Shih-Lun Wu

Ph.D. Student (incoming Fall '24), 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**

- Advisor: Dr. Cheng-Zhi Anna Huang

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

- **Citation count** (Google Scholar, as of 05/27/2024)**: 310+** total**, 200+** first-author | **GitHub stars: 400+**

Siebel Scholar, Class of 2024 | The Siebel Foundation 09.2023
- Awarded to **~85** graduate students worldwide for outstanding research & leadership (**$35K** prize money)

Winner (Research Org), Intern Project Showcase | Adobe Inc. 08.2023
- Won with the Music ControlNet and related music generation projects, against **200+** Adobe research interns

1st Prize, Automated Audio Captioning Challenge | DCASE 2023 06.2023
- Won by leveraging advanced encoder architecture & LLM supervision, surpassing runner-up by **1.2** points (**4%**)

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

[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] [project website]

[7] **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]

[6] **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]

[5] **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]

[4] **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] [project website]

[3] 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] [project website]

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

[1] **Shih-Lun Wu***, Ching-Yuan Bai*, Kai-Chieh Chang, Yi-Ting Shieh, Chao Huang, Chung-Wei Lin, Eunsuk Kang and Qi Zhu. "Efficient System Verification with Multiple Weakly-hard Constraints for Runtime Monitoring." *International Conference on Runtime Verification **(RV)*** 2020. (*: equal contribution) [pdf] [publisher page]

RESEARCH-FOCUSED WORK EXPERIENCE

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
- Demonstrated compositionality of proposed controls, and out-of-domain generalizability to user-specified controls
- Beat Meta's MusicGen by 49% on melody control, using 35x fewer params & 11x less training data (publication [8])

Graduate Research Assistant | Carnegie Mellon University 09.2022 ~ 05.2024

Watanabe's Audio & Voice Lab (WAVLab), Language Tech Institute. *Advisor: Dr. Shinji Watanabe*

- Achieved new SoTA on audio captioning task with ChatGPT mix-ups and LLM embedding supervision (publ. [7])
- Won ICASSP-23 Grand Challenge on spoken language understanding, utilizing Whisper model backbone [tech rep]
- Integrated OpenAI's Whisper model into the lab's 7000+ star ESPnet speech processing toolkit [GitHub PR]

Research Engineer | Taiwan AI Labs 08.2021 ~ 03.2022

Research Intern | Taiwan AI Labs 07.2020 ~ 07.2021

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

- Made a 3-stage model to generate well-structured music with recurring & developing content (some results in [5])
- Bridged Transformers and VAEs for fine-grained style transfer of arbitrarily long musical pieces, allowing users to exert bar-level controls such as harmonic and rhythmic intensities (publ. [4])
- Collaborated with researchers @ INRIA / Télécom Paris on positional encodings for O(n) Transformers (publ. [3])
- Developed a set of widely-used quantitative metrics to assess the quality of machine-generated music (publ. [2])

Undergraduate Research Assistant | National Taiwan University 02.2019 ~ 06.2020

Cyber-Physical Systems Lab, Dept. of CSIE. *Advisor: Dr. Chung-Wei Lin*

- Formulated the formal verification problem under multiple weakly-hard constraints on environmental faults
- Discovered and proved the mathematical properties between pairs of weakly-hard constraints
- Devised a lowest-cost-first heuristic using the properties, accelerating verification algorithm by up to 12x (publ. [1])

OTHER WORK EXPERIENCE

Software Engineering Intern | Asus Inc. 07.2019 ~ 08.2019

Cloud Infrastructure Team, Asus Intelligent Cloud Services (AICS) Center

- Developed a Kubernetes + Python (Flask) template for launching containerized, cloud-based ML solutions
- Integrated Azure Key Vault, Mutual TLS auth & Azure App Insights to the template to streamline model deployment

EXTRACURRICULAR ACTIVITIES & 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

- Conferences: **ICMLA** (2020), **ISMIR** (2021, 2022, 2023, 2024), **ICASSP** (2024)
- Journals: **TISMIR** (2021), **ACM Computing Surveys** (2023)

Research Mentor

- **Yi-Jen Shih** (2021~2022, now PhD student at UT Austin), **Fang-Duo Tsai** (2024~, MS student at NTU)

Teaching Assistant

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

SKILLS & QUALIFICATIONS

- Programming Languages & Infrastructure: *Python · C/C++ · JavaScript · ReactJS · LaTeX · Linux · Kubernetes*
- Machine Learning Frameworks*: PyTorch · Keras · Tensorflow · HuggingFace · PyTorch Lightning*
- Selected Coursework: **Straight A+'s** in the following courses (NTU & CMU)
  -- CS fundamentals: DS & Algo, Algo Design & Analysis, Formal Language & Automata, Linear Algebra
  -- ML-/DL-related: ML Techniques, Advanced NLP, Speech Recognition & Understanding, Multimodal ML