Chung-Ming **Chien**

Chicago, Illinois, United States

cmchien@ttic.edu | Homepage | LinkedIn | Google Scholar | +1 (415) 539-5179

Research Interests

Speech-text joint learning, self-supervised speech representations and units, speech generation, multi-modal learning, representation learning

Education

Toyota Technological Institute at Chicago (TTIC) — *Chicago, IL*

Ph.D. in Computer Science — *Sep. 2022 - Present*

- Advisor: Karen Livescu
- GPA: 4.0/4.0

National Taiwan University (NTU) — *Taipei, Taiwan*

M.S. in Computer Science and Information Engineering — *Sep. 2019 - Aug. 2021*

- Thesis: End-to-End Prosody Learning Frameworks for Multi-Speaker Speech Synthesis
- Advisors: Lin-shan Lee & Hung-yi Lee at Speech Processing Lab
- GPA: 4.02/4.3

National Taiwan University (NTU) — *Taipei, Taiwan*

B.S.E. in Electrical Engineering — *Sep. 2015 - Aug. 2019*

- GPA: 4.08/4.3; Ranked 25/256 (9%) with two Dean's List Awards

Experience

Speech and Language Group, TTIC — *Chicago, IL*

Graduate Student Researcher — *Sep. 2022 - Present*

- Advisor: Karen Livescu
- Revealed **word-level language structures** intrinsically encoded in self-supervised speech representations [arXiv]
- Discovered speech-text models with **text-to-speech transferability** which enables **zero-shot spoken language understanding** [ASRU'23]

FAIR (Fundamental AI Research) at Meta — *Menlo Park, CA*

Research Scientist Intern — *Jun. 2023 - Dec. 2023*

- Mentor: Andros Tjandra
- Worked on **universal speech generation** and additionally conditioned the model on **unnormalized/punctuated transcripts**

Hotpot.ai — *Remote*

Machine Learning Researcher — *Jun. 2022 - Aug. 2022*

- Researched on **text-to-image generaiton** by combining pre-trained word representations with **diffusion models**

World Quant LLC — *Taipei, Taiwan*

Quantitative Research Intern — *Jun. 2022 - Jul. 2022*

- Developed novel Alpha ideas and evaluated their performance with historical market data

Amazon Alexa — *Cambridge, UK*

Applied Scientist Intern — *Jul. 2021 - Nov. 2021*

- Mentors: Adam Gabryś and Jaime Lorenzo-Trueba
- Improved extremely **low-resource speaker-adaptive text-to-speech (TTS)** by modeling content and speaker information separately [ICASSP'22]
- Reduced the gap between synthesized and real speech by over 30%

Speech Processing Laboratory, NTU — *Taipei, Taiwan*

Student Researcher — *Sep. 2018 - Jul. 2021*

- Advisors: Lin-shan Lee and Hung-yi Lee
- Disentangled **speaker and phonetic information in self-supervised speech representations** for the task of voice conversion (VC) [InterSpeech'21]
- Proposed **SOTA zero-shot any-to-any VC** by learning **sub-phoneme alignments between utterances with Transformer attention** [ICASSP'21]
- Proposed **generative speaker embedding pre-training** for speech synthesis [ICASSP'21]
- Led a team to win the 2nd prize of the IEEE M2VoC Challenge on **low-resource voice cloning** [M2VoC Challenge]
- Built and maintained a state-of-the art TTS system **FastSpeech 2** [Github]
- Developed **hierarchical prosody modeling** in TTS [SLT'21]

Machine Learning and Estimation Theory Laboratory, NTU *Taipei, Taiwan*

Student Researcher *Feb. 2018 - Feb. 2019*

- Advisor: Pei-Yuan Wu
- Discovered a critical privacy leakage issue in a privacy-preserving support vector machine

Publications

† indicates equal contribution

Conference Proceedings

[1] Ju-Chieh Chou, **Chung-Ming Chien**, Wei-Ning Hsu, Karen Livescu, Arun Babu, Alexis Conneau, Alexei Baevski, and Michael Auli, "Toward Joint Language Modeling for Speech Units and Text," in *Findings of EMNLP*, 2023.

[2] **Chung-Ming Chien**, Mingjiamei Zhang, Ju-Chieh Chou, and Karen Livescu, "Few-Shot Spoken Language Understanding via Joint Speech-Text Models," in *ASRU*, 2023.

[3] Adam Gabryś, Goeric Huybrechts, Manuel Sam Ribeiro, **Chung-Ming Chien**, Julian Roth, Giulia Comini, Roberto Barra-Chicote, Bartek Perz, and Jaime Lorenzo-Trueba, "Voice Filter: Few-Shot Text-to-Speech Speaker Adaptation Using Voice Conversion as a Post-Processing Module," in *ICASSP*, 2022.

[4] Jheng-hao Lin, Yist Y. Lin, **Chung-Ming Chien**, and Hung-yi Lee, "S2VC: A Framework for Any-to-Any Voice Conversion with Self-Supervised Pretrained Representations," in *Interspeech*, 2021.

[5] **Chung-Ming Chien**, Jheng-Hao Lin, Chien-yu Huang, Po-chun Hsu, and Hung-yi Lee, "Investigating on Incorporating Pretrained and Learnable Speaker Representations for Multi-Speaker Multi-Style Text-to-Speech," in *ICASSP*, 2021.

[6] **Chung-Ming Chien**†, Yist Y. Lin†, Jheng-Hao Lin, Hung-yi Lee, and Lin-shan Lee, "Fragmentvc: Any-To-Any Voice Conversion by End-To-End Extracting and Fusing Fine-Grained Voice Fragments with Attention," in *ICASSP*, 2021.

[7] **Chung-Ming Chien** and Hung-yi Lee, "Hierarchical Prosody Modeling for Non-Autoregressive Speech Synthesis," in *SLT*, 2021.

Pre-Prints

[1] Ankita Pasad, **Chung-Ming Chien**, Shane Settle, and Karen Livescu, "What Do Self-Supervised Speech Models Know about Words?," *preprint arXiv:2307.00162*, 2023.

[2] Ju-Chieh Chou, **Chung-Ming Chien**, and Karen Livescu, "AV2Wav: Diffusion-Based Re-synthesis from Continuous Self-supervised Features for Audio-Visual Speech Enhancement," *preprint arXiv:2309.08030*, 2023.

Honors

Scholarship

2023 **Government Scholarship to Study Abroad**, Ministry of Education of Taiwan ($32,000 in 2 years) *Taiwan*

2020 **Advanced Speech Technologies Scholarship**, NTU EECS ($17,000) *Taipei, Taiwan*

2016 **NTUEE60 Scholarship**, NTU EE ($3,500) *Taipei, Taiwan*

Awards

2021 **2rd Place**, IEEE ICASSP 2021 M2VoC Challenge *Virtual*

2020 **Top 20 Finalist**, Trans Action Award *Taipei, Taiwan*

2019 **Cathay United Bank Special Award**, Make NTU *Taipei, Taiwan*

2016-2017 **Dean's List Awards (Two-Time)**, NTU EE *Taipei, Taiwan*

Leadership

2019-2020 **Captain**, NTU Baseball Varsity Team *Taipei, Taiwan*

Non-Academic

2023 **1st Place within UChicago-Affiliated Athletes**, J.P. Morgan Corporate Challenge 3.5-Mile Road Race *Chicago, IL*

2019&2021 **5th Places (Two-Time)**, University Baseball League of Taiwan (equivelent to NCAA Division III) *Taiwan*

2019 **Golden Medal, Men's Half-Iron Relay**, Yilan National Triathlon Championships *Yilan, Taiwan*

Service

2022 **Reviewer**, IEEE JSTSP

Talks

Nov. 2022 **Self-Supervised Pre-Trained Voice Conversion**, TTIC Student Workshop *Chicago, IL*

Nov. 2021 **Few-Shot Speaker Adaptive TTS by Learning from Non-Target Speakers**, Amazon Text-to-Speech Group *Cambridge, UK*

Aug. 2020 **Speech Synthesis in the Deep Learning Era**, AI Summer School 2020, NTU *Taipei, Taiwan*

Teaching

National Taiwan University

Taipei, Taiwan

Teaching Assistant

- EE5184 Machine Learning, Spring 2020 and Spring 2019, instructed by Hung-yi Lee
- EE4049 Speech Processing Project, Spring 2020 and Fall 2019, instructed by Lin-shan Lee
    - Led 26 undergraduate students to do research in speech and natural language processing
- EE4037 Digital Speech Processing, Fall 2019, instructed by Lin-shan Lee
- EE2011 Signals and Systems, Spring 2018, instructed by Lin-shan Lee

Projects

FastSpeech2

Open-Sourced Project *Jun., 2020*

- Open-sourced TTS project with **over 1.4k stars on Github**, supporting multiple languages and more than 100 speakers [Github]

TTS without T

Course Project *Jun., 2019*

- Applied discrete speech units unsupervisedly discovered by the multilabel-binary vectors (MBV) and the vector quantized variational autoencoder (VQ-VAE) to a VC task [Link]

Skills

Natural Languages	Mandarin (native), Taiwanese (native), English (fluent), German (basic)
Programming Languages	Python, C/C++, Shell Script, MATLAB, Verilog, HTML+CSS
Toolkits	PyTorch, MXNet, ESPnet, Kaldi, Git, LaTeX