

Chung-Ming Chien

CHICAGO, ILLINOIS, UNITED STATES

✉ cmchien@ttic.edu | 🏠 Homepage | 🔗 LinkedIn | 🎓 Google Scholar | 📞 +1 (415) 539-5179

Research Interests

Speech language models, speech generative models, self-supervised speech representation learning

Education

Toyota Technological Institute at Chicago (TTIC)

PH.D. IN COMPUTER SCIENCE

- Advisor: Karen Livescu
- Ph. D. candidate
- GPA: 4.0/4.0

Chicago, IL

Sep. 2022 - Present

National Taiwan University (NTU)

M.S. IN COMPUTER SCIENCE AND INFORMATION ENGINEERING

- 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

Taipei, Taiwan

Sep. 2019 - Aug. 2021

National Taiwan University (NTU)

B.S.E. IN ELECTRICAL ENGINEERING

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

Taipei, Taiwan

Sep. 2015 - Aug. 2019

Experience

Speech and Language Group, TTIC

GRADUATE STUDENT RESEARCHER

- Advisor: Karen Livescu
- Discovered **text-to-speech transferability** in speech-text models, which enables **zero-shot spoken language understanding** [\[ASRU'23\]](#)
- Revealed **word-level language structures** intrinsically encoded in self-supervised speech representations [\[TACL'24\]](#)
- Benchmarked **speech foundation models** on spoken language understanding tasks under various resource considerations [\[ACL'24\]](#)
- Conducted a comprehensive comparison of **SpeechLLM**'s capabilities on various speech tasks.

Chicago, IL

Sep. 2022 - Present

NVIDIA

SPEECH AI RESEARCH INTERN

- Mentors: Zhehuai Chen and Jason Li
- Augmented pre-trained Canary **LLMs** with **speech generation** capabilities for speech-to-speech translation and speech question answering

Santa Clara, CA

Jun. 2024 - Sep. 2024

FAIR (Fundamental AI Research) at Meta

RESEARCH SCIENTIST INTERN

- Mentors: Andros Tjandra and Wei-Ning Hsu
- Worked on the **Voicebox** project, enhancing **fine-grained controllability** of **speech generation** models under resource-limited scenarios. [\[InterSpeech'24\]](#)

Menlo Park, CA

Jun. 2023 - Dec. 2023

Hotpot.ai

MACHINE LEARNING RESEARCHER

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

Remote

Jun. 2022 - Aug. 2022

World Quant LLC

QUANTITATIVE RESEARCH INTERN

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

Taipei, Taiwan

Jun. 2022 - Jul. 2022

Amazon Alexa

APPLIED SCIENTIST INTERN

- 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%

Cambridge, UK

Jul. 2021 - Nov. 2021

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) ([InterSpeech21](#))
- Proposed **SOTA zero-shot any-to-any VC** by learning **sub-phoneme alignments between utterances with Transformer attention** ([ICASSP21](#))
- Proposed **generative speaker embedding pre-training** for speech synthesis ([ICASSP21](#))
- 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 ([SLT21](#))

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

JOURNAL ARTICLES

- [1] Ankita Pasad, **Chung-Ming Chien**, Shane Settle, and Karen Livescu, “What Do Self-Supervised Speech Models Know About Words?,” *Transactions of the Association for Computational Linguistics* 12 (Apr. 2024) pp. 372–391. 2024

CONFERENCE PROCEEDINGS

- [1] **Chung-Ming Chien**, Andros Tjandra, Apoorv Vyas, Matt Le, Bowen Shi, and Wei-Ning Hsu, “Learning Fine-Grained Controllability on Speech Generation via Efficient Fine-Tuning,” in *Interspeech*, 2024.
- [2] Siddhant Arora, Ankita Pasad, **Chung-Ming Chien**, Jionghao Han, Roshan Sharma, Jee-weon Jung, Hira Dharmyal, William Chen, Su-won Shon, Hung-yi Lee, Karen Livescu, and Shinji Watanabe, “On the Evaluation of Speech Foundation Models for Spoken Language Understanding,” in *Findings of ACL*, 2024.
- [3] Ju-Chieh Chou, **Chung-Ming Chien**, and Karen Livescu, “AV2WAV: Diffusion-Based Re-Synthesis from Continuous Self-Supervised Features for Audio-Visual Speech Enhancement,” in *ICASSP*, 2024.
- [4] 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.
- [5] **Chung-Ming Chien**, Mingjiamei Zhang, Ju-Chieh Chou, and Karen Livescu, “Few-Shot Spoken Language Understanding via Joint Speech-Text Models,” in *ASRU*, 2023, **Best Student Paper Award**.
- [6] 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.
- [7] 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.
- [8] **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.
- [9] **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.
- [10] **Chung-Ming Chien** and Hung-yi Lee, “Hierarchical Prosody Modeling for Non-Autoregressive Speech Synthesis,” in *SLT*, 2021.

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

- | | | |
|-----------|--|----------------|
| 2023 | Best Student Paper Award , ASRU (with Mingjiamei Zhang, Ju-Chieh Chou, and Karen Livescu) | Taipei, Taiwan |
| 2021 | 2rd Place , ICASSP 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&2024 **1st Place within UChicago-Affiliated Athletes (Two Straight Years)**, J.P. Morgan Corporate Challenge 3.5-Mile Road Race *Chicago, IL*
- 2019&2021 **5th Place (Two-Time)**, University Baseball League of Taiwan (equivalent to NCAA Division III) *Taiwan*
- 2019 **Golden Medal, Men's Half-Iron Relay**, Yilan National Triathlon Championships *Yilan, Taiwan*

Service

- 2024 **Organizer**, TTIC Student Workshop
- 2022-2025 **Reviewer**, IEEE JSTSP, ICLR, ICASSP

Talks

- Apr. 2024 **Few-Shot Spoken Language Understanding via Joint Speech-Text Models**, Midwest Speech and Language Days *Ann Arbor, MI*
- 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

Toyota Technological Institute at Chicago *Chicago, IL*

TEACHING ASSISTANT

- TTIC 31020 Introduction to Machine Learning, Winter 2024, instructed by Nathan Srebro

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 1.5k stars on Github**, supporting multiple languages and more than 100 speakers ([Github](#))

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