. Interspeech*, 2019, pp. 1541–1545

**Domestic Conferences**

1. W.-C. Huang, Y.-C. Wu, H.-T. Hwang, P. Lumban Tobing, T. Hayashi, K. Kobayashi, T. Toda, Y. Tsao, and H.-M. Wang, "Reducing mismatch of WaveNet vocoder for variational autoencoder based voice conversion," in *ASJ*, Mar 2019
2. W.-C. Huang, C.-C. Lo, H.-T. Hwang, Y. Tsao, and H.-M. Wang, "Wavenet vocoder and its applications in voice conversion," in *Proc. The 30th ROCLING Conference on Computational Linguistics and Speech Processing (ROCLING)*, Oct 2018

## MISCELLANEOUS WORKS

**Invited/visiting talks**

1. Erica Cooper, W.-C. Huang "From human ears to deep neural networks: automatic evaluation of synthetic speech and audio data," Joint Workshop VoicePersonae and ASVspoof, Tokyo, Nov. 2023
2. W.-C. Huang Erica Cooper, "The VoiceMOS Challenge 2022", IPSJ-SLP/SP (Japanese domestic conference), Mar. 2022

**Open-source software development**

1. ESPnet: End-to-end speech processing toolkit.
2. S3PRL: Self-Supervised Speech Pre-training and Representation Learning toolkit.

## SKILLS

| | |
|---|---|
| **Languages** | Mandarin: native<br>English: fluent (TOEIC 960, TOEFL 108)<br>Japanese: intermediate (JLPT N2) |
| **Programming languages** | Python, shell script, Matlab, C++ |
| **Deep learning frameworks** | PyTorch, TensorFlow |

Last updated: 2024/4/1