# Yuki Saito

Room #140, Eng. bldg. #6, 7-3-1 Hongo, Bunkyo, Tokyo 113–8656, Japan
☎ +81 3 5841 6904  •  ✉ yuuki_saito@ipc.i.u-tokyo.ac.jp
🖥 http://sython.org/ (updated: Dec. 2024)

## Research interests

Statistical machine learning, speech synthesis, voice conversion, and human computation.

## Education and research experience

### Education

**Ph.D. degree in Information Science and Technology**
*Graduate School of Information Science and Technology, The University of Tokyo, Japan.*  *2021*

**M.S. degree in Information Science and Technology**
*Graduate School of Information Science and Technology, The University of Tokyo, Japan.*  *2018*

**B.S. degree in Engineering**
*National Institution for Academic Degrees and Quality Enhancement of Higher Education, Japan*  *2016*

### Research experience

**Lecturer**
*Graduate School of Information Science and Technology, The University of Tokyo, Japan.*  *2024–current*
I work on research and development on universal spoken language processing technologies.

**Assistant Professor**
*Graduate School of Information Science and Technology, The University of Tokyo, Japan.*  *2023–2024*
I work on research and development on human-in-the-loop speech synthesis technologies.

**Project Research Associate**
*Graduate School of Information Science and Technology, The University of Tokyo, Japan.*  *2021–2023*
I work on research and development on acoustic information processing and voice conversion.

**Research Assistant**
*The University of Tokyo, Japan.*  *2019–2021*
I worked on research and development on stress-free, real-time, and full-band voice conversion based on perceptual models.

**Part-time Researcher**
*DeNA Co., Ltd., Japan.*  *2019–2021*
I worked on research and development on many-to-one voice conversion technology.

**Research Fellow (DC1)**
*Japan Society of the Promotion of Science (JSPS), Japan.*  *2018–2021*
I worked on research and development on speech synthesis based on statistical modeling of human perception.

## Languages

Japanese (native) and English (conversant)

## Competitive funds

**Human-in-The-Loop Adaptive Text-To-Speech based on Interactive Reinforcement Learning**
*ACT-X, JST (research representative), 4,500,000 JPY.*  *2023–2026*

**Research and Development on Automatic Audio Commentary Generation on Gameplay Videos**
*Grant-in-Aid for Young Scientists, JSPS (research representative), 3,600,000 JPY.*  *2022–2025*

**Research and Development on Sustainable Speech Synthesis Technology based on Continual Learning**
*Grant-in-Aid for Research Activity Start-up, JSPS (research representative), 2,400,000 JPY.*  *2021–2023*

+ 7 fund since 2018

## Teaching experience

**Applied Acoustics (Instructor)**
*Department of Mathematical engineering and information physics, The University of Tokyo, Japan.*  *2024–current*
I lectured fundamentals of spoken language processing techniques. (in Japanese)

**Digital Signal Processing and Acoustic Systems (Instructor)**
*Department of Mathematical engineering and information physics, The University of Tokyo, Japan.*  *2023*
I lectured basic signal processing techniques for audio and speech signals (in Japanese)

- **Advanced Signal Processing (Invited Guest Presenter)**
  *Graduate School of Information Science and Technology, The University of Tokyo, Japan.* *2022*
  I lectured recent development of deep-learning-based speech synthesis technologies. (in Japanese)
- + 4 lecture since 2016

# Publications

### Journal papers

1. <u>Yuki Saito</u>, Shinnosuke Takamichi, and Hiroshi Saruwatari, "Perceptual-similarity-aware deep speaker representation learning for multi-speaker generative modeling," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 29, pp. 1033–1048, Feb. 2021.
2. <u>Yuki Saito</u>, Shinnosuke Takamichi, and Hiroshi Saruwatari, "Vocoder-free text-to-speech synthesis incorporating generative adversarial networks using low-/multi-frequency STFT amplitude spectra," Computer Speech and Language, Vol. 58, pp. 347–363, Nov. 2019. **[Selected as "Most popular articles" of the Journal]**
3. <u>Yuki Saito</u>, Shinnosuke Takamichi, and Hiroshi Saruwatari, "Statistical parametric speech synthesis incorporating generative adversarial Networks," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 26, No. 1, pp. 84–96, Jan. 2018. **[2020 IEEE SPS Young Author Best Paper Award, IEEE SPS Japan Student Journal Paper Award, 34th TELECOM System Technology Award for Students from TAF]**
4. + 11 papers since 2017

### International conferences

1. Kaito Baba, Wataru Nakata, <u>Yuki Saito</u>, and Hiroshi Saruwatari, "The T05 system for The VoiceMOS Challenge 2024: Transfer learning from deep image classifier to naturalness MOS prediction of high-quality synthetic speech," Proc. SLT, pp. xxxx–xxxx, Macau, China, Dec. 2024. (ACCEPTED)
2. <u>Yuki Saito</u>, Takuto Igarashi, Kentaro Seki, Shinnosuke Takamichi, Ryuichi Yamamoto, Kentaro Tachibana, and Hiroshi Saruwatari, "SRC4VC: Smartphone-recorded corpus for voice conversion benchmark," Proc. INTERSPEECH, pp. 1825–1829, Kos, Greece, Sep. 2024.
3. <u>Yuki Saito</u>, Shinnosuke Takamichi, Eiji Iimori, Kentaro Tachibana, and Hiroshi Saruwatari, "ChatGPT-EDSS: empathetic dialogue speech synthesis trained from ChatGPT-derived context word embeddings," Proc. INTERSPEECH, pp. 3048–3052, Dublin, Ireland, Aug. 2023.
4. + 42 papers since 2014

### Corpora

1. <u>Yuki Saito</u>, Shinnosuke Takamichi, and Hiroshi Saruwatari, "SMASH corpus: a spontaneous speech corpus recording third-person audio commentaries on gameplay," available in, `https://ss-takashi.sakura.ne.jp/corpus/smash`, May 2020.
2. Shinnosuke Takamichi, Kentaro Mitsui, <u>Yuki Saito</u>, Tomoki Koriyama, Naoko Tanji, and Hiroshi Saruwatari, "JVS: Japanese multi-speaker and multi-style voice corpus," available in `https://sites.google.com/site/shinnosuketakamichi/research-topics/jvs_corpus`, Aug. 2019.
3. + 6 corpora since 2017

# Awards

1. IEEE Signal Processing Society Young Author Best Paper Award, Jun. 2021.
2. Travel Grant Award for INTERSPEECH2023, Aug. 2023.
3. The 12th IEEE Signal Processing Society Japan Student Journal Paper Award, Nov. 2018.
4. Spoken Language Processing Student Grant of ICASSP, Mar. 2017.
5. + 18 awards since 2014 (and 15 awards received by my collaborators)

# Patents

1. Kentaro Tachibana, <u>Yuki Saito</u>, Kei Akuzawa, "SPEECH PROCESSING APPARATUS AND SPEECH PROCESSING PROGRAM," JP2020190605, Filled in May 21.
2. Shinnosuke Takamichi, <u>Yuki Saito</u>, Takaaki Saeki, and Hiroshi Saruwatari, "VOICE CONVERSION DEVICE, VOICE CONVERSION METHOD, AND VOICE CONVERSION PROGRAM," JP2021032940, Filled in Aug. 19, 2019.
3. + 2 patents

# Misc.

1. I gave a research talk about our recent work of human-in-the-loop speech synthesis technologies, hosted by Prof. Haizhou Li at NUS, in Aug. 2022.
2. I was invited to Google Speech Technology Summit 2018 at Google London, UK, in May 2018.
3. A figure taken from our paper was used on the cover of IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 26, No. 1, Jan. 2018.