

Bochen Li

<http://www.ece.rochester.edu/~bli23>

+1(917)-969-5428

bli23@ur.rochester.edu

INTERESTS

- Computer audition, music information retrieval, machine learning, audio-visual scene analysis

EDUCATION

University of Rochester

- PhD Candidate, Electrical and Computer Engineering (ECE)
Diploma from Eastman Community School, Jazz Piano Theory

Rochester, NY, USA

December 2019 (*Expected*)

December 2019 (*Expected*)

University of Science and Technology of China

- Bachelor of Science, Electronic Engineering and Information Science

Hefei, China

June 2014

KEY PROJECTS

- Music query by unconstrained video: cross-modal retrieval constraining on emotions. [\[url\]](#)
- Expressive visual performance generation from MIDI stream. [\[url\]](#)
- Audio-visual analysis of music performances. [\[url\]](#)

RESEARCH/INDUSTRIAL EXPERIENCE

University of Rochester - Audio Information Research Lab

Rochester, NY, USA

Research Assistant

September 2014 - Present

o Audio-visual scene analysis (music performance):

- * Dataset creation. Created the URMP dataset, which consists of 44 multi-modal, multi-instrument, multi-track music ensemble performance videos, to facilitate research on audio-visual analysis of music performance. [\[url\]](#)
- * Source association. A system to identify the affiliation between the players and the audio/score tracks in music ensemble performance videos by analyzing the temporal correspondence between visual motions and tone events.
- * Multi-pitch analysis. A system to improve multi-pitch estimation and streaming results in ensemble performances by constraints from visually detected play/non-play labels on each individual player.
- * Polyphonic vibrato analysis. A system to detect and parameterize vibrato articulations by visually analyzing the left-hand rolling motion on the string board for string instrumentalists from ensembles.
- * Source separation: Proposed to enhance audio source separation results by recognizing the performance motion features from visual scenes.
- * Cross-modal localization. Developed a two-stream network with attention model to spatially localize the sound source in video frame, and temporally localize the visual events in sound clip.

o Computer audition:

- * Designed a real-time audio-score alignment (score following) system for piano music based on the hidden Markov model and an observation model that reduces sustained effect (e.g., play with sustained-pedal) in piano.
- * Implemented an online time warping algorithm for real-time audio-to-audio alignment. The system was built in Java interface (along with modules of multi-thread recording, sound event detection, etc.) to serve the lyrics automatic scrolling function in a Rochester annual choir concert from 2016 to 2019, four years in a row. [\[url\]](#)
- * Designed core algorithms for [Mango Future](#), which launched iOS/Android Apps for music learners. Key techniques include score-informed tuning/rhythm rating, sound enhancement, sound synthesis, automatic accompaniment, etc.

Bytedance - Speech/Audio/Music Intelligence Team

Palo Alto, CA, USA

Research Intern

February 2019 - May 2019

- Implemented the state-of-the-art music source separation method. It separates the original recording of a pop song into vocals, percussion, bass, and other components. [\[url\]](#)

- **Spotify - Music Intelligence (MiQ) Team** New York, NY, USA
Research Intern *June 2018 - August 2018*

- Music-query by video. Developed a two-stream network to learn the cross-modal distance between music and unconstrained videos via latent emotion space, which includes audio/video emotion tagging branches and cross-modal distance learning framework. The model recommends a music/playlist given user-uploaded video clip.

- **Yamaha - Music AI Team** Hamamatsu, Shizuoka, Japan
Research Intern *October 2017 - December 2017*

- Visual performance generation. Developed a system to learn the music context of the given MIDI data (music score) and generate expressive whole-body visual performance as pianist skeleton key points, using convolutional and recurrent neural networks.

- **Knowles Intelligent Audio - Speech Interface Team** Mountain View, CA, USA
Intern *May 2017 - August 2017*

- Performed the keyword spotting process including data augmentation, training, parameter tuning, and testing.
- Developed the framework for talker ID recognition based on the Gaussian mixture models (GMM).

TEACHING EXPERIENCE

- **Guest lectures**

- “Audio-visual analysis for music performance” given at the *Computer Audition* course. *Fall 2018*
- “Machine learning for audio signal processing” given at the *Audio Signal Processing* course. *Spring 2018*

- **Upward bound program**

- Assisted the teaching of the *Music and Math* summer mini-course for pre-college students. *Summer 2016*

- **Teaching assistant at University of Rochester for:**

- *Audio Signal Processing* *Spring 2018*
- *Circuits and Signals* *Spring 2015*
- *Intro to C/C++ Programming* *Fall 2014*

PUBLICATIONS / PATENTS

- **Bochen Li** and Aparna Kumar, “Systems, Methods & Computer Program Products for Associating Media Content Having Different Modalities”, *U.S. Patent 16/439,626*. June 2019. (Patent filed with Spotify USA Inc.)
- **Bochen Li**, Karthik Dinesh, Chenliang Xu, Gaurav Sharma, and Zhiyao Duan, “Online Audio-Visual Source Association for Chamber Music Performances”, *Transactions of the International Society for Music Information Retrieval*, 2(1), pp.29-42, 2019.
- **Bochen Li** and Aparna Kumar, “Query by Video: Cross-modal Music Retrieval”, in *Proc. International Society for Music Information Retrieval (ISMIR)*, 2019.
- **Bochen Li***, Xinzhaoy Liu*, Karthik Dinesh, Zhiyao Duan, and Gaurav Sharma, “Creating A Musical Performance Dataset for Multimodal Music Analysis: Challenges, Insights, and Applications”, *IEEE Transactions on Multimedia*, vol. 21, no. 2, pp. 522-535, 2019. (* *Equal contribution*)
- **Bochen Li**, Akira Maezawa, and Zhiyao Duan, “Skeleton Plays Piano: Online Generation of Pianist body Movements from MIDI Performance”, in *Proc. International Society for Music Information Retrieval (ISMIR)*, 2018.

- **Bochen Li** and Akira Maezawa, “MIDI2Pose: Online keyboard performance motion generation from performance data”, in *Proc. Information Processing Society of Japan*, 2018.
- Yapeng Tian, Jing Shi, **Bochen Li**, Zhiyao Duan, and Chenliang Xu, “Audio-Visual Event Localization in Unconstrained Videos”, in *Proc. European Conference on Computer Vision (ECCV)*, 2018.
- Xueyang Wang, Ryan Stables, **Bochen Li**, and Zhiyao Duan, “Score-aligned Polyphonic Microtiming Estimation”, in *Proc. International Conference on Audio, Speech and Signal Processing (ICASSP)*, 2018.
- **Bochen Li**, Karthik Dinesh, Gaurav Sharma, and Zhiyao Duan, “Video-based Vibrato Detection and Analysis for Polyphonic String Music”, in *Proc. International Society for Music Information Retrieval (ISMIR)*, 2017. (**Best Paper Nomination**)
- **Bochen Li**, Chenliang Xu, and Zhiyao Duan, “Audio-visual Source Association for String Ensemble Videos through Multi-modal Vibrato Analysis”, in *Proc. Sound and Music Computing Conference*, 2017. (**Best Paper Award**)
- **Bochen Li**, Karthik Dinesh, Zhiyao Duan, and Gaurav Sharma, “See and Listen: Score-informed Association of Sound Tracks to Players in Chamber Music Performance Videos”, in *Proc. International Conference on Audio Speech and Signal Processing (ICASSP)*, 2017.
- Karthik Dinesh*, **Bochen Li***, Xinzhao Liu, Zhiyao Duan, and Gaurav Sharma, Visually Informed Multi-pitch Analysis of String Ensembles, in *Proc. International Conference on Audio, Speech and Signal Processing (ICASSP)*, 2017. (* Equal contribution)
- **Bochen Li** and Zhiyao Duan, “An approach to score following for piano performances with sustained effect”, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 12, 2016.
- **Bochen Li** and Zhiyao Duan, “Score following for piano performances with sustain-pedal effects”, in *Proc. International Society for Music Information Retrieval (ISMIR)*, 2015.

AWARDS

- **Best Paper Nomination**, 18th International Society for Music Information Retrieval (ISMIR) October, 2017
- **Best Paper Award**, 14th Sound and Music Computing Conference (SMC) July, 2017

ACADEMIC SERVICE

- Reviewer for journals
 - *Digital Audio Processing: A Review Journal*
 - *EURASIP Journal on Audio Speech and Music Processing*
 - *IEEE Transactions on Multimedia*
 - *EURASIP Journal on Audio, Speech, and Music Processing*
 - *IEEE Transactions on Music Information Retrieval*
 - *IEEE/ACM Transactions on Audio, Speech and Language Processing*
 - *IEEE Transactions on Affective Computing*
 - *The Journal of Supercomputing*
 - *Journal of Scientific Programming*
- Reviewer for conferences
 - *ACM Symposium on Applied Perception (SAP)*
 - *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*
 - *International Society for Music Information Retrieval (ISMIR)*
 - *ACM International Conference on Multimedia* (Program Committee Member)
 - *IEEE Technical Committee on Multimedia Computing*
 - *IEEE International Symposium on Multimedia*
- Reviewer for book chapters
 - *Audio Source Separation and Speech Enhancement*