ZHIYAO DUAN TEACHING STATEMENT

The main reason that I joined the faculty of a major research university after completing my PhD is that it allows me to be a researcher and an educator at the same time. This is also a reason why I chose to come back to academia after a two-year sabbatical leave in industry. Teaching provides me the opportunity to introduce the latest research to classroom, to observe students grow intellectually, and to get inspiration from them. Integrating research and teaching is what I find the most interesting part of a faculty job.

My teaching philosophy can be summarized as three points:

- **Foster passion:** Interest is the best teacher. Fostering passion, in my view, is the most important goal of all teaching activities.
- **Promote active learning:** I deeply agree with the saying of "the purpose of teaching is to not teach". Students' active learning is the goal of education.
- **Tailor to students' aptitude**: Each student is unique. They come with different backgrounds and have different strengths and weaknesses. A good teacher should interact with students, pay attention to their feedback, and adjust teaching/learning agenda to their aptitude.

After my promotion to tenure in July 2020, I took a two-year sabbatical leave to work in industry. During this time, I did not teach formal classes at the University, but I continued supervising PhD, MS and undergraduate students. As those two years overlapped with COVID, much of the supervision was remote. Since Fall 2022, I have resumed classroom teaching. In particular, I developed a new course ECE 208/408 – The Art of Machine Learning for ECE and Engineering students, which turned out to be quite successful. I have also participated in co-teaching two modular courses. The following lists my classroom teaching activities. Teaching evaluation results are attached in the package:

- Fall 2022 – ECE 277/477 – Computer Audition
- Spring 2023 – ECE 208/408 – The Art of Machine Learning
- Fall 2023 - ECE 277/477 – Computer Audition (evaluation not available due to too few responses)
- Spring 2024 - ECE 208/408 – The Art of Machine Learning
- Spring 2024 – ECE 411 - Selected Topics in Augmented and Virtual Reality – Module on Acoustic Rendering and Computer Audition (6 weeks)
- Fall 2024 - ECE 277/477 – Computer Audition (evaluation not available due to too few responses)

Besides classroom teaching, since 2020, I have supervised or been supervising 11 PhD students, 1 visiting PhD student, 7 MS students, and 12 undergraduate students on research in my lab. I was/am the academic advisor of AME Class of 2022 and Class of 2026, each with about 20 students. I was the Co-PI of the NSF REU Site on Computational Methods for Understanding Music, Media, and Minds from 2021 to 2023, which hosted 10 undergraduate students from around the US each summer at the University of Rochester. Outside the University of Rochester, I was invited as the primary lecturer to teach at the 2023 Winter School on Speech and Audio Processing (WiSSAP) in IIT-Kanpur in India. Within a week, I gave multiple lectures on "Introduction to Music Information Retrieval", "Music Rhythm Analysis", "Music Pitch Analysis", and "Interactive Music Systems" to a group of about 50 students and faculty. The remainder of the document highlights some of these achievements.

Classroom Teaching

ECE 208/408 – The Art of Machine Learning (cross listed as TEE 408)

(Course website: https://hajim.rochester.edu/ece/sites/zduan/teaching/ece408/index.html)

This course was initiated by one of my PhD graduates, Andrea Cogliati, in Spring 2022 during my sabbatical leave, in response to the increasing demands of machine learning in the ECE department. I like the name of the course very much, as training and using machine learning models require a lot of know-hows, which on many aspects, are like an art on top of the mathematical rigor of model derivations. Therefore, I kept this name after I took this course over in Spring 2023. However, I completed re-

designed this course. As the first course to machine learning for ECE students, I designed the course to start with classical models such as nearest neighbors, decision trees, linear models, and support vector machines (SVM) as a good way to introduce important machine learning concepts such as inductive bias, overfitting, and cross validation. However, the course could not stop there because those models cannot meet students' need in learning state-of-the-art methods. Therefore, the course then moves on to neural networks, covering multi-layer perceptron (MLP), convolutional neural networks (CNN), recurrent neural networks (RNN) and transformers. The content is designed with a focus on supervised learning, but it also spends two weeks on unsupervised learning, dimensionality reduction, and reinforcement learning. It touches on some state-of-the-art topics such as diffusion models and flow matching.

One challenge I face is how to quickly bring students on track and keep their interests throughout the course. My approach is to have them practice the machine learning concepts and models on real-world ECE problems. Together with my first two TAs, Neil Zhang and Melissa Chen, I developed eight programming assignments toward solving a wide range of problems in circuit design, salary prediction, maternal health prediction, ultrasound image classification, melody generation, and image denoising. Furthermore, students need to work on a final project to experience the full cycle of research from project proposal to algorithm implementation, experimentation, paper writing and presentation.

The course has been successful. In Spring 2023, 28 students were enrolled. The overall course rating was 4.14 and the overall instructor rating was 4.29. In Spring 2024, 40 students were enrolled. The overall course rating was 4.85 and the overall instructor rating was 4.77. All ratings were out of 5.0.

ECE 277/477 – Computer Audition (cross listed as AME 277, CSC 264/464)

(Course website: https://hajim.rochester.edu/ece/sites/zduan/teaching/ece477/index.html)

I created this graduate course upon my arrival at UofR in 2013 with the aim of training students to do research in this field. In the first half of the semester, this course covers fundamental concepts including short-time Fourier transform (STFT), auditory models, audio features (e.g., pitch, rhythm, timbre) and audio modeling techniques (e.g., non-negative matrix factorization, hidden Markov models, and deep neural networks). Students work on six programming assignments to practice these concepts. In the second half of the semester, the course covers state-of-the-art research topics in the field, including multi-pitch analysis, source separation, source localization, instrument identification, speaker recognition, etc. Students work on a final project that allows them to experience the full cycle of research, from project proposal to algorithm implementation, experimentation, paper writing and presentation.

After I returned from sabbatical leave, I resumed teaching this course and made revisions on the lectures, assignments and grading schemes. I added new topics including Interactive Music Systems, Voice Conversion, Audio-Visual Scene Understanding, and Music Generation. Some of the topics were offered by my PhD students who conduct that research. The course continues to be successful overall, although the attendance of lectures has been lower compared to pre-COVID times. Therefore, I started checking attendance by taking short quizzes at the end of some lectures. This was an approach I learned from education literature that could encourage regular attendance and focus during lectures.

This course continues to provide opportunities for students to enter the door of research. Several final projects were further developed after the course, resulting in peer-reviewed publications for the students taking the course (marked by *). In particular, for some undergraduate students at the time (e.g., Yongyi Zang, Yutong Wen), the papers were their very first publications:

- Yongyi Zang*, You Zhang, and Zhiyao Duan, "Phase perturbation improves channel robustness for speech spoofing countermeasures," in *Proc. Interspeech*, 2023.
- Ge Zhu, Yutong Wen*, Marc-André Carbonneau, and Zhiyao Duan, "EDMSound: Spectrogram based diffusion models for efficient and high-quality audio synthesis," in *NeurIPS 2023 Workshop on Machine Learning for Audio*, 2023.
- Huiran Yu* and Zhiyao Duan, "Note-level transcription of choral music," in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2024.

Mentoring Student Research

One advantage that students have at a research university is the opportunity to get involved in research. Since my promotion to tenure, I continued mentoring student research at different levels. This effort even continued during my sabbatical leave. Most of my undergraduate and MS mentees conducted their first research project and published their first peer-reviewed paper with me. After graduation, several students went on to pursue a PhD in institutions like Leigh University, NYU, UIUC, and Georgia Tech. Several students who have not graduated yet are actively applying for PhD programs.

At the undergraduate level, I mentored 12 students, 6 of which have co-authored peer-reviewed publications with me. The complete list of publications with undergraduate students marked can be found in my CV. Here I would like to provide a few highlights that I am very proud of. Yongyi Zang was in AME Class of 2023. He started doing research with me by helping my PhD student, Christos Benetatos, in implementing a web-based system called BachDuet, which allows users to play counterpoint improvisation with an AI agent in real time. This work resulted in a cool system[1]. Yongyi later expanded his interest into another area, namely audio deepfake detection and worked together with another PhD student, Neil Zhang. We have co-authored 1 journal paper and 5 peer-reviewed conference papers, among which he was the first author or co-first author for 5 of the papers. Yutong Wen was in AME Class of 2024. He started doing research with me in his junior year on head-related transfer function (HRTF) personalization with a PhD student, Neil Zhang, and later also worked on diffusion models for sound effect generation with another PhD student, Ge Zhu. He published two peer-reviewed conference papers with me and won the Donald M. Barnard Prize from the ECE Department at graduation. Kyungbok Lee is in Computer Science Class of 2025. He has been doing research with me since the summer before his senior year. He has published one paper as the first author at the IEEE International Workshop on Multimedia Signal Processing and submitted another paper to IEEE International Conference on Image Processing (ICIP). All these students have received a University Undergraduate Research Presentation Award to present their work at the conferences.

At the MS student level, I have mentored 7 students since my tenure promotion in 2020, 4 of which have co-authored peer-reviewed publications with me. All these publications were their first academic papers, and this also made me very proud. These students came from different backgrounds including ECE, Computer Science, Data Science, and Tech Entrepreneurship, and some students participated in our research remotely from other institutions.

At the PhD level, I have been running a large lab even during my sabbatical leave. My most recent PhD graduate was Bochen Li in August 2020. His dissertation titled “Multi-modal Analysis for Music Performances” won a 2021 Outstanding PhD Dissertation Award at the University of Rochester (only one in Engineering). Currently, I have 11 PhD students, where 5 students are finishing their dissertations and expected to graduate in 2025. The remaining 6 students (1 in 6th year, 1 in the 5th year, 2 in second year, 2 in first year) are making satisfactory progress on their research.

In addition to mentoring students directly, I was also the Co-PI of the NSF Research Experience for Undergraduates (REU) site program on Computational Methods for Understanding Music, Media, and the Minds from 2021 to 2023. Each year, this program hosted 10 undergraduate students from around the US for 10 weeks at the University of Rochester to conduct research. As the Co-PI, I co-managed the program, including student admission, lab placement, orientation, faculty communication, career pathways talk series, and final presentation of student projects. I also participated as senior personnel in the NSF NRT program on Interdisciplinary Graduate Training in the Science, Technology, and Applications of Augmented and Virtual Reality, where I contributed to discussions of various educational aspects of this program. I also served as the faculty mentor of three practicum projects, each of which had two or three PhD students from different departments working together.

Managing a large lab requires not only effort but also skills. I found that a “semi-supervised” way is the most effective. For PhD students, I set clear expectations at different stages of their PhD but also give them the maximal freedom to explore within the bounds set by funding sources. For example, I expect students to tackle a small-scale well-defined problem and submit a conference paper in the first year, dig

[1] https://bachduet.com/

deeper and define the thesis topic in the second and third years, and graduate with contributions summarized in three journal papers in the fifth/sixth year. I organize three weekly project meetings, for music, speech, and general audio, respectively, and meet with students in subgroups. I find this an efficient and effective way to track student progress, provide timely feedback, and to create a collaborative environment among the students. For mentoring MS and undergraduate students, I use my PhD students as a proxy. This protects my own time, gains PhD students mentoring experiences, and ensures MS/undergraduate students to receive timely guidance.

Educational Outreach

Beyond the University of Rochester, I have been invited to teach at other educational institutions. At the pre-college level, I gave a lecture on “Towards Human-Computer Collaborative Music Making” to the AI Club of Sutherland High School in Pittsford, NY. At the college level, I was invited as the primary lecturer to teach at the 2023 Winter School on Speech and Audio Processing (WiSSAP) in IIT-Kanpur in India. Within a week, I gave multiple lectures on “Introduction to Music Information Retrieval”, “Music Rhythm Analysis”, “Music Pitch Analysis”, and “Interactive Music Systems” to a group of about 50 students and faculty. At conferences, I co-delivered tutorials on “Computer Assisted Music Making Systems” at ISMIR 2023, “A Brief History of Interactive Music System” at the CAAI International Conference on Artificial Intelligence (CICAI) in 2022, and “Audio-Visual Scene Understanding” at CVPR 2021.

Conclusions and Future Plan

My teaching philosophy – fostering passion, promoting active learning, and tailoring to students’ aptitude – have been manifested through my diverse teaching, mentoring and outreach activities. I feel satisfied when I see my students grow. I feel proud when I see my students succeed and surpass me. In the future, I wish to expand my course repertoire by teaching other existing courses and creating new courses. These new courses include a seminar course on Generative AI for the Arts to survey advances in generative AI and its impacts on arts. Another new course I consider teaching is Machine Learning for Spatial Audio to introduce the emerging research area at the intersection of acoustics, signal processing, and machine learning. On mentoring, I will continue improving my mentoring skills and learning from my students. On educational outreach, I will continue my efforts to teach at different levels for different audiences.