Li Li

AI Lab, Cyberagent, Inc.,

22F Shibuya Scramble Square, 2-24-12 Shibuya, Shibuya-ku, Tokyo, 150-6121, Japan

☎ +81 50-5530-9270 | ✉ lili-0805@ieee.org | ⊡ https://lili-0805.github.io/

RESEARCH INTERESTS

Audio signal processing, source separation, machine learning, and speech analysis

EDUCATION

MAR. 2021	Ph.D. degree in Engineering, Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba, Japan.
	Topic: Determined multichannel source separation based on signal independence
	Advisors: Prof. Shoji Makino and Dr. Hirokazu Kameoka
MAR. 2018	M.S. degree in Engineering, Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba, Japan.
	Topic: Monaural speech enhancement with non-negative matrix factorization
	Advisors: Prof. Shoji Makino and Dr. Hirokazu Kameoka
JULY 2014	B.E. degree in Engineering, Department of Information Management and Engineering, Shanghai University of Finance and Economics, China.

RESEARCH & WORK EXPERIENCES

JAN. 2023-CURRENT	Research scientist at AI Lab, Cyberagent, Inc., Japan.
APR. 2021-MAR. 2024	Postdoctoral researcher at Toda Lab, Nagoya University, Japan.
APR. 2021-DEC. 2022	Adjunct researcher at NTT communication Science Laboratories, NTT Corporation, Japan.
JUN. 2021-DEC. 2022	Researcher at TARVO Inc., Japan.
APR. 2018-MAR. 2021	Research fellow (DC1) of Japan Society of Promotion of Science (JSPS), Japan.
APR. 2018-MAR. 2021	Research intern at NTT Communication Science Laboratories, NTT Corporation, Japan.
JULY 2019-OCT. 2019	Internship at Applied Sciences Lab, Microsoft Corporation, USA.
	Topic: Geometrically constrained independent vector analysis for speech enhancement
JULY 2018-AUG. 2018	Research student at Toda Lab, Nagoya University, Japan.
	Topic: Singing voice modeling for singing style control
OCT. 2014-MAR. 2016	Research student at Kameoka Lab and Saruwatari Lab, University of Tokyo, Japan.

Awards

Mar. 2025	IEEE Signal Processing Society Japan Young Author Best Paper Award
Mar. 2024	The 19th Itakura Prize Innovative Young Researcher Award from The Acoustical Society of Japan (ASJ).
Mar. 2022	The 51th Awaya Prize Young Researcher Award, The Acoustical Society of Japan (ASJ).
Mar. 2021	President's Award of University of Tsukuba.
Dec. 2020	IEEE Signal Processing Society Japan Student Conference Paper Award.
Dec. 2019	The Best Student Presentation Award, IEICE Electroacoustics Symposium.
Nov. 2018	IEEE Signal Processing Society Tokyo Joint Chapter Student Award.
Mar. 2018	Chair Award of the Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba.
Sep. 2016	Best Student Presentation Award, The Acoustical Society of Japan (ASJ).

Fundings

May. 2019	Grants for Researchers Attending International Conferences, The Telecommunications Advancement Foundation.
2018-2021	Grant-in-Aid for JSPS Fellows, Japan Society of the Promotion of Science (JSPS).
Aug. 2017	Travel Support for Overseas Dispatch of Graduate Students, University of Tsukuba.
Mar. 2017	Travel Support for Overseas Dispatch of Graduate Students, University of Tsukuba.

Patents

9 issued patents, 1 patent application

Languages

Chinese (native), English (fluent), Japanese (fluent)

Publications

Journal papers

[1] **L. Li**, H. Kameoka, and S. Makino, "FastMVAE2: On improving and accelerating the fast variational autoencoder-based source separation algorithm for determined mixtures," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 31, pp. 96–110, Oct. 2022.

[2] **L. Li**, H. Kameoka, S. Inoue, and S. Makino, "FastMVAE: A fast optimization algorithm for the multichannel variational autoencoder method," *IEEE Access*, vol. 8, pp. 228740–228753, Dec. 2020.

[3] **L. Li**, H. Kameoka, and S. Makino, "Majorization-minimization algorithm for discriminative non-negative matrix factorization," *IEEE Access*, vol. 8, pp. 227399–227408, Dec. 2020.

[4] Y. Hashizume, **L. Li**, A. Miyashita, and T. Toda, "Learning separated representations for instrument-based music similarity," *APSIPA Transactions on Signal and Information Processing*, vol. 14, no. 1, e16, pp. 1–32, Jul. 2025.

[5] R. Wang, **L. Li**, and T. Toda, "Dual-channel target speaker extraction based on conditional variational autoencoder and directional information," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 3, pp. 1968–1979, Mar. 2024.

[6] M. Qin, **L. Li**, and S. Makino, "Deep complex-valued neural network-based triple-path mask and steering vector estimation for multichannel target speech separation," *Journal of Signal Processing*, vol. 27, no. 4, pp. 87–91, July, 2023.

[7] R. Takahashi, **L. Li**, S. Makino, and T. Yamada, "VMInNet: Interpolation of virtual microphones in optimal latent space explored by auto encoder," *Journal of Signal Processing*, vol. 25, no. 6, pp. 245–250, Nov. 2021.

[8] N. Murashima, H. Kameoka, **L. Li**, S. Shogo, and S. Makino, "Single-channel multi-speaker separation with variational autoencoder spectrogram model," *Journal of Signal Processing*, vol. 25, no. 4, pp. 145–149, Jul. 2021.

[9] S. Seki, H. Kameoka, **L. Li**, T. Toda, and K. Takeda, "Underdetermined Source Separation Based on Generalized Multichannel Variational Autoencoder," *IEEE Access*, vol. 7, no. 1, pp. 168104–168115, Nov. 2019.

[10] H. Kameoka, **L. Li**, S. Inoue, and S. Makino, "Supervised determined source separation with multichannel variational autoencoder," *Neural Computation*, vol. 31, no. 9, pp. 1–24, Sep. 2019.

[11] H. Kameoka, T. Higuchi, M. Tanaka, and **L. Li**, "Non-negative matrix factorization with basis clustering using cepstral distance regularization," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 26, no. 6, pp. 1025-1036, Jun. 2018.

PEER-REVIEWED INTERNATIONAL CONFERENCE PAPERS

[1] **L. Li** and S.Seki, "Exploring dual-mode training for real-time target speaker extraction," in Proc. *The 17th Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA2025)*, Singapore, Oct. 2025. (accepted)

[2] **L. Li** and S. Seki, "Improved remixing process for domain adaptation-based speech enhancement by mitigating data imbalance in signal-to-noise ratio," in Proc. *The 25th Annual Conference of the International Speech Communication Association (Interspeech2024)*, pp. 1710–1714, Kos, Sep. 2024.

[3] **L. Li** and Shogo Seki, "Remixed2Remixed: Domain adaptation for speech enhancement by Noise2Noise learning with remixing," in Proc. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2024)*, pp. 806–810, Seoul, Apr. 2024.

[4] **L. Li**, H. Kameoka, and S. Seki, "HBP: An efficient block permutation solver using Hungarian algorithm and spectrogram inpainting for multichannel audio source separation," in Proc. *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2022)*, pp. 516–520, Singapore, May 2022.

[5] **L. Li**, K. Koishida, and S. Makino, "Online directional speech enhancement using geometrially constrained independent vector analysis," in Proc. it The 21th Annual Conference of the International Speech Communication Association (Interspeech2020), pp. 61–65, Shanghai, Oct. 2020.

[6] **L. Li**, H. Kameoka, and S. Makino, "Determined audio source separation with multichannel star generative adversarial network," in Proc. *The 30th IEEE International Workshop on Machine Learning for Signal Processing (MLSP2020)*, Espoo, Sep. 2020.

[7] **L. Li** and K. Koishida, "Geometrically constrained independent vector analysis for directional speech enhancement," in Proc. *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2020)*, pp. 846–850, Barcelona, May 2020.

[8] **L. Li**, T. Toda, K. Morikawa, K. Kobayashi, and S. Makino, "Improving singing aid system for laryngectomees with statistical voice conversion and VAE-SPACE," in Proc. *20th International Society for Music Information Retrieval Conference (ISMIR2019)*, pp. 784–790, Delft, Nov. 2019.

[9] **L. Li**, K. Yamaoka, Y. Koshino, M. Matsumoto, and S. Makino, "Voice activity detection under high levels of noise using gated convolutional neural networks," in Proc. *International Congress on Acoustics (ICA2019)*, pp. 2862–2869, Aachen, Sep. 2019.

[10] **L. Li**, H. Kameoka, and S. Makino, "Fast MVAE: Joint separation and classification of mixed sources based on multichannel variational autoencoder with auxiliary classifier," in Proc. *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2019)*, pp. 546–550, Brighton, May 2019.

[11] **L. Li** and H. Kameoka, "Deep clustering with gated convolutional networks," in Proc. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2018)*, pp. 16–20, Calgary, Apr. 2018.

[12] **L. Li**, H. Kameoka, and S. Makino, "Mel-generalized cepstral regularization for discriminative non-negative matrix factorization," in Proc. *The 27th IEEE International Workshop on Machine Learning for Signal Processing (MLSP2017)*, Tokyo, Sep. 2017.

[13] **L. Li**, H. Kameoka, T. Toda, and S. Makino, "Speech enhancement using non-negative spectrogram models with mel-generalized cepstral regularization," in Proc. *The 18th Annual Conference of the International Speech Communication Association (Interspeech2017)*, pp. 1998–2002, Stockholm, Aug. 2017.

[14] **L. Li**, H. Kameoka, and S. Makino, "Discriminative non-negative matrix factorization with majorization-minimization," in Proc. *The 5th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA2017)*, pp. 141–145, San Francisco, Mar. 2017.

[15] **L. Li**, H. Kameoka, T. Higuchi, and H. Saruwatari, "Semi-supervised joint enhancement of spectral and cepstral sequences of noisy speech," in Proc. *The 17th Annual Conference of the International Speech Communication Association (Interspeech2016)*, pp. 3753–3757, San Francisco, Sep. 2016.

[16] S. Dang, **L. Li**, and S. Seki, "First Analyze Then Enhance: A Task-Aware System for Speech Separation, Denoising, and Dereverberation," in Proc. *The 26th Annual Conference of the International Speech Communication Association (Interspeech2025)*, pp. 3848–3852, Rotterdam, Aug. 2025.

[17] S. Seki and **L. Li**,, "Inference Efficient Source Separation Using Input-dependent Convolutions," in Proc. *The 16th annual conference of Asia-Pacific Signal and Information Processing Association (APSIPA2024)*, pp. 1–5, Macau, Dec. 2024.

[18] S. Ayano, **L. Li**, S. Seki, and D. Kitamura, "Audio spotforming using nonnegative tensor factorization with attractor-based regularization," in Proc. *The 2024 European Signal Processing Conference (EUSIPCO2024)*, pp. 121–125 Lyon, Aug. 2024.

[19] J. Wang, **L. Li**, and S. Makino, "Knowledge distillation with mask-based relationship for speech enhancement," in Proc. *RISP Internaonal Workshop on Nonlinear Circuits, Communicaions and Signal Processing (NCSP2023)*, pp. 29–32, Hawaii, Feb. 2023.

[20] M. Qin, **L. Li**, and S. Makino, "Deep complex-valued neural network-based triple-path mask and steering vector estimation for multi-channel target speech separation," in Proc. *RISP Internaonal Workshop on Nonlinear Circuits, Communicaions and Signal Processing (NCSP2023)*, pp. 33–36, Hawaii, Feb. 2023.

[21] K. Goto, T. Ueda, **L. Li**, S. Makino, and T. Yamada, "Investigation of partial update and source localization in geometrically constrained independent vector analysis with auxiliary function approach for moving source," in Proc. *RISP Internaonal Workshop on Nonlinear Circuits, Communicaions and Signal Processing (NCSP2023)*, pp. 426–429, Hawaii, Feb. 2023.

[22] Y. Hashizume, **L. Li**, and T. Toda, "Music similarity calculation of individiual instrumental sounds using metric learningm," in Proc. *The 14th annual conference of Asia-Pacific Signal and Information Processing Association (APSIPA2022)*, pp. 33–38, Chiang Mai, Nov. 2022.

[23] R. Wang, **L. Li**, and T. Toda, "Direction-aware target speaker extraction with a dual-channel system based on conditional variational autoencoders in underdetermined conditions," in Proc. *The 14th annual conference of Asia-Pacific Signal and Information Processing Association (APSIPA2022)*, pp. 347–353, Chiang Mai, Nov. 2022.

[24] S. Yamaji, T. Nakashima, N. Ono, **L. Li**, and H. Kameoka, "Encoder re-training with mixture signals on FastMVAE method," in Proc. *The 14th annual conference of Asia-Pacific Signal and Information Processing Association (APSIPA2022)*, pp 705–709, Chiang Mai, Nov. 2022.

[25] K. Goto, T. Ueda, **L. Li**, Takeshi Yamada, and S. Makino, "Accelerating online algorithm using geometrically constrained independent vector analysis with iterative source steering," in Proc. *The 14th annual conference of Asia-Pacific Signal and Information Processing Association (APSIPA2022)*, pp. 755–760, Chiang Mai, Nov. 2022.

[26] S. Furunaga, K. Goto, T. Ueda, **L. Li**, T. Yamada, and S. Makino, "Numerical investigation of weight parameters for geometrically constrained independent vector analysis using vectorwise coordinate descent or iterative source steering," in Proc. *The 17th International Workshop on Acoustic Signal Enhancement (IWAENC2022)*, F–06, Bamberg, Sep. 2022.

[27] K. Goto, T. Ueda, **L. Li**, T. Yamada, and S. Makino, "Geometrically constrained independent vector analysis with auxiliry function approach and iterative source steering," in Proc. *The 2022 European Signal Processing Conference (EUSIPCO2022)*, pp. 757–761, Belgrade, Aug. 2022.

[28] H. Kameoka, S. Seki, **L. Li**, and C. Watanabe, "AttentionPIT: Soft permutation invariant training for audio source separation with attention mechanism," in Proc. *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2022)*, pp. 706–710, Singapore, May 2022.

[29] S. Seki, H. Kameoka, and ,**L. Li** "Investigation and comparison of optimization methods for variational autoencoder-based underdetermined multichannel source separation," in Proc. *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2022)*, pp. 511–515, Singapore, May 2022.

[30] H. Segawa, **L. Li**, S. Makino, and T. Yamada, "Extension of virtual microphone technique to multiple real microphones and investigation of the impact of phase and amplitude interpolation on speech enhancement," in Proc. *The 13th annual conference of Asia-Pacific*

Signal and Information Processing Association (APSIPA2021), pp. 591–602, Tokyo, Dec. 2021.

[31] S. Nakaoka, **L. Li**, S. Makino, and T. Yamada, "Reducing algorithmic delay using low-overlap window for online Wave-U-Net," in Proc. *The 13th annual conference of Asia-Pacific Signal and Information Processing Association (APSIPA2021)*, pp. 1210–1214, Tokyo, Dec. 2021.

[32] S. Inoue, H. Kameoka, **L. Li**, and S. Makino, "SepNet: A deep separation matrix prediction network for multichannel audio source separation," in Proc. *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2021)*, pp. 191–195, Toronto, Jun. 2021.

[33] S. Nakaoka, **L. Li**, S. Inoue, and S. Makino, "Teacher-student learning for low-latency online speech enhancement using Wave-U-Net," in Proc. *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2021)*, pp. 661–665, Toronto, Jun. 2021.

[34] R. Takahashi, **L. Li**, S. Makino, and T. Yamada, "VMInNet: Interpolation of virtual microphones in optimal latent space explored by autoencoder," in Proc. *RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP2021)*, pp. 93–96, online, Mar. 2021.

[35] N. Murashima, H. Kameoka, **L. Li**, S. Seki, and S. Makino, "Single-channel multi-speaker separation via discriminative training of variational autoencoder spectrogram model," in Proc. *RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP2021)*, pp. 149–152, online, Mar. 2021.

[36] K. Goto, **L. Li**, R. Takahashi, S. Makino, and T. Yamada, "Study on geometrically constrained IVA with auxiliary function approach and VCD for in-car communication," in Proc. *The 12th annual conference of Asia-Pacific Signal and Information Processing Association (APSIPA2020)*, pp. 858–862, Auckland, Dec. 2020.

[37] R. Takahashi, K. Yamaoka, **L. Li**, S. Makino, T. Yamada, and M. Matsumoto, "Underdetermined multichannel speech enhancement using time-frequency-bin-wise switching beamformer and gated CNN-based time-frequency mask for reverberant environments," in Proc. *RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP2020)*, Hawaii, Feb. 2020.

[38] S. Seki, H. Kameoka, **L. Li**, T. Toda, and K. Takeda, "Generalized multichannel variational autoencoder for underdetermined source separation," in Proc. *The 2019 European Signal Processing Conference (EUSIPCO2019)*, pp. 1973–1977, A Coruña, Sep. 2019.

[39] K. Yamaoka, **L. Li**, N. Ono, S. Makino, and T. Yamada, "CNN-based virtual microphone signal estimation for MPDR Beamforming in underdetermined situations," in Proc. *The 2019 European Signal Processing Conference (EUSIPCO2019)*, pp. 1049–1053, A Coruña, Sep. 2019.

[40] S. Inoue, **L. Li**, H. Kameoka, and S. Makino, "Joint separation, dereverberation and classification of mixed sources using multichannel variational autoencoder with auxiliary classifier," in Proc. *International Congress on Acoustics (ICA2019)*, pp. 6988–6995, Aachen, Sep. 2019.

[41] S. Inoue, H. Kameoka, **L. Li**, S. Seki, and S. Makino, "Joint separation and dereverberation of reverberant mixtures with multichannel variational autoencoder," in Proc. *2019*

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2019), pp. 56–60, Brighton, May 2019.

ACADEMIC SERVICES

CONFERENCES SERVICES

- Registration co-chair of DCASE2024 Workshop in Oct. 2024

- Publicity co-chair of DCASE2020 Workshop in Nov. 2020

- Web administration of IWAENC2018 in Sep. 2020

- Session co-chairs (ICASSP2023, ICASSP2024, INTERSPEECH2025, Meetings of The Acoustical Society of Japan)

REVIEWERS

- IEEEACM Transactions on Audio, Speech and Language Processing

- IEEE Transactions on Signal Processing

- IEEE Signal Processing Letters

- IEEE Open Journal of Signal Processing

- EURASIP Journal on Audio, Speech, and Music Processing

- APSIPA Transactions on Signal and Information Processing

- Computer Speech & Language

- Journal of Computational and Applied Mathematics

- IEICE Transactions

- The Journal of the Acoustical Society of America

- International Conferences (ICASSP, INTERSPEECH, WASPAA, HSCMA, DCASE, AP-SIPA, etc)

OTHERS

- Technical Committee Mmember of ASJ Electroacoustics and IEICE Engineering Acoustics, 2024.06–