HIROFUMI INAGUMA

Research building #7, Room #417, Yoshidahonmachi, Sakyo-ku, Kyoto, 606-8501, Japan
Phone (office):(+81)-75-753-5992 ⋄ (mobile): (+81)-80-6967-1623
E-mail (office): inaguma [at] sap.ist.i.kyoto-u.ac.jp / (private): hiro.mhbc [at] gmail.com

RESEARCH INTERESTS

Automatic speech recognition (ASR)
- End-to-end speech recognition
- Online Streaming ASR

Speech translation
- End-to-end speech translation
- Multilingual translation
- Non-autoregressive modeling
- Knowledge distillation

EDUCATION

Ph.D. in Computer Science, Kyoto University, Kyoto, Japan *April 2018 - Present*
Department of Intelligence Science and Technology, Graduate School of Informatics
- Supervisor: Prof. Tatsuya Kawahara

M.E. in Computer Science, Kyoto University, Kyoto, Japan *April 2016 - March 2018*
Department of Intelligence Science and Technology, Graduate School of Informatics
- Thesis title: Joint Social Signal Detection and Automatic Speech Recognition based on End-to-End Modeling and Multi-task Learning
- Supervisor: Prof. Tatsuya Kawahara

B.E. in Computer Science, Kyoto University, Kyoto, Japan *April 2012 - March 2016*
- Supervisor: Prof. Tatsuya Kawahara

WORK EXPERIENCES

Microsoft Research, Redmond, WA, USA *July 2019 - October 2019*
Research Internship (3 months)
- Mentor: Yifan Gong, Jinyu Li, Yashesh, Gaur, and Liang Lu

Johns Hopkins University, Baltimore, MD, USA *July 2018 - September 2018*
Visiting student (2.5 months)
- Worked on end-to-end speech recognition and translation
- Participated in the JSALT workshop (topic: multilingual end-to-end speech recognition)
- Participated in the IWSLT2018 evaluation campaign on the end-to-end speech translation track
- Mentor: Prof. Shinji Watanabe

IBM research AI, Tokyo, Japan *September 2017 - November 2017*
Research Internship (2 months)
- Worked on end-to-end ASR systems
- Mentor: Gakuto Kurata, and Takashi Fukuda

Recruit Co.,Ltd., Tokyo, Japan *February 2017*
Research Intern (2 weeks)
- Worked on data analysis and image classification competition
- Won the 1st place

CyberAgent, Inc., Tokyo, Japan *September 2016 - October 2016*
Research Intern (1 months)
- Worked on the development of music recommendation systems

TECHNICAL STRENGTHS

Programming language	Python, Bash, C/C++, Java, LaTeX, PHP, Javascript, CSS, HTML
Software & Tools	ESPnet, Kaldi, Docker
Deep learning frameworks	Pytorch, Tensorflow, Chainer

LANGUAGE SKILL

Japanese (native), English (fluent)

AWARDS

14th IEEE Signal Processing Society (SPS) Japan Student Conference Paper Award, from IEEE Signal Processing Society (SPS) Tokyo Joint Chapter, December 2020
- Paper title: "*Minimum Latency Training Strategies for Streaming Sequence-to-Sequence ASR*"

Microsoft Research Asia Ph.D. Fellowship Award (top 12 phd students in Asia), from Microsoft Research Asia (MSRA), October 2019

Yamashita SIG Research Award, from Information Processing Society of Japan (IPSJ), March 2019
- Paper title: "*An End-to-End Approach to Joint Social Signal Detection and Automatic Speech Recognition*"

Yahoo! JAPAN award (best student paper), from SIG-SLP, June 2018

Full exemption from Repayment of Scholarship Loan for Students with Outstanding Results, from Japan Student Services Organization (JASSO), May 2018.

Research Fellowship for Young Scientists (DC1), from Japan Society for the Promotion of Science (JSPS), April 2018 - March 2021

Student award, from the Acoustical Society of Japan (ASJ), March 2018

Student award, from the 79th of National Convention of Information Processing Society of Japan (IPSJ), March 2017

ACADEMIC SERVICES

Reviewing

Interspeech: 2021

ICASSP: 2019, 2020, 2021

IWSLT: 2021

IEEE/ACM Transactions on Audio Speech and Language Processing: 2020, 2021

APSIPA Transactions on Signal and Information Processing: 2020

IEEE Signal Processing Letters: 2020, 2021

PUBLICATIONS (PREPRINT)

Hirofumi Inaguma and Tatsuya Kawahara, "Alignment Knowledge Distillation for Online Streaming Attention-based Speech Recognition, https://arxiv.org/abs/2103.00422.

Hirofumi Inaguma, Siddharth Dalmia, Brian Yan, and Shinji Watanabe, "Fast-MD: Fast Multi-Decoder End-to-End Speech Translation with Non-Autoregressive Hidden Intermediates.

PUBLICATIONS (REVIEW PAPER, FIRST AUTHOR)

[IWSLT2021 (system)] **Hirofumi Inaguma***, Brian Yan*, Siddharth Dalmia, Pengcheng Guo, Jiatong Shi, Kevin Duh, and Shinji Watanabe (*equal contribution), "ESPnet-ST IWSLT 2021 Offline Speech Translation System", 18th International Conference on Spoken Language Translation (IWSLT), 2021 (accepted), https://arxiv.org/abs/2107.00636.

[Interspeech2021] **Hirofumi Inaguma** and Tatsuya Kawahara, "StableEmit: Selection Probability Discount for Reducing Emission Latency of Streaming Monotonic Attention ASR", Interspeech, 2021 (accepted), https://arxiv.org/abs/2107.00635.

[Interspeech2021] **Hirofumi Inaguma** and Tatsuya Kawahara, "VAD-free Streaming Hybrid CTC/Attention ASR for Unsegmented Recording", Interspeech, 2021 (accepted), https://arxiv.org/abs/2107.07509

[NAACL-HLT2021] **Hirofumi Inaguma**, Tatsuya Kawahara, and Shinji Watanabe, "Source and Target Bidirectional Knowledge Distillation for End-to-end Speech Translation", NAACL-HLT, pp. 1872-1881, 2021, https://arxiv.org/abs/2104.06457

[DSLW2021] Shinji Watanabe, Florian Boyer, Xuankai Chang, Pengcheng Guo, Tomoki Hayashi, Yosuke Higuchi, Takaaki Hori, Wen-Chin Huang, **Hirofumi Inaguma**, Naoyuki Kamo, Shigeki Karita, Chenda Li, Jing Shi, Aswin Shanmugam Subramanian, and Wangyou Zhang, "The 2020 ESPnet update: new features, broadened applications, performance improvements, and future plans", IEEE Data Science and Learning Workshop (DSLW), 2021, https://arxiv.org/abs/2012.13006.

[ICASSP2021] **Hirofumi Inaguma**, Yosuke Higuchi, Kevin Duh, Tatsuya Kawahara, and Shinji Watanabe, "Orthros: Non-autoregressive End-to-end Speech Translation with Dual-decoder", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7488-7492, 2021, https://arxiv.org/abs/2010.13047. (acceptance rate: 1734/3610=48.0%)

[ICASSP2021] Yosuke Higuchi, **Hirofumi Inaguma**, Shinji Watanabe, Tetsuji Ogawa, and Tetsunori Kobayashi, "Improved Mask-CTC for Non-Autoregressive End-to-End ASR", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3655-3659, 2021, https://arxiv.org/abs/2010.13270.

[ICASSP2021] Pengcheng Guo, Florian Boyer, Xuankai Chang, Tomoki Hayashi, Yosuke Higuchi, **Hirofumi Inaguma**, Naoyuki Kamo, Chenda Li, Daniel Garcia-Romero, Jiatong Shi, Jing Shi, Shinji Watanabe, Kun Wei, Wangyou Zhang, and Yuekai Zhang, "Recent Developments on ESPnet Toolkit Boosted by Conformer", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5874-5878, 2021, https://arxiv.org/abs/2010.13956.

[Interspeech2020] **Hirofumi Inaguma**, Masato Mimura, and Tatsuya Kawahara, "CTC-synchronous Training for Monotonic Attention Model", the 21th Annual Conference of International Speech Communication Association (Interspeech), pp. 571-575, 2020, https://arxiv.org/abs/2005.04712. (acceptance rate: 47.0%)

[Interspeech2020] **Hirofumi Inaguma**, Masato Mimura, and Tatsuya Kawahara, "Enhancing Monotonic Multihead Atteniton for Streaming ASR", the 21th Annual Conference of International Speech Communication Association (Interspeech), pp. 2137-2141, 2020, https://arxiv.org/abs/2005.09394.

[Interspeech2020] Hayato Futami, **Hirofumi Inaguma**, Sei Ueno, Masato Mimura, Shinsuke Sakai, and Tatsuya Kawahara, "Distilling the Knowledge of BERT for Sequence-to-Sequence ASR", the 21th Annual Conference of International Speech Communication Association (Interspeech), pp. 3635-3639, 2020, https://arxiv.org/abs/2008.03822.

[Interspeech2020] Trung V. Dang, Tianyu Zhao, Sei Ueno, **Hirofumi Inaguma**, and Tatsuya Kawahara, "End-to-end speech-to-dialog-act recognition", the 21th Annual Conference of International Speech Communication Association (Interspeech), pp. 3910-3914, 2020, https://arxiv.org/abs/2004.11419.

[ACL2020 (demo)] **Hirofumi Inaguma**, Shun Kiyono, Kevin Duh, Shigeki Karita, Nelson Yalta, Tomoki Hayashi, and Shinji Watanabe, "ESPnet-ST: All-in-One Speech Translation Toolkit", the 58th Annual Meeting of the Association for Computational Linguistics (ACL): System Demonstrations, pp. 302-311, 2020, https://www.aclweb.org/anthology/2020.acl-demos.34/.

[ICASSP2020] **Hirofumi Inaguma**, Yashesh, Gaur, Liang Lu, Jinyu Li, and Yifan Gong, "Minimum latency training strategies for streaming sequence-to-sequence ASR", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6064-6068, 2020, https://arxiv.org/abs/2004.05009. (acceptance rate: 47%, **Oral**)

[ASRU2019] **Hirofumi Inaguma**, Kevin Duh, Tatsuya Kawahara, and Shinji Watanabe, "Multilingual End-To-End Speech Translation", IEEE Automatic Speech Recognition and Understanding Workshop (**ASRU**), pp. 570-577, 2019, https://arxiv.org/abs/1910.00254. (acceptance rate: 144/299=48.1%)

[ASRU2019] Shigeki Karita, Nanxin Chen, Tomoki Hayashi, Takaaki Hori, **Hirofumi Inaguma**, Ziyan Jiang, Masao Someki, Nelson Enrique Yalta Soplin, Ryuichi Yamamoto, Xiaofei Wang, Shnji Watanabe, Takenori Yoshimura, and Wangyou Zhang, "A Comparative Study on Transformer vs RNN in Speech Applications"', IEEE Automatic Speech Recognition and Understanding Workshop (**ASRU**), pp. 499-456, 2019, https://arxiv.org/abs/1909.06317.

[ICASSP2019] **Hirofumi Inaguma**, Jaejin Cho, Murali Karthick Baskar, Tatsuya Kawahara, and Shinji Watanabe, "Transfer Learning of Language-Independent End-to-End ASR with Language Model Fusion", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6096-6100, 2019, https://arxiv.org/abs/1811.02134. (acceptance rate: 1774/3815=46.5%)

[ICASSP2019] Jaejin Cho, Shinji Watanabe, Takaaki Hori, Murali Karthick Baskar, **Hirofumi Inaguma**, Jesus Villalba, and Najim Dehak, "Language Model Integration Based on Memory Control for Sequence to Sequence Speech Recognition", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6191-6195, 2019, https://arxiv.org/abs/1811.02162.

[SLT2018] **Hirofumi Inaguma**, Masato Mimura, Shinsuke Sakai, and Tatsuya Kawahara, "Improving OOV Detection and Resolution with External Language Models in Acoustic-to-Word ASR", IEEE Spoken Language Technology Workshop (SLT), pp. 212-218, 2018, http://sap.ist.i.kyoto-u.ac.jp/EN/bib/intl/INA-SLT18.pdf. (acceptance rate: 150/257=58.3%)

[SLT2018] Masato Mimura, Sei Ueno, **Hirofumi Inaguma**, Shinsuke Sakai, and Tatsuya Kawahara, "Leveraging Sequence-to-Sequence Speech Synthesis for Enhancing Acoustic-to-Word Speech Recognition", IEEE Spoken Language Technology Workshop (SLT), pp. 477-484, 2018, http://sap.ist.i.kyoto-u.ac.jp/lab/bib/intl/MIM-SLT18.pdf.

[IWSLT2018 (system)] **Hirofumi Inaguma**, Xuan Zhang, Zhiqi Wang, Adithya Renduchintala, Shinji Watanabe, and Kevin Duh, "The JHU/KyotoU Speech Translation System for IWSLT 2018", 15th International Conference on Spoken Language Translation (IWSLT), 2018.

[ICASSP2018] **Hirofumi Inaguma**, Masato Mimura, Koji Inoue, Kazuyoshi Yoshii, and Tatsuya Kawahara, "An End-to-End Approach to Joint Social Signal Detection and Automatic Speech Recognition", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.

6214-6218, 2018. (acceptance rate: 1406/2830=49.7%)

[ICASSP2018] Sei Ueno, **Hirofumi Inaguma**, Masato Mimura, and Tatsuya Kawahara, "Acoustic-to-Word Attention-Based Model Complemented with Character-level CTC-Based Model", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5804-5808, 2018, `http://www.sap.ist.i.kyoto-u.ac.jp/lab/bib/intl/UEN-ICASSP18.pdf`.

[Interspeech2017] **Hirofumi Inaguma**, Koji Inoue, Masato Mimura, and Tatsuya Kawahara, "Social Signal Detection in Spontaneous Dialogue Using Bidirectional LSTM-CTC", 18th Annual Conference of International Speech Communication Association (Interspeech), pp. 1691-1695, 2017. (acceptance rate: 799/1582=52.0%)

[ICMI2016 WS] **Hirofumi Inaguma**, Koji Inoue, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara, "Prediction of Ice-breaking between participants using prosodic features in the first meeting dialogue", International Conference Multimodal Interaction workshop on Advancements in Social Signal Processing for Multimodal Interaction (ASSP4MI), 2016.

(Last update: 07/05/2021)