ANNA MIN

anna.min1754@gmail.com(man20@mails.tsinghua.edu.cn) ◇ Personal Website ◇ Github ◇ Linkedin

EDUCATION

Bachelor of Engineering in Software Engineering, Tsinghua University - 2025 (Expected)
Major GPA: 3.72 /4.00, sophomore: 3.78, junior: 3.86

PUBLICATION

[1] **Anna Min***, Chenxu Hu*, Yi Ren, Hang Zhao. A Unit-based System and Dataset for Expressive Direct Speech-to-Speech Translation (*Interspeech 2024* [Paper])

[2] **Anna Min**, Ziyang Chen, Hang Zhao. Quantifying geometrical associations across multi-modal perception: An Information-theoretic Perspective (*NeurIPS2024 Wi@ML Workshop* [Paper])

MANUSCRIPT

[1] **Anna Min**, Chenxu Hu, Yi Ren, Hang Zhao. When End-to-End is Overkill: Rethinking Cascaded Speech-to-Text Translation (*in submission to ICASSP 2025, under review* [Paper])

[2] **Anna Min**, Ziyang Chen, Hang Zhao, Andrew Owens. Supervising Sound Localization using In-the-wild Ego-motion (*In submission to CVPR 2025, under review* [Paper], [Website])

HONORS AND COMPETITIONS

Tsinghua Academic Excellence Award (2/102) 2023
Tsinghua Research Excellence Award (2/102) 2022-2024
Tsinghua Spark Scientific Innovation Fellowship (50/3900) 2022
The first prize of 2021 National Student Mathematical Modeling Competition (ranked 89 out of 3000) 2021
National High School Mathematics Competition, Provincial Second Prize in Hubei Province 2018

SKILLS

Programming: Python, C, C++, MATLAB, Rust, Java, Javascript, LaTeX, Verilog, SQL, Docker
Framework & Tools: Pytorch, Fairseq, Soundspace, Qt
Languages: Chinese (Native), English (TOEFL 106 (R30+L27+W26+S23))

PROFESSIONAL SERVICE AND PRESENTATION

Reviewer in ICASSP 2025, WiML Workshop @ NeurIPS 2024
Invited presentation at MIT SLS group titled "Audio-visual Encodec: High-Fidelity Noise-Robust Audio-Visual Compression for Universal Speech Regeneration and Dubbing" Nov. 2024
Invited talk at MIT Machine Learning Tea Time titled "Multi-sensory Perception from Top to Down" Sep. 2024
Invited presentation at UMich titled "Supervising Sound Localization using In-the-wild Ego-motion" June. 2024

RESEARCH EXPERIENCE

Audio-visual Encodec: Reducing Audio Input Bandwidth with Vision Codec
Research Intern Aug. 2024 - Present
Advisor: Prof. Jim Glass, Massachusetts Institute of Technology

- Developed models to reduce high-dimensional time-series data into a simplified discrete format
- Distilled video quantizer from pre-trained audio encoded and discretized them to reduce the bandwidth of audio

Improving Score-entropy Discrete Diffusion
Research Intern Nov. 2024 - Present
Advisor: Prof. Stefano Ermon, Stanford University

- Utilized pretrained Codec models and Score-entropy Discrete Diffusion models for multi-modal generation
- Invested into faster sampling methods in a SDE form for discrete diffusion process

Supervising Sound Localization using In-the-wild Ego-motion
Visiting Research Intern July. 2023 - 2024
Advisor: Prof. Andrew Owens, University of Michigan

- Explored the ego-motion of visual cues from limited perspective in-the-wild videos to learn 360-degree spatial audio
- Introduced the first stereo dataset and benchmark for sound localization in the wild which is gathered from an extensive corpus of 8,000 hours of YouTube stereo sound videos

Fine-grained Emotion Transfer for Speech-to-Speech Translation in Expressive Video Dubbing
Research Intern Nov. 2022 - 2023
Advisor: Prof. Hang Zhao, Tsinghua University

- Constructed the first training set with aligned bilingual audio tracks with the same emotion from movies
- Used wavform to tokenized unit translation and HiFi-GAN-based networks for transferring pitches and rhythms
- Outperformed the baseline by a significant 20% improvement in emotional expression

Synchronized Video-to-Audio Generation for Multi-Style Videos [Media Coverage] Jan. 2024 - Feb. 2024
Research Intern, Pika
Host: Chenlin Meng, Stanford University

- Led a project on audio-visual synchronization generation, garnering 260,000 views on Twitter
- Implemented automated editing by integrating audio spectrum features and applied a latent-diffusion-based model to learn continuous audio representations from contrastive language-audio pretraining

SELECTED PROJECTS

Mini Database [Code] Feb. - May 2023

- Implemented a database that supports basic SQL queries with optimization, transactions, locks, and recovery

Android Chat App [Code] Feb. - May 2023

- Utilized Kotlin/Jetpack Compose along with Material3 to craft the Android frontend
- Built the backend using Django with Channels, enabling the implementation of both HTTP APIs and real-time WebSocket communication which support user searching, following, and chatting

FTP Server & Client [Code] Sep. 2022 - Jan. 2023

- Wrote an FTP server in C and FTP client UI in Python supporting most basic commands

Compiler from C++ to LLVM [Code] Sep. 2022 - Jan. 2023

- Translated C++ code to LLVM intermediate representation utilizing Python-Lex-Yacc and LLVMLite compilers
- Implemented error handling, preprocessing capabilities, multidimensional array operations, scope mechanisms

Mars Online Judge Platform [Code] [Website] [Media Coverage] Sep. 2022 - Jan. 2023
A web-based platform for online supplementary coding training and multiplayer quiz learning

- Implemented real-time answer battles utilizing Socket.IO, comprehensive review functionalities, and a sophisticated tracking system empowering administrators to closely monitor student progress
- Led backend and contributed to frontend development using Vue3 for the frontend and Flask for the backend HTTP API, along with python-lsp-server for backend language services
- Achieved a user base of over 3000 users and 1000 subscribers, launching both web and mini-program versions

Machine Learning Modeling Services [Code] July 2022 - Aug. 2022
One-stop solution providing model deployment online

- Implemented functionality to adapt to ONNX, PMML, Keras multiple machine learning models, enabling out-of-the-box use, load balancing, and complete testing
- Led back-end development, contributed to front-end development, and utilized Django, Docker, Kubernetes, Celery, Vue, JavaScript, and other technology stacks