Curriculum Vitae

Yu Tsao

Research Fellow (Professor)
Bio-Acoustic Signal Processing (Bio-ASP) Lab: http://bio-asplab.citi.sinica.edu.tw/
Research Center for Information Technology Innovation, Academia Sinica
128 Academia Rd., Sec. 2, Nankang District, Taipei 115, Taiwan
TEL: +886-2-2787-2390
Email: yu.tsao@citi.sinica.edu.tw
Website: http://www.citi.sinica.edu.tw/pages/yu.tsao/
Google Scholar: https://scholar.google.com/citations?hl=zh-TW&user=ZO5e5I4AAAAJ
Google Scholar Citation: 15262 (H-index: 58) until 2026/04/08

RESEARCH INTERESTS

- Speech Enhancement and Voice Conversion
- Assistive Oral Communication Technologies
- Biomedical Acoustic Signal Processing
- Deep Learning Algorithm Development

EDUCATION

GEORGIA INSTITUTE OF TECHNOLOGY, Atlanta, Georgia
Ph.D. in Electrical and Computer Engineering Aug. 2003-Dec. 2008
- Research Topic: Robust Speech Recognition, under advisor: Dr. Chin-Hui Lee
- Leadership Activities: President, Taiwanese Student Association (TSA): represented the school; assisted students and visiting scholars from Taiwan.

NATIONAL TAIWAN UNIVERSITY, Taipei, Taiwan
Master of Science in Electrical Engineering Sept. 1999-June 2001
Bachelor of Science in Electrical Engineering Sept. 1995-June 1999
- Graduate Research Topic: Rapid Speaker Adaptation, under advisor Dr. Lin-Shan Lee
- Leadership Activities: Chairman, Public Relations Department of the Student Association: organized events to cultivate strong relationships among members.

PROFESSIONAL EXPERIENCE

RESEARCH CENTER FOR INFORMATION TECHNOLOGY INNOVATION, ACADEMIA SINICA, Taipei, Taiwan
Researcher Fellow (Professor)/Deputy Director 2020/08–Present
Associate Researcher Fellow (Associate Professor) 2016/05–2020/08
Assistant Researcher Fellow (Assistant Professor) 2011/11–2016/04
- Research and develop speech signal processing algorithms.
- Develop devices of assistive oral communication technology.
- Derive acoustic signal processing for biomedical applications.
- Derive novel machine-learning algorithms for acoustic signal processing.

NATIONAL INSTITUTE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY, Kyoto, Japan
Expert Researcher April 2009-Sept. 2011
- Developed research to handle non-native accent issues for automatic speech recognition applications.
- Developed digital signal processing-based solutions to improve the performance of speech recognizers under real-world adverse conditions.
- Contributed to developing VoiceTra multilingual speech-to-speech translation application on iPhone.
- Carried out projects of field tests and dissemination of spoken dialog interface technologies.

HONORS

- **IEEE Consumer Technology Society, Chester Sall Awards (**corresponding author**) (2025)**
- **IEEE Signal Processing Society (SPS) Young Author Best Paper Award (**corresponding author**) (2021)**
- **Cough Diagnostic Algorithm for Tuberculosis DREAM Challenge-SubChallenge 2 First Place (2023)**
- **Clarity Machine Learning Challenges for Hearing Aids Challenge Gold Prize (2022)**
- Outstanding Elite Award, Chung Hwa Rotary Educational Foundation 2019-2020 (2019–present)
- Y. Z. Hsu Scientific Paper Award, Taiwan (2025)
- Outstanding Research Award, NSTC, Taiwan (2023)
- Distinguished Lecture Award, APSIPA (2018)
- Best Student Paper Award, ISCSLP 2018 (2018)

PROFESSIONAL ACTIVITIES

Invited Talks

- Survey Talk, Interspeech 2025, Title: Assistive Oral Communication Technology (2025/08)
- Survey Talk, Interspeech 2024, Title: Neural Speech Assessment (2024/08)
- Keynote Speech, AVSEC-3 Workshop (Satelite Event of Interspeech 2024), Title: Enhancing Oral Communication with AI Powered Assistive Technologies (2024/08)
- Keynote Speech, IEEE ICTS 2021, Title: Deep-learning-based Speech Enhancement with Its Application to Assistive Oral Communications Devices (2021/10).
- Invited Lecture APSIPA Japan Chapter, Title: Deep-learning-based Speech Enhancement with Its Application to Assistive Oral Communications Devices (2021/03).
- Keynote Speech in M3Oriental Workshop, ACM Multimedia Asia, Title: Utilizing Deep Learning for Speech Enhancement in Assistive Oral Communication Technologies, (12/2023).
- CTSoc Technical Talk, Title: Wearable Devices and Machine Learning Algorithms for Augmented Oral Communication Assistance, (2023/11).

Challenge Organizer

- Voice Detection Challenge in IEEE Big Data 2018. The challenge has attracted the participation of **109 teams from 27 different countries**.
- VoiceMOS Challenge, a special session in Interspeech 2022.
- The VoiceMOS Challenge 2023: Zero-shot Subjective Speech Quality Prediction for Multiple Domains, a special session in ASRU 2023.
- The VoiceMOS Challenge 2024: Beyond Speech Quality Prediction, a special session in SLT 2024.

International Services

- Chair of the IEEE CTSoc Audio/Video Systems and Signal Processing Technical Committee (AVS TC) (2025-present)
- Chair, Speech, Language, and Audio (SLA) Technical Committee, APSIPA (2020-2022)
- Distinguished Lecturer, APSIPA (2019-2021)
- Member, IEEE Speech and Language Processing Technical Committee Members (2021-present)

Editorship

- Senir Associate Editor, IEEE Signal Processing Letters (2025-present)
- Associate Editor, IEEE Transactions on Consumer Electronics, (2025/01–present)
- Associate Editor, APSIPA Transactions on Signal and Information Processing (2022- present)
- Associate Editor, IEEE Signal Processing Letters (2020-2025)
- Associate Editor, IEEE/ACM Transactions on Audio, Speech, and Language Processing (2019-2025)
- Co-lead Guest Editor of the Special Issue "Deep Multimodal Speech Enhancement and Separation", IEEE Journal of Selected Topics in Signal Processing, IEEE (2025/05)
- Guest Editor of the journal Biomedical Signal Processing and Control (Title: Advances in biomedical signal processing for communication disorders)

- Guest Editor of the Themed Series of APSIPA Transactions on Signal and Information Processing (Title: Advanced Acoustic, Sound and Audio Processing Techniques and Their Applications)

Conference Chairs

- General Chair, IEEE SLT 2026, Italy (2026/12)
- Tutorial Chair, IEEE GEM 2024, Italy (2024/06)
- General Chair, IEEE ASRU 2023, Taiwan (2023/12)
- General Chair, ROCLING 2017, Taiwan (2017/11)
- Sponsor Chair, TAAI 2017, Taiwan (2017/11)
- Program Chair, ROCLING 2016, Taiwan (2016/10)
- Organizer, SWS 2015, IIS, Academia Sinica, Taiwan (2015/03)

REPRESENTATIVE RESEARCH WORKS

Prof. Yu Tsao's research primarily focuses on "**Neural Speech Enhancement and Assessment for Assistive Oral Communication Technologies**", which can be further divided into three parts: **(1) Neural Speech Enhancement and Voice Conversion; (2) Neural Speech Assessment; (3) Assistive Oral Communication Technologies.**

(1) Neural Speech Enhancement and Voice Conversion (in total more than 6000 citations)
Selected Publications:

- Paper#1: Speech Enhancement Based on Deep Denoising Autoencoder, in *Proc. Interspeech 2013*, **Google Citation: 1192.**
- Paper#2: End-to-End Waveform Utterance Enhancement for Direct Evaluation Metrics Optimization by Fully Convolutional Neural Networks, *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 26(9), pp. 1570-1584, April 2018. (**2021 IEEE Signal Processing Society (SPS) Young Author Best Paper Award**). **Google Citation: 384**.
- Paper#3: MetricGAN: Generative Adversarial Networks based Black-box Metric Scores Optimization for Speech Enhancement, in *Proc. ICML 2019*, **Long Oral Presentation** with **Travel Grant**. **Google Citation: 498.**
- Paper#4: Audio-Visual Speech Enhancement Using Multimodal Deep Convolutional Neural Networks, *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2(2), pp. 117-128, April. 2018. **Google Citation: 347.**
- Paper#5: Raw waveform-based Speech Enhancement by Fully Convolutional Networks, in *Proc. APSIPA 2017*, **Google Citation: 299.**
- Paper#6: SNR-Aware Convolutional Neural Network Modeling for Speech Enhancement, in *Proc. Interspeech 2016*, **Google Citation: 239.**
- Paper#7: Complex spectrogram enhancement by convolutional neural network with multi-metrics learning, in *Proc. MLSP 2017*, **Google Citation: 237.**
- Paper#8: MetricGAN+: An Improved Version of MetricGAN for Speech Enhancement, in *Proc. Interspeech 2021*, **Google Citation: 365.**
- Paper#9: Conditional Diffusion Probabilistic Model for Speech Enhancement, in *Proc. ICASSP 2022*, **Google Citation: 334.**
- Paper#10: An Investigation of Incorporating Mamba for Speech Enhancement, in *Proc. SLT 2024*, **Google Citation: 127.**
- Paper#11: WaveCRN: An Efficient Convolutional Recurrent Neural Network for End-to-end Speech Enhancement, *IEEE Signal Processing Letters*, vol. 27, pp. 2149-2153, Nov. 2020, **Google Citation: 108.**
- Paper#12: Learning with Learned Loss Function: Speech Enhancement with Quality-Net to Improve Perceptual Evaluation of Speech Quality, *IEEE Signal Processing Letters*, vol. 27, pp. 26-30. Nov. 2019, **Google Citation: 104.**
- Paper#13: Ensemble Modeling of Denoising Autoencoder for Speech Spectrum Restoration, in *Proc. Interspeech 2014*, **Google Citation: 76.**
- Paper#14: Improving Perceptual Quality by Phone-Fortified Perceptual Loss using Wasserstein Distance for

Speech Enhancement, in *Proc. Interspeech 2020*, **Google Citation:94.**

- Paper#15: Experimental Study on Extreme Learning Machine Applications for Speech Enhancement, *IEEE Access*, vol. 5, pp. 25542-25554, Oct. 2017, **Google Citation: 58.**
- Paper#16: Noise Adaptive Speech Enhancement using Domain Adversarial Training, in *Proc. Interspeech 2019*, **Google Citation: 76.**
- Paper#17: MetricGAN-U: Unsupervised Speech Enhancement/dereverberation Based only on Noisy/reverberated Speech, in *Proc. ICASSP 2022*, **Google Citation: 68.**
- Paper#18: Time-domain Multi-modal Bone/air Conducted Speech Enhancement, *IEEE Signal Processing Letters* vol. 27, pp. 1035-1039, **Google Citation: 67.**
- Paper#19: Speech Enhancement Based on Denoising Autoencoder with Multi-branched Encoders, *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 28, pp. 2756-2769, October 2020, **Google Citation: 65**.
- Paper#20: Bone-conducted Speech Enhancement Using Deep Denoising Autoencoder, *Speech Communication* 104, pp. 106-112, **Google Citation: 65**.
- Paper#21: Generalized Maximum A Posteriori Spectral Amplitude Estimation for Speech Enhancement, *Speech Communication*, vol. 76, pp. 112-126, Feb. 2016, **Google Citation: 58.**
- Paper#22: Speech Enhancement using Segmental Nonnegative Matrix Factorization, in *Proc. ICASSP 2014*, **Google Citation: 58.**
- Paper#23: Multichannel Speech Enhancement by Raw Waveform-Mapping Using Fully Convolutional Networks, *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 28, pp. 1888-1900, Feb. 2020, **Google Citation: 70.**
- Paper#24: An Investigation of Spectral Restoration Algorithms for Deep Neural Networks Based Noise Robust Speech Recognition, in *Proc. Interspeech 2013*, **Google Citation: 51.**
- Paper#25: Voice Conversion From Unaligned Corpora Using Variational Autoencoding Wasserstein Generative Adversarial Networks, in *Proc. Interspeech 2017*, **Google Citation: 485.**
- Paper#26: Voice Conversion From Non-parallel Corpora Using Variational Auto-encoder, in *Proc.APSIPA 2016*, **Google Citation: 401.**
- Paper#27: 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 (4), pp. 468-479, April 2020, **Google Citation: 62.**

(2) Neural Speech Assessmnet (in total more than 1900 citations)
Selected Publications:

- Paper#1: ASVspoof 2019: A Large-scale Public Database of Synthetized, Converted and Replayed Speech, *Computer Speech and Language,* vol. 64, 101114, Nov. 2020, **Google Citation: 693**
- Paper#2: MOSNet: Deep Learning based Objective Assessment for Voice Conversion, in *Proc. Interspeech 2019*, **Google Citation: 447**
- Paper#3: Quality-Net: An End-to-End Non-intrusive Speech Quality Assessment Model based on BLSTM, in *Proc. Interspeech 2018*, **Google Citation: 252.**
- Paper#4: The VoiceMOS Challenge 2022, in *Proc. Interspeech 2022*, **Google Citation: 199.**
- Paper#5: Deep Learning-based Non-Intrusive Multi-Objective Speech Assessment Model with Cross-Domain Features, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 54-70., Sept. 2022, **Google Citation: 140.**
- Paper#6: STOI-Net: A Deep Learning based Non-intrusive Speech Intelligibility Assessment Model, in *Proc. APSIPA 2020*, **Google Citation: 64.**
- Paper#7: MBI-Net: A Non-Intrusive Multi-Branched Speech Intelligibility Prediction Model for Hearing Aids, in *Proc. Interspeech 2022*, **Google Citation: 37.**
- Paper#8: InQSS: A Speech Intelligibility and Quality Assessment Model using a Multi-task Learning Network, in *Proc. Interspeech 2022*, **Google Citation: 42.**
- Paper#9: Toward Automating Oral Presentation Scoring during Principal Certification Program using Audio-Video Low-level Behavior Profiles, *IEEE Transactions on Affective Computing*, vol. 10(4), pp. 552-567, Dec. 2019, **Google Citation: 13**.
- Paper#10: HASA-NET: A Non-Intrusive Hearing-Aid Speech Assessment Network, in *Proc. ASRU 2021*,

Google Citation: 20.

- Paper#11: SVSNet: An End-to-end Speaker Voice Similarity Assessment Model, *IEEE Signal Processing Letters*, 29, pp. 767-771, Feb. 2022, **Google Citation: 14**.

(3) Assistive Oral Communication Technologies (in total more than 1000 citations)
Selected Publications:

- Paper#1: Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach, *Journal of Voice*, vol 33(5), pp. 634-641, Sept. 2019. **Google Citation: 313.**
- Paper#2: A Deep Denoising Autoencoder Approach to Improving the Intelligibility of Vocoded Speech in Cochlear Implant Simulation, *IEEE Transactions on Biomedical Engineering*, vol. 64(7), pp. 1568-1578, July, 2017. **Google Citation: 151.**
- Paper#3: Deep learning–based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients, *Ear and Hearing*, vol. 4, pp. 795-809, July 2018, **Google Citation: 107.**
- Paper#4: A Smartphone-Based Multi-Functional Hearing Assistive System to Facilitate Speech Recognition in the Classroom, *IEEE Access*, vol. 5, pp. 10339-10351, June 2017, **Google Citation: 57.**
- Paper#5: Joint Dictionary Learning-based Non-Negative Matrix Factorization for Voice Conversion to Improve Speech Intelligibility After Oral Surgery, *IEEE Transactions on Biomedical Engineering*, vol. 64 (11), pp. 2584-2594, Nov. 2017, **Google Citation: 60**.
- Paper#6:A Mobile Phone–Based Approach for Hearing Screening of School-Age Children: Cross-Sectional Validation Study, *JMIR mHealth and uHealth*, vol. 7(4), e12033, April 2019, **Google Citation: 50**.
- Paper#7: Generative Adversarial Networks for Unpaired Voice Transformation on Impaired Speech*,* in *Proc. Interspeech 2018*, **Google Citation: 39.**
- Paper#8: Enhancing Intelligibility of Dysarthric Speech Using Gated Convolutional-based Voice Conversion System, in *Proc. Interspeech 2020*, **Google Citation: 43.**
- Paper#9: Demographic and Symptomatic Features of Voice Disorders and Their Potential Application in Classification Using Machine Learning Algorithms, *Folia Phoniatrica et Logopaedica*, vol. 70 (3-4), 174-182, **Google Citation: 27.**
- Paper#10: Combining Acoustic Signals and Medical Records to Improve Pathological Voice Classification, *APSIPA Transactions on Signal and Information Processing*, vol. 8, e14, June 2019. **Google Citation: 32.**
- Paper#11: A Smartphone-Based Remote Microphone Hearing Assistive System Using Wireless Technologies, *IEEE Systems Journal* vol. *12*(1), pp. 20-29, Oct. 2015, **Google Citation: 24.**
- Paper#12: Improving the Intelligibility of Speech for Simulated Electric and Acoustic Stimulation Using Fully Convolutional Neural Networks*, IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 29, pp. 184-195, Dec. 2020. **Google Citation: 31.**
- Paper#13: Continuous Speech for Improved Learning Pathological Voice Disorders, *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 3, pp. 25-33, Feb. 2022, **Google Citation: 39.**
- Paper#14: A Study of Joint Effect on Denoising Techniques and Visual Cues to Improve Speech Intelligibility in Cochlear Implant Simulation, *IEEE Transactions on Cognitive and Developmental*, vol. 13(4), pp. 984-994, Dec. 2021. **Google Citation: 23**.
- Paper#15: Ensemble and Multimodal Learning for Pathological Voice Classification, *IEEE Sensors Journal (Letters)*, vol. 5 (7), June 2021, **Google Citation: 17.**
- Paper#16: Mandarin Electrolaryngeal Speech Voice Conversion with Sequence-to-Sequence Modeling, in *Proc. ASRU 2021*, **Google Citation: 17.**