

(Last updated on July 1, 2020)

## INTERESTS

Computer Audition, Music Information Retrieval, Audio-Visual Analysis, Machine Learning.

## CURRENT APPOINTMENT

**University of Rochester** – Rochester, NY, USA

*Associate Professor, Department of Electrical and Computer Engineering (primary)*

Jul. 2020 – present

Department of Computer Science (secondary)

Goergen Institute for Data Science (affiliated)

*(On Sabbatical Leave from University of Rochester from June 2020 to July 2021)*

**Kwai Inc.** – Bellevue, WA, USA

*Principal Research Scientist, Kwai Seattle AI Lab*

Jun. 2020 – present

## EDUCATION

**Northwestern University** - Evanston, IL, USA

August 2013

Ph.D., Department of Electrical Engineering and Computer Science

Thesis: *Computational Music Audio Scene Analysis*

Advisor: Bryan Pardo

**Tsinghua University** - Beijing, China

July 2008

Master of Science, Department of Automation

Thesis: *Research on Polyphonic Music Pitch Estimation*

Advisor: Changshui Zhang

**Tsinghua University** - Beijing, China

July 2004

Bachelor of Science, Department of Automation

Thesis: *Constructing an Assistant Training System for Long Jump*

Advisor: Changshui Zhang

## PROFESSIONAL EXPERIENCE

**University of Rochester** – Rochester, NY, USA

*Assistant Professor, Department of Electrical and Computer Engineering (primary)*

Jul. 2013 – Jun. 2020

*Assistant Professor, Department of Computer Science (secondary)*

Jun. 2016 – Jun. 2020

*Assistant Professor, Goergen Institute for Data Science (affiliated)*

Jun. 2016 – Jun. 2020

**Ohio State University** - Columbus, OH, USA

Feb. 2013 – Mar. 2013

*Visiting Researcher, Department of Computer Science and Engineering*

- Investigated application of deep learning in speech and audio signal processing

Advisor: DeLiang Wang

**Northwestern University** - Evanston, IL, USA

Sep. 2008 – Jun. 2013

*Research Assistant, Department of Electrical Engineering and Computer Science*

- Developed machine learning algorithms towards audio information retrieval applications, e.g. multi-pitch

estimation and tracking of music and speech, audio-score alignment, source separation, etc.

Advisor: Bryan Pardo

**Adobe Systems** - San Francisco, CA, USA Jun. 2011 - Dec. 2011

*Research Intern, [Advanced Technology Labs \(ATL\)](#)*

- Invented an online machine learning algorithm for real-time semi-supervised source separation, with an application on real-time speech enhancement in non-stationary noise environments

Advisors: Gautham J. Mysore and Paris Smaragdis

**Microsoft Research Asia** - Beijing, China Jul. 2007 - Apr. 2008

*Research Intern, [Speech Group](#)*

- Designed algorithms for music tagging and tonality classification for an automatic music recommendation system

Advisor: Lie Lu

**Stanford University** - Stanford, CA, USA Apr. 2007 - Jun. 2007

*Visiting Researcher, [Center for Computer Research in Music and Acoustics \(CCRMA\)](#)*

- Implemented and compared audio signal processing algorithms for extracting guitar excitation signals

Advisor: Julius O. Smith III

**Tsinghua University** - Beijing, China Sep. 2005 - Mar. 2007

*Research Assistant, [State Key Laboratory of Intelligent Technology and Systems](#)*

- Developed machine learning algorithms towards audio information retrieval applications, e.g. multi-pitch estimation and source separation

Advisor: Changshui Zhang

**NTP CO., LTD** - Shenzhen, Guangdong, China Jul. 2003 - Aug. 2003

*Software and Hardware Developer*

- Developed and tested a motor control system

## RESEARCH FUNDING

**NRT-HDR: Interdisciplinary Graduate Training in the Science, Technology, and Applications of Augmented and Virtual Reality** 09/01/2019 – 08/31/2024

National Science Foundation – NRT (\$1,560,000)

PI: Mujdat Cetin, Co-PIs: Jannick Rolland, Michele Rucci, Zhen Bai

Senior Personnel: Zhiyao Duan, Ross Maddox, Andrew White, Chenliang Xu, and Yuhao Zhu

**REU Site: Computational Methods for Understanding Music, Media, and Minds** 03/01/2020 – 02/28/2023

National Science Foundation – REU Site (\$405,000)

PI: Ajay Anand, Co-PI: Zhiyao Duan

**End-to-End Speaker Verification** 09/01/2019 – 08/31/2020

Voice Biometric Group – Unrestricted Gift

PI: Zhiyao Duan (\$37,531)

**Visually Informed Music Generation** 01/01/2020 – 12/31/2020

Kwai Inc. – Unrestricted Gift

PI: Zhiyao Duan (\$25,000)

<b>Music Understanding and Generation</b> ByteDance – Unrestricted Gift PI: Zhiyao Duan (\$50,000)	10/15/2019 – 10/14/2020
<b>CAREER: Human-Computer Collaborative Music Making</b> National Science Foundation – CISE IIS-CHS core program PI: Zhiyao Duan (\$499,219)	06/01/2019 – 05/31/2024
<b>BIGDATA: F: Audio-Visual Scene Understanding</b> National Science Foundation – Big Data Science & Engineering PI: Chenliang Xu (\$349,999), Co-PI: Zhiyao Duan (\$300,000)	09/01/2017 – 08/31/2021
<b>Real-Time Synthesis of a Virtual Talking Face from Acoustic Speech</b> University of Rochester AR/VR Pilot Funding (\$50,000) PIs: Ross Maddox, Zhiyao Duan, and Chenliang Xu	07/01/2017 – 06/30/2018
<b>Adding High-quality Spatial Audio to 3D-VR-360 Recordings for Live Streaming and Building a VR Video Database</b> University of Rochester AR/VR Pilot Funding (\$69,800) PIs: Zhiyao Duan, Ming-Lun Lee, and Matthew Brown	07/01/2017 – 06/30/2018
<b>Development and Evaluation of an Evidence-Based Mobile Health Caregiver Intervention for FASD</b> National Institute of Health (\$1,504,884) PIs: Christie Petrenko and Cristiano Tapparello; Co-Is: Heather Olson, Wendi Heinzelman, and Zhiyao Duan	07/01/2017 – 05/31/2022
<b>Algorithms for Query by Example of Audio Databases</b> National Science Foundation – CISE III core program PI: Zhiyao Duan (\$299,775), Co-PI: Bryan Pardo (\$199,996)	09/01/2016 – 08/31/2019
<b>Predicting Adverse Events from Cardiac Signals using Deep Neural Networks</b> University of Rochester Goergen Institute for Data Science Collaborative Pilot Award Program in Health Analytics PI: Mina Attin (\$26,995), Co-PI: Zhiyao Duan (\$19,701)	08/22/2016 – 08/21/2017

## TEACHING

### Tutorials

[2] Audiovisual Music Processing, co-presented with Slim Essid, Bochen Li, and Sanjeel Parekh International Society for Music Information Retrieval conference (ISMIR), Delft, The Netherlands	Nov. 2019
[1] Automatic Music Transcription, co-presented with Emmanouil Benetos International Society for Music Information Retrieval conference (ISMIR), Malaga, Spain	Oct. 2015

### Courses Designed

[5] Music and Math, Pre-College Level <i>Instructor</i> , Upward Bound Program, University of Rochester, Rochester, NY, USA	Summer 2016, 2017, 2019
[4] ECE 477: Computer Audition, Grad Level <i>Instructor</i> , University of Rochester, Rochester, NY, USA	Fall 2014, 2015, 2017-2019

[3] Y0250421: Computer Audition, Grad Level Summer, 2015  
*Instructor*, Tsinghua University, Beijing, China

[2] ECE 272/472: Audio Signal Processing, Undergrad/Grad Level Spring 2014-2020  
*Instructor*, University of Rochester, Rochester, NY, USA

[1] ECE 492: Computer Audition and Its Applications in Music, Grad Level Fall 2013  
*Instructor*, University of Rochester, Rochester, NY, USA

## Courses Involved

[7] CSC 294: AR/VR Interaction Design Fall 2019  
*Guest Lecturer*, University of Rochester, Rochester, NY, USA

- Designed and gave a lecture on Computer Audition and Music Interaction

[6] CSC 249/449: Machine Vision Spring 2018  
*Guest Lecturer*, University of Rochester, Rochester, NY, USA

- Designed and gave a lecture on Multi-Modal Music Scene Understanding

[5] CSC 412: Human Computer Interaction Fall 2013  
*Guest Lecturer*, University of Rochester, Rochester, NY, USA

- Designed and gave a lecture on Music Interaction

[4] EECS 349: Machine Learning Fall 2010, 2011, 2012  
*Teaching Assistant and Guest Lecturer*, Northwestern University, Evanston, IL, USA

- Designed and gave lectures on Ensemble Learning, Memory-based Learning, Gaussian Mixture Models, and Expectation-Maximization;
- Designed homework problems on the above topics and decision trees
- Held office hours; graded homework, exams and final projects

[3] Introduction to Artificial Intelligence Fall 2007  
*Teaching Assistant*, Tsinghua University, Beijing, China

- Held office hours, graded homework and final projects

[2] Object-Oriented Computer Programming (Visual C++) Fall 2007  
*Lab Instructor*, Tsinghua University, Beijing China

- Led weekly lab sessions
- Mentored students on final projects; graded homework and final projects

[1] Fundamentals of Computer Programming (C++) Spring 2006

- Led weekly lab sessions
- Mentored students on final projects; graded homework and final projects

## Visiting Scholar Hosting

- Hongjuan Zhang (Shanghai University, China, December 2016 – December 2017)
- Jun Zhou (Southwest University, China, September 2014 – August 2015)

## Doctoral Thesis Supervising

## Current

- Mojtaba Heydari (expected June 2024)
- Frank Cwitkowitz (expected June 2024)
- You (Neil) Zhang (expected June 2024)
- Christos Benetatos (expected June 2023)
- Ge Zhu (expected June 2023)
- Yujia Yan (expected June 2022)
- Fei Jiang (expected July 2021), visiting student from Beijing Institute of Technology
- Nan Jiang (expected August 2020), visiting student from Tsinghua University
- Bochen Li (expected February 2020)

## Graduated

- Yichi Zhang (graduated December 2019)
- Sefik Emre Eskimez (graduated August 2019), co-supervised with Prof. Wendi Heinzelman
- Rui Lu (graduated July 2019), visiting student from Tsinghua University
- Andrea Cogliati (graduated December 2017)

## Doctoral Thesis Reading

- Sahar Hashemgeloogardi (ECE, November 2019)
- Jianbo Yuan (CS, September 2019)
- Sarah Smith (ECE, May 2019)
- Priyanga Gunarathne (Simon Business School, May 2018)
- Xiaochang Peng (CS, May 2018)
- Chen Wang (ECE, December 2017)
- Ahmed Elliethy (ECE, February 2017)
- Dave Anderson (ECE, January 2017)
- Gang Ren (ECE, November 2015)
- He Ba (ECE, February 2015)
- Na Yang (ECE, March 2015)

## Master's Student Advising

Mingqing Yun, ECE, U Rochester – Multi-talker Localization in Reverberant Environments	01/18-05/19
Isaac Mosebrook, ECE, U Rochester – Parkinson's Disease Diagnosis from Voice	09/18-05/19
Hilary Mogul, ECE, U Rochester – Beamforming with Ambisonics	Fall 2018
Yufei Zhang, ECE, U Rochester – Audio-Visual Pitch Estimation	Summer 2018
Yiming Zhao, ECE, U Rochester – Automatic Lyrics Display System Design	06/17-05/18
Zhuohuang Zhang, ECE, U Rochester – Speaker Diarization	09/16-05/17
Iris Yuping Ren, ECE, U Rochester – Singing Tutoring System	09/15-12/17
Jonathan Downing, ECE, U Rochester – Thesis on Joint Separation and Dereverberation	01/16-08/16
Jay Biernat, ECE, U Rochester – Music Visualization	Summer 2016
Marko Stamenovic, ECE, U Rochester – Music Recommendation	Summer 2016
Xinzhao Liu, ECE, U Rochester – Thesis on Audio-Visual Music Performance Analysis	05/15-05/16
Hanqing Wen, ECE, U Rochester – Music Onset Detection	06/15-05/16
Zejin Li, ECE, U Rochester – Audio-Visual Guitar Transcription	Summer 2015
Yuhui Chen, ECE, U Rochester – Audio-Visual Guitar Transcription	Summer 2015
Shumin Xu, ECE, U Rochester – Music Universe	09/14-05/15
David Heid, ECE, U Rochester – Vibrato Analysis	01/15-05/15
Andrew Trahan, ECE, U Rochester – Thesis on Drum Kit Transcription	Spring 2014

Jonathan Springer, EECS, Northwestern University – Bird Species Recognition Fall 2012  
Jesse Bowman, EECS, Northwestern University – Real-time Multi-pitch Estimation for Guitars 07/2018-06/2011

## Undergraduate Research Advising

Mingrui Yuan, EE, Tsinghua University, Visiting Student	Summer 2019
Hangyu Li, EE, Beihang University, Visiting Student	Summer 2019
Yinghao Ma, Math, Peking University, Visiting Student	Summer 2019
Daniel Dopp, Kentucky University, NSF REU Site on Music, Media and Mind	Summer 2019
Nick Creel, Marlboro College, NSF REU Site on Music, Media and Mind	Summer 2019
Yiting Zhang, ECE, University of Rochester, Xerox Engineering Fellowship	Summer 2018 – Spring 2019
Peizhe Gao, CS, Beihang University, Visiting Student	Summer 2018
Junyi Fan, China University of Geosciences, Visiting Student	Spring and Summer 2018
Zhihan Zhou, Math, Zhejiang University, Visiting Student	Summer 2017
Andrew Smith, CS, University of Central Florida, NSF REU Site on Music, Media and Mind	Summer 2017
Arlen Fan, ECE, University of Rochester, NSF REU Site on Music, Media and Mind	Summer 2017
Ayumi Yuasa, ECE, University of Rochester, NSF REU Student	Summer 2017
Ryan Bhular, ECE University of Rochester, NSF REU Student	Summer 2017
Meixiao Han, ECE, University of Rochester, Undergraduate Researcher	Summer 2017
Yukun Chen, ECE, University of Rochester, Undergraduate Researcher	Summer and Fall 2016
Karan Vombatkere, ECE, University of Rochester, Xerox Engineering Fellowship	Summer 2016
Steven Belitzky, ECE, University of Rochester, Discover Grant Researcher	Summer 2016
Almas Abdibayev, CS, Nazarbayev University, Visiting Student	Summer 2015
Ibrahim Akbar, ECE, University of Rochester, Undergraduate Researcher	Spring 2015
Haowen Pan, ECE, University of Rochester, Xerox Engineering Fellowship	Summer 2014
Prem Seetharaman, EECS, Northwestern University	Winter 2012
Jiawei Lyu, Automation, Tsinghua University	Spring 2008

## PUBLICATIONS

### Book Chapters

- [2] Bryan Pardo, Antoine Liutkus, **Zhiyao Duan**, Gaël Richard, “Applying source separation to music,” in *Audio Source Separation and Speech Enhancement*, eds. E. Vincent, T. Virtanen, S. Gannot. Wiley, pp. 347-376, 2018.
- [1] Bryan Pardo, Zafar Rafii, and **Zhiyao Duan**, “Audio source separation in a musical context,” in *Springer Handbook of Systematic Musicology*, ed. Rolf Bader. Springer-Verlag Berlin Heidelberg, pp. 285-298, 2018.

### Journal Publications

- [23] Sefik Emre Eskimez, Ross Maddox, Chenliang Xu, and **Zhiyao Duan**, “Noise-resilient training method for face landmark generation from speech,” *IEEE/ACM Transactions on Audio Speech and Language Processing*, accepted with minor revision, 2019.
- [22] Bochen Li, Karthik Denish, 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*, vol. 2, no. 1, pp.29–42, 2019.
- [21] Rui Lu, **Zhiyao Duan**, and Changshui Zhang, “Audio-visual deep clustering for speech separation,” *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 27, no. 11, pp. 1697-1712, 2019.
- [20] Sefik Emre Eskimez, Kazuhito Koishida, and **Zhiyao Duan**, “Adversarial training for speech super-resolution,” accepted by *IEEE Journal of Selected Topics in Signal Processing*.

- 
- [19] **Zhiyao Duan\***, Slim Essid\*, Cynthia C. S. Liem\*, Gaël Richard\*, “Audio-visual analysis of music performances,” *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 63-73, 2019. (\* authors in alphabetical order)
- [18] Emmanouil Benetos\*, Simon Dixon\*, **Zhiyao Duan\***, and Sebastian Ewert\*, “Automatic music transcription: an overview,” *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 20-30, 2019. (\* authors in alphabetical order)
- [17] Yichi Zhang, Bryan Pardo, and **Zhiyao Duan**, Siamese style convolutional neural networks for sound search by vocal imitation, accepted by *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 27, no. 2, pp. 429-441, 2019.
- [16] Bochen Li\*, Xinzhao Liu\*, Karthik Dinesh, **Zhiyao Duan**, and Gaurav Sharma, “Creating a multi-track classical music performance dataset for multi-modal music analysis: challenges, insights, and applications,” *IEEE Transactions on Multimedia*, vol., 21, no. 2, pp. 522-535, 2019. (\* equal contribution)
- [15] Rui Lu, **Zhiyao Duan**, and Changshui Zhang, “Listen and look: audio-visual matching assisted speech source separation”, *IEEE Signal Processing Letters*, vol. 25, no. 9, 2018.
- [14] Sefik Emre Eskimez, Peter Soufleris, **Zhiyao Duan**, and Wendi Heinzelman, “Front-end speech enhancement for commercial speaker verification systems,” *Speech Communication*, vol. 99, no. pp. 101-113, 2018.
- [13] Shiwei Yu, Hongjuan Zhang, and **Zhiyao Duan**, “Singing voice separation by low-rank and sparse spectrogram decomposition with pre-learned dictionaries,” *Journal of the Audio Engineering Society*, vol. 65, no. 5, pp. 377-388, 2017.
- [12] Andrea Cogliati, **Zhiyao Duan**, and Brendt Wohlberg, “Piano transcription with convolutional sparse lateral inhibition,” *IEEE Signal Processing Letters*, vol. 24, no. 4, pp. 392-396, 2017.
- [11] David Temperley, Iris Ren, and **Zhiyao Duan**, “Mediant mixture and ‘blue notes’ in rock: An exploratory study,” accepted by *Music Theory Online*, 2017.
- [10] Na Yang, Jianbo Yuan, Yun Zhou, Ilker Demirkol, **Zhiyao Duan**, Wendi Heinzelman, and Melissa Sturge-Apple, “Enhanced multiclass SVM with thresholding fusion for speech-based emotion classification,” *International Journal of Speech Technology*, vol. 20, no. 1, pp. 27-41, 2017. DOI: 10.1007/s10772-016-9364-2.
- [9] Bochen Li and **Zhiyao Duan**, “An approach to score following for piano performances with the sustained effect,” *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 24, no. 12, pp. 2425-2438, 2016.
- [8] Andrea Cogliati, **Zhiyao Duan**, and Brendt Wohlberg, “Context-dependent piano music transcription with convolutional sparse coding,” *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 24, no. 12, pp. 2218-2230, 2016.
- [7] Yichi Zhang and **Zhiyao Duan**, “Supervised and unsupervised sound retrieval by vocal imitation,” *Journal of Audio Engineering Society*, vol. 64, no. 7/8, pp. 533-543, 2016.
- [6] Francisco J. Rodriguez-Serrano, **Zhiyao Duan**, Pedro Vera-Candeas, Bryan Pardo, and Julio J. Carabias-Orti, “Online score-informed source separation with adaptive instrument models,” *Journal of New Music Research*, vol., 44, no. 2, pp., 83-96, 2015. DOI: 10.1080/09298215.2014.989174.
- [5] Zafar Rafii, **Zhiyao Duan**, and Bryan Pardo, “Combining rhythm-based and pitch-based methods for background and melody separation,” *IEEE Trans. Audio Speech Language Process.*, vol. 22, no. 12, pp. 1884-1893, 2014.
- [4] **Zhiyao Duan**, Jinyu Han, and Bryan Pardo, “Multi-pitch streaming of harmonic sound mixtures,” *IEEE Trans. Audio Speech Language Process.*, vol. 22, no. 1, pp. 138-150, 2014.
- [3] **Zhiyao Duan** and Bryan Pardo, “Soundprism: an online system for score-informed source separation of music audio,” *IEEE Journal of Selected Topics in Signal Processing.*, vol. 5, no. 6, pp. 1205-1215, 2011.
- [2] **Zhiyao Duan**, Bryan Pardo, and Changshui Zhang, “Multiple fundamental frequency estimation by modeling

spectral peaks and non-peak regions,” *IEEE Trans. Audio Speech Language Process.*, vol. 18, no. 8, pp. 2121-2133, 2010.

[1] **Zhiyao Duan**, Yungang Zhang, Changshui Zhang, and Zhenwei Shi, “Unsupervised single-channel music source separation by average harmonic structure modeling,” *IEEE Trans. Audio Speech Language Process.*, vol. 16, no. 4, pp. 766-778, 2008.

## Peer-reviewed Conference Publications

[48] Christodoulos Benetatos, Joseph VanderStel, and **Zhiyao Duan**, “BachDuet: A deep learning system for human-machine counterpoint improvisation,” in *Proc. International Conference on New Interfaces for Musical Expression (NIME)*, 2020.

[47] Sefik Emre Eskimez, Ross Maddox, Chenliang Xu, and **Zhiyao Duan**, “End-to-end generation of talking faces from noisy speech,” in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[46] Yichi Zhang, Junbo Hu, Yiting Zhang, Bryan Pardo, and **Zhiyao Duan**, “Vroom!: A search engine for sounds by vocal imitation queries,” in *Proc. ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR)*, 2020. (oral presentation)

[45] Nan Jiang, Sheng Jin, **Zhiyao Duan**, and Changshui Zhang, “RL-Duet: Online music accompaniment generation using deep reinforcement learning,” accepted by *AAAI*, 2020. (oral/poster presentation)

[44] Lele Chen, Ross K. Maddox, **Zhiyao Duan**, and Chenliang Xu, “Hierarchical cross-modal talking face generation with dynamic pixel-wise loss,” in *Proc. CVPR*, 2019.

[43] Bongjun Kim, Madhav Ghei, Bryan Pardo, and **Zhiyao Duan**, “Vocal Imitation Set: a dataset of vocally imitated sound events using the AudioSet ontology,” in *Proc. of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE)*, 2018, pp. 148-152.

[42] Yapeng Tian, Jing Shi, Bochen Li, **Zhiyao Duan**, and Chenliang Xu, “Audio-visual event localization in unconstrained videos,” accepted by *European Conference on Computer Vision (ECCV)*, 2018, pp. 247-263.

[41] Lele Chen, Zhiheng Li, Ross Maddox, **Zhiyao Duan**, and Chenliang Xu, “Lip movements generation at a glance,” accepted by *European Conference on Computer Vision (ECCV)*, 2018, pp. 520-535.

[40] Bochen Li, Akira Maezawa, and **Zhiyao Duan**, “Skeleton plays piano: online generation of pianist body movements from MIDI performance,” accepted by *International Society for Music Information Retrieval Conference (ISMIR)*, 2018, pp. 218-224 (oral/poster presentation).

[39] Yujia Yan, Ethan Lustig, Joseph Vanderstel, and **Zhiyao Duan**, “Part-invariant model for music generation and harmonization,” accepted by *International Society for Music Information Retrieval Conference (ISMIR)*, 2018, pp. 204-210. (oral/poster presentation)

[38] Sefik Emre Eskimez, Ross K. Maddox, Chenliang Xu, and **Zhiyao Duan**, “Generating talking face landmarks from speech,” in *Proc. International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, 2018. (poster presentation)

[37] Zhihan Zhou, Yichi Zhang, and **Zhiyao Duan**, “Joint speaker diarization and recognition using convolutional and recurrent neural networks,” accepted by *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2496-2500. (poster presentation)

[36] Xueyang Wang, Ryan Stables, Bochen Li, and **Zhiyao Duan**, “Score-aligned polyphonic microtiming estimation,” accepted by *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 361-365. (poster presentation)

[35] Sefik Emre Eskimez, **Zhiyao Duan**, and Wendi Heinzelman, “Unsupervised learning approach to feature analysis for automatic speech emotion recognition,” accepted by *IEEE International Conference on Acoustics,*



*Speech and Signal Processing (ICASSP)*, 2018, pp. 5099-5103. (poster presentation)

[34] Yichi Zhang and **Zhiyao Duan**, “Visualization and interpretation of Siamese style convolutional neural networks for sound search by vocal imitation,” accepted by *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018. (oral presentation)

[33] Rui Lu, **Zhiyao Duan**, and Changshui Zhang, “Multi-scale recurrent neural network for sound event detection,” accepted by *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 131-135. (oral presentation)

[32] Lele Chen, Sudhanshu Srivastava, **Zhiyao Duan**, and Chenliang Xu, “Deep cross-modal audio-visual generation,” accepted by *ACM Multimedia Thematic Workshops*, 2017. (poster presentation)

[31] Yichi Zhang and **Zhiyao Duan**, “IMINET: convolutional semi-siamese networks for sound search by vocal imitation,” accepted by IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2017. (poster presentation)

[30] Rui Lu, **Zhiyao Duan**, and Changshui Zhang, “Metric learning based data augmentation for environmental sound classification,” accepted by IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2017. (oral presentation)

[29] Bochen Li, Karthik Dinesh, Gaurav Sharma, and **Zhiyao Duan**, “Video-based vibrato detection and analysis for polyphonic string music,” accepted by International Society for Music Information Retrieval Conference (ISMIR), 2017. (oral presentation) (**best paper nomination**)

[28] Andrea Cogliati and **Zhiyao Duan**, “A metric for music notation transcription accuracy,” accepted by International Society for Music Information Retrieval Conference (ISMIR), 2017. (poster presentation)

[27] Bochen Li, Chenliang Xu, and **Zhiyao Duan**, “Audio-visual source association for string ensembles through multi-modal vibrato analysis,” in *Proc. 14th Sound and Computing Conference (SMC)*, 2017. (oral presentation) (**best paper award**)

[26] 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,” accepted by *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017. (oral presentation)

[25] Karthik Dinesh\*, Bochen Li\*, Xinzhao Liu, **Zhiyao Duan**, and Gaurav Sharma, “Visually informed multi-pitch analysis of string ensembles,” accepted by *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017. (\* equal contribution) (poster presentation)

[24] Rui Lu, Kailun Wu, **Zhiyao Duan**, and Changshui Zhang, “Deep ranking: triplet MatchNet for music metric learning,” accepted by *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017. (oral presentation)

[23] Sefik Emre Eskimez, Melissa Sturge-Appley, **Zhiyao Duan**, and Wendi Heinzelman, “WISE: web-based interactive speech emotion classification,” accepted by 4th Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), 2016. (oral presentation)

[22] Andrea Cogliati, David Temperley, and **Zhiyao Duan**, “Transcribing human piano performances into music notation,” in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2016. (poster presentation)

[21] Sefik Emre Eskimez, Kenneth Imade, Na Yang, Melissa Sturge-Apple, **Zhiyao Duan**, and Wendi Heinzelman, “Emotion classification: How does an automated system compare to naive human coders?,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016. (oral presentation)

[20] Yichi Zhang and **Zhiyao Duan**, “IMISOUND: An unsupervised system for sound query by vocal imitation,”

in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016. (oral presentation)

[19] Andrea Cogliati, **Zhiyao Duan**, Brendt Wohlberg, “Piano music transcription with fast convolutional sparse coding,” in *Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2015. (poster presentation)

[18] Yichi Zhang and **Zhiyao Duan**, “Retrieving sounds by vocal imitation recognition,” in *Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2015. (poster presentation)

[17] Jun Zhou, Shuo Chen, and **Zhiyao Duan**, “Rotational reset strategy for online semi-supervised NMF-based speech enhancement for long recordings,” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2015. (poster presentation)

[16] Bochen Li and **Zhiyao Duan**, “Score following for piano performances with sustain-pedal effects,” in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2015, pp. 469-475. (poster presentation)

[15] Andrea Cogliati and **Zhiyao Duan**, “Piano music transcription modeling note temporal evolution,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 429-433. (poster presentation)

[14] **Zhiyao Duan** and David Temperley, “Note-level music transcription by maximum likelihood sampling,” in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2014, pp. 181-186. (oral presentation)

[13] **Zhiyao Duan**, Bryan Pardo, Laurent Daudet, “A novel cepstral representation for timbre modeling of sound sources in polyphonic mixtures,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 7495-7499. (poster presentation)

[12] Jonathan Springer, **Zhiyao Duan** and Bryan Pardo, “Approaches to multiple concurrent species bird song recognition,” in the *2nd International Workshop on Machine Listening in Multisource Environments (CHIME)*, 2013. (poster presentation)

[11] **Zhiyao Duan**, Gautham Mysore and Paris Smaragdis, “Speech enhancement by online non-negative spectrogram decomposition in non-stationary noise environments,” in *Proc. InterSpeech*, 2012, Portland, Oregon. (oral presentation)

[10] **Zhiyao Duan**, Gautham Mysore and Paris Smaragdis, “Online PLCA for real-time semi-supervised source separation,” in *Proc. International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, LNCS 7191, pp. 34-41, 2012. (oral presentation)

[9] **Zhiyao Duan** and Bryan Pardo, “Aligning semi-improvised music audio with its lead sheet,” in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2011, pp. 513-518. (poster presentation)

[8] **Zhiyao Duan** and Bryan Pardo, “A state space model for online polyphonic audio-score alignment,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 197-200. (poster presentation)

[7] **Zhiyao Duan**, Jinyu Han and Bryan Pardo, “Song-level multi-pitch tracking by heavily constrained clustering,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2010, pp. 57-60. (oral presentation)

[6] **Zhiyao Duan**, Jinyu Han, and Bryan Pardo, “Harmonically informed multi-pitch tracking,” in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2009, pp. 333-338. (oral presentation)

[5] **Zhiyao Duan**, Lie Lu, and Changshui Zhang, “Collective annotation of music from multiple semantic categories,” in *Proc. International Conference on Music Information Retrieval (ISMIR)*, 2008, pp. 237-242. (poster presentation)

presentation)

[4] **Zhiyao Duan**, Lie Lu, and Changshui Zhang, “Audio tonality mode classification without tonic annotations,” in *Proc. International Conference on Multimedia & Expo (ICME)*, 2008, pp. 1361-1364. (poster presentation)

[3] **Zhiyao Duan**, Changshui Zhang, “A maximum likelihood approach to multiple fundamental frequency estimation from the amplitude spectrum peaks,” in *Music, Brain and Cognition (MBC) workshop in the Twenty-first Annual Conference on Neural Information Processing Systems (NIPS)*, 2007. (spotlight and poster presentation)

[2] **Zhiyao Duan**, Dan Zhang, Changshui Zhang, and Zhenwei Shi, “Multi-pitch estimation based on partial event and support transfer,” in *Proc. International Conference on Multimedia & Expo (ICME)*, 2007, pp.216-219. (poster presentation)

[1] Nelson Lee, **Zhiyao Duan**, and Julius O. Smith, “Excitation signal extraction for guitar tones,” in *Proc. International Computer Music Conference (ICMC)*, 2007, pp. 450-457.

## Patents

[2] Andrea Cogliati, **Zhiyao Duan**, and Brendt Wohlberg, “Context-dependent piano music transcription with convolutional sparse coding,” U.S. Patent 9779706, issued in September 2017.

[1] Gautham J. Mysore, Paris Smaragdis, and **Zhiyao Duan**, “Online Source Separation,” U.S. Patent US 2013/0121506A1.

## INVITED TALKS

[23] *Towards Human-Computer Collaborative Music Making*  
Invited Keynote Talk, Midwest Music and Audio Day – Bloomington, NY Jun. 2019

[22] *Computer Audition and Its Potential Application in Digital Health*  
International Symposium for Digital Health – Hong Kong Jun. 2019

[21] *Audio Information Research Lab Overview*  
North East Music Information Special Interest Group (NEMISIG) – Brooklyn, NY Feb. 2019  
North East Music Information Special Interest Group (NEMISIG) – Rochester, NY Feb. 2017  
North East Music Information Special Interest Group (NEMISIG) – Philadelphia, PA Feb. 2015  
North East Music Information Special Interest Group (NEMISIG) – Ithaca, NY Feb. 2014

[20] *Audio-Visual Analysis of Music Performances: State and Beyond*  
Music and Audio Research Laboratory (MARL), New York University – Brooklyn, NY Feb. 2019

[19] *Some Thoughts about Singing*  
Dagstuhl Seminar 19052 – Dagstuhl, Germany Jan. 2019

[18] *Deep Learning for Health and Wellbeing Being Diagnosis*  
World University Network Meeting on Digital Health – Sydney, Australia Dec. 2018

[17] *Toward Machine Musicianship*  
Upstate New York Sound Meetup – Ithaca, NY Aug. 2018  
Birmingham City University – Birmingham, UK Aug. 2018

- [16] *Multimodal Music Scene Analysis*  
Tencent AI Lab – Seattle, WA May 2018  
State University of New York at Fredonia – Fredonia, NY Mar. 2018  
SUSTC, Dept. of Computer Science – Shenzhen, China May 2017  
Fudan University, School of Computer Science – Shanghai, China May 2017
- [15] *Teaching Machines to Listen*  
Upstate New York Sound Meetup – Rochester, NY Aug. 2017  
USTC, School of Computer Science and Technology – Hefei, China May 2017
- [14] *Transcribing Piano Music in the Time Domain into Music Notation*  
Joint Meeting of the Acoust. Society of America and Acoust. Society of Japan – Honolulu, HI Dec. 2016
- [13] *Towards Complete Music Notation Transcription of Piano*  
Western New York Image and Signal Processing Workshop (WNYISPW) – Rochester, NY Nov. 2016
- [12] *The Machine Musicianship: Automatic Music Transcription*  
Beihang University, Image Processing Center – Beijing, China Nov. 2016
- [11] *Enriching Sound Interactions through Computer Audition*  
Shanghai Jiao Tong University, Dept. of Computer Science and Engineering – Shanghai, China May 2017  
Peking University, Advanced Data & Signal Processing Laboratory – Shenzhen, China May 2017  
Indiana University Bloomington, Department of Computer Science – Bloomington, IN Sep. 2016
- [10] *Retrieving Sounds through Vocal Imitation*  
The 3<sup>rd</sup> Rochester Interdisciplinary Audio Engineering Symposium (RIAES) – Rochester, NY Aug. 2016  
University of Rochester Goergen Institute for Data Science Symposium - Rochester, NY June 2018
- [9] *Computational Music Scene Analysis*  
RIT, Center for Applied and Computational Mathematics – Rochester, NY Mar. 2016  
Shanghai University, Department of Mathematics – Shanghai, China Mar. 2016
- [8] *Tutorial on Automatic Music Transcription*, co-presented with Emmanouil Benetos  
International Society for Music Information Retrieval conference (ISMIR) – Malaga, Spain Oct. 2015
- [7] *Computational Music Audio Scene Analysis*  
Auditory Attention and Scene Analysis workshop and summer school – Delmenhorst, Germany Jul. 2014
- [6] *Note-Level Music Transcription by Maximum Likelihood Sampling*  
International Audio Labs Erlangen – Erlangen, Germany Jul. 2014  
1<sup>st</sup> Rochester Interdisciplinary Audio Engineering Symposium (RIAES) – Rochester, NY Jun. 2014
- [5] *Combining Data-driven and Knowledge-driven Models for Automatic Music Transcription*  
2<sup>nd</sup> Midwest Music Information Retrieval Gathering (MMIRG) – Evanston, IL Jun. 2014
- [4] *Transcribing the Pitch Content of Polyphonic Music Audio*  
IEEE Signal Processing Society Rochester Chapter IEEE Day Seminar – Rochester, NY Oct. 2013
- [3] *Computer Audition: Analyzing Complex Auditory Scenes*

---

University of Rochester, Department of Electrical and Computer Engineering – Rochester, NY	Apr. 2013
The Ohio State University, Department of Computer Science and Engineering – Columbus, OH	Mar. 2013
Northwestern University, Department of EECS – Evanston, IL	Jan. 2013
[2] <i>Music Audio Scene Analysis Informed by a Score</i>	
Ohio State University, Department of Computer Science and Engineering – Columbus, OH	May 2012
Northwestern University, Department of EECS – Evanston, IL	May 2012
[1] <i>An Approach to Multi-Pitch Tracking of Polyphonic Music</i>	
Dolby Laboratories – Beijing, China	Dec. 2011
Tsinghua University, Department of Automation – Beijing, China	Dec. 2011
Peking University, Institute of Computer Science and Technology – Beijing, China	Dec. 2011
Stanford University, Center for Computer Research in Music and Acoustics – Stanford, CA	Aug. 2011

## HONORS AND AWARDS

NSF CAREER Award	Mar. 2019
Best Paper Nomination at ISMIR 2017	Oct. 2017
Best Paper Award in the 2017 Sound and Music Computing (SMC) Conference	Jul. 2017
Terminal Year Fellowship in Northwestern University	2012-13
Chinese Government Award for Outstanding Self-Financed Students Abroad	Jun. 2011
Walter P. Murphy Fellowship in Northwestern University	2008-09
Second-Class Scholarship for Academic Excellent Students of Tsinghua University	2002-03
Third-Class Scholarship for Academic Excellent Students of Tsinghua University	2001-02
Third-Class Scholarship for Academic Excellent Students of Tsinghua University	2000-01
Machine Learning Summer School at Purdue University Scholarship	Jun. 2011
Student Travel Grant for International Society for Music Information Retrieval conference (ISMIR)	2008, 2010
Excellent Intern in Microsoft Research Asia (MSRA)	Apr. 2008
Champion and Best Control Scheme Prize, Tsinghua University Electronic Design Competition	Dec. 2002

## ACADEMIC SERVICE

### University-wide

Decanal Review Committee for Hajim School of Engineering and Applied Sciences	2020
Faculty Search Committee of the Department of ECE	2015-17
Steering Committee of the Faculty Council of the College of Arts, Sciences and Engineering	2015-17
ECE Department Graduate Admissions Committee	2014-18
AME Major Advisor for the Class of 2018	2014-18
AME Major Advisor for the Class of 2022	2018-22
Hajim School Outstanding PhD Dissertation Award Committee	2014-18
Robert L. And Mary L. Sproull University Fellowships Committee	2017

### Chairing

Chair – 2017 North East Music Informatics Special Interest Group (NEMISIG) Workshop	2017
Publications Chair - International Society for Music Information Retrieval (ISMIR) Conference	2017
Session Chair - International Society for Music Information Retrieval (ISMIR) Conference	2015

### Program committee

Interspeech	2020
IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)	2015, 17, 19
International Society for Music Information Retrieval (ISMIR) Conference	2014-20
ACM International Conference on Multimedia (ACM MM)	2013-15
IEEE Western New York Image and Signal Processing Workshop (WNYISPW)	2014-18
IEEE Workshop on Broadcast and User-generated Content Recognition and Analysis (BRUREC)	2013
IEEE Western New York Image Processing Workshop (WNYIPW)	2013

## Reviewer for journals

IEEE Transactions on Audio, Speech and Language Processing, IEEE Transactions on Image Processing, IEEE Transactions on Human Machine Systems, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Multimedia, IEEE Journal of Selected Topics in Signal Processing, IEEE Multimedia, IEEE Signal Processing Magazine, IEEE Signal Processing Letters, ACM Transactions on Intelligent Systems and Technology, ACM Transactions on Multimedia Computing Communications and Applications, EURASIP Journal on Audio, Speech, and Music Processing, EURASIP Journal on Advances in Signal Processing, Elsevier Computer Science Review, Elsevier Computer Communications, Elsevier Journal on Computer Methods and Programs in Biomedicine, Elsevier Speech Communication, Journal of New Music Research, Music Perception, Neural Processing Letters, Journal of Audio Engineering Society.

## Reviewer for conferences

ACM Multimedia, AES (Audio Engineering Society) Conference on semantic Audio, Audio Mostly, EUSIPCO (European Signal Processing Conference), DAFx (International Conference on Digital Audio Effects), ICASSP (IEEE International Conference on Acoustics, Speech, and Signal Processing), ICME (IEEE International Conference on Multimedia & Expo), Interspeech, ISCA Tutorial and Research Workshops on Statistical and Perceptual Audition (SAPA), ISM (IEEE International Symposium on Multimedia), ISMIR (International Society for Music Information Retrieval conference), WASPAA (IEEE Workshop on Applications of Signal Processing to Audio and Acoustics).

## PROFESSIONAL MEMBERSHIPS

IEEE (Institute of Electrical and Electronics Engineers)	2009-20
• Signal Processing Society	
ACM (Association for Computing Machinery)	2020
AES (Audio Engineering Society)	2013-19
ISCA (International Speech Communication Association)	2012