Shota Horiguchi

horiguchi@ieee.org | shota.horiguchi.wk@hitachi.com | linkedin.com/in/shota-horiguchi |

EDUCATION

University of Tsukuba   Ibaraki, Japan
Doctor of Information Science and Technology   *Apr. 2022 – Present*
  - Supervisor: Assoc. 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

Hitachi, Ltd. Research and Development Group   Tokyo, Japan
Senior Researcher   *October 2021 – Present*
Researcher   *April 2017 – September 2021*
  - Meeting transcription using asynchronous distributed microphones (2019–Present)
  - 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

PROJECTS

Meeting transcription using distributed microphones   April 2019 – Present
  - Developed a meeting transcription framework based on distributed microphones, e.g., smartphones or tablets [9].
  - Developed its component technologies including speech separation [7] and speaker diarization [4–6, 8].
  - Joined the Seventh Frederick Jelinek Memorial Summer Workshop (JSALT) as a part-time member, especially for discussion and technical support about speaker diarization [6].
  - Won the second prize in the Third DIHARD Speech Diarization Challenge (DIHARD III).

Multi-modal recognition and human robot interaction   April 2017 – March 2019
  - Developed multi-modal (especially audio-visual) functions for Hitachi's humanoid robot named EMIEW3, including multi-modal response obligation detection [10] and cross-modal biometric search between faces and voices [11].

Personalized image recognition   April 2014 – March 2017
  - Developed personalized image classifiers, especially for daily life-logging service [3].
  - Investigated better image representations for more general settings using convolutional neural networks [2].

PUBLICATIONS

Journal Articles

[1] **S. Horiguchi**, Y. Fujita, S. Watanabe, Y. Xue, and P. García, "Encoder-decoder based attractors for end-to-end neural diarization", *IEEE/ACM Trans. on Audio, Speech and Language Processing*, vol. 30, pp. 1497–1507, 2022.

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

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

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

[4] **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)*, 2022, pp. 7332–7336.

[5] **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)*, 2021, pp. 98–105.

[6] **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)*, 2021, pp. 7188–7192.

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

[8] **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)*, 2020, pp. 269–273.

[9] **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)*, 2020, pp. 344–348.

[10] **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)*, 2019, pp. 4180–4184.

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

[12] **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)*, 2016, pp. 3668–3672.

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

[13] Y. Takashima, **S. Horiguchi**, S. Watanabe, P. Garcia, and Y. Kawaguchi, "Updating only encoders prevents catastrophic forgetting of end-to-end ASR models", in *INTERSPEECH*, 2022.

[14] 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)*, 2022.

[15] 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)*, 2022.

[16] 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)*, 2022, pp. 221–225.

[17] 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 *INTERSPEECH*, 2021, pp. 3116–3120.

[18] 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 *INTERSPEECH*, 2021, pp. 3096–3110.

[19] 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)*, 2021, pp. 841–848.

[20] 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)*, 2021, pp. 849–856.

[21] 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)*, 2020, pp. 9687–9693.

[22] 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)*, 2019, pp. 31–38.

[23] 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)*, 2019, pp. 296–303.

[24] N. Kanda, **S. Horiguchi**, R. Takashima, Y. Fujita, K. Nagamatsu, and S. Watanabe, "Auxiliary interference speaker loss for target-speaker speech recognition", in *INTERSPEECH*, 2019, pp. 236–240.

[25] 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 *INTERSPEECH*, 2019, pp. 1248–1252.

[26] Y. Fujita, N. Kanda, **S. Horiguchi**, K. Nagamatsu, and S. Watanabe, "End-to-end neural speaker diarization with permutation-free objectives", in *INTERSPEECH*, 2019, pp. 4300–4304.

[27] 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)*, 2019, pp. 6630–6634.

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

[29] 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)*, 2016, pp. 71–75.

Preprints
[30] **S. Horiguchi**, S. Watanabe, P. Garcia, Y. Takashima, and Y. Kawaguchi, *Online neural diarization of unlimited numbers of speakers*, arXiv:2206.02432.

Invited Talk
- "Face-Voice Matching Using Cross-Modal Embeddings," in MIRU, July 29–Aug. 1, 2020 (in Japanese).

AWARDS

- ITE Outstanding Research Presentation Award, The Institute of Image Information and Television Engineers, 2017.

COMPETITIONS

- 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 Transactions on Pattern Recognition and Machine Intelligence, IEEE Transactions on Multimedia, IEEE Transactions on Audio, Speech and Language Processing, IEEE Transactions on Neural Networks and Learning Systems, Elsevier Neural Networks, Elsevier Computer Speech & Language
Conferences/Workshops: ICASSP (2019–2022), ICML (2020), MLSP (2021), EUSIPCO (2021–2022) DCASE (2020–2021), APSIPA (2021), RO-MAN (2022)

LANGUAGE SKILLS

Japanese (native), English (business)