APSIPA ASC 2020 – TUTORIALS

Timing is in New Zealand Time

	Morning Tutorials (12:00 – 15.00)	Research Area	Name	Institution
T1	Theory and Practice of Voice Conversion	Voice Conversion	Berrak Sisman, Yu Tsao and Haizhou Li	Singapore University of Technology and Design, **Singapore,** Research Center for Information Technology Innovation, **Taiwan** , NUS, **Singapore**
T2	Cochlear Implant Sound Processing and Perception	Cochlear sound processing	Brett Swanson	Cochlear Ltd., **Australia**
T3	Tensor Representations in Signal Processing and Machine Learning	Tensor Processing	Tatsuya Yokota	Nagoya Institute of Technology, **Japan**
	Afternoon Tutorials (16.00 – 19.00)			
T4	Action and Event Understanding: From Theory to Applications	Multimedia analysis and understanding technologies	Weiyao Lin	Shanghai Jiao Tong University, **China**
T5	Beyond Noise Cancelling Headphones: Recent Advances and Challenges in the Practical Implementation of Multichannel Active Control of Noise (MCANC) in Large Spaces.	Active Noise Cancellation in spatial audio	Woon-Seng, GAN, Dongyuan, SHI and Bhan, LAM	Nanyang Technological University, **Singapore**
T6	Machine Learning Based Behavioural Analysis for Cybersecurity of IoT Devices	Cybersecurity of IoT Devices	Hassan Gharakheili and Vijay Sivaraman	UNSW, **Australia**

T1	**Title:** Theory and Practice of Voice Conversion **Presenters:** Berrak Sisman (Singapore University of Technology and Design, Singapore), Yu Tsao, (Research Center for Information Technology Innovation, Taiwan), Haizhou Li (National University of Singapore, Singapore)
T1	**Abstract:** Human voice carries the information of speaker identity. Voice conversion (VC) converts one speaker's voice to sound like that of another. Since 1980s, significant progress has been made with the evolution of algorithms from statistical modeling to deep learning, that brings about the improvement of voice quality by leaps and bounds. Voice conversion has enabled many applications such as personalized speech synthesis, spoofing attacks, and dubbing of movies now a day. Addressing the needs from real-world voice conversion applications, the research community has stepped up the effort which is evident from the increased number of publications in the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), INTERSPEECH, APSIPA ASC and various workshops organized by IEEE and International Speech Communication Association. The number of related articles published in IEEE/ACM Transactions on T-ASLP, Speech Communication, and APSIPA Transactions on Signal and Information Processing is also in a steady rise. Since 2016, Voice Conversion Challenges have been organized for 3 times that attracted a growing research community. In APSIPA ASC 2018, a tutorial on Voice Conversion: Challenges and Opportunities was presented, that was before neural vocoders become widely adopted. This tutorial provides perspectives on the latest development of voice conversion. The proposed voice conversion tutorial will consist of 7 sections, 1) an introduction to voice conversion; 2) history of voice conversion; 3) statistical modelling and exemplar-based techniques; 4) deep learning techniques; 5) voice conversion applications; 6) future research directions; and 7) conclusions.
T1	 **Biography:** Professor Sisman is currently an Assistant Professor at Singapore University of Technology and Design (SUTD). She is also an Affiliated Researcher at the National University of Singapore (NUS). Before joining SUTD, she was a visiting researcher at Columbia University, New York, United States in 2020. She obtained her PhD from National University of Singapore in January 2020 fully funded by A*STAR Graduate Academy under Singapore International Graduate Award. She was an exchange PhD student at the University of Edinburgh and a visiting scholar at The Centre for Speech Technology Research (CSTR), University of Edinburgh in 2019. She was attached to RIKEN Advanced Intelligence Project, Japan in 2018. Her research is focused on machine learning, signal processing, speech synthesis and voice conversion. She was awarded APSIPA PhD Forum Best Presentation Award for her presentation titled "Limited Data Voice Conversion from Sparse Representation to GANs and WaveNet" in 2018, Hawaii, United States. She has served as the General Coordinator of the Student Advisory Committee (SAC) of the International Speech Communication Association (2017-2019), and she is now serving as an advisor to ISCA SAC.
T1	 **Biography:** Professor Tsao received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1999 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2008. From 2009 to 2011, he was a Researcher with the National Institute of Information and Communications Technology, Tokyo, Japan, where he engaged in research and product development in automatic speech recognition for multilingual speech-to-speech translation. He is currently an Associate Research Fellow with the Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan. His research interests include speech and speaker recognition, acoustic and language modelling, audio coding, and bio-signal processing. He is currently an Associate Editor of the IEEE/ACM Transactions on Audio, Speech, and Language Processing and IEICE transactions on Information and Systems. Dr. Tsao received the Academia Sinica Career Development Award in 2017, National Innovation Award in 2018 and 2019, and Outstanding Elite Award, Chung Hwa Rotary Educational Foundation 2019-2020.
T1	 **Biography:** Professor Li is currently a Professor at the Department of Electrical and Computer Engineering, National University of Singapore (NUS). His research interests include automatic speech recognition, speaker and language recognition, natural language processing, and neuromorphic computing. Prior to joining NUS, he taught in the University of Hong Kong (1988–1990) and South China University of Technology (1990–1994). He was a Visiting Professor at CRIN in France (1994–1995), Research Manager at the Apple-ISS Research Centre (1996–1998), Research Director in Lernout & Hauspie Asia Pacific (1999–2001), Vice President in InfoTalk Corp.

	Ltd. (2001–2003), and the Principal Scientist and Department Head of Human Language Technology in the Institute for Infocomm Research, Singapore (2003-2016). He was the recipient of National Infocomm Awards 2002, Institution of Engineers Singapore (IES) Prestigious Engineering Achievement Award 2013 and 2015, President's Technology Award 2013, and MTI Innovation Activist Gold Award 2015 in Singapore. He was named one of the two Nokia Visiting Professors in 2009 by Nokia Foundation, IEEE Fellow in 2014 for leadership in multilingual, speaker and language recognition, ISCA Fellow in 2018 for contributions to multilingual speech information processing, and Bremen Excellence Chair Professor in 2019. Dr. Li is a member of ACL, ACM, and IEICE.

T2	**Title:** Cochlear Implant Sound Processing and Perception **Presenter:** Brett A. Swanson (Cochlear Ltd, Sydney, Australia)
T2	**Abstract**: A cochlear implant restores a sense of hearing to a person with severe to profound deafness. Most cochlear implant recipients achieve good speech perception under good listening conditions, but the two big challenges are speech perception in noisy conditions, and musical pitch perception. Multiple microphones and adaptive beam forming techniques can provide large improvements in intelligibility when the speech is spatially separated from the noise. When the speech and noise come from the same direction, noise reduction algorithms seek to identify and exploit specific noise types such as impulsive noise, wind noise, and stationary noise. Because the dynamic range of electrical stimulation in a cochlear implant is much less than the dynamic range of normal hearing, some form of compression or automatic gain control is necessary. A simple model that calculates how the signal-to-noise ratio is modified by the time-varying gains can make useful predictions of speech intelligibility under a variety of processing algorithms. Modest improvements in pitch perception have recently been demonstrated by a new sound processing algorithm named OPAL, which estimates the fundamental frequency of the incoming sound and applies corresponding amplitude modulation to the appropriate bands. However, cochlear implant pitch perception remains substantially worse than in normal hearing, most likely due to our inability to reproduce the spatio-temporal neural firing pattern evoked by resolved harmonics in normal hearing. The next generation of cochlear implants, with electrode arrays that are positioned closer to the neural tissue, and which can produce more focused stimulation patterns, may provide improvements.
T2	 **Biography:** Dr Brett Swanson received a BE (Electrical) from the University of New South Wales in 1985, and a PhD from the University of Melbourne in 2008. He joined Cochlear Ltd in Sydney in 1992, and led the firmware development for the company's first sound processor that utilized digital signal processing. He was a key developer of the sound processing algorithms that today are used by more than 100,000 cochlear implant recipients worldwide. His PhD thesis was entitled "Pitch perception with cochlear implants". He has run clinical studies to evaluate speech and pitch perception with novel processing in the areas of filterbanks, multiband automatic gain control, spatial and temporal stimulation patterns, and multipolar focused stimulation. He has co-supervised three PhD students to successful completion. He has contributed to the design of several generations of cochlear implant systems, and has 10 US patents.

T3	**Title:** Tensor Representations in Signal Processing and Machine Learning **Presenter:** Tatsuya Yokota (Nagoya Institute of Technology, Japan)
T3	**Abstract**: Tensor is a general name of multi-dimensional array. For example, a vector and a matrix are respectively the first order and the second order tensors, and these are generalized as N-th order tensors. Many kinds of data can be naturally represented as tensors such as multi-media data (audio, image, video), bio-medical data (EEG, MEG, CT, MRI, PET), remote sensing data (hyper-spectral), and others (wireless communications, social data, consumer data, financial data). Tensor representation has benefits when the data have low-rank and smooth structures. These tensor data structures are often utilized for dimensionality reduction, representation learning, source separation, denoising, and completion. The techniques of higher order tensorization, which re-represent the data as more beneficial styles (e.g., time-frequency analysis), has also attractive attentions in recent year. In addition, tensors are also utilized for representing weight parameters in deep learning. Tensor decomposition/factorization is now quite important in wide range of signal processing and machine learning studies. The objective of this tutorial is for (1) looking back the basics of tensor methods in signal processing and machine learning, (2) introducing the various applications of tensor methods in signal processing and machine learning, and (3) finally explaining a selected emerging topic of Hankel tensor representations.
T3	 **Biography:** Professor Tatsuya Yokota received the Ph.D. degree in engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 2014. From 2011 to 2014, he was a Junior Research Associate with the Laboratory for Advanced Brain Signal Processing (ABSP), RIKEN Brain Science Institute, Japan. From 2014 to 2016, he was a Research Scientist with the Laboratory for ABSP. He is currently an Assistant Professor with the Department of Computer Science, Nagoya Institute of Technology, Japan, and a Visiting Research Scientist with the Tensor Learning Unit in RIKEN AIP. His research interests include matrix/tensor decomposition, signal/image processing, and pattern recognition.

<table>
<tr><td>T4</td><td>Title: Action and Event Understanding: From Theory to Applications
Presenter: Weiyao Lin (Shanghai Jiao Tong University,China)</td></tr>
<tr><td>T4</td><td>Abstract: The developments of modern intelligent city, social media, and other intelligent media services highly rely on the advancement of human-centric multimedia analysis technologies. Intelligent multimedia understanding is one of the essential technologies for visual analysis which requires many human-centered and event-driven visual understanding tasks such as human action recognition & localization, event detection & understanding, and spatial-temporal analysis of object relations. Recently, due to the fast progress and widespread applications of computer vision, significant advances in deep learning techniques, action and event understanding has attracted enormous attention from both research communities and industries. Many action & event understanding methods, especially for deep learning architectures have been proposed, demonstrating breathtaking performance in various tasks and successful application in industries.

Nevertheless, with the rapid growth of video data scale, increasingly diverse & complex video contents, and higher demands on more intelligent & reliable services, new challenges such as complex action & event modeling, unsupervised or sparse-labeled video data handling, efficient processing of huge video data, and densely description of actions & events are being issued in front of us. In this tutorial, we will cover the current state-of-the-art theories and technologies on action and event understanding. After discussing the basic theories and development history of action & event understanding, we will cover a series of milestone techniques for handling various action & event understanding tasks. We will also provide theoretic analysis on SOTA techniques, discuss the possible solutions to emerging challenges, and share our experience in the technical design on various applications.</td></tr>
<tr><td>T4</td><td>
Biography: Professor Weiyao Lin received the B.E. degree from Shanghai Jiao Tong University, China, in 2003, the M.E. degree from Shanghai Jiao Tong University, China, in 2005, and the Ph.D degree from the University of Washington, Seattle, USA, in 2010, all in electrical engineering. He is currently a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. He has authored or coauthored 100+ technical papers on top journals/conferences including TPAMI, IJCV, CVPR, and ICCV. He holds 18 patents and has 10+ under reviewing patents. His research interests include multimedia content understanding, computer vision, video/image compression, and video/image processing applications.Dr. Lin served as an associate editor for IEEE Trans. Image Processing, IEEE Trans. Circuits & Systems for Video Technology, IEEE Trans. Intelligent Transportation Systems. He is an organizing committee chair of International Conference on Image and Graphics (ICIG) 2017, an area chair of ICPR’20, BMVC'19, ICIP'19, ACM MM'18, ICME'2018, and VCIP'2017, and an organizer of 6+ workshops or challenges in ACM MM, ICCV, ECCV, ICME, BigMM, and VCIP. He is a member of the IEEE Multimedia Signal Processing Technical Committee (MMSP TC), a member of the IEEE Multimedia Systems and Applications Technical Committee (MSA TC), and a member of the IEEE Visual Signal Processing and Communication Technical Committee (VSPC TC). He received the Multimedia Rising Star award in ICME'2019, the outstanding Area Chair award in ICME'2018, and the Best Associate Editor award of the month for IEEE Access. He is a senior member of IEEE.</td></tr>
</table>

T5	**Title:** Beyond Noise Cancelling Headphones: Recent Advances and Challenges in the Practical Implementation of Multichannel Active Control of Noise (MCANC) in Large Spaces. **Presenters:** Woon-Seng, GAN, Dongyuan, SHI and Bham, LAM (Nanyang Technological University, Singapore)
	Abstract: The implementation of ANC has been limited to confined spaces such as headphones and automobiles. Owing to practical limitations, extending the control zone to large spaces is no trivial task. In this tutorial, we will present the physical basis and recent advancements of the MCANC technique to extend the active control zone. In Part I, the principles of ANC are discussed. Next, the role of ANC in combating urban noise pollution and recent innovations are discussed. Some foundations on the physics and signal processing for active control will be briefly introduced. Part II focuses on the physical aspects of ANC of the open MCANC windows. The design requirements are firstly detailed, followed by the optimization of control sources using numerical methods. The physical limits of active and passive control for the open window is presented. In situ measurements and reproduction techniques of the primary disturbance are discussed to ensure sufficient real-world fidelity. Part III explores the hardware implementation and recent advances in signal processing of ANC. We will revisit MCANC and its recent developments. Next, we introduce an FPGA architecture to implement MCANC. Developments in signal processing to deal with practical implementation challenges such as (1) coping with output saturation, (2) “virtual” projection of the quiet zone, and (3) selection of pre-trained fixed-filters for varying primary disturbances, are presented. Finally, Part IV summarizes this tutorial and highlights the challenges in practical implementation. Future trends in ML and DL and the role of IoT in MCANC is briefly discussed.
T5	 **Biography:** Professor Woon-Seng Gan received his BEng (1st Class Hons) and PhD degrees, both in Electrical and Electronic Engineering from the University of Strathclyde, UK in 1989 and 1993 respectively. He is currently a Professor of Audio Engineering and Director of the Smart Nation Lab in the School of Electrical and Electronic Engineering in Nanyang Technological University. From 2011-2014, he also served as the Head of the Information Engineering Division in the School of Electrical and Electronic Engineering in Nanyang Technological University. He also served as the Director of Center for Inforcomm Technology from 2015-2019. His research has been concerned with the connections between the physical world, signal processing and sound control, which resulted in the practical demonstration and licensing of spatial audio algorithms, directional sound beam, and active noise control for headphones. He is a Fellow of the Audio Engineering Society(AES), a Fellow of the Institute of Engineering and Technology(IET), and a Senior Member of the IEEE. He served as an Associate Editor of the IEEE/ACM Transaction on Audio, Speech, and Language Processing (TASLP; 2012-15) and was presented with an Outstanding TASLP Editorial Board Service Award in 2016. He also served as the Associate Editor for the IEICE transaction (2014-2016) on Fundamentals of Electronics, Communications and Computer Sciences(Japan). He is currently serving as the Senior Area Editor of the IEEE Signal Processing Letters (2020-); Associate Technical Editor of the Journal of Audio Engineering Society (JAES; 2013-); Editorial member of the Asia Pacific Signal and Information Processing Association (APSIPA; 2011-) Transaction on Signal and Information Processing; Associate Editor of the EURASIP Journal on Audio, Speech and Music Processing (2007-).
T5	 **Biography**: Dr Dongyuan Shi received his B.ENG and M.S degrees in the University of Electronic Science and Technology of China (UESTC), in 2010 and 2013, respectively. He worked as a research fellow of the Applied Micro-fluidicology Laboratory of Nanyang Technology University to design the cell CMOS image capture system and the framework of the image processing algorithm from 2013-2014. In 2020, he received his Ph.D. degree in electrical and electronic engineering at Nanyang Technological University (NTU), Singapore. His Ph.D. work has been published in the Journal of the Acoustical Society of America, IEEE Transactions on Very Large Scale Integration (VLSI) system, IEEE Signal Processing Letters, Elsevier Mechanical Systems and Signal Processing, and ICASSP conferences. He is currently a member of the IEEE and Signal Processing Society (SPS). His research interests include advanced digital signal processing, adaptive array processing, high-speed real-time digital system implementation, and active noise control.

T5	 **Biography**: Dr Bhan Lam received his B.Eng. (Hons) and Ph.D. degrees in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2013 and 2019, respectively. He is currently a research fellow at the same university. In 2015, he was a visiting postgrad in the Signal Processing and Control group at the Institute of Sound and Vibration Research, University of Southampton, UK. His current research includes the acoustics of active noise control, soundscape, and signal processing for active control. He is currently on the reviewer board of MDPI Applied Sciences, and has published 12 peer-reviewed journal articles in Elsevier Building and Environment, Elsevier Mechanical Systems and Signal Processing, Elsevier Applied Acoustics, IEEE Transactions on Very Large Scale Integration (VLSI) system, IEEE Signal Processing Letters, Elsevier Journal of Sound and Vibration and AIP Journal of the Acoustical Society of America, and MDPI Applied Sciences

T6	**Title:** Machine Learning-Based Behavioural Analysis for Cybersecurity of IoT Devices **Presenter:** Hassan Habibi Gharakheili, Vijay Sivaraman (UNSW, Sydney, Australia)
T6	**Abstract:** IoT devices, sourced from a diversity of vendors and deployed in large numbers for residential and industrial purposes, create cyber-security vulnerabilities at unprecedented scale for the Internet ecosystem. In this tutorial, we present how privacy and security risks of IoT devices can be systematically evaluated by demonstrating real-life threats to typical users posed by cyber attackers. Next, a behavioural analysis of IoT network traffic is presented that leads to development of machine learning-based models (multi-class) for classifying IoT devices based on their network activities. Lastly, flow-level models (one-class) are trained for detecting anomalous patterns in network traffic of individual devices.
T6	 **Biography:** Dr Hassan Habibi Gharakheili received his B.Sc. and M.Sc. degrees of Electrical Engineering from the Sharif University of Technology in Tehran, Iran in 2001 and 2004 respectively, and his Ph.D. in Electrical Engineering and Telecommunications from the University of New South Wales (UNSW Sydney) in Australia in 2015. He is now a lecturer at UNSW Sydney. His current research interests include programmable networks, learning-based networked systems, and data analytics in computer systems
T6	 **Biography**: Professor Vijay Sivaraman received his B. Tech. from the Indian Institute of Technology in Delhi, India, in 1994, his M.S. from North Carolina State University in 1996, and his Ph.D. from the University of California at Los Angeles in 2000. He has worked at Bell-Labs as a student Fellow, in a Silicon Valley start-up manufacturing optical switch-routers, and as a Senior Research Engineer at the CSIRO in Australia. He is now a Professor at the University of New South Wales in Sydney, Australia. His research interests include Software Defined Networking, network architectures, and cyber-security particularly for IoT networks