

November 21st, 2023, Hitotsubashi Hall 101-103

National Center of Sciences building, Hitotsubashi, Chiyoda-ku, Tokyo, Japan

The VoicePersonae Workshop

– VoicePersonae overview and highlights –

10:00-11:00 AM: Opening - Overview of the VoicePersonae Project

Speaker: Prof. Junichi Yamagishi (National Institute of Informatics, Japan)

11:00-11:15 AM: Coffee break (15 min) @ Room 101-103

11:15-11:50 AM: Voice privacy Processing

Speaker: Prof. Jean-Francois Bonastre (Inria & Avignon University, France)

12:00-12:30 PM: Voice Biometrics & DeepFake Detection

Speaker: Prof. Nicholas Evans (EURECOM, France)

12:30-2:00 PM: Lunch Break

[Lunch map](#)

– Theme 1: Speech generative models and related topics –

2:00-2:30 PM: From DSP and DNN to DNN/DSP: neural speech waveform models and its applications in speech and music audio waveform modelling

Speaker: Dr. Xin Wang (NII, Japan)

Abstract: The waveform generation model is indispensable to speech and music audio generation. The VoicePersonae project started when the deep neural network (DNN) was introduced to produce the waveform, an approach deviating from traditional digital signal processing (DSP) approaches. Combining DNN and DSP, the VoicePersonae project produced pioneering DSP plus DNN models, such as the neural source-filter waveform model and the GAN-excited linear prediction waveform model. This presentation will explain those outcomes, highlight the technical novelties, and summarise models from the research community that follow the DSP plus DNN (a.k.a differential DSP) approach.

References

- [1] Wang, Xin, Shinji Takaki, and Junichi Yamagishi. 2019. "Neural Source-Filter-Based Waveform Model for Statistical Parametric Speech Synthesis." In Proc. ICASSP, 5916-5920.
- [2] Wang, Xin, and Junichi Yamagishi. 2019. "Neural Harmonic-plus-Noise Waveform Model with Trainable Maximum Voice Frequency for Text-to-Speech Synthesis." In Proc. SSW, 1-6.
- [3] Wang, Xin, Shinji Takaki, and Junichi Yamagishi. 2020. "Neural Source-Filter Waveform Models for Statistical Parametric Speech Synthesis." IEEE/ACM Transactions on Audio, Speech, and Language Processing 28: 402-415.
- [4] Juvela, Lauri, Bajibabu Bollepalli, Junichi Yamagishi, and Paavo Alku. 2019. "GELP: GAN-Excited Linear Prediction for Speech Synthesis from Mel-Spectrogram." In Proc. Interspeech, 694-699.
- [5] Wang, Xin, and Junichi Yamagishi. 2020. "Using Cyclic Noise as the Source Signal for Neural Source-Filter-Based Speech Waveform Model." In Proc. Interspeech, 1992-1996.
- [6] Zhao, Yi, Xin Wang, Lauri Juvela, and Junichi Yamagishi. 2020. "Transferring Neural Speech Waveform Synthesizers to Musical Instrument Sounds Generation." In Proc. ICASSP, 6269-6273.

2:30-3:00 PM: End-to-End Speech Synthesis and Its Entertainment Applications: Rakugo Modelling & Musical Instrument Sound Modelling

Speaker: Dr. Shuhei Kato (Revcomm, Japan) & Dr. Yusuke Yasuda (Nagoya University, Japan)

Abstract: Text-to-speech synthesis is a technique for converting text into speech. Recent developments in deep learning have led to the emergence of end-to-end speech synthesis, a technique that converts text directly into speech. This method enables the generation of synthesised speech that is indistinguishable from

human speech. This method is also highly versatile and can handle not only text-to-speech conversion, but also score-to-music conversion within the same framework. However, from the perspective of entertaining people, it is extremely difficult to develop synthesis technology that extends to professional human performers and musicians. In this project, we have improved end-to-end speech synthesis to realise Rakugo speech synthesis and guitar music synthesis. In this presentation, we will present the results of our research to realise synthesis technology to entertain people.

3:00-3:30 PM: From Human Ears to Deep Neural Networks: Automatic Evaluation of Synthetic Speech and Audio Data

Speaker: Dr. Erica Cooper (National Institute of Informatics, Japan) & Mr. Wen-Ching Huang (Nagoya University, Japan)

Abstract: Evaluation of experimental speech synthesis systems is largely conducted by human listening tests, such as the widely-used mean opinion score (MOS) test. These tests can be time-consuming and costly, precluding rapid experimental iteration; however, existing objective metrics fall short of correlating well with human judgments or generalizing to unseen conditions. We will present the efforts of the VoicePersonae project to address this problem, including the collection of large-scale public datasets and the development of strong baseline MOS predictors by applying modern techniques such as self-supervised learning based models for speech. We will also present the experiences and lessons learned from two years of running the VoiceMOS Challenge, an international shared task for MOS prediction.

References

- [1] Lo, Chen-Chou, Szu-Wei Fu, Wen-Chin Huang, Xin Wang, Junichi Yamagishi, Yu Tsao, and Hsin-Min Wang. 2019. "MOSNet: Deep learning based objective assessment for voice conversion." In Proc. Interspeech, 1541-1545.
- [2] Cooper, Erica, and Junichi Yamagishi. 2021. "How do Voices from Past Speech Synthesis Challenges Compare Today?" In Proc. 11th ISCA Speech Synthesis Workshop, 183-188.
- [3] Cooper, Erica, Wen-Chin Huang, Tomoki Toda, and Junichi Yamagishi. 2022. "Generalization Ability of MOS Prediction Networks." In Proc. ICASSP, 8442-8446.
- [4] Huang, Wen-Chin, Erica Cooper, Junichi Yamagishi, and Tomoki Toda. 2022. "LDNet: Unified Listener Dependent Modeling in MOS Prediction for Synthetic Speech." In Proc. ICASSP, 896-900.
- [5] Huang, Wen-Chin, Erica Cooper, Yu Tsao, Hsin-Min Wang, Tomoki Toda, and Junichi Yamagishi. 2022. "The VoiceMOS Challenge 2022." In Proc. Interspeech, 4536-4540.
- [6] Cooper, Erica, and Junichi Yamagishi. 2023. "Investigating Range-Equalizing Bias in Mean Opinion Score Ratings of Synthesized Speech." In Proc. Interspeech, 1104-1108.

[7] Cooper, Erica, Wen-Chin Huang, Yu Tsao, Hsin-Min Wang, Tomoki Toda, and Junichi Yamagishi. 2023. "The VoiceMOS Challenge 2023: Zero-shot Subjective Speech Quality Prediction for Multiple Domains." Workshop on Automatic Speech Recognition and Understanding (to appear).

[8] Yadav, Hemant, Erica Cooper, Junichi Yamagishi, Sunayana Sitaram, and Rajiv Ratn Shah. 2023. "Partial Rank Similarity Minimization Method for Quality MOS Prediction of Unseen Speech Synthesis Systems in Zero-Shot and Semi-supervised Setting." Workshop on Automatic Speech Recognition and Understanding (to appear).

3:30-3:45 PM: Coffee break (15 min) @ room 101-103

– Theme 2: Speech security and deepfake detection–

3:45-4:15 PM: From Artefacts to Insights: A Topical Analysis of Voice Biometric Security

Speaker: Prof. Massimiliano Todisco (EURECOM, France) & Dr. Hemlata Tak

Abstract: In recent years, voice biometric security has gained significant attention for its potential to offer secure and convenient authentication. However, the emergence of generative AI or deepfake technology has raised concerns as it enables the creation of highly realistic synthetic media, including speech recordings. Speech deepfakes pose a serious threat to voice biometric security systems, allowing unauthorized access and impersonation. This presentation explores the interaction between voice biometrics and security, especially in the face of the challenges posed by deepfakes. To tackle these challenges, we present advanced deepfake detection techniques that leverage graph neural networks (GNNs) and self-supervised learning based front-ends. GNNs have demonstrated promising results in diverse domains such as natural language processing and computer vision. By exploiting the inherent graph structure of voice data, GNNs effectively capture intricate relationships between acoustic features and identify distinctive cues indicative of spoofing attacks. This talk sheds light on the potential of self-supervised learning and GNN-based solutions to enhance the robustness and reliability of voice biometric systems, ensuring the integrity and authenticity of voice-based authentication processes.

In the last part of the talk, we will discuss *Malafide*, a universal adversarial attack against voice biometrics. Through convolutional noise and specialized filters, *Malafide* is able to compromise spoofing countermeasures, posing significant threats even in black-box settings.

References

- [1] Evans, N. et al. Spoofing and countermeasures for automatic speaker verification. in *Proc. Interspeech* 925-929, 2013.
- [2] Sahidullah, M. et al. Introduction to voice presentation attack detection and recent advances. in *Handbook of Biometric Anti-Spoofing* 321-361, 2019.
- [3] Todisco, M. et al. "ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection." *Interspeech*, 2019.
- [4] Todisco, M. et al. "Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification." *Computer Speech Lang.* 45, 516-535, 2017.
- [5] Kamble, M. R. et al. Advances in anti-spoofing: from the perspective of ASVspoof challenges. *APSIPA Trans. Signal Inf. Process.* 9, e2, 2020.
- [6] Yamagishi, J. et al., "ASVspoof 2021: accelerating progress in spoofed and deepfake speech detection", *Proc. 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge*, 2021.
- [7] Wu, Z. et al. "A comprehensive survey on graph neural networks." *IEEE transactions on neural networks and learning systems*, 32.1, 2020.
- [8] Velickovic, P. et al., "Graph Attention Networks", in *Proc. International Conference on Learning Representations (ICLR)*, 2018.
- [9] Mohamed, A. et al., "Self-Supervised Speech Representation Learning: A Review", *IEEE Journal of Selected Topics in Signal Processing* 16 (6): 1179-1210. 2022.
- [10] Jung, J. et al., "AASIST: Audio anti-spoofing using integrated spectro-temporal graph attention networks", in *Proc. ICASSP*, 2022.
- [11] Wang, X. et al., "Investigating self-supervised front ends for speech spoofing countermeasures.", in *Proc. Speaker Odyssey*, 2022.
- [12] Tak, H. et al., "Automatic speaker verification spoofing and deepfake detection using wav2vec 2.0 and data augmentation", in *Proc. Speaker Odyssey*, 2022.
- [13] Liu, X. et al., "ASVspoof 2021: Towards Spoofed and Deepfake Speech Detection in the Wild", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31, 2023.
- [14] Panariello, M. et al. "Malafide: a novel adversarial convolutive noise attack against deepfake and spoofing detection systems." in *Proc. Interspeech*, 2023.

4:15-4:45 PM: Partial Spoof Detection: Countermeasures for the Detection of Short Fake Speech Segments Embedded in an Utterance

Speaker: Ms. Lin Zhang (National Institute of Informatics, Japan)

Abstract: Automatic speaker verification is susceptible to various manipulations and spoofing, such as text-to-speech synthesis, voice conversion, replay, tampering, adversarial attacks, and so on. We consider a new spoofing scenario called *Partial Spoof* (PS) in which synthesized or transformed {speech} segments are embedded into a bona fide utterance. While existing countermeasures (CMs) can detect fully spoofed utterances, there is a need for their adaptation or extension to the PS scenario.

We propose various improvements to construct a significantly more accurate CM that can detect and locate short-generated spoofed {speech} segments at finer temporal resolutions. First, we introduce newly developed self-supervised pre-trained models as enhanced feature extractors. Second, we extend our PartialSpoof database by adding segment labels for various temporal resolutions. Since the short spoofed {speech} segments to be embedded by attackers are of variable length, six different temporal resolutions are considered, ranging from as short as 20 ms to as large as 640 ms. Third, we propose a new CM that enables the simultaneous use of the segment-level labels at different temporal resolutions as well as utterance-level labels to execute utterance- and segment-level detection at the same time. We also show that the proposed CM is capable of detecting spoofing at the utterance level with low error rates in the PS scenario as well as in a related logical access (LA) scenario. The equal error rates of utterance-level detection on the PartialSpoof database and ASVspoof 2019 LA database were 0.77% and 0.90%, respectively.

4:45-5:15 PM: From Deepfake Detection and Segmentation to Restoration: Protection of Facial Biometrics from Manipulative and Generative AI

Speaker: Dr. Huy H.Nguyen (National Institute of Informatics, Japan) & Dr. C.C. Chang (NII, Japan)

Abstract: With the rapid advancement of technology, image manipulation has evolved to a sophisticated level, enabling effortless alteration of a subject's identity, facial features, movements, and even the creation of entirely fabricated images or videos, popularly known as deepfakes. The widespread use of social media has facilitated the generation and dissemination of vast amounts of data, encompassing personal information, news articles, images, and videos. Unfortunately, this environment has also made it easier for malicious actors to create and widely distribute deepfakes to large audiences. In response to this emerging threat, we have pioneered comprehensive countermeasures against deepfakes, focusing on detection, segmentation, and restoration. Our detection methods assess the probability of a face being a deepfake, while segmentation provides explainable information on manipulated regions. Conversely, our innovative approach to restoration, termed "cyber vaccination," involves adding imperceptible but resilient latent noises to original images before publication. These noises are crucial in restoring original images from manipulated ones. Through extensive experimentation, we have demonstrated the effectiveness of our proposed methods in deepfake detection, segmentation, and restoration.

References

[1] Afchar, Darius, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. "MesoNet: a Compact Facial Video Forgery Detection Network." In *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1-7. IEEE, 2018.

- [2] Nguyen, Huy H., T. Ngoc-Dung Tieu, Hoang-Quoc Nguyen-Son, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. "Modular Convolutional Neural Network for Discriminating between Computer-Generated Images and Photographic Images." In *Proceedings of the 13th international conference on availability, reliability and security*, pp. 1-10. 2018.
- [3] Nguyen, Huy H., Junichi Yamagishi, and Isao Echizen. "Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos." In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2307-2311. IEEE, 2019.
- [4] Nguyen, Huy H., Junichi Yamagishi, and Isao Echizen. "Capsule-Forensics Networks for Deepfake Detection." In *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, pp. 275-301. Cham: Springer International Publishing, 2022.
- [5] Tolosana, Ruben, Christian Rathgeb, Ruben Vera-Rodriguez, Christoph Busch, Luisa Verdoliva, Siwei Lyu, Huy H. Nguyen et al. "Future Trends in Digital Face Manipulation and Detection." In *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, pp. 463-482. Cham: Springer International Publishing, 2022.
- [6] Le, Trung-Nghia, Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. "Robust Deepfake on Unrestricted Media: Generation and Detection." In *Frontiers in Fake Media Generation and Detection*, pp. 81-107. Singapore: Springer Nature Singapore, 2022.
- [7] Sun, YuYang, ZhiYong Zhang, Isao Echizen, Huy H. Nguyen, ChangZhen Qiu, and Lu Sun. "Face Forgery Detection Based on Facial Region Displacement Trajectory Series." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 633-642. 2023.
- [8] Nguyen, Huy H., Fuming Fang, Junichi Yamagishi, and Isao Echizen. "Multi-Task Learning for Detecting and Segmenting Manipulated Facial Images and Videos." In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1-8. IEEE, 2019.
- [9] Le, Trung-Nghia, Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. "Openforensics: Large-Scale Challenging Dataset for Multi-Face Forgery Detection and Segmentation in-the-Wild." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10117-10127. 2021.
- [10] Chang, Ching-Chun, Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. "Cyber Vaccine for Deepfake Immunity." *IEEE Access* (2023).
- [11] Huang, Rong, Fuming Fang, Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. "Security of Facial Forensics Models Against Adversarial Attacks." In *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 2236-2240. IEEE, 2020.

5:15-5:30 PM: Coffee break (15 min) @ room 101-103

– Theme 3: Speech privacy and speaker anonymization –

5:30-6:00 PM: Preserving Speaker's Privacy and Utility: The VoicePrivacy Progress and VoxCeleb2 Database anonymization

Speaker: Dr. Natalia Tomashenko (Avignon University, France) & Dr. Xiaoxiao Miao (Singapore Institute of Technology, Singapore)

Abstract: Speech contains a lot of personal, private information (e.g., age, gender, personality traits, health and emotional state, socio-economic status, geographical background, etc.) which can be associated to the speaker's identity using automatic speaker recognition or metadata. Thus, collection, processing, and storage of speech data poses serious privacy risks. Formed in 2020, the VoicePrivacy initiative is spearheading the effort to develop privacy preservation solutions for speech technology.

In the first part, we provide an overview of privacy preservation solutions for speech data, with a focus on voice anonymization. We present the VoicePrivacy challenge design including the voice anonymization task and evaluation metrics, and discuss findings of the first two challenge editions.




In the second part, we address the legal and ethical concerns that led to the withdrawal of the VoxCeleb2 ASV dataset by creating a privacy-friendly synthetic VoxCeleb2 dataset. Specifically, we employ the state-of-the-art speaker anonymization techniques to anonymize authentic VoxCeleb2 dataset and evaluate the quality of the generated speech in terms of privacy, utility, and fairness. We also discuss the challenges of using synthetic data for the downstream task of speaker verification.

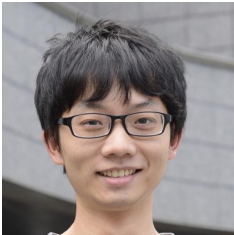




References

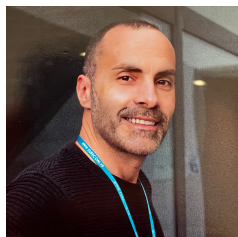
- [1] The VoicePrivacy 2020 Challenge: Results and findings. Natalia Tomashenko, Xin Wang, Emmanuel Vincent, Jose Patino, Brij Mohan Lal Srivastava, Paul-Gauthier Noé, Andreas Nautsch, Nicholas Evans, Junichi Yamagishi, Benjamin O'Brien, Anaïs Chanclu, Jean-François Bonastre, Massimiliano Todisco, Mohamed Maouche. Computer Speech and Language 2022
- [2] The VoicePrivacy 2022 Challenge evaluation plan. Natalia Tomashenko, Xin Wang, Xiaoxiao Miao, Hubert Nourtel, Pierre Champion, Massimiliano Todisco, Emmanuel Vincent, Nicholas Evans, Junichi Yamagishi, Jean-François Bonastre
- [3] Speaker anonymisation using the McAdams coefficient. Jose Patino, Natalia Tomashenko, Massimiliano Todisco, Andreas Nautsch, Nicholas Evans. Interspeech 2021
- [4] Privacy and utility of x-vector based speaker anonymization. Brij M. L. Srivastava, Mohamed Maouche, Md Sahidullah, Emmanuel Vincent, Aurélien Bellet, Marc Tommasi, Natalia Tomashenko, Xin Wang, Junichi Yamagishi. IEEE/ACM Transactions on Audio, Speech, and Language Processing
- [5] Towards a unified assessment framework of speech pseudonymisation. Paul-Gauthier Noé, Andreas Nautsch, Nicholas Evans, Jose Patino, Jean-François Bonastre, Natalia Tomashenko, Driss Matrouf. Computer Speech & Language, 2022.
- [6] Speaker anonymization using orthogonal Householder neural network. Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi, Natalia Tomashenko. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023.
- [7] SynVox2: Towards a privacy-friendly VoxCeleb2 dataset. Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi, Nicholas Evans, Massimiliano Todisco, Jean-François Bonastre, Mickael Rouvier, 2023.

6:00-6:10 PM: Closing

Speaker Bibliographies

Speaker portrait	Speaker info
	<p>Junichi Yamagishi (Senior member, IEEE) received a Ph.D. degree from the Tokyo Institute of Technology (Tokyo Tech), Tokyo, Japan, in 2006. From 2007-2013, he was a research fellow in the Centre for Speech Technology Research (CSTR) at the University of Edinburgh, UK. He was appointed Associate Professor at the National Institute of Informatics, Japan, in 2013. He is currently a Professor at NII, Japan. His research topics include speech processing, machine learning, signal processing, biometrics, digital media cloning, and media forensics. He served previously as co-organizer for the bi-annual ASVspoof Challenge and the bi-annual Voice Conversion Challenge. He also served as a member of the IEEE Speech and Language Technical Committee (2013-2019), an Associate Editor of the IEEE/ACM Transactions on Audio Speech and Language Processing (2014-2017), a chairperson of ISCA SynSIG (2017-2021), and a Senior Area Editor of the IEEE/ACM TASLP (2019-2023). He is currently a PI of JST-CREST and ANR supported VoicePersonae project and a board member of IEEE Signal Processing Society Education Board.</p>
	<p>Nicholas Evans (Member, IEEE) is a Professor at EURECOM, France, where he heads research in Audio Security and Privacy. He received his Ph.D. from the University of Wales Swansea, UK in 2004. He is a co-founder of the community led, ASVspoof, SASV, and VoicePrivacy challenge series. He participated in the EU FP7 Tabula Rasa and EU H2020 OCTAVE projects, both involving antispoofing. Today, his team is leading the EU H2020 TReSPAsS-ETN project, a training initiative in security and privacy for multiple biometric characteristics. He co-edited the second and third editions of the Handbook of Biometric Anti-Spoofing, is currently co-editing the Handbook of Privacy and Security Matters in Biometric Technologies, served previously on the IEEE Speech and Language Technical Committee and serves currently as an associate editor for the IEEE Trans. on Biometrics, Behavior, and Identity Science.</p>
	<p>Jean-François Bonastre is Senior Researcher (Directeur de Recherche) at Inria Defense & Security where he acts as Research head and a professor of computer science at LIA, the computer science laboratory of Avignon University. He received his PhD on automatic speaker recognition in 1994 and his "Habilitation à Diriger les Recherches" (HDR) in 2000. He spent a sabbatical year at Panasonic Speech Technology Laboratory (Santa Barbara, California, USA) in 2002-2003. He was vice-president of Avignon University from 2008 to 2015 and an auditor of the 26th Mediterranean session of IHEDN (Institut des Hautes Etudes de la Défense Nationale) in 2015. He was the director of the LIA and a member of the Scientific Committee of the Montreal Computer Research Center (CRIM) from 2016 to 2020. He was the President of the International Speech Communication Association (ISCA) from 2011 to 2013 and the President of the Association Francophone de la Communication Parlée from 2000 to 2004. He is one of the founders of ISCA's Special Interest Group "Speaker and Language Characterization" (SPLC). He was awarded "ISCA Fellow Member" in 2021, and "Fellow member of the Asia-Pacific Artificial Intelligence Association" in 2023. He is also an IEEE Senior Member and was elected member of the IEEE Speech and Language Technical Committee and IEEE Biometrics Council. Jean-François Bonastre is member of the scientific committee of the main speech conferences and is "Technical Program Chair" of Interspeech 2024.</p>

	<p>Xin Wang is a project associate professor at the National Institute of Informatics (NII), Japan. He received the Ph.D. degree from SOKENDAI/NII, Japan, in 2018. Before that, he received M.S. and B.E degrees from the University of Science and Technology of China and University of Electronic Science and Technology of China in 2015 and 2012, respectively. His research interests include statistical speech synthesis, speech security, and machine learning. He is a co-organizer of the ASVspoof (2019, 2021, ASVspoof5) and VoicePrivacy (2020, 2022) challenges. He is a JST PRESTO Researcher from 2023 October.</p>
	<p>Shuhei Kato received B.E. and Master of Information Science and Technology degrees from The University of Tokyo, Japan, in 2011 and 2013, respectively. He received a Ph.D. degree from The Graduate University for Advanced Studies, SOKENDAI, Japan, in 2021. Since 2020, he has worked as a research engineer for RevComm Inc., Japan. His research interests include speech processing, especially speech synthesis.</p>
	<p>Yusuke Yasuda received the B.S. and M.S. degree at Waseda University, Japan, in 2012 and 2014, respectively. He received the Ph.D degree at graduate university for advanced studies, SOKENDAI, Japan, in 2021. He was a Research Assistant with the National Institute of Informatics, Japan from 2018 to 2021. Since 2021, he has been an Assistant Project Professor with the Information Technology Center, Nagoya University. He received the 2nd Yoshida Award and the 20th Best Student Paper Award from the Acoustical Society of Japan. His research interests include statistical machine learning and speech synthesis.</p>
	<p>Erica Cooper received a B.Sc. degree and M.Eng. degree both in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2009 and 2010, respectively. She received a Ph.D. degree in computer science from Columbia University, New York, NY, USA, in 2019. She is currently a Project Assistant Professor with the National Institute of Informatics, Chiyoda, Tokyo, Japan. Her research interests include statistical machine learning and speech synthesis. Dr. Cooper's awards include the 3rd Prize in the CSAW Voice Biometrics and Speech Synthesis Competition, the Computer Science Service Award from Columbia University, and the Best Poster Award in the Speech Processing Courses in Crete. She has served as a co-organizer for the VoiceMOS Challenge in 2022 and 2023.</p>
	<p>Wen-Chin Huang received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 2018 and the M.S. degree from Nagoya University, Nagoya, Japan in 2021. He was a Research Assistant with the Institute of Information Science, Academia Sinica, Taipei, Taiwan from 2017 to 2019. He is currently a Ph.D. candidate at Nagoya University, Nagoya, Japan. He was the recipient of the Best Student Paper Award in ISCSLP2018, the 16th IEEE Signal Processing Society Japan Student Best Paper Award, the 2023 Outstanding Graduate Student Award of Nagoya University, and the research fellowship for young scientists (DC1) from the Japan Society for the Promotion of Science in 2021. He was a co-organizer of the Voice Conversion Challenge 2020, the Singing Voice Conversion Challenge 2023, and VoiceMOS Challenge 2022 and 2023. His research focuses on deep learning applications to speech processing, with a main focus in voice conversion and speech quality assessment.</p>



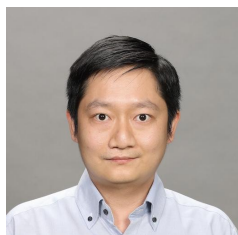
Massimiliano Todisco (Member, IEEE) is currently an Associate Professor (MCF) with EURECOM, France. He received the Ph.D. degree in sensorial and learning systems engineering from the University of Rome Tor Vergata, Italy, in 2012. He is currently the principal investigator and coordinator for TReSPAsS-ETN, a H2020 Marie Curie ITN, RESPECT, a French-German funded project and P-SPIKE a French funded project. He co-organises the ASVspooF and VoicePrivacy challenge series. His research delves into speech processing, voice biometrics, fake audio and speech detection, and algorithms that bolster privacy. In 2020, he was the recipient of the International Speech Communication Association (ISCA) Award for the best article published in the journal 'Computer Speech and Language' during the quinquennium 2015–2019 for his contribution to speaker verification anti-spoofing. He co-edits the first edition of the Handbook of Privacy and Security Matters in Biometric Technologies.



Hemlata Tak received her Ph.D. degree from Sorbonne University, France. She received her Master's degree in 2018 from DA-IICT, Gandhinagar, India. She co-organized the inaugural edition of the Spoof-Aware Speaker Verification (SASV) Challenge 2022. She is also a co-organiser of the ASVspooF 5 Challenge. Her research interests include voice biometrics, audio deepfake detection and anti-spoofing.



Lin Zhang (Student Member, IEEE) received an M.S. degree from Tianjin University, China, in 2020. She served as a Research Assistant with the SMIIP Lab at Duke Kunshan University in 2020 and Visitor with Speech@FIT at Brno University of Technology in 2023. She is currently working toward the Ph.D. degree with the SOKENDAI/National Institute of Informatics, Japan. Her research interests include anti-spoofing, speaker recognition, speaker diarization, and machine learning.



Huy H. Nguyen received the Ph.D. from the Graduate University for Advanced Studies, SOKENDAI, Japan, in 2022. He is a Specially Appointed Assistant Professor with the National Institute of Informatics, Tokyo, Japan. He received the Best Student Award from the National Institute of Informatics, Japan (2020), the Telecom Interdisciplinary Research Award from the Telecommunication Advancement Foundation, Japan (2023), the Best Paper Award (WIFS 2017), Excellent Paper Awards (three journal papers) from the Institute of Electronics, Information and Communication Engineers (2023), Japan. He serves as a reviewer for several conferences (NeurIPS, ICML, WACV, ICME, ACL RR, APSIPA ASC) and journals (IEEE [Access, TIP, TIFS], IEEE/CAA JAS, ACM TOMM, Elsevier [PRLETTERS, EAAI], EURASIP JIVP, IEICE). His research interests include security and privacy in biometrics & machine learning, and disinformation.



Ching-Chun Chang received his PhD in Computer Science from the University of Warwick, UK, in 2019. He participated in a short-term scientific mission supported by European Cooperation in Science and Technology Actions at the Faculty of Computer Science, Otto-von-Guericke-Universität Magdeburg, Germany, in 2016. He was granted the Marie-Curie fellowship and participated in a research and innovation staff exchange scheme supported by Marie Skłodowska-Curie Actions at the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, USA, in 2017. He was a Visiting Scholar with the School of Computer and Mathematics, Charles Sturt University, Australia, in 2018, and with the School of Information Technology, Deakin University, Australia, in 2019. He was a Research Fellow with the Department of Electronic Engineering, Tsinghua University, China, in 2020. He is currently a Postdoctoral Fellow with the National Institute of Informatics, Japan. His research interests include steganography, forensics, biometrics, cybersecurity, computer vision, computational linguistics and artificial intelligence.



Natalia Tomashenko is a researcher at the University of Avignon, France. She received the Ph.D. degree in computer science from the University of Le Mans, France. Her research interests focus on statistical machine learning for speech and language processing with application to automatic speech and speaker recognition, spoken language understanding, machine translation, and speech privacy. She has been a lead co-organizer of the 1st and 2nd VoicePrivacy Challenges, guest editor for the Computer Speech & Language journal, co-chair of the 2nd Symposium on Security and Privacy in Speech Communication, and organiser of other scientific events.



Xiaoxiao Miao is a postdoctoral researcher at the National Institute of Informatics (NII), Japan until October 2023. She will be an assistant professor at the Singapore Institute of Technology (SIT) in November 2023. She received the Ph.D. degree from the Institute of Acoustics, Chinese Academy of Sciences/University Chinese Academy of Sciences, in 2021. Her research interests include speaker and language recognition, speech security, and machine learning. She is a co-organizer of the latest VoicePrivacy challenge.