Takaaki Saeki

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

RESEARCH INTERESTS

Spoken language processing: Speech synthesis, speech recognition, speech representation learning
Natural language processing: Language modeling, multilinguality
Deep learning: Sequential modeling, multimodal learning, self-supervised representation 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*
- 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 New York, USA
Research Intern *May 2023 – Aug. 2023*
- Researching on multilingual speech representation learning.
- Advisor: Dr. Bhuvana Ramabhadran and Dr. Andrew Rosenberg

Carnegie Mellon University Pittsburgh, USA
Visiting Scholar *Oct. 2022 – Jan. 2023*
- Researched on speech synthesis.
- Advisor: Prof. Shinji Watanabe

Google Tokyo, Japan
Student Researcher (Intern) *Apr. 2022 – Sep. 2022*
- Researched on multilingual speech-text representation learning.
- Advisor: Dr. Heiga Zen

LINE Corporation Tokyo, Japan
Part-Time Researcher *Mar. 2021 – Mar. 2022*
- Worked on text-to-speech synthesis at Voice Team.
- Advisor: Kentaro Tahchibana and Ryuichi Yamamoto

Preferred Networks, Inc. Tokyo, Japan
Research Intern *Aug. 2021 – Sep. 2021*
- Researched on application of deep learning to music creation at Music Team.
- Advisor: Daiki Higurashi and Dr. Brian Vogel

NEC Data Science Research Laboratories Kanagawa, Japan
Research Intern *Aug. 2019 – Sep. 2019*
- Worked on speech enhancement.
- Advisor: Dr. Osamu 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.
- Advisor: Naoya Osugi

Delight Inc. Tokyo, Japan
Part-Time Engineeer *May 2017 – Apr. 2019*
- Developed simulation software of 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

Yamashita SIG Research Award | *Information Processing Society of Japan (IPSJ)*	*July 2022*
Best Paper Award | *The Institute of Electronics, Information and Communication Engineers (IEICE)*	*Apr. 2022*
Ranked 1st Place in 10/16 Metrics | *VoiceMOS Challenge 2022*	*Mar. 2022*
Telecom System Technology Award for Students | *The Telecommunication Advancement Foundation*	*Mar. 2022*
Best Student Presentation Award | *Acoustical Society of Japan*	*Oct. 2021*
IEICE ISS Best Student Poster Award | *IEICE Speech Committee 2021*	*Mar. 2021*
Award for Excellence (2nd place) | *Natural Language Processing Hackathon at Recruit Co., Ltd.*	*Feb. 2019*

Grants and Scholarships

Google East Asia Student Travel Grants (100,000 JPY)	*Sep. 2022*
Research Fellowship for Young Scientists (DC2) (2,400,000 JPY/year)	*Apr. 2022 – Mar. 2024*
UTokyo-TOYOTA Study Abroad Scholarship in AI Field (220,000 JPY/month)	*Mar. 2022*
Tobitate Study Abroad Initiatives (160,000 JPY/month)	*July 2021*
TOYOTA/Dwango AI Scholarship (1,200,000 JPY/year)	*Apr. 2021 – Mar. 2022*

Publications (Journal Paper)

Takaaki Saeki, Shinnosuke Takamichi, Tomohiko Nakamura, Naoko Tanji, and Hiroshi Saruwatari,
"SelfRemaster: Self-Supervised Speech Restoration with Analysis-by-Synthesis Approach Using Channel Modeling,"
IEEE Access, Dec. 2023. (Accepted)

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.
(**Best Paper Award**)

Takaaki Saeki, Shinnosuke Takamichi, and Hiroshi Saruwatari,
"Incremental Text-to-Speech Synthesis Using Pseudo Lookahead with Large Pretrained Language Model,"
*IEEE Signal Processing Letters (**SPL**)*, vol. 28, pp. 857–861, Apr. 2021.
(**The 37th Telecom System Technology Award for Students from TAF**)

Publications (Peer-Reviewed Conference Paper)

Takaaki Saeki, Gary Wang, Nobuyuki Morioka, Isaac Elias, Kyle Kastner, Andrew Rosenberg, Bhuvana Ramabhadran, Heiga Zen, Françoise Beaufays, Hadar Shemtov,
"Extending Multilingual Speech Synthesis to 100+ Languages without Transcribed Data,"
*IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**)*, Apr. 2024. (Accepted)

Kentaro Seki, Shinnosuke Takamichi, **Takaaki Saeki**, and Hiroshi Saruwatari,
"Diversity-based Core-Set Selection for Text-to-Speech with Linguistic and Acoustic Features,"
*IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**)*, Apr. 2024. (Accepted)

Xinjian Li, Shinnosuke Takamichi, **Takaaki Saeki**, William Chen, Sayaka Shiota, Shinji Watanabe,
"YODAS: Youtube-Oriented Dataset for Audio and Speech,"
*IEEE Automatic Speech Recognition and Understanding Workshop (**ASRU**)*, Dec. 2023.

Yuta Matsunaga, **Takaaki Saeki**, Shinnosuke Takamichi, and Hiroshi Saruwatari,
"Improving Robustness of Spontaneous Speech Synthesis with Linguistic Speech Regularization and
Pseudo-Filled-Pause Insertion,"
*Speech Synthesis Workshop (**SSW**)*, Aug. 2023.

Takaaki Saeki, Soumi Maiti, Xinjian Li, Shinji Watanabe, Shinnosuke Takamichi, Hiroshi Saruwatari,
"Learning to Speak from Text: Zero-Shot Multilingual Text-to-Speech with Unsupervised Text Pretraining,"
*International Joint Conference on Artificial Intelligence (**IJCAI**)*, Aug. 2023.

Soumi Maiti, Yifan Peng, **Takaaki Saeki**, Shinji Watanabe,
"SpeechLMScore: Evaluating Speech Generation Using Speech Language Model,"
*IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**)*, June 2023.

Dong Yang, Tomoki Koriyama, Yuki Saito, **Takaaki Saeki**, Detai Xin, and Hiroshi Saruwatari,
"Duration-Aware Pause Insertion Using Pre-trained Language Model for Multi-speaker Text-to-Speech,"
*IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**)*, June 2023.

Kentaro Seki, Shinnosuke Takamichi, **Takaaki Saeki**, and Hiroshi Saruwatari,
"Text-to-Speech Synthesis from Dark Data with Evaluation-in-the-Loop Data Selection,"
*IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**)*, June 2023.
(**IEEE SPS Travel Grant for IEEE ICASSP 2023**)

Takaaki Saeki, Heiga Zen, Zhehuai Chen, Nobuyuki Morioka, Gary Wang, Yu Zhang, Ankur Bapna, Andrew
Rosenberg, and Bhuvana Ramabhadran,
"Virtuoso: Massive Multilingual Speech-Text Joint Semi-Supervised Learning for Text-to-Speech,"
*IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**)*, June 2023. (**Oral**)

Yoshifumi Nakano, **Takaaki Saeki**, Shinnosuke Takamichi, Katsuhito Sudoh, and Hiroshi Saruwatari,
"vTTS: Visual-Text to Speech,"
*IEEE Spoken Language Technology Workshop (**SLT**)*, pp. 936–942, Jan. 2023.

Yuta Matsunaga, **Takaaki Saeki**, Shinnosuke Takamichi, and Hiroshi Saruwatari,
"Empirical Study Incorporating Linguistic Knowledge on Filled Pauses for Personalized Spontaneous Speech
Synthesis,"
*Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (**APSIPA
ASC**)*, pp. 1898–1903, June 2022.

Takaaki Saeki, Shinnosuke Takamichi, Tomohiko Nakamura, Naoko Tanji, and Hiroshi Saruwatari,
"SelfRemaster: Self-Supervised Speech Restoration with Analysis-by-Synthesis Approach Using Channel
Modeling,"
*Annual Conference of the International Speech Communication Association (**Interspeech**)*, pp.4406–4410,
Sep. 2022. (**Oral**)
(**Google East Asia Student Travel Grants**)

Takaaki Saeki, Kentaro Tachibana, Ryuichi Yamamoto,
"DRSpeech: Degradation-Robust Text-to-Speech Synthesis with Frame-Level and Utterance-Level Acoustic
Representation Learning,"
*Annual Conference of the International Speech Communication Association (**Interspeech**)*, pp. 793–797. Sep.
2022.

Takaaki Saeki*, Detai Xin*, Wataru Nakata*, Tomoki Koriyama, Shinnosuke Takamichi, and Hiroshi
Saruwatari (* Equal Contribution),
"UTMOS: UTokyo-SaruLab System for VoiceMOS Challenge 2022,"

*Annual Conference of the International Speech Communication Association (**Interspeech**)*, pp. 4521–4525,
Sep. 2022. (**Oral**)
(**Ranked 1st Place in 10/16 Metrics**)

Yuta Matsunaga, **Takaaki Saeki**, Shinnosuke Takamichi, and Hiroshi Saruwatari,
"Personalized Filled-Pause Generation with Group-Wise Prediction Models,"
*Language Resources and Evaluation Conference (**LREC**)*, June 2022.

Naoki Kimura, Zixiong Su, **Takaaki Saeki**, and Jun Rekimoto,
"SSR7000: A Synchronized Corpus of Ultrasound Tongue Imaging for End-to-End Silent Speech Recognition,"
in *Language Resources and Evaluation Conference (**LREC**)*, June 2022.

Takaaki Saeki, Shinnosuke Takamichi, and Hiroshi Saruwatari,
"Low-latency Incremental Text-to-Speech Synthesis with Distilled Context Prediction Network,"
*IEEE Automatic Speech Recognition and Understanding Workshop (**ASRU**)*, pp. 749-756, Dec. 2021.

Takaaki Saeki, Yuki Saito, Shinnosuke Takamichi, and Hiroshi Saruwatari,
"Real-Time, Full-Band, Online DNN-Based Voice Conversion System Using a Single CPU,"
*Annual Conference of the International Speech Communication Association (**Interspeech**)*, pp. 1021–1022,
Oct. 2020. (Show & Tell)

Naoki Kimura, Zixiong Su, and **Takaaki Saeki**,
"End-to-End Deep Learning Speech Recognition Model for Silent Speech Challenge,"
*Annual Conference of the International Speech Communication Association (**Interspeech**)*, pp. 1025–1026,
Oct. 2020. (Show & Tell)

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,"
*IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**)*, pp. 7784–7788, May
2020.

PUBLICATIONS (PREPRINT)

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,"
arXiv preprint arXiv:2112.09323, Dec. 2021.

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,"
arXiv preprint arXiv:2110.07840, Oct. 2021.

ACADEMIC SERVICES

Reviewing

ICASSP: 2023, 2024

Interspeech: 2023

IEEE Signal Processing Letters: 2023

IEEE/ACM Transactions on Audio Speech and Language Processing: 2023

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)

Miscellaneous Work

Research Talk at Sphinx Lunch in Carnegie Mellon University   ONLINE
"Self-Supervised Speech Restoration for Historical Audio"   *Jan. 2023*

Research Talk at Google Tokyo   ONLINE
"Self-Supervised Speech Restoration for Historical Audio"   *June 2022*

Exhibition at Sainokuni Buisiness Arena 2021   ONLINE
"Research on Stress-Free, Real-Time, and Full-Band Voice Conversion Based on Perceptual Model"   *Jan. 2021*

Exhibition at CEATEC 2020   ONLINE
"Research on Stress-Free, Real-Time, and Full-Band Voice Conversion Based on Perceptual Model"   *Oct. 2020*

(Last update: 12/18/2023)