

# KAI ZHEN

<http://kaizhen.us> • [kaizhen723@gmail.com](mailto:kaizhen723@gmail.com)

---

## POSITIONS HELD

### Amazon.com, Inc.

- Applied Scientist
  - Alexa Speech, Pittsburgh, PA Apr. 2021 – present
- Applied Scientist Intern
  - Alexa Speech, Pittsburgh, PA Summer 2020
    - Supervisors: Athanasios Mouchtaris, Hieu Duy Nguyen, Feng-Ju (Claire) Chang
    - Project: Network Compression for On-Device ASR Solutions

### Indiana University

Aug. 2015 – Mar. 2021

- Research Assistant: Audio Signal Analysis/Synthesis Technology Based on Machine Learning
  - Took the ownership in multiple machine learning research projects
  - Published in leading machine learning and speech processing conferences and journals
  - Contributed to 5 US patents as an inventor
- Associate instructor in Department of Computer Science and Intelligent Systems Engineering

### LinkedIn Corporation

- Machine Learning & Relevance Intern
  - Ads-AI Group, Mountain View, CA Summer 2019
    - Supervisors: Sara Smoot, Lijun Peng, Hiroto Udagawa
    - Project: Ads Response Rate Prediction with BERT Enriched Semantic Features
  - Company Standardization Group, New York City, NY Summer 2018
    - Supervisors: Xiaoqiang Luo, Deirdre Hogan
    - Project: Relevance Ranking via Non-Categorical User Inputs for LinkedIn Resume Builder

---

## EDUCATION

### Ph.D., dual major in Computer Sciences and Cognitive Science

May. 2021

- Indiana University, Bloomington, United States
- Committee: Minje Kim (chair, IU Intelligent Systems Engineering), Robert Goldstone (co-chair, IU Cognitive Science), Donald Williamson (IU Computer Science), and Shen Yi (U. of Washington, Speech and Hearing Sciences)
- Dissertation: "Neural Waveform Coding: Scalability, Efficiency and Psychoacoustic Calibration"  
<Winner of the Outstanding Research Award (IU Cognitive Science)>

### M.S., major in Computer Science

Jul. 2015

- Tsinghua University, Beijing, China

### B.S., major in Software Engineering (Summa Cum Laude)

Jul. 2012

- Xidian University, Xi'an, China

---

## PROFESSIONAL ACTIVITIES

### Conference Reviewer

- IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP): 2019 - 2023
- ISCA Interspeech: 2022 - 2023
- EURASIP European Signal Processing Conference (EUSIPCO): 2022 - 2023
- IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA): 2021 - 2023
- IEEE International Conference on Data Mining (ICDM): 2020
- Association for the Advancement of Artificial Intelligence (AAAI): 2017 - 2018

## Journal Reviewer

- European Association for Signal Processing (EURASIP) Journal on Audio, Speech, and Music Processing
  - IEEE MultiMedia
  - Speech Communication
- 

## PUBLICATIONS

### International Journal Articles

- [J002] **Kai Zhen**, Jongmo Sung, Mi Suk Lee, Seungkwon Beack, and Minje Kim, "Scalable and Efficient Neural Speech Coding: A Hybrid Design", *IEEE/ACM Transactions on Audio, Speech, and Language Processing (IEEE/ACM TASLP)*, **30** (2021): 12-25.
- [J001] **Kai Zhen**, Mi Suk Lee, Jongmo Sung, Seungkwon Beack, and Minje Kim, "[Psychoacoustic Calibration of Loss Functions for Efficient End-to-End Neural Audio Coding](#)," *IEEE Signal Processing Letters (SPL)* **27** (2020): 2159-2163.

### Referred International Conference Proceedings

- [C008] Martin Radfar, Paulina Lyskawa, Brandon Trujillo, Yi Xie, **Kai Zhen**, Jahn Heymann, Denis Filimonov, Grant Strimel, Nathan Susanj, Athanasios Mouchtaris, "[Conmer: Streaming Conformer with no self-attention for interactive voice assistants](#)," In *Proc. Annual Conference of the International Speech Communication Association (Interspeech)*, Dublin, Ireland, August 21-24, 2023.
- [C007] **Kai Zhen**, Martin Radfar, Hieu Duy Nguyen, Nathan Susanj, Grant Strimel, Athanasios Mouchtaris, "[Sub-8-bit Quantization for On-Device Speech Recognition: A Regularization-Free Approach](#)", *IEEE Workshop on Spoken Language Technology (IEEE SLT)*, Doha, Qatar, January 9-12, 2023.
- [C006] **Kai Zhen**, Hieu Duy Nguyen, Raviteja Chinta, Nathan Susanj, Athanasios Mouchtaris, Tariq Afzal, and Ariya Rastrow, "[Sub-8-Bit Quantization Aware Training for 8-Bit Neural Network Accelerator with On-Device Speech Recognition](#)," in *Proceedings of Annual Conference of the International Speech Communication Association (Interspeech)*, Incheon, Korea, September 18-22, 2022.
- [C005] **Kai Zhen**, Hieu Duy Nguyen, Feng-Ju (Claire) Chang, Athanasios Mouchtaris, "[Sparsification via Compressed Sensing for Automatic Speech Recognition](#)," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toronto, ON, Canada, June 6-12, 2021.
- [C004] Haici Yang, **Kai Zhen**, Seungkwon Beack, Minje Kim, "[Source-Aware Neural Speech Coding for Noisy Speech Compression](#)," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toronto, ON, Canada, June 6-12, 2021.
- [C003] **Kai Zhen**, Mi Suk Lee, Jongmo Sung, Seungkwon Beack, and Minje Kim, "[Efficient And Scalable Neural Residual Waveform Coding with Collaborative Quantization](#)," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Barcelona, Spain, May 4-8, 2020.
- [C002] **Kai Zhen**, Mi Suk Lee, Minje Kim. "[A Dual-Staged Context Aggregation Method towards Efficient End-To-End Speech Enhancement](#)," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Barcelona, Spain, May 4-8, 2020.
- [C001] **Kai Zhen**, Jongmo Sung, Mi Suk Lee, Seungkwon Beack, and Minje Kim, "[Cascaded Cross-Module Residual Learning towards Lightweight End-to-End Speech Coding](#)," in *Proceedings of Annual Conference of the International Speech Communication Association (Interspeech)*, Graz, Austria, September 15-19, 2019.

### Patents

- [P005] Kim, Minje, Mi Suk Lee, Seung Kwon Beack, Jongmo Sung, Tae Jin Lee, Jin Soo Choi, and **Kai Zhen**. "[Apparatus and method for speech processing using a densely connected hybrid neural network](#)." U.S. Patent Application 17/308,800, filed November 11, 2021.
- [P004] Mi Suk Lee, Seung Kwon Beack, Jongmo Sung, Tae Jin Lee, Jin Soo Choi, Minje Kim, **Kai Zhen**, "[Method and apparatus for processing audio signal](#)," *US Patent App. 17/156,006*, 2021.
- [P003] Minje Kim, **Kai Zhen**, Mi Suk Lee, Seung Kwon Beack, Jongmo Sung, Tae Jin Lee, Jin Soo Choi. "[Residual coding method of linear prediction coding coefficient based on collaborative quantization, and computing device for performing the method](#)." U.S. Patent Application 17/098,090, filed May 13, 2021.

- [P002] Mi Suk Lee, Jongmo Sung, Minje Kim, **Kai Zhen**, "[Audio signal encoding method and audio signal decoding method, and encoder and decoder performing the same](#)," U.S. Patent Application No. 16/543,095
- [P001] Minje Kim, Aswin Sivaraman, **Kai Zhen**, Jongmo Sung, et al, "[Audio signal encoding method and apparatus and audio signal decoding method and apparatus using psychoacoustic-based weighted error function](#)", US Patent Application, US 2019 / 0164052 A1.

## Peer Reviewed Workshops & Forums

- [W005] Kai Zhen, Martin Radfar, Hieu Duy Nguyen, Nathan Susanj, Grant Strimel, Athanasios Mouchtaris. General Quantization for On-Device ASR. *Amazon Machine Learning Conference (AMLC)*, 2022.
- [W004] **Kai Zhen**, Hieu Duy Nguyen, Feng-Ju (Claire) Chang, Athanasios Mouchtaris. Network Sparsification for On-Device ASR. *Amazon Machine Learning Conference (AMLC) Workshop on Network Inference Optimization*, 2020.
- [W003] **Kai Zhen**, Aswin Sivaraman, Jongmo Sung, Minje Kim. [On Psychoacoustically Weighted Cost Functions Towards Resource-efficient Deep Neural Networks for Speech Denoising](#). *The 7th Annual Midwest Cognitive Science Conference*, 2018.
- [W002] Peter Miksza, Kevin Watson, **Kai Zhen**, Sanna Wager, Minje Kim. Relationships between experts' subjective ratings of jazz improvisations and computational measures of melodic entropy. *The Improvising Brain III: Cultural Variation and Analytical Techniques Symposium*, Atlanta, GA, in Feb, 2017.
- [W001] **Kai Zhen** and David Crandall. [Finding egocentric image topics through convolutional neural network based representations](#) (extended abstract). In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshop on Egocentric Computer Vision*, 2016.

---

## HONORS, AWARDS & SCHOLARSHIP

- |   |                      |
|---|----------------------|
| <b>Outstanding Research Award</b>   | Apr. 2021            |
| • Given by <a href="#">Cognitive Science Program at Indiana University</a>                      |                      |
| <b>Top-Rated Intern Poster</b>  | Aug. 2020            |
| • <a href="#">Among 17 interns receiving the highest rate out of more than 180 participants</a> |                      |
| <b>Summa Cum Laude</b>  | Jul. 2012            |
| • Graduate with honor from Xidian University  |                      |
| <b>China National Scholarship</b>   | Nov. 2010, Nov. 2011 |
| • For the effort on maintaining top-tier GPA and mathematical contest in modeling (MCM)         |                      |

---

## INVITED TALKS

- [T003] [Microsoft Research Talks](#), September, 2020 [[Video link](#)]
- [T002] IU Hearing Sciences Seminar, March, 2019
- [T001] IU Grey Matters, Graduate and Post-doc Colloquium, March, 2019

---

## TEACHING & TUTORING

### Graduate Level

- "Machine Learning for Signal Processing" (ENGR-E 599, ISE IU), Fall 2017
- "Elements of Artificial Intelligence" (CSCI-B 551, CS IU), Fall 2016
- "Computer Vision" (CSCI-B 657, CS IU), Spring 2016
- "Data Structures" (ENGR-E 599, ISE IU), Fall 2015

### Undergraduate Level

- "Introduction of Artificial Intelligence" (CSCI-B 351, CS IU), Spring 2017

---

## CRAFTSMANSHIP

**Deep Learning/Artificial Intelligence (over 5 years experience)**

- TensorFlow, PyTorch, etc;
- recommendation, feature learning, autoregressive modeling, recognition, etc

#### **Audio Signal Processing (over 5 years experience)**

- bitrate efficient and scalable audio/speech coding, speech enhancement;
- subjective/objective audio quality assessment;
- psychoacoustic models and optimization skills;
- end-to-end speech recognition (RNN-Transducer).

#### **Machine Learning (over 5 years experience)**

- regression (GLMix) and classification (decision trees, SVM);
  - dimension reduction (PCA/ICA/NMF/ISOMAP);
  - clustering analysis (k-means, GMM);
  - topic modeling (LDA).
-