(Updated on Dec. 31, 2025)

Shota Horiguchi

horiguchi@ieee.org | shota.horiguchi@ntt.com | shota-horiguchi.github.io | linkedin.com/in/shota-horiguchi

Education

University of Tsukuba — Ibaraki, Japan
Ph.D. in Computer Science — *Apr. 2022 – Mar. 2023*
- Thesis title: Study on overlap-aware speaker diarization and its applications
- Supervisor: Prof. Takeshi Yamada

The University of Tokyo — Tokyo, Japan
Master of Information Science and Technology — *Apr. 2015 – May 2017*
- Thesis title: Personalized object recognition
- Supervisor: Prof. Kiyoharu Aizawa

The University of Tokyo — Tokyo, Japan
Bachelor of Engineering — *Apr. 2011 – May 2015*
- Supervisor: Prof. Kiyoharu Aizawa

Work Experiences

NTT, Inc., Human Informatics Laboratories — Kanagawa, Japan
Research Specialist — *February 2024 – Present*
- Speaker recognition in multi-talker environments (2024–Present)

Hitachi, Ltd. Research and Development Group — Tokyo, Japan
Senior Researcher — *October 2021 – January 2024*
Researcher — *April 2017 – September 2021*
- Meeting transcription using asynchronous distributed microphones (2019–2023)
- Multimodal recognition and interaction for humanoid robots (2017–2019)

Yahoo Japan Corporation — Tokyo, Japan
Research & Development Intern — *August, October – December 2015*
- Click through rates prediction

Publications

Journal Articles (first author)

[1] **S. Horiguchi**, S. Watanabe, P. Garcia, Y. Takashima, and Y. Kawaguchi, "Online neural diarization of unlimited numbers of speakers using global and local attractors", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 706–720, Jan. 2023.

[2] **S. Horiguchi**, Y. Fujita, S. Watanabe, Y. Xue, and P. García, "Encoder-decoder based attractors for end-to-end neural diarization", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 1493–1507, Mar. 2022.

[3] **S. Horiguchi**, D. Ikami, and K. Aizawa, "Significance of softmax-based features in comparison to distance metric learning-based features", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 5, pp. 1279–1285, May 2020.

[4] **S. Horiguchi**, S. Amano, M. Ogawa, and K. Aizawa, "Personalized classifier for food image recognition", *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2836–2848, Oct. 2018.

Journal Articles (co-author)

[5] N. Kamo, N. Tawara, A. Ando, T. Kano, H. Sato, R. Ikeshita, T. Moriya, **S. Horiguchi**, K. Matsuura, A. Ogawa, A. Plaquet, T. Ashihara, T. Ochiai, M. Mimura, M. Delcroix, T. Nakatani, T. Asami, and S. Araki, "Microphone array geometry independent multi-talker distant ASR: NTT system for the DASR task of the CHiME-8 challenge", *Computer Speech & Language*, vol. 95, p. 101 820, Jan. 2026.

Conference & Workshop Proceedings (Peer-reviewed, first author)

[6] **S. Horiguchi**, N. Tawara, T. Ashihara, A. Ando, and M. Delcroix, "Can we really repurpose multi-speaker ASR corpus for speaker diarization?", in *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Dec. 2025.

[7] **S. Horiguchi**, T. Ashihara, M. Delcroix, A. Ando, and N. Tawara, "Mitigating non-target speaker bias in guided speaker embedding", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Aug. 2025, pp. 5208–5212.

[8] **S. Horiguchi**, A. Ando, N. Tawara, and M. Delcroix, "Pretraining multi-speaker identification for neural speaker diarization", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Aug. 2025, pp. 1608–1612.

[9] **S. Horiguchi**, T. Moriya, A. Ando, T. Ashihara, H. Sato, N. Tawara, and M. Delcroix, "Guided speaker embedding", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2025.

[10] **S. Horiguchi**, A. Ando, T. Moriya, T. Ashihara, H. Sato, N. Tawara, and M. Delcroix, "Recursive attentive pooling for extracting speaker embeddings from multi-speaker recordings", in *IEEE Spoken Language Technology Workshop (SLT)*, Dec. 2024, pp. 1219–1226.

[11] **S. Horiguchi**, K. Dohi, and Y. Kawaguchi, "Streaming active learning for regression problems using regression via classification", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2024, pp. 4955–4959.

[12] A. Ito and **S. Horiguchi**, "Spoofing attacker also benefits from large-scale self-supervised models", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Aug. 2023, pp. 5346–5350.

[13] **S. Horiguchi**, Y. Takashima, S. Watanabe, and P. García, "Mutual learning of single- and multi-channel end-to-end neural diarization", in *IEEE Spoken Language Technology Workshop (SLT)*, Jan. 2023, pp. 620–625.

[14] **S. Horiguchi**, Y. Takashima, P. García, S. Watanabe, and Y. Kawaguchi, "Multi-channel end-to-end neural diarization with distributed microphones", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, pp. 7332–7336.

[15] **S. Horiguchi**, P. García, S. Watanabe, Y. Xue, Y. Takashima, and Y. Kawaguchi, "Towards neural diarization for unlimited numbers of speakers using global and local attractors", in *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Dec. 2021, pp. 98–105.

[16] **S. Horiguchi**, P. García, Y. Fujita, S. Watanabe, and K. Nagamatsu, "End-to-end speaker diarization as post-processing", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2021, pp. 7188–7192.

[17] **S. Horiguchi**, Y. Fujita, and K. Nagamatsu, "Block-online guided source separation", in *IEEE Spoken Language Technology Workshop (SLT)*, Jan. 2021, pp. 236–242.

[18] **S. Horiguchi**, Y. Fujita, S. Watanabe, Y. Xue, and K. Nagamatsu, "End-to-end speaker diarization for an unknown number of speakers with encoder-decoder based attractors", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Oct. 2020, pp. 269–273.

[19] **S. Horiguchi**, Y. Fujita, and K. Nagamatsu, "Utterance-wise meeting transcription system using asynchronous distributed microphones", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Oct. 2020, pp. 344–348.

[20] **S. Horiguchi**, N. Kanda, and K. Nagamatsu, "Multimodal response obligation detection with unsupervised online domain adaptation", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Sep. 2019, pp. 4180–4184.

[21] **S. Horiguchi**, N. Kanda, and K. Nagamatsu, "Face-voice matching using cross-modal embeddings", in *ACM International Conference on Multimedia (ACMMM)*, Oct. 2018, pp. 1011–1019.

[22] **S. Horiguchi**, K. Aizawa, and M. Ogawa, "The log-normal distribution of the size of objects in daily meal images and its application to the efficient reduction of object proposals", in *IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, pp. 3668–3672.

Conference & Workshop Proceedings (Peer-reviewed, co-author)

[23] N. Tawara and **S. Horiguchi**, "Target-speaker voice activity detection with chunk-level speaker queries", in *IEEE International Conference on Acoustics, Speech and Signal Processing Workshop (ICASSPW)*, May 2026.

[24] T. Ashihara, M. Delcroix, T. Ochiai, K. Matsuura, and **S. Horiguchi**, "Analysis of semantic and acoustic token variability across speech, music, and audio domains", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Aug. 2025.

[25] K. Fujita, **S. Horiguchi**, and Y. Ijima, "Voice impression control in zero-shot TTS", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Aug. 2025.

[26] N. Tawara, A. Ando, **S. Horiguchi**, and M. Delcroix, "Multi-channel speaker counting for EEND-VC-based speaker diarization on multi-domain conversation", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2025.

[27] A. Plaquet, N. Tawara, M. Delcroix, **S. Horiguchi**, A. Ando, and S. Araki, "Mamba-based segmentation model for speaker diarization", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2025.

[28] T. Moriya, **S. Horiguchi**, M. Delcroix, R. Masumura, T. Ashihara, H. Sato, K. Matsuura, and M. Mimura, "Alignment-free training for transducer-based multi-talker ASR", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2025.

[29] T. Ashihara, T. Moriya, **S. Horiguchi**, J. Peng, T. Ochiai, M. Delcroix, K. Matsuura, and H. Sato, "Investigation of speaker representation for target-speaker speech processing", in *IEEE Spoken Language Technology Workshop (SLT)*, Dec. 2024, pp. 433–440.

[30] A. Ando, T. Moriya, **S. Horiguchi**, and R. Masumura, "Factor-conditioned speaking-style captioning", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Sep. 2024, pp. 782–786.

[31] H. Sato, T. Moriya, M. Mimura, **S. Horiguchi**, T. Ochiai, T. Ashihara, A. Ando, K. Shinayama, and M. Delcroix, "SpeakerBeam-SS: Real-time target speaker extraction with lightweight Conv-TasNet and state space modeling", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Sep. 2024, pp. 5033–5037.

[32] T. V. Ho, **S. Horiguchi**, S. Watanabe, P. Garcia, and T. Sumiyoshi, "Synthetic data augmentation for ASR with domain filtering", in *Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Nov. 2023, pp. 1760–1765.

[33] Y. Okamoto, K. Shimonishi, K. Imoto, K. Dohi, **S. Horiguchi**, and Y. Kawaguchi, "CAPTDURE: Captioned sound dataset of single sources", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Aug. 2023, pp. 1683–1687.

[34] Y. Takashima, **S. Horiguchi**, S. Watanabe, P. Garcia, and Y. Kawaguchi, "Updating only encoders prevents catastrophic forgetting of end-to-end ASR models", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Sep. 2022, pp. 2218–2222.

[35] T. Morishita, G. Morio, **S. Horiguchi**, H. Ozaki, and N. Nukaga, "Rethinking Fano's inequality in ensemble learning", in *International Conference on Machine Learning (ICML)*, Jul. 2022, pp. 15 976–16 016.

[36] N. Yamashita, **S. Horiguchi**, and T. Homma, "Improving the naturalness of simulated conversations for end-to-end neural diarization", in *The Speaker and Language Recognition Workshop (Odyssey)*, Jun. 2022, pp. 133–140.

[37] Y. Okamoto, **S. Horiguchi**, M. Yamamoto, K. Imoto, and Y. Kawaguchi, "Environmental sound extraction using onomatopoeic words", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, pp. 221–225.

[38] Y. Xue, **S. Horiguchi**, Y. Fujita, Y. Takashima, S. Watanabe, P. Garcia, and K. Nagamatsu, "Online streaming end-to-end neural diarization handling overlapping speech and flexible numbers of speakers", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Sep. 2021, pp. 3116–3120.

[39] Y. Takashima, Y. Fujita, **S. Horiguchi**, S. Watanabe, P. Garcia, and K. Nagamatsu, "Semi-supervised training with pseudo-labeling for end-to-end neural diarization", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Sep. 2021, pp. 3096–3110.

[40] Y. Xue, **S. Horiguchi**, Y. Fujita, S. Watanabe, P. Garcia, and K. Nagamatsu, "Online end-to-end neural diarization with speaker-tracing buffer", in *IEEE Spoken Language Technology Workshop (SLT)*, Jan. 2021, pp. 841–848.

[41] Y. Takashima, Y. Fujita, S. Watanabe, **S. Horiguchi**, P. Garcia, and K. Nagamatsu, "End-to-end speaker diarization conditioned on speech activity and overlap detection", in *IEEE Spoken Language Technology Workshop (SLT)*, Jan. 2021, pp. 849–856.

[42] K. Ito, Q. Kong, **S. Horiguchi**, T. Sumiyoshi, and K. Nagamatsu, "Anticipating the start of user interaction for service robot in the wild", in *IEEE International Conference on Robotics and Automation (ICRA)*, Jun. 2020, pp. 9687–9693.

[43] N. Kanda, **S. Horiguchi**, Y. Fujita, Y. Xue, K. Nagamatsu, and S. Watanabe, "Simultaneous speech recognition and speaker diarization for monaural dialogue recordings with target-speaker acoustic models", in *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Dec. 2019, pp. 31–38.

[44] Y. Fujita, N. Kanda, **S. Horiguchi**, Y. Xue, K. Nagamatsu, and S. Watanabe, "End-to-end neural speaker diarization with self-attention", in *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Dec. 2019, pp. 296–303.

[45] N. Kanda, **S. Horiguchi**, R. Takashima, Y. Fujita, K. Nagamatsu, and S. Watanabe, "Auxiliary interference speaker loss for target-speaker speech recognition", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Sep. 2019, pp. 236–240.

[46] N. Kanda, C. Boeddeker, J. Heitkaemper, Y. Fujita, **S. Horiguchi**, K. Nagamatsu, and R. Haeb-Umbach, "Guided source separation meets a strong ASR backend: Hitachi/Paderborn University joint investigation for dinner party scenario", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Sep. 2019, pp. 1248–1252.

[47] Y. Fujita, N. Kanda, **S. Horiguchi**, K. Nagamatsu, and S. Watanabe, "End-to-end neural speaker diarization with permutation-free objectives", in *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Sep. 2019, pp. 4300–4304.

[48] N. Kanda, Y. Fujita, **S. Horiguchi**, R. Ikeshita, K. Nagamatsu, and S. Watanabe, "Acoustic modeling for distant multi-talker speech recognition with single- and multi-channel branches", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 6630–6634.

[49] M. Tamura, **S. Horiguchi**, and T. Murakami, "Omnidirectional pedestrian detection by rotation invariant training", in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2019, pp. 1989–1998.

[50] S. Amano, **S. Horiguchi**, K. Aizawa, K. Maeda, M. Kubota, and M. Ogawa, "Food search based on user feedback to assist image-based food recording systems", in *International Workshop On Multimedia Assisted Dietary Management (MADiMa)*, Oct. 2016, pp. 71–75.

Preprints

[51] A. Plaquet, N. Tawara, M. Delcroix, **S. Horiguchi**, A. Ando, S. Araki, and H. Bredin, *Dissecting the segmentation model of end-to-end diarization with vector clustering*, arXiv:2506.11605, Jun. 2025.

[52] H. Namba, **S. Horiguchi**, M. Hamamoto, and M. Egi, *Thresholding data shapley for data cleansing using multi-armed bandits*, arXiv:2402.08209, Feb. 2024.

[53] Y. Fujita, S. Watanabe, **S. Horiguchi**, Y. Xue, and N. Kenji, *End-to-end neural diarization: Reformulating speaker diarization as simple multi-label classification*, arXiv:2003.20966, Feb. 2020.

[54] Y. Fujita, S. Watanabe, **S. Horiguchi**, Y. Xue, J. Shi, and N. Kenji, *Neural speaker diarization with speaker-wise chain rule*, arXiv:2006.01796, Jun. 2020.

Invited Talk

- "Speaker Diarization: A Key to Solving Cocktail Party Problem," in IEEE RO-MAN Workshop, Aug. 28, 2023.
- "Face-Voice Matching Using Cross-Modal Embeddings," in MIRU, July 29–Aug. 1, 2020 (in Japanese).

Awards

- IEEE SPS Young Author Best Paper Award, 2025.
- IEEE SPS Japan Young Auhor Best Paper Award, 2025.
- Honorable Mention Award at IEEE Spoken Language Technology Workshop, 2024.
- Itakura Prize Innovative Young Researcher Award, The Acoustical Society of Japan, 2023.
- ITE Outstanding Research Presentation Award, The Institute of Image Information and Television Engineers, 2017.

Competitions

- 2nd prize in the 8th CHiME Speech Separation and Recognition Challenge (CHiME-8) Task 1, 2024.
- 2nd prize in the Third DIHARD Speech Diarization Challenge (DIHARD III), 2021 (as a lead author).
- 2nd prize in the 5th CHiME Speech Separation and Recognition Challenge (CHiME-5), 2018.

Patents and Applications

- US Patent 11,107,476 B2 "Speaker estimation method and speaker estimation device," August 31, 2021.
- Contributed to six patent applications at Hitachi.

Professional Activities/Service

2022 Session chair (ICASSP, in-person conference at Singapore)

Review Experiences

Journals: IEEE/ACM Transactions on Audio, Speech, and Language processing, IEEE Transactions on Multimedia, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Pattern Recognition and Machine Intelligence, IEEE Transactions on Affective Computing, Computer Speech & Language, Speech Communication, EURASIP Journal on Audio, Speech and Language Processing, Neural Networks
Conferences/Workshops: ICASSP, EUSIPCO, ASRU, SLT, WASPAA, DCASE, MLSP, MMSP, APSIPA ASC, ACMMM, RO-MAN, ICML

Language Skills

Japanese (native), English (business)