Longshen Ou

Singapore | oulongshen@u.nus.edu | +65 8791 0291 | oulongshen.xyz | linkedin.com/in/longshen-ou

Career Interests

- **Desired Roles:** AI researcher or engineer in natural language processing, audio/speech processing, multimodal content analysis, music information retrieval, or machine learning.
- **Availability:** Singapore-based, part-time roles, starting from July 2025.

Education

National University of Singapore, PhD in Computer Science (last year in program) Jan 2022 – Jan 2026

- **Research Focus:** Member of the Sound and Music Computing Lab, advised by Prof. Wang Ye. Research focuses on audio content analysis, lyric transcription/translation/generation, and symbolic music generation, with emphasis on self-supervised learning (SSL), transfer learning, and integrating controllability into SSL models.
- **Relevant Courses:** NNs and Deep Learning, AI Planning and Decision Making, Sound and Music Computing.

Harbin Institute of Technology, China, Bachelor's Degree in Computer Science Aug 2017 – Jul 2021

- **Relevant Courses:** Visual and Audio Information Understanding, Pattern Recognition and Deep Learning, Sentiment Analysis, Data Mining.

University of California, San Diego, United States, Exchange Program Jul 2019 – Dec 2019

- **Relevant Courses:** Discrete Signal Processing, Introduction to Machine Learning.

Professional Experience

Researcher (Internship), Sony Computer Science Laboratories – Tokyo, Japan Aug 2024 – Nov 2024
Mentor: Dr. Taketo Akama

- Designed a tokenization scheme for Whisper to represent melody note sequences, enabling efficient transcription of musical elements.
- Fine-tuned Whisper for lyric and melody transcription using frame-level and end-to-end approaches, achieving performance comparable to state-of-the-art models.
- Implemented semi-supervised techniques, including auxiliary transposition-equivariant loss, pseudo-labeling, noisy student model, and FixMatch, to enhance transcription performance.
- Applied diverse data augmentation strategies (time shift, time stretch, SpecAugment, and pitch shift) to improve the model's ability to learn music-specific concepts and handle noisy labels.

Research Intern, Yamaha Corporation – Hamamatsu, Japan May 2024 – Aug 2024
Mentor: Dr. Yu Takahashi

- Initiated the task of lead instrument detection from multitrack music, creating high-quality annotated datasets and strong baseline models for segment- and frame-level analysis in both single- and multi-track settings.
- Designed a lead instrument detection framework leveraging self-supervised learning and a track-wise attention mechanism, achieving superior performance compared to existing models (e.g., SVM- and CRNN-based approaches) while demonstrating generalizability to unseen instruments and new domains.
- Evaluated multiple model designs, highlighting the advantages of the proposed track-wise attention mechanism and track classification approach.

Research Intern, DASLab, Harvard University – Remote Aug 2020 – Sep 2020
Advisor: Prof. Stratos Idreos

- Contributed to the research *More or Less: When and How to Build Convolutional Neural Network Ensembles*, accepted at *ICLR 2021*.

Back End Developer (Internship), Tencent – Shenzhen, China Jul 2020 – Aug 2020

- Enhanced a distributed online analytical processing system for user engagement data analysis.
- Accelerated query processing speed by 120x through a distributed storage caching strategy.
- **Skills:** Docker, Kubernetes, OLAP, Hadoop, Spark, Presto, Tencent Cloud, Go.

Publications

(* indicates equal contribution)

Lead Instrument Detection from Multitrack Music Apr 2025
Longshen Ou, Yu Takahashi, Ye Wang
2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2025)

- First work on lead instrument detection from multitrack music, providing annotated datasets and benchmark models. Work done during Yamaha internship.

DNA Storage Toolkit: End-to-End DNA Data Storage Codec and Simulator May 2024
Puru Sharma, Gary Goh, Bin Gao, ***Longshen Ou***, Dehui Lin, Deepak Sharma, Djordje Jevdjic
2024 IEEE Int'l Symposium on Performance Analysis of Systems and Software (ISPASS 2024)

- Developed a simulation framework for DNA-based storage systems using a sequence-to-sequence attention-based RNN, accurately modeling information alterations during read/write operations in wet lab scenarios.
- Outperformed baseline approaches on the Microsoft Nanopore dataset, significantly reducing reliance on costly and time-intensive wet lab experiments, enabling more efficient future research.

Lyric Transcription and Music Transcription from Multimodal Singing May 2024
Xiangming Gu, ***Longshen Ou***, Wei Zeng, Jianan Zhang, Nicholas Wong, Ye Wang
ACM Transactions on Multimedia Computing, Communications and Applications (TOMM 2024)

- Developed a unified framework for automatic lyric and melody transcription from multimodal singing, integrating both audio and video modalities, and extended the N20EM dataset for melody transcription tasks.
- Adapted self-supervised learning models from speech to singing, achieving state-of-the-art performance on lyric transcription and melody transcription tasks across benchmark datasets.
- Enhanced noise robustness using video modalities under challenging SNR environments.
- Introduced novel tasks of lyric lipreading and note lipreading, enabling language and music information extraction solely through video inputs.

Songs Across Borders: Singable and Controllable Neural Lyric Translation Oct 2023
Longshen Ou, Xichu Ma, Min-Yen Kan, Ye Wang
The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)

- Identified key constraints from translatology studies and formulated the singable lyric translation problem as a constrained generation task.
- Developed an automatic lyric translation method with user-specified control over length, rhyme, and word boundary to maintain singability across languages.
- Proposed a rhyme ranking mechanism that achieves higher quality and controllability for paragraph-level translations.
- Achieved 95%+ accuracy in controlling key aspects with minimal trade-offs in quality, enhancing singability by 75% compared to baseline methods.

Transfer Learning of wav2vec 2.0 for Automatic Lyric Transcription Dec 2022
Longshen Ou*, Xiangming Gu*, Ye Wang
Proceedings of the 23rd International Society for Music Information Retrieval Conf. (ISMIR 2022)

- Proposed a lyric transcription solution leveraging the similarities between spoken and singing voices, utilizing transfer learning with wav2vec 2.0 by pretraining and fine-tuning on speech data before domain adaptation.
- Extended the original Connectionist Temporal Classification (CTC) model into a hybrid CTC/attention model, achieving improved convergence and more accurate decoding.
- Achieved state-of-the-art results on benchmark ALT datasets, including DSing, DALI, and Jamendo, with an average 25% relative WER reduction compared to prior methods.
- Demonstrated effectiveness in low-resource ALT scenarios by achieving SOTA results on the DSing test split with less than one-tenth of the labeled singing data.

MM-ALT: Multimodal Automatic Lyric Transcription (*Oral, Top Paper Award*) Oct 2022
Xiangming Gu*, ***Longshen Ou****, Danielle Ong, Ye Wang
Proceedings of the 30th ACM International Conference on Multimedia (ACM Multimedia 2022)

- Curated the first dataset for multimodal lyric transcription (N20EM), laying the foundation for future research.
- Enhanced lyric transcription by integrating video and inertial measurement unit (IMU) signals, significantly improving noise robustness and overall system performance.
- Proposed RCA, a novel feature fusion method that combines self-attention and cross-attention mechanisms to improve multimodal singing feature integration.
- Demonstrated the efficacy of IMU signals in detecting singing activities, showcasing their potential in multimodal music processing.

Exploring Transformer's Potential on Automatic Piano Transcription May 2022
Longshen Ou, Ziyi Guo, Emmanouil Benetos, Jiqing Han, Ye Wang
2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2022)

- Investigated the application of Transformer networks across multiple piano transcription subtasks, including multi-pitch detection, onset and offset detection, and velocity estimation.
- Demonstrated the superiority of Transformers in velocity detection on MAESTRO and MAPS datasets, achieving significant improvements in both frame-level and note-level metrics.
- Enhanced a high-resolution piano transcription system by integrating a Transformer-based velocity branch, which improved overall transcription performance.

Preprints

Unlocking Potential in Pre-Trained Music Language Models for Versatile Multi-Track Music Arrangement Aug 2024
Longshen Ou, Jingwei Zhao, Ziyu Wang, Gus Xia, Ye Wang
arXiv:2408.15176 (2024)

- Proposed an efficient tokenizer (REMI-z) that reduces sequence length, lowers bar-level entropy, and improves note-level perplexity in unconditional training, enhancing task-specific performance, and is extensible to other music generation tasks.
- Introduced a unified seq2seq fine-tuning framework leveraging generative pre-trained music language models, supporting diverse arrangement tasks, ensuring coherence, and potentially applicable for other conditional music generation tasks.
- Achieved state-of-the-art performance on band arrangement, piano reduction, and drum arrangement, outperforming task-specific models in both objective and subjective evaluations.
- Demonstrated through ablation studies that generative pretraining significantly enhances seq2seq transformations, and is especially critical for processing long segments.

LOAF-M2L: Joint Learning of Wording and Formatting for Singable Melody-to-Lyric Generation Jul 2023
Longshen Ou, Xichu Ma, Ye Wang
arXiv:2307.02146 (2023)

- Developed a hybrid training strategy combining unsupervised and supervised learning, where self-supervised pretraining improved length-aware text generation, and supervised fine-tuning enabled melody constraint interpretation for enhanced singable lyric generation.
- Proposed a novel training objective leveraging limited paired data, decoding implicit formatting guidelines from music–lyric pairs to improve condition-aware generation and enhance music–lyric compatibility.
- Validated the model's effectiveness through objective metrics and subjective evaluations, demonstrating superior text fluency, music–lyric compatibility, and overall quality compared to previous M2L models and unsupervised-only baselines.

Teaching Assistant Experience

CS 4347 Sound and Music Computing, National University of Singapore 2023 Fall, 2022 Fall

- Designed and prepared all course assignments, covering both conceptual questions and coding tasks.
- Developed course project topics with detailed guidelines to support students in understanding core concepts.
- Created comprehensive tutorial materials and conducted weekly tutorial sessions to facilitate student learning.
- Evaluated assignments and project submissions, providing constructive feedback to students.

CS 4248 Natural Language Processing, National University of Singapore 2023 Spring

- Delivered tutorials to explain course concepts and help students understand core NLP techniques.
- Offered research direction and technical advice to project teams during weekly consultations.
- Graded final project submissions and contributed to setting final examination questions.

Engineering Projects

REMI-z: Efficient Tokenizer and Multitrack Data Structure for Symbolic Music

- Developed a tool to convert symbolic music between MIDI, piano roll, and token sequence formats, using the proposed efficient REMI-z tokenization scheme.
- Designed a hierarchical data structure for multitrack music, defining song, bar, track, and note-level operations, enabling automated symbolic music manipulation and batch processing.
- **Tools Used:** Python, miditoolkit, pretty_midi

GuitarFret: Guitar Fretboard Simulator

- Built a guitar fretboard simulator with an intuitive UI to visualize note names across the fretboard.
- Added features for marking pressed positions, chord roots, and sound simulation, facilitating guitar theory learning, arrangement, and score analysis.
- **Tools Used:** Python, PyQt5, fluidsynth

SimpleGuru: Automatic Ranking House Rent in Singapore

- Developed a web crawler to collect and analyze rental data from PropertyGuru, ranking listings using a customizable scoring system.
- Ranking criteria included structural attributes (rent, room/house size), geolocation metrics (commute time, proximity to key locations), and text-based filters (gender restrictions, air conditioning availability).
- **Tools Used:** Python, Selenium, Google Maps API

Honors and Awards

SoC Research Incentive Award Oct 2023
Issued by the School of Computing, National University of Singapore

Research Achievement Award (2022/2023) May 2023
Issued by the School of Computing, National University of Singapore

Top Paper Award (Top 2% of accepted full papers) Nov 2022
Issued by ACM Multimedia 2022

Honor Degree of Bachelor of Engineering Jun 2020
Conferred by Harbin Institute of Technology Honors School

Third Prize Oct 2018
Sogou Innovative Practice Project for College Students

Skills

Proficient in Deep Learning Coding

- Deep learning frameworks: Huggingface, Lightning, SpeechBrain, Fairseq, PyTorch, TensorFlow.
- Programming languages: Python, Java, C, Shell, JavaScript, Go, SQL, Verilog.

Professional-level Music Performance

- Over 20 years of violin performance experience. Proficient in percussion fingerstyle guitar.
- Held a solo concert at Harbin Institute of Technology, China, in Oct 2018.

Proficient in Music Production

- Extensive experience in composing and arranging for violin and guitar.
- Skilled in music recording, audio editing, multitrack mixing/mastering, and MIDI.
- Familiar with video editing tools.