

Xinlei Niu

 xinlei.niu.au@gmail.com

 linkedin.com/xinlei-niu

 github.com/XinleiNIU

Education

Australian National University <i>Doctor of Philosophy, Engineering and Computer Science</i>	2022 - Expected 2026 Canberra, Australia
• Thesis Title: Generative Audio Synthesis in Sound, Speech, and Music	
• Research Topic: Audio Synthesis, Generative Model, Bayesian Machine Learning	
Australian National University <i>Master of Machine Learning and Computer Vision (with Commendation)</i>	2019 - 2021 Canberra, Australia
• Master Thesis: Acoustic Scene Classification with Attention-based Neural Networks	
Hong Kong Baptist University <i>Bachelor's degree in Statistics (Second Class Honours, Division I)</i>	2014 - 2018 Hong Kong SAR
• Bachelor Thesis: Forecast Chinese Stock Index by Using Recurrent Neural Networks	

Experience

Amazon.com, Inc. <i>Applied Science Intern</i>	Jan. 2026 – Jul. 2026 Melbourne, Australia
• Incoming	
Dolby Laboratories <i>Research Internship</i>	May. 2025 – Sep. 2025 Sydney, Australia
• Video-to-Audio with Environmentally Aware Speech.	
SonyAI, Sony Group Corporation <i>Research Internship</i>	Jan. 2025 – Apr. 2025 Tokyo, Japan
• Text-Guided Music Editing and Personalized Music Editing.	
Australian National University <i>Tutor</i>	Mar. 2022 – Jun. 2024 Canberra, Australia
• COMP/ENGN6528 Computer Vision for Semester 1	
• ENGN8501/COMP8539 Advanced Topics of Computer Vision for Semester 2	
Murdoch University <i>Research Assistant</i>	Mar. 2021 – Jul. 2021 Remote
• Developed a system to batch process unformatted log files imported from 10EaZy, involving aggregate files into a database automatically and calculating statistical features.	
• Implement platform: MATLAB.	
Artisan Technologies CO.,Ltd. <i>Deep Learning Developer Intern</i>	Jul. 2018 – Dec. 2018 GuangZhou, China
• Developed a colour-matching neural network model to predict dyestuff concentration.	
• Implement platform: SQL, Python, TensorFlow.	

Publications

1. **Niu, X., & Martin, C. P.** "Spatial-temporal-class attention network for acoustic scene classification." 2022 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2022. (**Oral presentation**)
2. **Niu, X., Zhang, J., Walder, C., & Martin, C. P.** "SoundLoCD: An Efficient Conditional Discrete Contrastive Latent Diffusion Model for Text-to-Sound Generation." ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024. (**Oral presentation, Demo**)
3. **Niu, X., Walder, C., Zhang, J., & Martin, C. P.** Latent Optimal Paths by Gumbel Propagation for Variational Bayesian Dynamic Programming. In Forty-first International Conference on Machine Learning (ICML), 2024. (**Code**)

4. Niu, X., Zhang, J., & Martin, C. P. HybridVC: Efficient Voice Style Conversion with Text and Audio Prompts. Proc. Interspeech 2024, 4368-4372, doi: 10.21437/Interspeech.2024-46. InterSpeech 2024. ([Demo](#))
5. Niu, X., Cheuk, K. W., Zhang, J., Murata, N., Lai, C. H., Mancusi, M., ... & Mitsufuji, Y. (2025) SteerMusic: Enhanced Musical Consistency for Zero-shot Text-Guided and Personalized Music Editing. arXiv (cs.SD), <https://arxiv.org/pdf/2504.10826.pdf>. ([Accepted by AAAI2026](#), [Oral presentation](#), [Demo](#), [Code](#))
6. Niu, X., Ma, J., Harper-Harris, D., Zhang, X., Martin, C. P., & Zhang, J. (2025). Beyond Video-to-SFX: Video to Audio Synthesis with Environmentally Aware Speech. arXiv preprint arXiv:2509.15492. ([In submission](#), [Demo](#))
7. Niu, X., Zhang, J., & Martin, C. P. (2024) SoundMorpher: Perceptually-Uniform Sound Morphing with Diffusion Model. arXiv (cs.SD), <https://arxiv.org/abs/2410.02144>. ([In submission](#), [Demo](#), [Code](#))

Honors & Awards

2021 ANU PhD Scholarship and HDR Fee Remission Merit

2018 Simulated Investment Competition organized by Minmetals Securities Co., Ltd. (Simulated Investment Competition Ranked 5)

2017 Contemporary Undergraduate Mathematical Contest in Modelling Excellence Award in 2017 (CUMCM 2017 Excellence Award)

2016 Microsoft Office Specialist (MOS) Expert for Office Excel 2010

Technical Skills

Statistics: Machine Learning, Statistical Data Analysis, Bayesian Data Analysis, Variational Inference

Programming Languages: Python, C, Matlab, HTML/CSS, Javascript.

Audio Processing: Text-to-Speech, Singing Voice Synthesis, Text-to-Audio, Voice Conversion, Sound Morphing, Acoustic Scene Classification

Others: Microsoft Excel, SPSS, R, SQL

Service

Reviewer: Review for ACM Multimedia 2024, ICLR 2025, AAAI2026

Guest presenter: MIQ Reading group at Spotify.