WEN-CHIN HUANG

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

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

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

ACADEMIC ACTIVITIES

Organizer/Committee Member

1. Organizing Committee, The Singing Voice Conversion Challenge 2025
2. Challenge session organizing committee, "The AudioMOS Challenge 2025" at ASRU 2024
3. Special session organizing committee, "The VoiceMOS Challenge 2024" at SLT 2024
4. Special session organizing committee, "The VoiceMOS Challenge 2023" at ASRU 2023
5. Special session organizing committee, "The Singing Voice Conversion Challenge 2023" at ASRU 2023
6. Special session organizing committee, "The VoiceMOS Challenge" at Interspeech 2022
7. Organizing Committee, Voice Conversion Challenge 2020

Session Chairs

1. Vice chair of Speech A/B, Acoustic Society Japan research meeting, Sep 2024 - Mar. 2025

Reviewer

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

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

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

FUNDINGS

· **Grant-in-Aid for Scientific Research (B) (Co-I)**, Title: Universal, explainable and extensible automatic evaluation of synthesized speech, PI: Dr. Erica Cooper (NICT, Japan), total: 1,440,000 yen, Japan Society for the Promotion of Science, Apr, 2025 - Mar, 2029

· **Audiobox Responsible Generation Grant**, $50,000 USD Unrestricted Research Gift from Meta, received: Nov. 2024

· **Google Research Grant**, Unrestricted Research Gift from Google, received: Sep. 2024

· **Research Fellowship for Young Scientists (DC1)**, from Japan Society for the Promotion of Science (JSPS), April 2021 - March 2024

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

· **INTERSPEECH 2019 Travel grant**, from ISCA, Sep. 2019

INVITED/VISITING TALKS

· W.-C. Huang, Erica Cooper, Jiatong Shi, "Automatic Quality Assessment for Speech and Beyond", Tutorial session, INTERSPEECH, Rotterdam, The Netherlands, Aug. 2025

· W.-C. Huang, "Automatic Quality Assessment for Speech and Beyond", Conversational AI Reading Group, MILA, online, May 2025

· W.-C. Huang, "Progress and Visions in Deep Learning based Voice Conversion", IPSJ-SLP/SP (Japanese domestic conference), Oct. 2024

· W.-C. Huang, "Fundamentals, Prospectives and Challenges in Deep-learning based Voice Conversion," Research Center for Information Technology Innovation (CITI), Academia Sinica, Aug. 2024

· 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

· W.-C. Huang, Erica Cooper, "The VoiceMOS Challenge 2022", IPSJ-SLP/SP (Japanese domestic conference), Mar. 2022

COURSES

· Computer Science Experiments (a), School of Informatics, Nagoya University, 2025

· Programming 2, School of Informatics, Nagoya University, 2024

· Probability and Statistics, School of Informatics, Nagoya University, 2024

PUBLICATIONS

Journals/Letters

1. W.-C. Huang, Y.-C. Wu, and T. Toda, "Multi-speaker text-to-speech training with speaker anonymized data," *IEEE Signal Processing Letters*, vol. 31, pp. 2995–2999, 2024
2. L. P. Violeta, D. Ma, W.-C. Huang, and T. Toda, "Pretraining and Adaptation Techniques for Electrolaryngeal Speech Recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 2777–2789, 2024
3. S.-W. Yang, H.-J. Chang, Z. Huang, A. T. Liu, C.-I. Lai, H. Wu, J. Shi, X. Chang, H.-S. Tsai, W.-C. Huang, T.-H. Feng, P.-H. Chi, Y. Y. Lin, Y.-S. Chuang, T.-H. Huang, W.-C. Tseng, K. Lakhotia, S.-W. Li, A. Mohamed, S. Watanabe, and H.-Y. Lee, "A large-scale evaluation of speech foundation models," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 2884–2899, 2024
4. "A review on subjective and objective evaluation of synthetic speech," *Acoustical Science and Technology*, vol. 45, no. 4, p. 161
5. 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

6. 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
7. 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
8. 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
9. 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. K. Nishizawa, R. Yamamoto, W.-C. Huang, and T. Toda, "Investigating factors related to the naturalness of synthesized unison singing," in *Proc. ICASSP*, 2025, pp. 1–5
2. W.-C. Huang, S.-W. Fu, E. Cooper, R. E. Zezario, T. Toda, H.-M. Wang, J. Yamagishi, and Y. Tsao, "The VoiceMOS Challenge 2024: Beyond Speech Quality Prediction," in *Proc. ASRU*, 2024
3. B. M. Halpern, T. Tienkamp, W.-C. Huang, L. P. Violeta, T. Rebernik, S. de Visscher, M. Witjes, M. Wieling, D. Abur, and T. Toda, "Quantifying the Effect of Speech Pathology on Automatic and Human Speaker Verification," in *Proc. Interspeech*, 2024, pp. 3015–3019
4. L. P. Violeta, W.-C. Huang, D. Ma, R. Yamamoto, K. Kobayashi, and T. Toda, "Electrolaryngeal Speech Intelligibility Enhancement through Robust Linguistic Encoders," in *Proc. ICASSP*, 2024, pp. 10 961–10 965
5. W.-C. Huang, L. Violeta, S. Liu, J. Shi, and T. Toda, "The Singing Voice Conversion Challenge 2023," in *Proc. ASRU*, 2023
6. 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
7. 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
8. 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
9. W.-C. Huang and T. Toda, "Evaluating Methods for Ground-Truth-Free Foreign Accent Conversion," in *Proc. APSIPA ASC*, 2023
10. 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
11. 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
12. W.-C. Huang, D. Markovic, A. Richard, I. D. Gebru, and A. Menon, "End-to-End Binaural Speech Synthesis," in *Proc. Interspeech*, 2022
13. W.-C. Huang, E. Cooper, Y. Tsao, H.-M. Wang, J. Yamagishi, and T. Toda, "The VoiceMOS Challenge 2022," in *Proc. Interspeech*, 2022
14. L. P. Violeta, W.-C. Huang, and T. Toda, "Investigating Self-supervised Pretraining Frameworks for Pathological Speech Recognition," in *Proc. Interspeech*, 2022
15. 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
16. "S3PRL-VC: A Voice Conversion Toolkit based on S3PRL," `https://github.com/unilight/s3prl-vc`, 2023
17. 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

18. 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
19. E. Cooper, W.-C. Huang, T. Toda, and J. Yamagishi, "Generalization ability of MOS prediction networks," in *Proc. ICASSP*, 2022
20. 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
21. 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
22. 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
23. 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
24. 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
25. 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
26. 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
27. 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
28. 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
29. 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
30. 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
31. T. Hayashi, W.-C. Huang, K. Kobayashi, and T. Toda, "Non-autoregressive sequence-to-sequence voice conversion," in *Proc. ICASSP*, 2021, pp. 7068–7072
32. 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
33. 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
34. 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
35. 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
36. 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
37. 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
38. 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

39. 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
40. 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
41. 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
42. 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

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

Open-source software development

- "SHEET: Speech Human Evaluation Estimation Toolkit," `https://github.com/unilight/sheet`, 2024
- "S3PRL-VC: A Voice Conversion Toolkit based on S3PRL," `https://github.com/unilight/s3prl-vc`, 2023
- "seq2seq-vc: A Sequence-to-sequence Voice Conversion Toolkit." `https://github.com/unilight/seq2seq-vc`, 2023

Skills

- **Languages**: Mandarin (native), English (fluent; TOEIC 960, TOEFL 108), Japanese (business; JLPT N1)
- **Programming languages**: Python, shell script, C++
- **Deep learning frameworks**: PyTorch, TensorFlow, JAX

Last updated: 2025/04/26