

ZHEPEI WANG

Adobe Inc.
345 Park Ave,
San Jose, CA 95110

Phone: +1 (909) 480-6613 (US)
+86 13662609293 (CN)
Email: zhepeiw03@gmail.com
Homepage: <http://zhepeiw.github.io>

Scholar:  LinkedIn:  Twitter:  GitHub: 

Education

Ph.D., Computer Science August 2018 - December 2023
University of Illinois Urbana-Champaign, Champaign, IL, USA

- Division: Artificial Intelligence
- GPA: 3.97/4.00
- Dissertation: *Data-Efficient Approaches for Audio Classification and Separation* ([pdf](#)) ([slides](#))
 - Proposed the first supervised and self-supervised continual learning approaches to train a sound classifier that can incrementally recognize new sound classes
 - Introduced network designs and training frameworks that enable adaptive computation and better generalization for sound event detection
 - Integrated semi-supervised learning to improve the performance of separation models for music and speech applications
- Advisor: Prof. Paris Smaragdis

B.S., Computer Science August 2014 - May 2018
Harvey Mudd College, Claremont, CA, USA

- GPA: 3.93/4.00
- Graduate with High Distinction and Honor in Computer Science

Interests

machine learning, audio and speech processing, self-supervised learning, continual learning, multi-modal learning, generative AI

Professional Experience

Adobe Research, Research Scientist/Engineer May 2024 - present

- Data enhancement for text-to-music synthesis
- Supervisor: Dr. Nicholas J. Bryan

Amazon Web Services Chime SDK, Applied Scientist September 2023 - May 2024

- Real-time low-compute speech enhancement for Amazon Voice Focus using distance-based features without enrollment utterances from users

- The enrollment-free model reaches 97% of the perceptual quality score of the model using enrollment speech
- Inconsistency detection with prompting-based methods from passages retrieved from enterprise database with user queries for Amazon Q

Amazon Web Services Chime SDK, Scientist intern

Summer 2022

- The first unified framework for real-time low-complexity personalized and non-personalized speech enhancement for Amazon Voice Focus
- The unified framework reduces 50% of the memory usage while reaching 99.8% of the enhancement performance of the task-specific models
- Publication in proceedings of ICASSP 2023
- Supervisors: Dr. Ritwik Giri, Dr. Michael M. Goodwin

Amazon Web Services Chime SDK, Scientist intern

Summer 2021

- Offline personalized speech enhancement for Amazon Personalized Voice Focus
- Semi-supervised target speaker extraction using speaker identity cues
- Extraction performance outperforms the previous state-of-the-art supervised approach by 12.6%
- Supervisor: Dr. Ritwik Giri

Amazon Web Services AI, Scientist intern

Summer 2020

- Semi-supervised singing voice separation and data augmentation
- Separation performance outperforms the previous state-of-the-art by 18.5%
- Publication in proceedings of ICASSP 2021
- Supervisor: Dr. Ritwik Giri

Tencent AI Lab, Research intern

Summer 2019

- Score-to-sound singing synthesis using a neural vocoder based on Text-to-Speech (TTS) pipelines
- Supervisor: Dr. Shiyin Kang

Tencent AI Platform Department, ML Engineering intern

Summer 2018

- Implemented object detection and image segmentation networks including SSD and Mask-RCNN using Tensorflow
- Supervisor: Dr. Xiaolong Zhu

Amazon Prime Now - HMC Clinic, Project leader

2017 - 2018

- Report: *Image-Text Classification to Correct the Amazon PrimeNow Search Experience* ([pdf](#)) ([poster](#))
- Designed a system that automatically detects mismatches between product images and text descriptions with deep learning
- Designed and implemented a workflow for training a deep neural network to determine the similarity between a pair of images
- Supervisor: Prof. Yekaterina Kharitonova

- HMC Music Information Retrieval Lab, Student researcher** 2017 - 2018
- Live song identification using supervised deep learning and unsupervised machine learning methods
 - Supervisor: Prof. Timothy J. Tsai

Honors and Awards

- Outstanding Reviewer Recognition, ICASSP** 2023
Awarded to the reviewers with outstanding contributions (220/4445)
- Saburo Muroga Endowed Fellowship** (\$ 6,740) 2018 - 2019
Awarded to outstanding graduate students in computer science
- Outstanding Clinic Individual Award** 2018
Awarded to top-performing students in the senior capstone project (4/100)
- Harvey S. Mudd Merit Award** (\$10,000 per academic year) 2014 - 2018
Awarded to students with superior academic achievement

Reviewer Experience

- International Conference of Acoustics, Speech and Signal Processing (ICASSP) 2021 - 2024
- International Conference on Learning Representations (ICLR) 2024
- Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) 2021, 2023
- Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE) 2022, 2023
- IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP) 2023

Skills

- **Programming Languages:** Python, C, C++, Unix Shell Scripting, Matlab
- **Machine Learning, Deep Learning:** PyTorch, TensorFlow, MxNet, SpeechBrain, NumPy, scikit-learn
- **Software:** Git, Docker, AWS, Slurm, JupyterLab, Vim, Tmux

Peer-Reviewed Publications

F. Paissan, L. D. Libera, **Z. Wang**, P. Smaragdis, M. Ravanelli, and Y. C. Sübakan, "Audio Editing with Non-Rigid Text Prompts", *In Interspeech*, Sept. 2024 ([pdf](#))([demo](#))

Z. Wang, Y. C. Sübakan, K. Subramani, J. Wu, T. Tavares, F. Ayres, and P. Smaragdis, "Unsupervised Improvement of Audio-Text Cross-Modal Representations", *In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2023 ([pdf](#)) ([code](#)) ([poster](#)) ([talk](#))

Z. Wang, R. Giri, D. Shah, J.-M. Valin, M. Goodwin, and P. Smaragdis, "A Framework for Unified Real-time Personalized and Non-Personalized Speech Enhancement", *In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, June. 2023 ([pdf](#)) ([poster](#)) ([talk](#))

Z. Wang, Y. C. Sübakan, X. Jiang, J. Wu, E. Tzinis, M. Ravanelli, and P. Smaragdis, "Learning Representations for New Sound Classes With Continual Self-Supervised Learning", *In IEEE Signal Processing Letters*, vol. 29, pp. 2607-2611, 2022 ([pdf](#)) ([code](#)) ([poster](#))

E. Tzinis, **Z. Wang**, X. Jiang, and P. Smaragdis, "Compute and Memory Efficient Universal Sound Source Separation", *In Journal of Signal Processing Systems*, vol. 9, no. 2, pp. 245-259, 2022 ([pdf](#))

S. Yuan, **Z. Wang**, U. Isik, R. Giri, J.-M. Valin, M. Goodwin, and A. Krishnaswamy, "Improved Singing Voice Separation with Chromagram-Based Pitch-Aware Remixing", *In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May. 2022 ([pdf](#))

Z. Wang, J. Casebeer, A. Clemmitt, E. Tzinis, and P. Smaragdis, "Sound Event Detection with Adaptive Frequency Selection", *In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2021, ([pdf](#)) ([code](#)) ([poster](#)) ([talk](#)) Nomination of Best Paper Award

E. Tzinis, J. Casebeer, **Z. Wang**, and P. Smaragdis, "Separate But Together: Unsupervised Federated Learning for Speech Enhancement from Non-IID Data", *In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2021 ([pdf](#))

Z. Wang, R. Giri, U. Isik, J.-M. Valin, and A. Krishnaswamy, "Semi-Supervised Singing Voice Separation with Noisy Self-Training", *In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, June. 2021 ([pdf](#)) ([poster](#)) ([talk](#))

E. Tzinis, **Z. Wang**, and P. Smaragdis, "Sudo rm -rf: Efficient Networks for Universal Audio Source Separation", *In IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Sept. 2020 ([pdf](#))

E. Tzinis, S. Venkataramani, **Z. Wang**, Y. C. Sübakan, and P. Smaragdis, "Two-Step Sound Source Separation: Training on Learned Latent Targets", *In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May. 2020 ([pdf](#))

Z. Wang, Y. C. Sübakan, E. Tzinis, P. Smaragdis, and L. Charlin, "Continual Learning of New Sound Classes Using Generative Replay", *In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2019 ([pdf](#))

J. Casebeer[‡], **Z. Wang**[‡], and P. Smaragdis, "Multi-View Networks for Multi-Channel Audio Classification," *In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May. 2019 ([pdf](#)) ([poster](#))

[‡] Equal Contribution

Preprints

T. Tavares, F. Ayres, **Z. Wang**, and P. Smaragdis, "On Class Separability Pitfalls in Audio-Text Contrastive Zero-Shot Learning", *arXiv preprint arXiv:2408.13068*, Aug. 2024 ([pdf](#))

Z. Wang, R. Giri, S. Venkataramani, U. Isik, J.-M. Valin, P. Smaragdis, M. Goodwin, and A. Krishnaswamy, "Semi-Supervised Time Domain Target Speaker Extraction with Attention", *arXiv preprint arXiv:2206.09072*, June. 2022 ([pdf](#)) ([code](#))