

Hieu-Thi Luong

contact@hieuthi.com | <https://www.hieuthi.com> | <https://github.com/hieuthi>

Skills

Programming Languages: Python, Javascript, C/C++

Machine Learning and Data Analysis: pytorch, numpy, pandas, matplotlib

Research and Other Skills: academic paper writing, experimental design and implementation, project management

Language Proficiency: Vietnamese (native), English (proficient, C2), Japanese (beginner, A2)

Experience

Research Fellow, Nanyang Technological University – Singapore, Singapore Jan 2024 – Present

- Managed multinational research collaborations, coordinating global teams to meet project milestones
- Generated high-impact research results, such as academic papers, datasets, and application systems, creating value for diverse stakeholders including principal investigators, funders, and students
- Established a foundational framework for Partially Fake Speech Detection research by creating realistic samples and a competitive detection system to accelerate further R&D efforts
- Engaged with the research community and the industry on AI advancements, focusing on Speech Processing, Speech Synthesis, Voice Cloning, and Deep Fake Detection

Project Researcher, National Institute of Informatics – Tokyo, Japan Oct 2020 – Mar 2023

- Developed novel methods for audio/sound generation with a focus on human vocalizations to enhance the expressive capabilities of human-machine interaction systems.
- Designed and developed a pioneering TTS/VC hybrid Voice Cloning technology, yielding one granted patent and advancing the state-of-the-art in speech processing.
- Utilized Signal Processing and Speech Processing toolkits to tackle diverse problems and challenges

Research Assistant, National Institute of Informatics – Tokyo, Japan Nov 2017 – Aug 2020

- Pioneered a novel Machine Learning methodology for multi-lingual and cross-lingual generative AI models, enabling bootstrapping of low-resource language systems and advancing global speech technology accessibility.
- Architected innovative Deep Learning models for multi-speaker Text-to-Speech and Voice Conversion systems, enabling advanced speech synthesis capabilities.
- Engineered and implemented high-quality Japanese Text-to-Speech models, addressing key technical challenges such as data imbalance and unique linguistic features like pitch accent.

Research Assistant, VNUHCM - University of Science – Ho Chi Minh City, Vietnam Sep 2014 – July 2017

- Developed and published VIVOS, the first open-license Vietnamese Speech Corpus, accelerating R&D breakthroughs in Vietnamese Speech Processing Systems.
- Established cutting-edge Automatic Speech Recognition (ASR) and Text-to-Speech baselines for the Vietnamese language, addressing its unique characteristics, including tones and pitch.
- Supported lead instructors in instructional planning and student support for the Speech Processing course.

Side Projects

The Open-Source Joplin Plugin Development Project note-taking-joplin

- Multiple open-source plugins developed for Joplin to enhance the note-taking app user experience.

The Vietnamese technology, educational, lifestyle blog - 1/3 espresso 1p3espresso.com

- A self-hosted Vietnamese blog on technology, engineering, and education, sharing knowledge and experiences.

Education

The Graduate University for Advanced Studies, SOKENDAI – PhD in Multidisciplinary Science - Informatics Sept 2020

VNUHCM - University of Science – MS in Computer Science Dec 2016

VNUHCM - University of Science – BS in Information Technology Sept 2014

Patents

- JP2020027168A, Hieu-Thi Luong, and Junichi Yamagishi. "Learning device, learning method, voice synthesis device, voice synthesis method and program." Issued Jul 29, 2022.

Honors & awards

- Singapore Academies South-East Asia Fellowship (SASEAF) for a 2-year Postdoctoral Research Fellowship (January 2024 - December 2025)
- Japanese Government (Monbukagakusho: MEXT) Scholarship for 3-year PhD Research (October 2017 - September 2020)

Publications

Luong, H. T., Li, H., Zhang, L., Lee, K. A., & Chng, E. S. (2025, April). **LlamaPartialSpoof: An LLM-Driven Fake Speech Dataset Simulating Disinformation Generation**. 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. (to appear)

Kuzmin, N., Luong, H. T., Yao, J., Xie, L., Lee, K. A., & Chng, E. S. (2024, September). **NTU-NPU System for Voice Privacy 2024 Challenge**. Proceedings of the 4th Symposium on Security and Privacy in Speech Communication (pp. 72-79).

Luong, H. T., Truong, D. T., Lee, K. A., & Chng, E. S. (2024, December). **Room Impulse Responses help attackers to evade Deep Fake Detection**. 2024 IEEE Spoken Language Technology Workshop (SLT). IEEE.

Truong, D. T., Tao, R., Nguyen, T., Luong, H. T., Lee, K. A., & Chng, E. S. (2024, September). **Temporal-channel modeling in multi-head self-attention for synthetic speech detection**. Interspeech Conference 2023 (pp. 537-541). International Speech Communication Association.

Luong, H. T. & Yamagishi, J. (2023, June). **Controlling Multi-Class Human Vocalization Generation via a Simple Segment-based Labeling Scheme**. Interspeech Conference 2023 (pp. 4379-4383). International Speech Communication Association.

Luong, H. T. & Yamagishi, J. (2021, October). **Laughnet: synthesizing laughter utterances from waveform silhouettes and a single laughter example**. arXiv preprint arXiv:2110.04946.

Luong, H. T. & Yamagishi, J. (2021, August). **Preliminary study on using vector quantization latent spaces for TTS/VC systems with consistent performance**. Proceedings of the 11th ISCA Speech Synthesis Workshop (SSW 11) (pp. 136-141).

Luong, H. T. & Yamagishi, J. (2020, October). **Latent linguistic embedding for cross-lingual text-to-speech and voice conversion**. Proceedings of the Joint Workshop for the Blizzard Challenge and Voice Conversion Challenge 2020 (pp. 150-154).

Luong, H. T. & Yamagishi, J. (2020, October). **Nautilus: a versatile voice cloning system**. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 2967-2981.

Luong, H. T. (2020, September). **Deep learning based voice cloning framework for a unified system of text-to-speech and voice conversion**. [Doctoral thesis, The Graduate University for Advanced Studies, SOKENDAI]. The Graduate University for Advanced Studies, SOKENDAI

Luong, H. T. & Yamagishi, J. (2019, December). **Bootstrapping non-parallel voice conversion from speaker-adaptive text-to-speech**. 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) (pp. 200-207). IEEE.

Luong, H. T. & Yamagishi, J. (2019, June). **A unified speaker adaptation method for speech synthesis using transcribed and untranscribed speech with backpropagation**. arXiv preprint arXiv:1906.07414.

Luong, H. T., Wang, X., Yamagishi, J., & Nishizawa, N. (2019, September). **Training multi-speaker neural text-to-speech systems using speaker-imbalanced speech corpora**. Interspeech Conference 2019 (pp. 1303-1307). International Speech Communication Association.

Luong, H. T. & Yamagishi, J. (2018, December). **Scaling and bias codes for modeling speaker-adaptive DNN-based speech synthesis systems**. 2018 IEEE Spoken Language Technology Workshop (SLT) (pp. 610-617). IEEE.

Zhao, Y., Takaki, S., Luong, H. T., Yamagishi, J., Saito, D., & Minematsu, N. (2018, September). **Wasserstein GAN and waveform loss-based acoustic model training for multi-speaker text-to-speech synthesis systems using a WaveNet vocoder**. IEEE access, 6, 60478-60488.

Luong, H. T., & Yamagishi, J. (2018, September). **Multimodal Speech Synthesis Architecture for Unsupervised Speaker Adaptation**. Interspeech Conference 2018 (pp. 2494-2498). International Speech Communication Association.

Luong, H. T., Wang, X., Yamagishi, J., & Nishizawa, N. (2018, September). **Investigating Accuracy of Pitch-accent Annotations in Neural Network-based Speech Synthesis and Denoising Effects**. Interspeech Conference 2018 (pp. 37-41). International Speech Communication Association.

Luong, H. T., Takaki, S., Henter, G. E., & Yamagishi, J. (2017, March). **Adapting and controlling DNN-based speech synthesis using input codes**. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4905-4909). IEEE.

Luong, H. T., & Vu, H. Q. (2016, December). **A non-expert Kaldi recipe for Vietnamese speech recognition system**. Proceedings of the Third International Workshop on Worldwide Language Service Infrastructure and Second Workshop on Open Infrastructures and Analysis Frameworks for Human Language Technologies (WLSI/OIAF4HLT2016) (pp. 51-55).