Shih-Lun Wu

Ph.D. Student, Dept. of Electrical Engineering and Computer Science
Massachusetts Institute of Technology (MIT), Cambridge, MA, United States
Email: slseanwu@mit.edu | Homepage: https:/slseanwu.github.io | Google Scholar | GitHub | LinkedIn

EDUCATION

Doctor of Philosophy (Ph.D.) | Massachusetts Institute of Technology 08.2024 ~
in **Electrical Engineering and Computer Science (EECS)**

- Advisor: Dr. Cheng-Zhi Anna Huang
- Research areas: **Music & Audio Generation, LLM Adaptation, Preference Tuning**

Master of Science (M.Sc.) | Carnegie Mellon University 08.2022 ~ 05.2024
in **Language Technologies**

- Cumulative QPA – **4.10/4.00**
- Research areas: **Music & Audio Processing, Generative Models, Multimodal Learning**
- Advisors: Dr. Chris Donahue, Dr. Shinji Watanabe

Bachelor of Science (B.Sc.) | National Taiwan University 09.2017 ~ 06.2021
in **Computer Science** (with minor in **Economics**)

- Cumulative GPA – Overall: **4.28/4.30**, Major: **4.28/4.30**, Rank: **1/176**
- Research areas: **Symbolic Music Generation, Formal Verification**
- Advisors: Dr. Yi-Hsuan Yang, Dr. Chung-Wei Lin

HONORS & RECOGNITION

- **Citations** (Google Scholar, as of 09/15/2025): **700+** total, **450+** first-author | **GitHub stars: 500+**

MIT T.S. Lin Graduate Fellow | MIT Office of Grad Education 09.2025
- Selected through MIT institute-wide competition for full-year (AY25-26) graduate fellowship (**$110K** award)

Siebel Scholar, Class of 2024 | The Siebel Foundation 09.2023
- Selected as one of 83 graduate students worldwide for outstanding research & leadership (**$35K** award)

Winner (Research Org), Intern Project Showcase | Adobe Inc. 08.2023
- Secured top prize among **200+** research interns for developing Music ControlNet and related projects

1st Prize, Automated Audio Captioning Challenge | DCASE 2023 06.2023
- Won with advanced encoder architecture & LLM supervision, outperforming runner-up by **4%** (publication [6])

1st Prize (Ssu-Nien Fu's Award), Best Bachelor's Thesis | National Taiwan University 06.2021
- Awarded to **6** out of **3500+** graduating students for exceptional undergrad research [thesis] [defense slides]

SELECTED PUBLICATIONS

[12] **Shih-Lun Wu**, Ge Zhu, Juan-Pablo Caceres, Anna Huang, and Nicholas J. Bryan. "Stemphonic: All-at-once Flexible Multi-Stem Music Generation." *Under Review*, 2025. [pdf] [demo]

[11] **Shih-Lun Wu**, Yoon Kim, and Anna Huang. "MIDI-LLM: Adapting Large Language Models for Text-to-MIDI Music Generation." *Under Review*, 2025. [pdf] [live demo]

[10] **Shih-Lun Wu**, Aakash Lahoti, Arjun Desai, Karan Goel, Chris Donahue, and Albert Gu. "Towards Codec-LM Co-design for Neural Codec Language Models." *Student Research Workshop at Conf. Nations of the Americas Chapter of the Assoc. for Computational Linguistics (**NAACL-SRW**)* 2025. (**Best Paper**) [pdf] [code]

[9] Fang-Duo Tsai, **Shih-Lun Wu**, Weijaw Lee, Sheng-Ping Yang, Bo-Rui Chen, Hao-Chung Cheng, and Yi-Hsuan Yang. "MuseControlLite: Multifunctional Music Generation with Lightweight Conditioners." *International Conference on Machine Learning **(ICML)*** 2025. [pdf] [code][colab] [demo]

[8] **Shih-Lun Wu**, Chris Donahue, Shinji Watanabe, and Nicholas J. Bryan. "Music ControlNet: Multiple Time-varying Controls for Music Generation." *IEEE/ACM Transactions on Audio, Speech, & Language Processing (**TASLP**)* 2024. [pdf] [tl;dr] [demo] [code (3rd-party)]

[7] Fang-Duo Tsai, **Shih-Lun Wu**, Haven Kim, Bo-Yu Chen, Hao-Chung Cheng, and Yi-Hsuan Yang. "Audio Prompt Adapter: Unleashing Music Editing Abilities for Text-to-Music with Lightweight Finetuning." *Int. Society for Music Information Retrieval Conference **(ISMIR)*** 2024. [pdf] [code] [demo]

[6] **Shih-Lun Wu**, Xuankai Chang, Gordon Wichern, Jee-weon Jung, François Germain, Jonathan Le Roux, and Shinji Watanabe. "Improving Audio Captioning Models with Fine-grained Audio Features, Text Embedding Supervision, and LLM Mix-up Augmentation." *Int. Conf. on Acoustics, Speech, & Signal Processing (**ICASSP**)* 2024. (**Oral paper**) [pdf] [DCASE challenge results]

[5] **Shih-Lun Wu**, Yi-Hui Chou, and Liangze Li. "Listener Model for the PhotoBook Referential Game with CLIPScores as Implicit Reference Chain." *Annual Meeting of the Assoc. for Computational Linguistics **(ACL)*** 2023. [pdf] [code]

[4] **Shih-Lun Wu** and Yi-Hsuan Yang. "Compose & Embellish: Well-structured Piano Performance Generation via A Two-Stage Approach." *Int. Conf. on Acoustics, Speech, & Signal Processing **(ICASSP)*** 2023. (**Oral paper**) [pdf] [code]

[3] **Shih-Lun Wu** and Yi-Hsuan Yang. "MuseMorphose: Full-song and Fine-grained Music Style Transfer with One Transformer VAE." *IEEE/ACM Transactions on Audio, Speech, & Language Processing **(TASLP)*** 2023. [pdf] [code] [demo]

[2] Antoine Liutkus, Ondřej Cífka, **Shih-Lun Wu**, Umut Simsekli, Yi-Hsuan Yang, and Gaël Richard. "Relative Positional Encoding for Transformers with Linear Complexity." *International Conference on Machine Learning **(ICML)*** 2021. (**Long talk**, acceptance rate: **3.0%**) [pdf] [code] [presentation video] [demo]

[1] **Shih-Lun Wu** and Yi-Hsuan Yang. "The Jazz Transformer on the Front Line: Exploring the Shortcomings of AI-Composed Music through Quantitative Measures." *Int. Society for Music Information Retrieval Conference **(ISMIR)*** 2020. [pdf] [code] [poster] [presentation video]

RESEARCH WORK EXPERIENCE

Research Scientist/Engineer Intern | Adobe Research — 06.2025 ~ 09.2025

Music AI Group. *Supervisor: Dr. Nick Bryan*

- Invented *Stemphonic*, a diffusion-/flow-based model for 1-pass variable multi-stem music generation (publ. [12])
- Enabled composer-like workflows with accompaniment generation and stem-wise temporal activity controls
- Accelerated multi-stem workflow by ~50% while improving generation quality compared to existing methods

Research Intern | Cartesia AI Inc. — 05.2024 ~ 08.2024

Technical Staff. *Supervisors: Dr. Albert Gu, Dr. Arjun Desai*

- Developed real-time neural audio codecs for state-space model (SSM)-based text-to-speech (TTS) (publ. [10])
- Reduced time-to-first-audio (TTFA) latency by 40% (95ms → 57ms) and improved intelligibility (WER) by 10%

Research Scientist/Engineer Intern | Adobe Research — 05.2023 ~ 12.2023

Audio AI Lab. *Supervisors: Dr. Nick Bryan, Dr. Gautham Mysore*

- Invented *Music ControlNet*, enabling precise melody, dynamics, rhythm controls for diffusion text-to-music models
- Beat Meta's MusicGen by 49% on melody control, using 35x fewer params & 11x less training data (publ. [8])

Research Engineer | Taiwan AI Labs — 07.2020 ~ 03.2022

AI Music Team. *Supervisor: Dr. Yi-Hsuan Yang*

- Bridged Transformers and VAEs for fine-grained style transfer (rhythm & harmony) of long music pieces (publ. [3])
- Collaborated with researchers @ INRIA / Télécom Paris on positional encodings for O(n) Transformers (publ. [2])
- Developed a widely used suite of quantitative metrics to assess the quality of machine-generated music (publ. [1])

EXTRACURRICULARS & SERVICE

Pianist, Violist, & Director of General Affairs — 09.2018 ~ 06.2021

Symphony Orchestra, National Taiwan University

- Participated actively in concerts [playlist] and handled procurement, musical scores, and transportation affairs

Peer Reviewer

- **ISMIR** (2021~2025), **TISMIR** (2021, 2025), **ACM Computing Surveys** (2023), **ICASSP** (2024), **TASLP** (2024)

Research Mentor

- **Yi-Jen Shih** (2021~2022, now PhD @ UT Austin), **Fang-Duo Tsai** (2024~, PhD @ NTU), **Matthew Quispe** (2025~, MIT UROP)

Teaching Assistant

- **Algorithms Design and Analysis** (NTU, Fall 2019)

SKILLS & QUALIFICATIONS

- Programming Languages & Infrastructure: *Python · C/C++ · JavaScript · LaTeX · Linux · Slurm · Anaconda*
- Machine Learning Frameworks*: PyTorch · Keras · Tensorflow · HuggingFace · PyTorch Lightning*
- CS fundamentals courses: DS & Algo, Algo Design & Analysis, Formal Language & Automata, Linear Algebra
- ML/DL-related courses: ML Techniques, Advanced NLP, Speech Recog & Understdg, TinyML