

ZHEPEI WANG

Last updated: March 20, 2026

Adobe Inc.
601 Townsend St,
San Francisco, CA 94103

Phone: +1 (909) 480-6613 (US)
+86 13662609293 (CN)
Email: zhepei03@gmail.com
Homepage: <http://zhepei.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 Engineer May 2024 - present

- Data enhancement for text-to-music synthesis
- Multi-modal large language model for music understanding

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, <i>Student researcher</i>	2017 - 2018
<ul style="list-style-type: none"> • 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	

Professional Service

Data Competition Chair, IEEE International Workshop on Machine Learning for Signal Processing (MLSP) 2025

Reviewer Experience

International Conference of Acoustics, Speech and Signal Processing (ICASSP)	2021 - 2024, 2026
IEEE International Workshop on Machine Learning for Signal Processing (MLSP)	2025
IEEE Signal Processing Letters	2025
IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP)	2023, 2025
International Conference on Machine Learning (ICML)	2025
Interspeech	2025
International Society for Music Information Retrieval (ISMIR)	2025
Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)	2021 - 2025
International Conference on Learning Representations (ICLR)	2024
JMIR Biomedical Engineering (JBME)	2024
TC3 Workshop on Artificial Neural Networks in Pattern Recognition (ANNPR)	2024
Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)	2022, 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

- I. Bukey, **Z. Wang**, C. Donahue, and N. J. Bryan. “Rethinking Music Captioning with Music Metadata LLMs”. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2026 ([pdf](#)).
- J. Casebeer, G. Zhu, **Z. Wang**, and N. J. Bryan. “A Generative-First Neural Audio Autoencoder”. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2026 ([pdf](#)).
- R. Giri, M. M. Goodwin, A. Krishnaswamy, M. U. Isik, J.-M. Valin, **Z. Wang**, S. Venkataramani, and P. Smaragdis. “Semi-supervised training of a machine learning model for target speaker audio enhancement”. U.S. Patent 12,531,067. Inc. Amazon Technologies. Jan. 20, 2026 ([patent](#)).
- R. Giri, **Z. Wang**, D. Shah, J.-M. Valin, and M. M. Goodwin. “Unified audio suppression model”. U.S. Patent Application 2025/0111857. Inc. Amazon Technologies. Apr. 3, 2025 ([application](#)).
- T. Tavares, F. Ayres, **Z. Wang**, and P. Smaragdis. “On Class Separability Pitfalls In Audio-Text Contrastive Zero-Shot Learning”. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2025 ([pdf](#)).
- F. Paissan, L. D. Libera, **Z. Wang**, P. Smaragdis, M. Ravanelli, and Y. C. Sübakan. “Audio Editing with Non-Rigid Text Prompts”. *Interspeech*. 2024 ([pdf](#)) ([demo](#)).
- 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”. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2023 ([pdf](#)) ([poster](#)) ([talk](#)).
- Z. Wang**, Y. C. Sübakan, K. Subramani, J. Wu, T. Tavares, F. Ayres, and P. Smaragdis. “Unsupervised Improvement of Audio-Text Cross-Modal Representations”. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. 2023 ([pdf](#)) ([code](#)) ([poster](#)) ([talk](#)).
- E. Tzinis, **Z. Wang**, X. Jiang, and P. Smaragdis. “Compute and Memory Efficient Universal Sound Source Separation”. *Journal of Signal Processing Systems* 9.2 (2022), pp. 245–259 ([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”. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2022 ([pdf](#)).
- 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”. *IEEE Signal Processing Letters* 29 (2022), pp. 2607–2611 ([pdf](#)) ([code](#)) ([poster](#)).
- E. Tzinis, J. Casebeer, **Z. Wang**, and P. Smaragdis. “Separate But Together: Unsupervised Federated Learning for Speech Enhancement from Non-IID Data”. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. 2021 ([pdf](#)).

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

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

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

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

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

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

Preprints

S. Kumar, P. Seetharaman, K. Chen, O. Nieto, J. Su, **Z. Wang**, R. Kumar, D. Manocha, N. J. Bryan, Z. Jin, and J. Salamon. “Timestamped Audio Captioning”. *arXiv preprint arXiv:2602.15766* (2026) ([pdf](#)) ([demo](#)).

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* (2022) ([pdf](#)) ([code](#)).

Patents and Patent Applications

R. Giri, M. M. Goodwin, A. Krishnaswamy, M. U. Isik, J.-M. Valin, **Z. Wang**, S. Venkataramani, and P. Smaragdis. “Semi-supervised training of a machine learning model for target speaker audio enhancement”. U.S. Patent 12,531,067. Inc. Amazon Technologies. Jan. 20, 2026 ([patent](#)).

R. Giri, **Z. Wang**, D. Shah, J.-M. Valin, and M. M. Goodwin. “Unified audio suppression model”. U.S. Patent Application 2025/0111857. Inc. Amazon Technologies. Apr. 3, 2025 ([application](#)).