Chih-Sheng Chen, Ph.D.

Manager | AI & Speech Processing Specialist
Hsinchu, Taiwan • LinkedIn • Technical Blog

Summary

Speech AI engineer and R&D manager with 14+ years of experience (since 2012), from fixed-point voice wake-up at MediaTek to leading an AI speech algorithm team at intelliGo. Together with a team of 5–6 Ph.D./M.S. engineers, we design, compress and deploy neural speech models (noise reduction, echo cancellation/suppression, beamforming, bandwidth extension) on our own AI chip under tight power and latency budgets. Our work runs in Nintendo Switch 2 GameChat, in Microsoft Teams (Open Office) certified boomless headsets, and in millions of other devices.

Beyond speech at work, my research interests include model compression, LLM inference acceleration, reinforcement learning and generative models. I also keep a technical blog (100+ posts) on quantization, pruning and generative models (diffusion, flow matching), with the underlying math in stochastic processes, Bayesian methods and optimal transport.

Experience

intelliGo Technology Inc. — Hsinchu, Taiwan
2019-08 – Present

Manager (2023-01 – Present)

- Lead the algorithm team (5–6 Ph.D./M.S. engineers) developing NN-based speech front-end and audio algorithms: noise reduction, AEC/AES (acoustic echo suppression), shotgun/bubble beamforming (2–5 mics), DoA/DoV/DoK, speaker diarization and bandwidth extension (BWE).
- Nintendo Switch 2 GameChat: led the team that developed the speech front-end models (NR, AEC, AES, far-field pickup) and deployed them on our own chip, achieving high audio quality at low power.
- Boomless headset: led the team that developed "bubble mic" AI beamforming — only 3 microphones are used for beamforming — which passed Microsoft Teams (Open Office) certification with pickup quality comparable to a boom mic.
- Laptops: our team developed "shotgun" 3-mic beamforming, using a third microphone to resolve the front/back mirror-pickup ambiguity of 2-mic screen arrays.
- Led development of GAN-based bandwidth extension (BWE), deployed in a mass-produced mobile phone product.
- Accelerated OpenAI Whisper for streaming inference through quantization and pruning.
- Technologies shipped in millions of devices across laptops, TWS earbuds, speakerphones and a game console.
- Work closely with the FW team to deploy algorithms on chip; plan the technology roadmap; recruit; turn Sales/Marketing customer feedback into technical goals; run regular AI paper and engineering reviews.

Deputy Technical Manager (2021-01 – 2022-12)

- Led development of NN-based NR, AEC, beamforming and DoA; worked with the FW team on chip deployment; planned the roadmap; recruited talent.

Staff Engineer (2019-08 – 2020-12)

- Built a very low-MIPS/memory ASR and command-recognition system: acoustic-model training and quantization, a decoding graph with user-defined command grammar, and a fixed-point decoder deployed on the company's AI chip. Tracked the latest ASR research and the voice-UI roadmap.

Lenovo Research (AI Lab) — Beijing, China
2017-08 – 2019-07

Researcher

- Developed a deep-learning keyword trigger combined with speaker verification.
- Implemented acoustic echo cancellation and beamforming to make the trigger and cloud ASR more robust in noise.
- Co-inventor of three Chinese patents from this period (see Patents).

MediaTek — Hsinchu, Taiwan
2013-09 – 2017-07

Senior R&D Engineer

- Played a key role in building a very low-MIPS, low-memory voice wake-up system from scratch, deployed on several MediaTek chips including Helio X20/X25 and adopted by the Meizu Pro 6 smartphone.
- Designed and implemented the wake-up algorithm (training and decoding) with fixed-point optimization, integrated with CM4 and HAL.
- Trained acoustic models with Kaldi and built HMM/DNN toolchains.
- Worked on ML projects such as ultrasound gesture recognition, dynamic-programming pronunciation decoding and audio event detection (e.g., baby-crying detection).
- Automated and optimized the performance-testing flow (C++, bash/perl); applied ML and dimensionality reduction to classify and visualize problems.
- Co-inventor of a U.S. patent on low-latency keyword spotting (see Patents).

Altek Corporation — Hsinchu, Taiwan
2012-09 – 2013-08

R&D Engineer

- Developed low-MIPS fixed-point LPC feature extraction (Levinson–Durbin recursion) and a fixed-point single-channel noise-reduction method.

Skills

- **Audio & speech AI:** Noise reduction • AEC / AES • Beamforming (NN-based 2–5 mics; DS/LCMV/MVDR/CGMM-MVDR) • Direction of voice/keyword (DoV/DoK) • Speaker diarization and verification • Bandwidth extension • ASR (Kaldi, WeNet, Whisper) • Keyword spotting
- **Model compression & deployment:** PTQ (AIMET, OBQ, OCTAV, GPTQ, AWQ, SmoothQuant) • QAT (LSQ+, EWGS, MAD, OCTAV) • Pruning (Movement, L0 regularization, LoRAPrune) • Streaming-graph conversion with OP fusion • NAS
- **Training & generative models:** SWA and SAM for training stability • GAN • Transformer • Diffusion and flow matching
- **Tools:** Python • PyTorch • PyTorch Lightning + Hydra • C/C++ (basic) • Kaldi • git
- **Leadership:** Managing 5–6 Ph.D./M.S. engineers • cross-team projects (8–9 people) • technology roadmap • hiring • working with Sales/Marketing and FW teams • 0→1 concept to 1→100 mass production

Education

- **Ph.D., Computer Science**, National Tsing Hua University (NTHU) (2005–2011) — Multimedia Information Retrieval Lab; advisor: Prof. Jyh-Shing Roger Jang
- **Visiting Scholar, School of ECE (CSIP)**, Georgia Institute of Technology (2010–2011) — Supervised by Prof. Chin-Hui Lee
- **B.S. & M.S., Computer & Information Science**, National Chiao Tung University (NCTU) (M.S. 2003–2005) — M.S. at the Automatic Information Processing Lab

Selected publications

1. Sung-Feng Huang, Chia-ping Chen, Zhi-Sheng Chen, Yu-Pao Tsai, Hung-yi Lee. "Personalized Lightweight Text-to-Speech: Voice Cloning with Adaptive Structured Pruning." ICASSP, 2023.
2. Wei-Tsung Kao, Yuan-Kuei Wu, Chia-Ping Chen, Zhi-Sheng Chen, Yu-Pao Tsai, Hung-Yi Lee. "On the Efficiency of Integrating Self-supervised Learning and Meta-learning for User-defined Few-shot Keyword Spotting." IEEE SLT, 2022.
3. Zhi-Sheng Chen, Jyh-Shing Roger Jang, Chin-Hui Lee. "A Kernel Framework for Content-Based Artist Recommendation System in Music." IEEE Transactions on Multimedia, 2011.
4. M. Wu, Z. Chen, J. R. Jang, J. Ren, Y. Li, C. Lu. "Combining Visual and Acoustic Features for Music Genre Classification." ICMLA, 2011.

5. Jia-Min Ren, Zhi-Sheng Chen. “On the Use of Sequential Patterns Mining as Temporal Features for Music Genre Classification.” ICASSP, 2010.
6. Zhi-Sheng Chen, Jyh-Shing Roger Jang. “On the Use of Anti-Word Models for Audio Music Annotation and Retrieval.” IEEE Transactions on Audio, Speech, and Language Processing, 2009.

Patents

Co-inventor on all patents below. Issue dates are grant/publication dates.

- US 12143782 B2 — Microphone system (2024-11-12; Intelligo Technology Inc.)
- US 12219329 B2 — Beamforming method and microphone system in boomless headset (2025-02-04; Intelligo Technology Inc.)
- US 12231844 B2 — Microphone system and beamforming method (2025-02-18; Intelligo Technology Inc.)
- US 12256202 B2 — Stereo enhancement system and stereo enhancement method (2025-03-18; Intelligo Technology Inc.)
- TWI831513 B — Beamforming method and microphone system in boomless headset (2024-02-01; Intelligo Technology Inc.)
- TWI835246 B — Microphone system and beamforming method (2024-03-11; Intelligo Technology Inc.)
- TWI861569 B — Microphone system (2024-11-11; Intelligo Technology Inc.)
- TWI870681 B — Stereo enhancement system and stereo enhancement method (2025-01-21; Intelligo Technology Inc.)
- CN 108986815 B — Voice control method and device and electronic equipment (2021-11-16; Lenovo (Beijing))
- CN 110334240 B — Information processing method and system, first device and second device (2021-10-22; Lenovo (Beijing))
- CN 109978145 B — Processing method and device (2021-09-14; Lenovo (Beijing))
- US 10032449 B2 — Keyword spotting system for achieving low-latency keyword recognition by using multiple dynamic programming tables reset at different frames of acoustic data input and related keyword spotting method (2018-07-24; MediaTek Inc.)

Languages: Chinese (native), English (professional working)