Yi-Chiao Wu

+1-650-644-5368 | yichiao.wu@g.sp.m.is.nagoya-u.ac.jp | https://bigpon.github.io/ | Google scholar

EDUCATION

2017-2021 **Nagoya University** **Nagoya, Japan**
- Ph.D. in Graduate School of Informatics (Advisor: Tomoki Toda)
- Thesis: Incorporating prior knowledge on speech production mechanism into neural speech waveform generation

2005-2011 **National Chiao Tung University** **Hsinchu, Taiwan**
- M.S. and B.S. in Communication Engineering (Advisor: Yih-Ru Wang)
- Thesis: Speaker recognition system for intelligent home robot

WORK EXPERIENCE

2022-2025 **FAIR/Reality Lab Research, Meta (Facebook)** **NYC/Pittsburgh, USA**

Senior Research Scientist
- Lead the speech separation task of SAM Audio (Open-sourced unified general audio separation model)
- Lead the pseudo-labeling data pipeline and fine-tuning processes of SAM Audio.
- Lead the audio tokenizer, DACVAE, development for AudioBox (T2S/A), MovieGen (TV2A), and SAM Audio.
- Develop the long-form audio generation and mono2stereo modules for MovieGen.
- Organize AudioMOS challenge 2025 track 2 for general audio aesthetic assessment.
- Develop and open-source the first non-intrusive general audio quality assessment model, audiobox aesthetic.
- Develop and open-source a SOTA real-time flow-based general audio codec, FlowDec.
- Collect and open-source a full-band (48kHz) expressive anechoic human voice dataset, EARS.
- Develop phase-preserving and unseen data robust speech codecs, ScoreDec and ComplexDec.
- Develop and open-source a high-fidelity, low-bitrate, streamable audio codec, AudioDec.
- Develop a real-time, two-way communication system with spatial binaural rendering on a Meta VR headset.

2021-2021 **Institute of Information Science, Academia Sinica** **Taipei, Taiwan**

Postdoc
- Develop a non-intrusive speech assessment network, HASANet, for hearing aids.

2017-2021 **Graduate School of Informatics, Nagoya University** **Nagoya, Japan**

Researcher and Research Assistant
- Proposed a neural-post-filter for improving low-cost Text-To-Speech (TTS) systems.
- Proposed a pitch-dependent structure, QPNet/QPPWG, for the real-time speech generation model.
- Develop a baseline system and release the source code for Voice Conversion Challenge 2020.
- Propose a collapsed speech detection and suppression method for the WaveNet vocoder.
- Got an overall performance ranking of 2/12 in Voice Conversion Challenge 2018.

2019 **National Institute of Information and Communications Technology** **Kyoto, Japan**

Summer Intern
- Reduce 30% training time and model size of WaveGlow with the depthwise CNN.

2015-2017 **Institute of Information Science, Academia Sinica** **Taipei, Taiwan**

Research Assistant
- Combined manifold learning techniques with an exemplar-based speaker voice conversion system.
- Integrated exemplar-based post-filtering methods with neural-based speech enhancement systems.

2013-2015 **Da Vinci Innovation Lab, ASUS** **Taipei, Taiwan**

Software R&D Engineer
- Develop ASUS Zenbo robot’s text-independent speaker recognition engine.
- Design and implemented a speaker recognition mobile application on an in-house Android platform.

2012-2013 **Multimedia BU II, Realtek** **Hsinchu, Taiwan**

System Designer
- Work with IC component designers to design and verify TV audio systems on TV chips.
- Port audio drivers for five mass-production projects (for TOSHIBA, SONY, Skyworth, etc.).

RESEARCH INTEREST & SKILLS

- Text/visual to audio/speech/music generation, audio codec, speech vocoder, voice conversion, speech enhancement
- Programming: PyTorch, TensorFlow, Python, MATLAB, C/C++, JAVA, UNIX shell script