

Takaaki Saeki

saefrospace@gmail.com | takaaki-saeki.github.io | linkedin.com/in/takaaki-saeki

RESEARCH INTERESTS

Spoken language processing: Speech synthesis, speech translation, multimodal speech processing

Other: Natural language processing, signal processing, deep learning

EDUCATION

The University of Tokyo

Tokyo, Japan

Ph.D. Degree in Graduate School of Information Science and Technology

Apr. 2021 – Present

- Supervisor: Prof. Hiroshi Saruwatari

The University of Tokyo

Tokyo, Japan

Master's Degree in Graduate School of Information Science and Technology

Apr. 2019 – Mar. 2021

- Thesis Title: "Real-Time, Full-Band, High-Quality Neural Voice Conversion with Sub-Band Modeling and Data-Driven Phase Estimation"
- Supervisor: Prof. Hiroshi Saruwatari

The University of Tokyo

Tokyo, Japan

Bachelor's Degree in Aerospace Engineering

Apr. 2015 – Mar. 2019

- Supervisor: Prof. Kimiya Komurasaki

EXPERIENCE

Google Research

Tokyo, Japan

Student Researcher (Intern)

Apr 2022 – Present

- Researching on spoken language processing
- Advisor: H. Zen

The University of Tokyo

Tokyo, Japan

Research Assistant

May. 2020 – Mar. 2022

- Researched on streaming speech processing
- Advisor: H. Saruwatari and S. Takamichi

LINE Corporation

Tokyo, Japan

Part-Time Researcher

Mar. 2021 – Mar. 2022

- Designing research proposals and implementing new text-to-speech synthesis methods.
- Advisor: K. Tahchibana and R. Yamamoto

Preferred Networks, Inc.

Tokyo, Japan

Research Intern

Aug. 2021 – Sep. 2021

- Researched on application of machine learning to music creation
- Advisor: D. Higurashi and B. Vogel

NEC Data Science Research Laboratories

Kanagawa, Japan

Research Intern

Aug. 2019 – Sep. 2019

- Worked on speech enhancement.
- Advisor: O. Hoshuyama

Recruit Holdings Co., Ltd.

Tokyo, Japan

Data Engineering Intern & Part-Time Engineer

Feb. 2019 – June 2019

- Worked on data analysis and developed recommendation engine.
- Lectured a training course of data analysis for newcomers.
- Advisor: N. Osugi

Delight Inc.

Tokyo, Japan

Part-Time Engineer

May 2017 – Apr. 2019

- Developed simulation software of fluid dynamics.
- Read research papers and implemented algorithms of thermo-fluid dynamics.

TECHNICAL SKILLS

Languages: Python, C/C++, Shell, LaTeX, SQL (BigQuery), HTML/CSS

Developer Tools: Git, Docker, Amazon Web Service, Google Cloud Platform, VS Code

Libraries: NumPy, pandas, PyTorch, Tensorflow, Chainer

LANGUAGE SKILLS

Japanese (native) and English (fluent)

HONORS AND AWARDS

Ranked 1st Place in Several Metrics among 22 Teams <i>VoiceMOS Challenge 2022</i>	<i>Mar. 2022</i>
Telecom System Technology Award for Students <i>The Telecommunication Advancement Foundation</i>	<i>Mar. 2022</i>
Best Student Presentation Award <i>Acoustical Society of Japan</i>	<i>Oct. 2021</i>
IEICE ISS Best Student Poster Award <i>IEICE Speech Committee 2021</i>	<i>Mar. 2021</i>
Award for Excellence (2nd place) <i>Natural Language Processing Hackathon at Recruit Co., Ltd.</i>	<i>Feb. 2019</i>

GRANTS AND SCHOLARSHIPS

Research Fellowship for Young Scientists (DC2) (2,400,000 JPY/year)	<i>Apr. 2022 – Mar. 2024</i>
UTokyo-TOYOTA Study Abroad Scholarship in AI Field (220,000 JPY/month)	<i>Mar. 2022</i>
Tobitate Study Abroad Initiatives (160,000 JPY/month)	<i>July 2021</i>
TOYOTA/Dwango AI Scholarship (1,200,000 JPY/year)	<i>Apr. 2021 – Mar. 2022</i>

PUBLICATIONS (JOURNAL PAPER)

1. **Takaaki Saeki**, Yuki Saito, Shinnosuke Takamichi, and Hiroshi Saruwatari, “Real-time Full-Band Voice Conversion with Sub-Band Modeling and Data-Driven Phase Estimation of Spectral Differentials,” *IEICE Transactions on Information and Systems*, vol. E104-D, no.7, pp. 1002–1016, Jul. 2021.
2. **Takaaki Saeki**, Shinnosuke Takamichi, and Hiroshi Saruwatari, “Incremental Text-to-Speech Synthesis Using Pseudo Lookahead with Large Pretrained Language Model,” *IEEE Signal Processing Letters*, vol. 28, pp. 857–861, Apr. 2021.
(The 37th Telecom System Technology Award for Students from TAF)
(To be presented at *ICASSP2022*)

PUBLICATIONS (PEER-REVIEWED CONFERENCE PAPER)

1. **Takaaki Saeki**, Shinnosuke Takamichi, and Hiroshi Saruwatari, “Low-latency Incremental Text-to-Speech Synthesis with Distilled Context Prediction Network,” in *Proc. ASRU*, Dec. 2021.
2. **Takaaki Saeki**, Yuki Saito, Shinnosuke Takamichi, and Hiroshi Saruwatari, “Real-Time, Full-Band, Online DNN-Based Voice Conversion System Using a Single CPU,” in *Proc. INTERSPEECH*, pp. 1021–1022, Shanghai, China, Oct. 2020.
3. Naoki Kimura, Zixiong Su, and **Takaaki Saeki**, “End-to-End Deep Learning Speech Recognition Model for Silent Speech Challenge,” in *Proc. INTERSPEECH*, pp. 1025–1026, Shanghai, China, Oct. 2020..
4. **Takaaki Saeki**, Yuki Saito, Shinnosuke Takamichi, and Hiroshi Saruwatari, “Lifter Training and Sub-Band Modeling for Computationally Efficient and High-Quality Voice Conversion Using Spectral Differentials,” in *Proc. ICASSP*, pp. 7784–7788, Barcelona, Spain, May 2020.

PUBLICATIONS (PREPRINT)

1. **Takaaki Saeki**, Kentaro Tachibana, Ryuichi Yamamoto,
“DRSpeech: Degradation-Robust Text-to-Speech Synthesis with Frame-Level and Utterance-Level Acoustic Representation Learning,”
in *arXiv preprint arXiv:2203.15683*, Mar. 2022.
2. **Takaaki Saeki**, Shinnosuke Takamichi, Tomohiko Nakamura, Naoko Tanji, and Hiroshi Saruwatari,
“SelfRemaster: Self-Supervised Speech Restoration with Analysis-by-Synthesis Approach Using Channel Modeling,”
in *arXiv preprint arXiv:2203.12937*, Mar. 2022.
3. Yoshifumi Nakano, **Takaaki Saeki**, Shinnosuke Takamichi, Katsuhito Sudoh, and Hiroshi Saruwatari,
“vTTS: Visual-Text to Speech,”
in *arXiv preprint arXiv:2203.14725*, Mar. 2022.
4. Yuta Matsunaga, **Takaaki Saeki**, Shinnosuke Takamichi, and Hiroshi Saruwatari,
“Personalized Filled-Pause Generation with Group-Wise Prediction Models,”
in *arXiv preprint arXiv:2203.09961*, Mar. 2022.
5. Shinnosuke Takamichi, Ludwig Kürzinger, **Takaaki Saeki**, Sayaka Shiota, and Shinji Watanabe,
“JTubeSpeech: Corpus of Japanese Speech Collected from YouTube for Speech Recognition and Speaker Verification,”
in *arXiv preprint arXiv:2112.09323*, Dec. 2021.
6. Tomoki Hayashi, Ryuichi Yamamoto, Takenori Yoshimura, Peter Wu, Jiatong Shi, **Takaaki Saeki**, Yooncheol Ju, Yusuke Yasuda, Shinnosuke Takamichi, and Shinji Watanabe,
“ESPnet2-TTS: Extending the Edge of TTS Research,”
in *arXiv preprint arXiv:2110.07840*, Oct. 2021.

EXHIBITIONS

Sainokuni Buisiness Arena 2021	ONLINE
<i>“Research on Stress-Free, Real-Time, and Full-Band Voice Conversion Based on Perceptual Model”</i>	<i>Jan. 2021</i>
CEATEC 2020	ONLINE
<i>“Research on Stress-Free, Real-Time, and Full-Band Voice Conversion Based on Perceptual Model”</i>	<i>Oct. 2020</i>

PATENTS (SELECTED)

Shinnosuke Takamichi, Yuki Saito, **Takaaki Saeki**, and Hiroshi Saruwatari,
“Voice Conversion System, Method, and Program,”
International Patent Application No. PCT/JP2021/4367. (Submitted)

Shinnosuke Takamichi, Yuki Saito, **Takaaki Saeki**, and Hiroshi Saruwatari,
“Voice Conversion System, Method, and Program,”
International Patent Application No. PCT/JP2020/031122. (Submitted)

(Last update: 2022/04/04)