

# 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
- Language modeling
- Transfer learning, semi-supervised training
- Online/Streaming ASR
- Multilingual ASR

### **Speech translation**

- End-to-end speech translation
- Multilingual translation

## 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

---

<b>Programming language</b>	Python, Bash, C/C++, Java, LaTeX, PHP, Javascript, CSS, HTML
<b>Software &amp; Tools</b>	ESPnet, Kaldi, Docker
<b>Deep learning frameworks</b>	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

---

**Reviewer:** ICASSP2021, ICASSP2020, ICASSP2019, APSIPA Transactions on Signal and Information Processing, IEEE Signal Processing Letters

## PUBLICATIONS (REVIEW PAPER, FIRST AUTHOR)

---

**Hirofumi Inaguma**, Yosuke Higuchi, Kevin Duh, Tatsuya Kawahara, and Shinji Watanabe, "Orthros: Non-autoregressive End-to-end Speech Translation with Dual-decoder", International Conference on Acoustics, Speech and Signal Processing (**ICASSP**), 2021, <https://arxiv.org/abs/2010.13047>. (Acceptance rate:  $1734/3610=48.0\%$ )

**Hirofumi Inaguma**, Masato Mimura, and Tatsuya Kawahara, "CTC-synchronous Training for Monotonic Attention Model", the 21th Annual Conference of International Speech Communication Association (**Interspeech**), 2020, <https://arxiv.org/abs/2005.04712>. (Acceptance Rate: 47.0%)

**Hirofumi Inaguma**, Masato Mimura, and Tatsuya Kawahara, "Enhancing Monotonic Multihead Attention for Streaming ASR", the 21th Annual Conference of International Speech Communication Association (**Interspeech**), 2020, <https://arxiv.org/abs/2005.09394>.

**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, 2020, <https://www.aclweb.org/anthology/2020.acl-demos.34/>.

**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**), 2020, <https://arxiv.org/abs/2004.05009>. (Acceptance rate: 47%, **Oral**)

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

**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**), 2019, <http://sap.ist.i.kyoto-u.ac.jp/EN/bib/intl/INA-ICASSP19.pdf>. (Acceptance rate: 1774/3815=46.5%)

**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**), 2018, <http://sap.ist.i.kyoto-u.ac.jp/EN/bib/intl/INA-SLT18.pdf>. (Acceptance rate: 150/257=58.3%)

**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.

**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", International Conference on Acoustics, Speech and Signal Processing (**ICASSP**), 2018. (Acceptance rate: 1406/2830=49.7%)

**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**), 2017. (Acceptance rate: 799/1582=52.0%)

**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.

## PUBLICATIONS (REVIEW PAPER, CO-AUTHOR)

---

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

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, Yuekai Zhang, "Recent Developments on ESPnet Toolkit Boosted by Conformer", International Conference on Acoustics, Speech and Signal Processing (**ICASSP**), 2021, <https://arxiv.org/abs/2010.13956>.

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

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

Shigeki Karita, Nanxin Chen, Tomoki Hayashi, Takaaki Hori, **Hirofumi Inaguma**, Ziyang Jiang, Masao Someki, Nelson Enrique Yalta Soplín, Ryuichi Yamamoto, Xiaofei Wang, Shinji Watanabe, Takenori Yoshimura, and Wangyou Zhang, "A Comparative Study on Transformer vs RNN in Speech Applications", IEEE Automatic Speech Recognition and Understanding Workshop (**ASRU**), 2019, <https://arxiv.org/abs/1909.06317>.

Jaejin Cho, Shinji Watanabe, Takaaki Hori, Murali Karthick Baskar, **Hirofumi Inaguma**, Jesus Villalba, 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**), 2019, <https://arxiv.org/abs/1811.02162>.

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**), 2018, <http://sap.ist.i.kyoto-u.ac.jp/lab/bib/intl/MIM-SLT18.pdf>.

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**), 2018, <http://www.sap.ist.i.kyoto-u.ac.jp/lab/bib/intl/UEN-ICASSP18.pdf>.

(Last update: 01/30/2021)