Issue 19

April 2019

APSIPA NEWSLETTER

Asia-Pacific Signal and Information Processing Association Newsletter

Greetings!

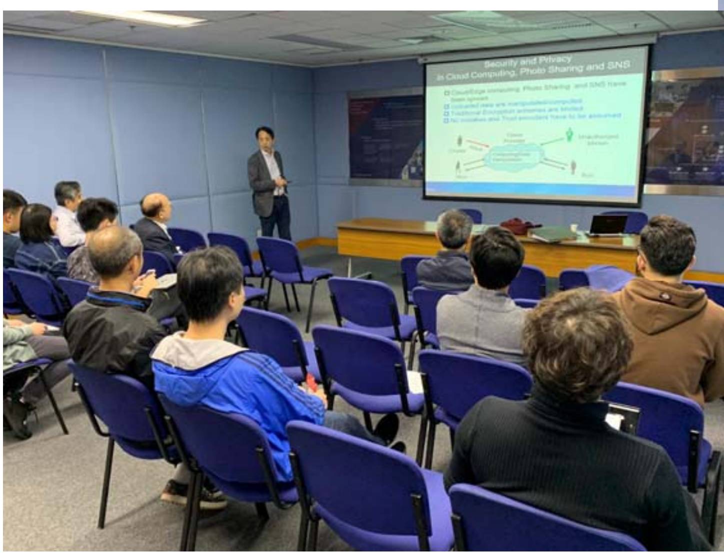
Welcome to the April issue of the APSIPA Newsletter. The APSIPA President Prof. Kiya has visited the Headquarters in March this year. A good discussion between Prof Kiya and Headquarters officers has been done. Prof. Kiya has also given a seminar entitled "Progress and Challenges in Compressible and Learnable Image Encryption for Untrusted Cloud Environments" which was well received by the audience. Following the announcement of the eight 2019-2020 Distinguished Lecturers by Prof. Woon-Seng Gan early this year, we would like to share with members about their biographies and the possible distinguished lectures topics. In particular, we are happy to have two detailed sharing from Prof. Yu Tsao and Prof. Xiao-Lei Zhang about their research works. One relates to the multilayer bootstrap networks while another is about speech processing.

Don't forget that APSIPA Annual Summit and Conference will be held in Lanzhou, China from 18 to 21 November 2019. The submission deadline for proposal of special sessions and regular session papers are May 1, 2019 and June 1, 2019 respectively. More information is available at http://www.apsipa2019.org/.

Enjoy reading this issue!

Bonnie Law

APSIPA Newsletter EiC



In this issue

2019-2020 Distinguished Lecturers	Page 2
Sharing from DL: Prof. Yu Tsao	Page 6
Sharing from DL: Prof. Xiao-Lei Zhang	Page 11
Call-for-papers in ATSIP	Page 14
Latest Articles from ATSIP	Page 15
APSIPA membership	Page 17
Summary of links	Page 18
APSIPA Who's Who	Page 19

APSIPA Distinguished Lecturers for 2019-2020

The Distinguished Lecturer (DL) Program is established to serve the communities by organizing lectures given by distinguished experts. It is an educational program that promotes the research and development of signal and information processing in Asia-Pacific region. During the two-year term, the APSIPA Distinguished Lecturers will be serving as ambassadors of APSIPA and outreaching to audience in Asia-Pacific region as well as worldwide. For more information about the DL Program, please contact Professor Woon-Seng Gan who is the Vice President of Institutional Relations and Education Program. This year, we have eight newly selected Distinguished Lecturers.

Prof. Chia-Ping Chen

I am a Professor in the Department of Computer Science and Engineering of National Sun Yat-sen University in Taiwan. I work in the areas of speech recognition, speaker recognition, emotion recognition, deep learning and machine intelligence. I have Bachelor and Master degrees in Physics from National Taiwan University and National Tsing-hua University in Taiwan, and PhD degree in Electrical Engineering from University of Washington at Seattle in US. Further information can be found at http://slpl.cse.nsysu.edu.tw/cpchen/research.

Tentative Distinguished Lecture Titles:

- Deep learning and machine intelligence
- Speaker recognition and spoof detection
- Beyond classification: detection and tracking

Prof. Christian Ritz

Christian graduated with a Bachelor of Electrical Engineering and a Bachelor of Mathematics (both in 1999) and a PhD in Electrical Engineering (in 2003) all from the University of Wollongong, Australia. His PhD research focused on very low bit rate coding of wideband speech signals. Since 2003, Christian has held a position within the School of Electrical, Computer and Telecommunications Engineering at the University of Wollongong where he is currently a Professor. Concurrently, he is also the Associate Dean (International) for the Faculty of Engineering and Information Sciences, with responsibility for managing the Faculty's international strategy including significant transnational programs and partnerships in China, Hong Kong, Dubai, Singapore and Malaysia. Christian is the deputy director of the Centre for Signal and Information Processing (CSIP) and leads the audio, speech and acoustics signal processing research of the centre. He is actively involved in several projects including microphone array signal processing for the enhancement of directional sound, acoustic scene classification, sound field reproduction and control using loudspeaker arrays and visual object classification using machine learning. His research has been from the Australian government and industry. For more information see: https://scholars.uow.edu.au/display/christian_ritz.

Tentative Distinguished Lecture Titles:

- Multizone Sound Field reproduction incorporating perceptual quality constraints
- Perceptual control of speech sound fields
- Novel microphone arrays for speech applications: vector sensors, B-format microphones and co-prime arrays

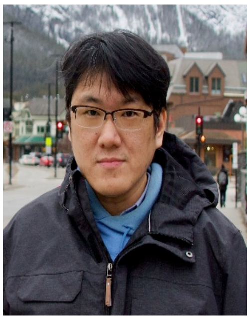
Prof. Tomoki Toda

Tomoki Toda is a Professor of the Information Technology Center at Nagoya University, Japan. He received the B.E. degree from Nagoya University in 1999 and the M.E. and Ph.D. degrees from the Graduate School of Information Science, NAIST, Japan, in 2001 and 2003, respectively. He was a Research Fellow of JSPS in the Graduate School of Engineering, Nagoya Institute of Technology, Aichi, Japan, from 2003 to 2005. He was then an Assistant Professor (2005-2011) and an Associate Professor (2011-2015) at the Graduate School of Information Science, NAIST. His research interests include statistical approaches to speech, music, and environmental sound processing. He published over 60 journal papers and 250 conference papers in this research area. He has served as an Associate Editor of the IEEE Signal Processing Letters since 2016 and of EURASIP Journal on Audio, Speech, and Music Processing since 2013. He is one of the organizers for Voice Conversion Challenge. He was a member of the Speech and Language Technical Committee of the IEEE SPS from 2007 to 2009, and from 2014 to 2016. He was a coordinating area chair for INTERSPEECH 2014. He received more than 10 paper and achievement awards including the IEEE SPS 2009 Young Author Best Paper Award, the 2013 EURASIP-ISCA Best Paper Award (Speech Communication Journal), and the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology, the Young Scientists' Prize.

Tentative Distinguished Lecture Titles:

- Recent trend of voice conversion techniques
- Speech waveform modeling for advanced voice conversion
- Augmented speech production based on real-time voice conversion

Prof. Yu Tsao



Yu Tsao received the B.S. and M.S. degrees in electrical engineering from National Taiwan University in 1999 and 2001, respectively. He received his Ph.D. degree in electrical and computer engineering from Georgia Institute of Technology in 2008. From 2009 to 2011, he was a researcher with the National Institute of Information and Communications Technology, Japan, where he was involved 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 modeling, audio-coding, and bio-signal processing. He received the Academia Sinica Career Development Award in 2017 and National Innovation Award in 2018.

Tentative Distinguished Lecture Titles:

- Artificial Intelligence for Assistive hearing and Speaking Technologies
- Deep Learning-based Speech Enhancement to Improve Intelligibility for Cochlear Implant Recipients
- Deep Learning-based Speech Enhancement with Direct Evaluation Metrics Optimization

Prof. Shinji Watanabe

Shinji Watanabe is an Associate Research Professor at Johns Hopkins University, Baltimore, MD, USA. He received his B.S., M.S. PhD (Dr. Eng.) Degrees in 1999, 2001, and 2006, from Waseda University, Tokyo, Japan. He was a research scientist at NTT Communication Science Laboratories, Kyoto, Japan, from 2001 to 2011, a visiting scholar in Georgia institute of technology, Atlanta, GA in 2009, and a Senior Principal Research Scientist at Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA from 2012 to 2017. His research interests include automatic speech recognition, speech enhancement, spoken language understanding, and machine learning for speech and language processing. He has been published more than 150 papers in top journals and conferences, and received several awards including the best paper award from the IEICE in 2003. He served as an Associate Editor of the IEEE Transactions on Audio Speech and Language Processing, and is a member of several technical committees including the IEEE Signal Processing Society Speech and Language Technical Committee and APSIPA Speech, Language, and Audio Technical Committee.

Tentative Distinguished Lecture Titles:

- Far-Field Speech Processing
- End-to-End Speech Processing

Prof. Kok Sheik Wong

KokSheik is an Associate Professor attached to the School of Information Technology at Monash University Malaysia. He received the B.S. and M.S. degrees in both Computer Science and Mathematics from Utah State University, USA, in 2002 and 2005, respectively. In 2009, he received the Doctor of Engineering degree from Shinshu University, Japan, under the scholarship of Monbukagakusho. He was with Multimedia University from 2009 to 2010, and University of Malaya from 2010 to 2016. His research interests include multimedia signal processing, data hiding, multimedia perceptual encryption, joint encryption and data-hiding method, as well as their applications. He is a senior member of IEEE and a member of APSIPA. He was a General Co-Chair for APSIPA ASC 2017 (Kuala Lumpur).

In 2015, his student's thesis received the Best Ph.D thesis award from the IEEE Signal Processing Society, Malaysia Section. Currently, he serves as an associate editor of the Journal of Information Security and Applications (JISA), IIEEJ Transactions on Image Electronics and Visual Computing, and Malaysian Journal of Computer Science. He is also a member in Multimedia Security and Forensics (MSF) Technical Committee.

Tentative Distinguished Lecture Titles:

- Information hiding in any digital signal
- Encryption of compressed multimedia content
- JEDI—Joint Encryption and Data Insertion

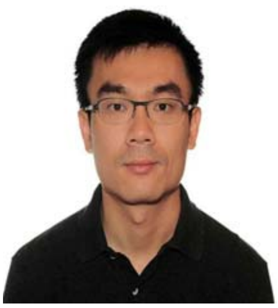
Prof. Xiao-Lei Zhang

Xiao-Lei Zhang is currently a Full Professor with the Center for Intelligent Acoustics and Immersive Communications, and the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China. He received the Ph.D. degree in information and communication engineering from Tsinghua University, Beijing, China, and was a Postdoctoral Researcher with the Perception and Neurodynamics Laboratory, The Ohio State University. His research interests include audio and speech signal processing, machine learning, statistical signal processing, and artificial intelligence. He has published over 30 journal articles and conference papers in IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE/ACM Transaction on Audio, Speech, and Language Processing, Neural Networks, IEEE Transactions on Cybernetics, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, ICASSP, Interspeech, etc. and co-edited a text book in statistics. He received the first-class Beijing Science and Technology Award. He serves/served as an associate editor of five international journals including Neural Networks and EURASIP Journal on Audio, Speech, and Music Processing.

Tentative Distinguished Lecture Titles:

- Recent research and development of deep learning based speech separation
- Metric learning based speaker recognition: Research and trends
- Multilayer bootstrap networks: Building billions of billions of hierarchical trees efficiently for nonlinear dimensionality reduction

Prof. Jiantao Zhou



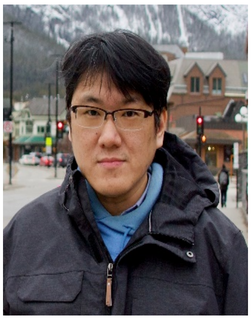
Jiantao Zhou (M'11) received the B.Eng. degree in electronic engineering from the Dalian University of Technology, Dalian, China, in 2002, the M.Phil. degree in radio engineering from Southeast University, Nanjing, China, in 2005, and the Ph.D. degree in electronic and computer engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2009. He held various research positions with the University of Illinois at Urbana–Champaign, the Hong Kong University of Science and Technology, and the McMaster University. He is currently an Associate Professor with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, China. He holds four granted U.S. patents and two granted Chinese patents. His research interests include multimedia security and forensics, multimedia signal processing, artificial intelligence, and big data. He has coauthored two papers that received the Best Paper Award at the IEEE Pacific-Rim Conference on Multimedia in 2007 and the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo in 2016. He is an Associate Editor for the IEEE Transactions on Image Processing.

Tentative Distinguished Lecture Titles:

- When Image Forensics/Anti-Forensics meets Adversarial Machine Learning
- Robust Subspace Representation with Independent and Piecewise Identically Distributed Noise Modeling
- Recent Progress on Image Noise Modeling and Estimation

Sharing from DL: Prof. Yu Tsao

Research Center for Information Technology Innovation
Academia Sinica, Taipei, Taiwan.



1. **AI for assistive speech communication technologies:** The proportional increase in the elderly population and the inappropriate use of portable audio devices have led to a rapid increase in incidents of hearing loss. Untreated hearing loss can cause feelings of loneliness and isolation in the elderly and may lead to learning difficulties in students. Over the past few years, our group has investigated the application of machine learning and signal processing algorithms in FM assistive hearing systems [1, 2], hearing aids [3, 4], and cochlear implants (CIs) [5-7] to improve speech communication in hearing-impaired patients and the subsequent enhancement in their quality of life. In addition to assistive listening devices, we have also investigated the development of machine learning-based assistive speaking devices to enhance intelligibility in individuals with speech and language disorders [8]. Oral cancer ranks in the top five of all cancers in Taiwan. To treat the oral cancer, surgical processes are often required to have parts of the patients' articulators removed. Because of the removal of parts of the articulator, a patient's speech may be distorted and difficult to understand. To overcome this problem, we propose two voice conversion (VC) approaches: the first one is the joint dictionary training non-negative matrix factorization (JD-NMF), and the second one is the end-to-end generative adversarial network (GAN)-based unsupervised VC model [9]. Experimental results show that both approaches can be applied to convert the distorted speech signals to the ones with improved intelligibility.

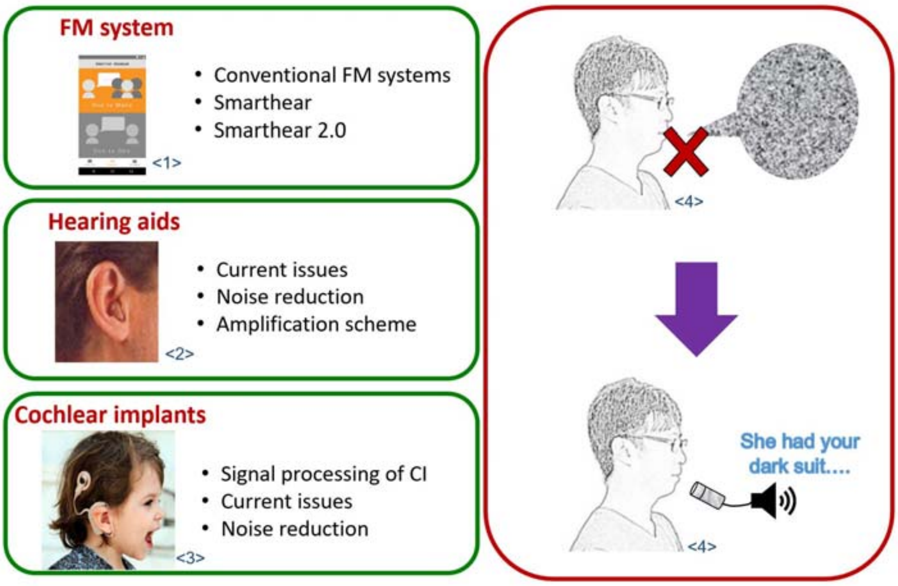


Fig. 1. AI for assistive listening (FM, HA, and CI) and speaking technologies.

2. **Deep learning-based speech signal processing:** In this investigation, we focused on deriving novel deep learning-based algorithms for denoising [10], dereverberation [11], and channel compensation [12] on speech signals. The goal is to enhance the speech signals in order to achieving improved human-human and human-machine communication efficacy. We investigated approaches to enhance speech intelligibility [13] and quality [14] to facilitate higher speech recognition rates and improved communication. We also proposed end-to-end waveform enhancement to directly improve the intelligibility and quality of speech. In addition, we have developed a novel integrated deep and ensemble learning algorithm (IDEA) [15] and an environment-adaptive algorithm (based on the domain adversarial training criterion) [16] for speech signal processing to address possible mismatched issues in real-world applications.

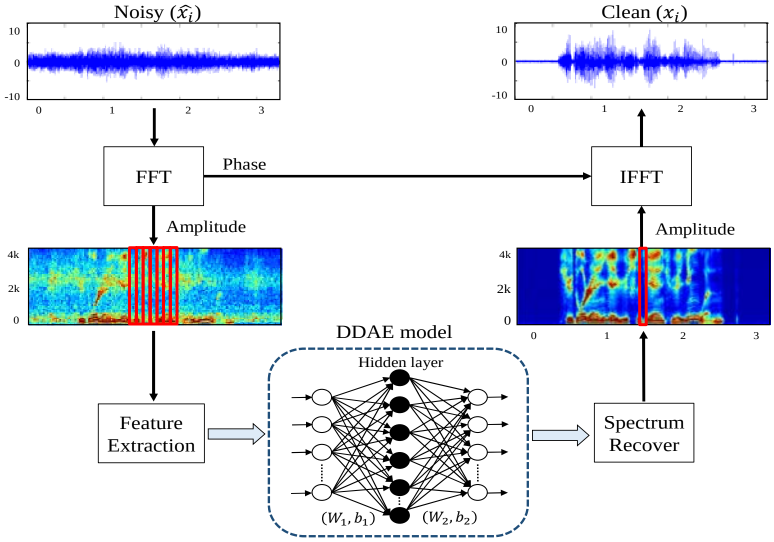


Fig. 2. Deep learning-based speech signal processing.

3. **Multimodal speech signal processing:** Communication can be verbal or nonverbal. Verbal communication includes speaking and listening. The speaker relays verbal information while the listener focuses on auditory signals and the visual cues for speech recognition. The visual signals may include articulatory movements, facial expressions, and co-speech gestures of the speaker, which constitutes the nonverbal part of the communication process. In various applied speech technologies, it has been shown that audio-visual integration can assist in human information exchange and the development of human-computer interfaces. We have conducted studies on the incorporation of audio and video information to facilitate improved speech signal processing performance. Currently, we have developed novel algorithms by fusing the audio and visual information for emotion recognition [17], oral presentation scoring [18], and speech enhancement [19].

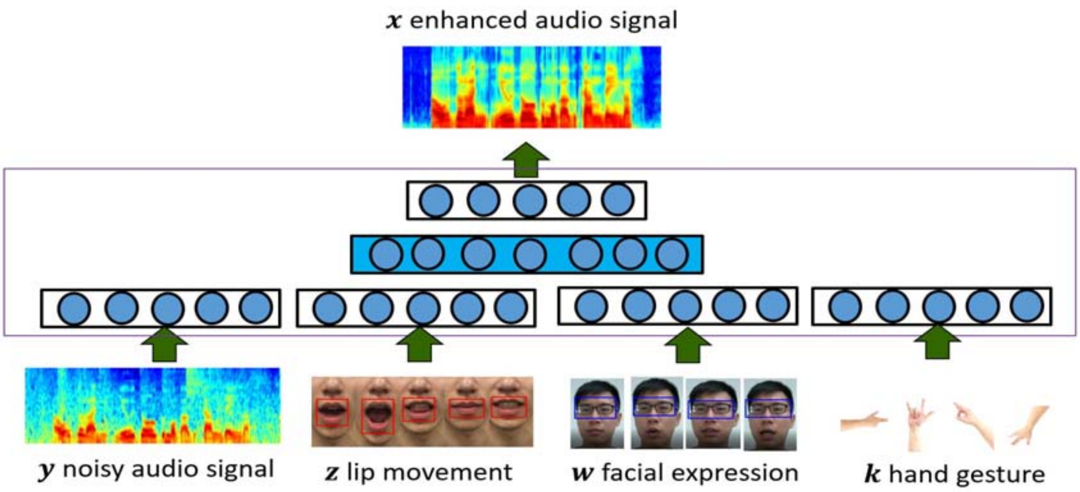


Fig. 3. The proposed multi-modal speech signal processing framework.

4. **Increasing Compactness of Deep Learning based Speech Enhancement Models:** Most recent studies on deep learning-based speech enhancement (SE) have focused on improving the denoising performance. However, successful SE application requires the achievement of a balance between denoising performance and computational cost in real scenarios. We have investigated two approaches to effectively compress deep learning models so that the SE can be performed at edge sides. These approaches are model pruning and parameter quantization. In model pruning, a computation-performance optimization (CPO) algorithm was developed [20] for the removal of redundant channels in a neural network, as shown in Fig. 4. For parameter quantization, we proposed an exponent-only floating point quantized neural network (EOFP-QNN) to compress the model and enhance inference efficiency [21]. Both the model pruning and parameter quantization techniques can significantly reduce model size and increase inference efficiency with an acceptable drop in performance .

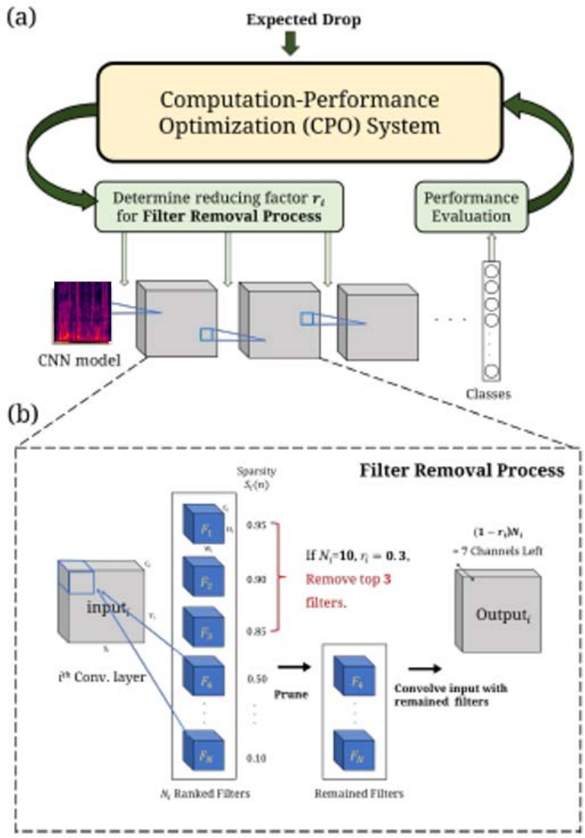


Fig. 4. Model compression based on the CPO algorithm [20].

5. **<u>Speech Enhancement with Direct Evaluation Metric Optimization</u>:** During the training process for an SE model, an objective function is used to optimize the model parameters. In the existing literature, there is an inconsistency between the model optimization criterion and the evaluation criterion for enhanced speech. For example, in the measurement of intelligibility, most of the evaluation metrics are based on short-time objective intelligibility (STOI) measure, while a frame based mean square error (MSE) between the enhanced speech and clean reference is widely used in the process of optimizing the model. Due to this inconsistency, there is no guarantee that the trained model will facilitate optimal performance in different applications [13]. We therefore investigated several algorithms with the aim of directly optimizing model parameters based on evaluation metrics including STOI [13], perceptual evaluation of speech quality (PESQ) [14], and automatic speech recognition (ASR) [22]. Reinforcement learning and GAN-based methods have also been exploited to facilitate optimization given that some evaluation metrics are complex and not differentiable. Experimental results show that by using the same specific evaluation metric as the objective function, the SE model can be trained to yield superior performance compared to the MSE to achieve a desired outcome.

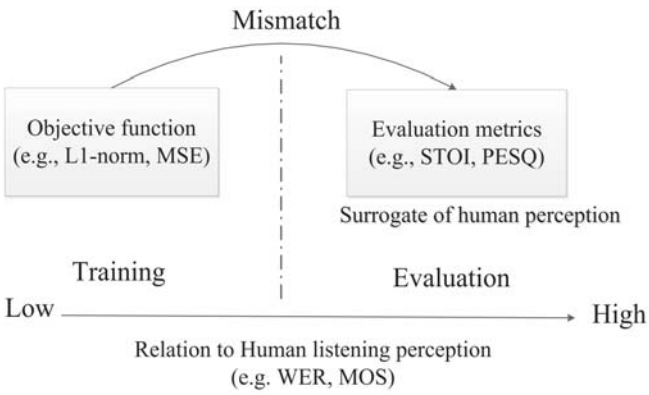


Fig. 5. Mismatch between training objective function and evaluation metrics [13].

6. **Multimodal Pathological Voice Classification:** We have conducted research on pathological voice classification based on medical records [23] and voice signals [24]. The results showed that voice disorders can be accurately identified using voice signals and medical records when advanced signal processing and machine learning methods are utilized. Based on the voice data, we organized a pathological Voice Detection Challenge in IEEE Big Data 2018 [25], which attracted 109 participating teams from 27 different countries. More recently, we investigated the combination of acoustic signals and medical records and derived a multimodal deep learning model. The proposed model consists of two stages: the first stage processes acoustic features and medical data individually and the second stage integrates the outputs from the first stage to perform classification. The proposed multimodal deep learning frameworks were evaluated using 589 samples collected from Far Eastern Memorial Hospital, consisting of three categories of vocal disease, i.e. glottic neoplasm, phonotraumatic lesions, and vocal paralysis. We obtained promising experimental results compared to systems that use only acoustic signals or medical records .

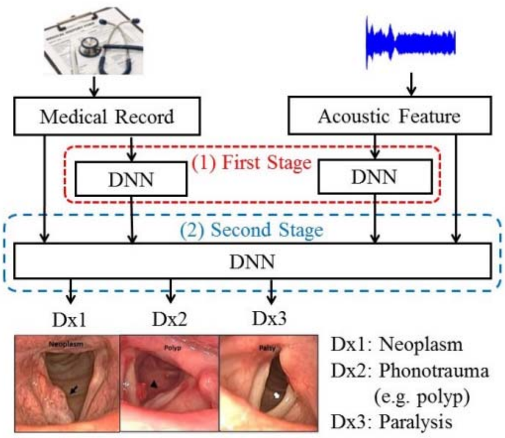


Fig. 6. Pathological voice classification based on voice signals and medical records.

Related Publications:

[1] A. Chern, Y.-H. Lai, Y.-p. Chang, **Y. Tsao**, R. Y. Chang, and H.-W. Chang, "A Smartphone-Based Multi-Functional Hearing Assistive System to Facilitate Speech Recognition in the Classroom," IEEE Access, vol. 5, pp. 10339-10351, 2017 (This paper has been selected as a Featured Article in IEEE Access).

[2] Y.-C. Lin, Y.-H. Lai, H.-W. Chang, **Y. Tsao**, Y.-p. Chang, and R. Y. Chang, "A Smartphone-Based Remote Microphone Hearing Assistive System Using Wireless Technologies," IEEE Systems Journal, vol. 12(1), pp. 20-29, 2018.

[3] Y.-T. Liu, R. Y. Chang, **Y. Tsao**, and Y.-p. Chang, "A New Frequency Lowering Technique for Mandarin-speaking Hearing Aid Users," in Proc. GlobalSIP 2015.

[4] Y.-T. Liu, **Y. Tsao**, and R. Y. Chang, "Nonnegative Matrix Factorization-based Frequency Lowering Technology for Mandarin-speaking Hearing Aid Users," in. Proc. ICASSP 2016.

[5] Y.-H. Lai, **Y. Tsao**, X. Lu, F. Chen, Y.-T. Su, K.-C. Chen, Y.-H. Chen, L.-C. Chen, P.-H. Li, and C.-H. Lee, "Deep Learning based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients," Ear and Hearing, vol. 39(4), pp. 795-809, 2018.

[6] Y.-H. Lai, F. Chen, S.-S. Wang, X. Lu, **Y. Tsao**, and C.-H. Lee, "A Deep Denoising Autoencoder Approach to Improving the Intelligibility of Vocoded Speech in Cochlear Implant Simulation," IEEE Transactions on Biomedical Engineering, vol. 64(7), pp. 1568-1578, 2017.

[7] Y.-H. Lai, **Y. Tsao**, and F. Chen, "Effects of Adaptation Rate and Noise Suppression on the Intelligibility of Compressed-Envelope Based Speech," PLoS ONE, vol. 10.1371, journal.pone.0133519, 2015.

[8] S.-W. Fu, P.-C. Li, Y.-H. Lai, C.-C. Yang, L.-C. Hsieh, and **Y. Tsao**, "Joint Dictionary Learning-based Non-Negative Matrix Factorization for Voice Conversion to Improve Speech Intelligibility After Oral Surgery," IEEE Transactions on Biomedical Engineering, vol. 64 (11), pp. 2584-2594, 2016.

[9] L.-W. Chen, H.-Y. Lee, and **Y. Tsao**, "Generative Adversarial Networks for Unpaired Voice Transformation on Impaired Speech," http://arxiv.org/abs/1810.12656

[10] X. Lu, **Y. Tsao**, S. Matsuda, and C. Hori, C, “Speech Enhancement Based on Deep Denoising Autoencoder,” in Proc. Interspeech 2013.
[11] W.-J. Lee, S.-S. Wang, F. Chen, X. Lu, S.-Y. Chien, and **Y. Tsao**, “Speech Dereverberation Based on Integrated Deep and Ensemble Learning Algorithm,” in Proc. ICASSP 2018.
[12] H.-P. Liu, **Y. Tsao**, Y., and C.-S. Fuh, “Bone-Conducted Speech Enhancement Using Deep Denoising Autoencoder,” Speech Communication, vol. 104, pp. 106-112, 2018.
[13] S.-W. Fu, T.-W. Wang, **Y. Tsao**, X. Lu, and H. Kawai, “End-to-End Waveform Utterance Enhancement for Direct Evaluation Metrics Optimization by Fully Convolutional Neural Networks,” IEEE Transactions on Audio, Speech and Language Processing, vol. 26(9), pp. 1570-1584, 2018.
[14] S.-W. Fu, **Y. Tsao**, H.-T. Hwang, and H.-M. Wang, “Quality-Net: An End-to-End Non-intrusive Speech Quality Assessment Model based on BLSTM,” in Proc. Interspeech 2018.
[15] X, Lu, **Y. Tsao**, S, Matsuda and C, Hori, “Ensemble Modeling of Denoising Autoencoder for Speech Spectrum Restoration,” pp. 885-889, in Proc. Interspeech 2014.
[16] C.-F. Liao, **Y. Tsao**, H.-y. Lee, and H.-M. Wang, “Noise Adaptive Speech Enhancement using Domain Adversarial Training,” arXiv preprint arXiv:1807.07501, 2018.
[17] W.-C. Chen, P.-T. Lai, **Y. Tsao**, and C.-C. Lee, “Multimodal Arousal Rating using Unsupervised Fusion Technique,” in Proc. ICASSP 2015.
[18] S.-W. Hsiao, H.-C. Sun, M.-C. Hsieh, M.-H. Tsai, **Y. Tsao**, and C.-C. Lee, “Toward Automating Oral Presentation Scoring during Principal Certification Program using Audio-Video Low-level Behavior Profiles,” IEEE Transactions on Affective Computing, in press.
[19] J.-C. Hou, S.-S. Wang, Y.-H. Lai, **Y. Tsao**, H.-W. Chang, and H.-M. Wang, “Audio-visual Speech Enhancement using Multimodal Deep Convolutional Neural Networks,” IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2(2), pp. 117-128, 2018.
[20] C.-T. Liu, T.-W. Lin, Y.-H. Wu, Y.-S. Lin, H. Lee, **Y. Tsao**, and S.-Y. Chien, “Computation-Performance Optimization of Convolutional Neural Networks with Redundant Filter Removal,” IEEE Transactions on Circuits and Systems I 2018.
[21] Y.-T. Hsu, Y.-C. Lin, S.-W. Fu, **Y. Tsao**, and T.-W. Kuo, “A Study on Speech Enhancement using Exponent-only Floating Point Quantized Neural Network (EOFP-QNN),” in Proc. SLT 2018.
[22] Y.-L. Shen, C.-Y. Huang, S.-S. Wang, **Y. Tsao**, H.-M. Wang, and T.-S. Chi, “Reinforcement Learning Based Speech Enhancement for Robust Speech Recognition,” to appear in ICASSP 2019.
[23] S.-Y. Tsui, **Y. Tsao**, C.-W. Lin, S.-H. Fang, and C.-T. Wang, “Demographic and Symptomatic Features of Voice Disorders and Their Potential Application in Classification using Machine Learning Algorithms,” Folia Phoniatrica et Logopaedica 2018.
[24] C.-T. Wang, F.-C. Lin, J.-Y. Chen, M.-J. Hsiao, S.-H. Fang, Y.-H. Lai, and **Y. Tsao**, “Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach,” Journal of Voice, 2018.
[25] https://femh-challenge2018.weebly.com/

Image sources:

<1> https://www.citi.sinica.edu.tw/papers/yu.tsao/6047-F.pdf

<2> https://www.dizziness-and-balance.com/disorders/hearing/hearing-aids/hearing_aid.html

<3> https://www.hearingchoices.com.au/report/cochlear-implants-and-hearing-aids/

<4> https://www.citi.sinica.edu.tw/papers/yu.tsao/4889-F.pdf

Sharing from DL: Prof. Xiao-Lei Zhang

Center for Intelligent Acoustics and Immersive Communications,
School of Marine Science and Technology,
Northwestern Polytechnical University, Xi'an, China.

Title: Multilayer Bootstrap Networks

Section 1: Introduction

Principal component analysis (PCA) is a simple and widely used unsupervised dimensionality reduction method, which finds a coordinate system in the original Euclidean space that the linearly uncorrelated coordinate axes (called principal components) describe the most variances of data. Because PCA is insufficient to capture highly-nonlinear data distributions, many nonlinear dimensionality reduction methods are explored.

Nonlinear dimensionality reduction methods can be categorized roughly to three classes. The first class is kernel methods. It first projects data to a kernel-induced feature space, and then conducts PCA or its variants in the new space. Examples include kernel PCA, Isomap, locally linear embedding (LLE), Laplacian eigenmaps, etc.. The second class is probabilistic models. It assumes that data are generated from an underlying probability function, and takes the posterior parameters as the feature representation, such as latent Dirichlet allocation. The third class is autoassociative neural networks. It learns a piecewise-linear coordinate system explicitly by backpropagation, and uses the output of the bottleneck layer as the new representation.

Although ensemble learning, which was triggered by random resampling, is a large family of supervised machine learning, it is not very prevalent in unsupervised dimensionality reduction. Recently, we have investigated deeply into this direction, and proposed multilayer bootstrap networks (MBN) which is a very simple deep ensemble learning method. We present the algorithm in detail as follows:

Section 2: Algorithm

Section 2.1: Network structure

MBN contains multiple hidden layers and an output layer (Fig. 1). Each hidden layer consists of a group of mutually independent *k*-centroids clusterings; each *k*-centroids clustering has *k* output units, each of which indicates one cluster; the output units of all *k*-centroids clusterings are concatenated as the input of their upper layer. The output layer is PCA.

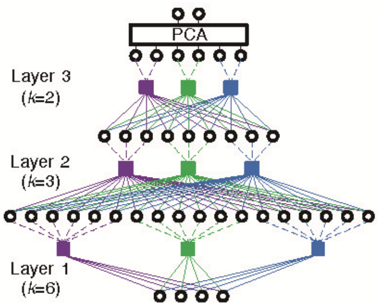


Figure 1 Network structure. The dimension of the input data for this demo network is 4. Each colored square represents a *k*-centroids clustering. Each layer contains 3 clusterings. Parameters *k* at layers 1, 2, and 3 are set to 6, 3, and 2 respectively. The outputs of all clusterings in a layer are concatenated as the input of their upper layer.

The network is gradually narrowed from bottom up, which is implemented by setting parameter k as large as possible at the bottom layer and be smaller and smaller along with the increase of the number of layers until a predefined smallest k is reached.

Section 2.2: Optimization algorithm

MBN is trained layer-by-layer from bottom up. For training each layer given a d-dimensional input data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ either from the lower layer or from the original data space, we simply need to focus on training each k-centroids clustering, which consists of the following steps:

- **Random sampling of features.** The first step randomly selects $\hat{d}$ dimensions of $\boldsymbol{X}$ ($\hat{d} \leq d$) to form a subset of $\boldsymbol{X}$, denoted as $\widehat{\boldsymbol{X}} = \{\widehat{\boldsymbol{x}}_1, \dots, \widehat{\boldsymbol{x}}_n\}$.
- **Random sampling of data.** The second step randomly selects k data points from $\widehat{\boldsymbol{X}}$ as the k centroids of the clustering, denoted as $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_k\}$.
- **One-nearest-neighbor learning**. The new representation of an input $\widehat{\boldsymbol{x}}$ produced by the current clustering is an indicator vector $\boldsymbol{h}$ which indicates the nearest centroid of $\widehat{\boldsymbol{x}}$. For example, if the second centroid is the nearest one to $\widehat{\boldsymbol{x}}$, then $\boldsymbol{h} = [0,1,0, \dots, 0]^T$. The similarity metric between the centroids and $\widehat{\boldsymbol{x}}$ at the bottom layer is customized, e.g. the squared Euclidean distance $\operatorname{argmin}_{i=1}^{k} ||\boldsymbol{w}_i - \widehat{\boldsymbol{x}}||^2$, and set to $\operatorname{argmax}_{i=1}^{k} \boldsymbol{w}_i^T \widehat{\boldsymbol{x}}$ at all other hidden layers.

Section 2.3: Fundamentals

MBN builds a gradually narrowed multilayer nonlinear network from bottom up for unsupervised nonlinear dimensionality reduction. Each layer of the network is a nonparametric density estimator. It consists of a group of k-centroids clusterings. Each clustering randomly selects data points with randomly selected features as its centroids, and learns a one-hot encoder by one-nearest-neighbor optimization.

Geometrically, the nonparametric density estimator at each layer projects the input data space to a uniformly-distributed discrete feature space, where the similarity of two data points in the discrete feature space is measured by the number of the nearest centroids they share in common. The multilayer network gradually reduces the nonlinear variations of data from bottom up by building a vast number of hierarchical trees implicitly on the original data space. Theoretically, the estimation error caused by the nonparametric density estimator is proportional to the correlation between the clusterings, both of which are reduced by the randomization steps. For the detailed explanation, please refer to [1].

Section 3: Results

The data set of the MNIST digits contains 10 handwritten integer digits ranging from 0 to 9. It consists of 60,000 training images and 10,000 test images. Each image has 784 dimensions. It is a highly nonlinear and non-Gaussian data set, which is well-known for the machine learning community. We use MBN to reduce MNIST to 2-dimensional subspace, where MBN adopts the default parameter setting without tuning. For details about the parameter setting, please refer to [1]. The visualization result is shown in Fig. 2. After applying k-means clustering to a 10-dimensional subspace of MNIST, the clustering accuracy of MBN on MNIST reaches 96.64%. If we fine-tune the hyperparameters, the result may be further improved.

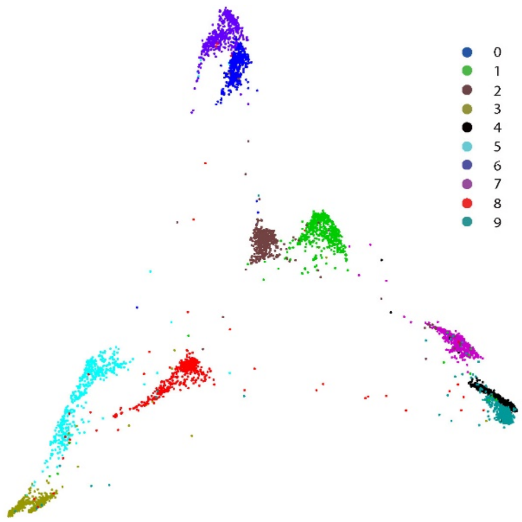


Figure 2. Visualization of MNIST produced by MBN. For clarity, only 500 images per digit are drawn.

Section 4: Conclusions

MBN is composed of two novel components: (i) each layer of MBN is a nonparametric density estimator by random resampling. It estimates the density of data correctly without any model assumption. It is exponentially more powerful than a single *k*-centroids clustering. Its estimation error is proven to be small and controllable. (ii) The network is a deep ensemble model. It essentially reduces the nonlinear variations of data by building a vast number of hierarchical trees on the data space. It can be trained as many layers as needed with both large-scale and small-scale data.

To our knowledge, MBN is the simplest nonlinear dimensionality reduction methods in mathematics, which contains only one equation besides the output layer: finding the one-nearest neighbor. It is a deep learning method beyond neural networks. It is also an unsupervised ensemble learning method for dimensionality reduction that reaches comparable performance to the representative algorithms of kernel methods, probabilistic models, and neural networks. It investigated deeply the research direction of ensemble learning which is not so prevalent compared to the other three classes of methods.

MBN performs robustly with a wide range of parameter settings. Its time and storage complexities scale linearly with the size of training data. It supports parallel computing naturally. Empirical results demonstrate its efficiency at the training stage and its effectiveness in a number of applications including density estimation, data visualization, clustering, and document retrieval.

References

[1] Xiao-Lei Zhang. Multilayer bootstrap networks. *Neural Networks*, volume 103, pages 29-43, 2018.

Call for Papers: Special Issue for APSIPA Transactions on Signal and Information Processing (ATSIP)

Topic of the special issue: **Hyperspectral and Multiview Image Processing**

Brief description:

In recent years, hyperspectral imaging has experienced evolutional advances and shown its great potential in wide applications due to its high spectral recognition and new techniques in computer vision and image processing. Indeed, many progresses have been made in multiview image processing (that with multi-cameras or video-cameras) and hyperspectral image processing. There is an increasing trend to combine both hyperspectral images and multiview images for more complex tasks such as target detection, recognition and classification where both geometric and attribute information are required. Thus, the special issue on hyperspectral and multiview image processing is proposed for APSIPA Transactions on Signal and Information Processing.

The topics of this special issue include:

- Hyperspectral image processing (enhancement, detection, classification, unmixing),
- Multiview image processing (3D scanning and acquisition, depth estimation, 3D reconstruction),
- Spatial and spectral image processing (registration, fusion, understanding, etc.) and related contents.

Editor in Chief APSIPA T-SIP

Tatsuya Kawahara, Professor, School of Informatics, Kyoto University, Japan

Editors of the special issue:

Mingyi HE, myhe@nwpu.edu.cn, Northwestern Polytechnical University, China
Qian DU, du@ece.msstate.edu, Mississippi State University, USA
Hongdong LI, hongdong.li@anu.edu.au, Australia National University, Australia

Tentative schedule of submission and publication:

Submission deadline:	April 30, 2019
Publication date:	October 30, 2019

Papers are published upon acceptance, regardless of the Special Issue publication date.

Please refer to the journal website for paper submission procedure.
http://journals.cambridge.org/sip

Latest Articles from APSIPA Transactions on Signal and Information Processing (ATSIP)

- **Neutral-to-emotional voice conversion with cross-wavelet transform F0 using generative adversarial networks**
  - Zhaojie Luo, Jinhui Chen, Tetsuya Takiguchi, Yasuo Ariki
  - DOI: https://doi.org/10.1017/ATSIP.2019.3
  - Published online: 04 March 2019, e10

- **Checkerboard artifacts free convolutional neural networks**
  - Yusuke Sugawara, Sayaka Shiota, Hitoshi Kiya
  - DOI: https://doi.org/10.1017/ATSIP.2019.2
  - Published online: 19 February 2019, e9

- **Learning from past mistakes: improving automatic speech recognition output via noisy-clean phrase context modeling**
  - Prashanth Gurunath Shivakumar, Haoqi Li, Kevin Knight, Panayiotis Georgiou
  - DOI: https://doi.org/10.1017/ATSIP.2018.31
  - Published online: 01 February 2019, e8

- **Grayscale-based block scrambling image encryption using YCbCr color space for encryption-then-compression systems**
  - Warit Sirichotedumrong, Hitoshi Kiya
  - DOI: https://doi.org/10.1017/ATSIP.2018.33
  - Published online: 01 February 2019, e7

- **Analysis and generation of laughter motions, and evaluation in an android robot**
  - Carlos Toshinori Ishi, Takashi Minato, Hiroshi Ishiguro
  - DOI: https://doi.org/10.1017/ATSIP.2018.32
  - Published online: 25 January 2019, e6

Most Read Articles from ATSIP

https://www.cambridge.org/core/journals/apsipa-transactions-on-signal-and-information-processing/most-read

- **A tutorial survey of architectures, algorithms and applications for deep learning**
  - Li Deng
  - DOI: https://doi.org/10.1017/atsip.2013.9
  - Published online: 22 January 2014, e2

- **Dimensionality reduction of visual features for efficient retrieval and classification**
  - Petros T. Boufounos, Hassan Mansour, Shantanu Rane, Anthony Vetro
  - DOI: https://doi.org/10.1017/ATSIP.2016.14
  - Published online: 12 July 2016, e14

- **Use cases and challenges in telecom big data analytics**
  - Chung-Min Chen
  - DOI: https://doi.org/10.1017/ATSIP.2016.20
  - Published online: 12 December 2016, e19

- **The artificial intelligence renaissance: deep learning and the road to human-level machine intelligence**
  - Kar-Han Tan, Boon Pang Lim
  - DOI: https://doi.org/10.1017/ATSIP.2018.6
  - Published online: 23 July 2019, e6

- **Unified access in licensed and unlicensed bands in LTE-A Pro and 5G**
  - Boon Loong Ng, Hongbo Si, Aris Papasakellariou, Jianzhong Charlie Zhang
  - DOI: https://doi.org/10.1017/ATSIP.2017.7
  - Published online: 12 July 2017, e6

Most Cited Articles from ATSIP

https://www.cambridge.org/core/journals/apsipa-transactions-on-signal-and-information-processing/most-cited

- **A tutorial survey of architectures, algorithms and applications for deep learning**
  - Li Deng
  - DOI: https://doi.org/10.1017/atsip.2013.9
  - Published online: 22 January 2014, e2

- **An overview on video forensics**
  - Simone Milani, Marco Fontani, Paolo Bestagini, Mauro Barni, Alessandro Piva, Marco Tagliasacchi, Stefano Tubaro
  - DOI: https://doi.org/10.1017/ATSIP.2012.2
  - Published online: 28 August 2012, e2

- **Recent advances on active noise control: open issues and innovative applications**
  - Yoshinobu Kajikawa, Woon-Seng Gan, Sen M. Kuo
  - DOI: https://doi.org/10.1017/ATSIP.2012.4
  - Published online: 28 August 2012, e3

- **Survey on audiovisual emotion recognition: databases, features, and data fusion strategies**
  - Chung-Hsien Wu, Jen-Chun Lin, Wen-Li Wei
  - DOI: https://doi.org/10.1017/ATSIP.2014.11
  - Published online: 11 November 2014, e12

- **Survey on securing data storage in the cloud**
  - Chun-Ting Huang, Lei Huang, Zhongyuan Qin, Hang Yuan, Lan Zhou, Vijay Varadharajan, C.-C. Jay Kuo
  - DOI: https://doi.org/10.1017/ATSIP.2014.6
  - Published online: 23 May 2014, e7

APSIPA Membership

Membership Benefits:

- Links to highly qualified people within the organization to develop research, technology, teaching, and career
- Discount fee on APSIPA conferences
- Reduced subscription fee for APSIPA journals
- Access to information about the international activities in signal and information processing such as conferences, continuing education, short courses, seminars, distinguished lecture series, student internships, scholarships, job listings, publication venues, and mentorships

To motivate APSIPA members to participate in APSIPA conferences, the registration for the 11th APSIPA conference implies an automatic renewal of APSIPA membership up to the end of December 2020.

You may join as:

- Student Membership: Student members are those who are enrolled full time in universities, institutes, or any accredited degree
- Full Membership: Full members are individuals interested in being part of the APSIPA mission to excel signal and information processing field. They are eligible to vote, hold positions in APSIPA association, and contribute to serve as editorial board and program committee members in APSIPA journals and conferences
- **Life Membership**: Full members may choose to subscribe as life members pending on paying the discount fee of life membership. Early-bird registration fee is available for life members at all times when registering for APSIPA ASC

Type of membership	Fees in US$	Fees in HK$
Student Membership	10 (per annual)	78 (per annual)
Full Membership	30 (per annual)	234 (per annual)
Life Membership	300 (a one-off fee)	2340 (a one-off fee)

Act Today! Join us at: http://www.apsipa.org/reg.asp

Summary of Links

- APSIPA ASC 2019: http://www.apsipa2019.org
- APSIPA Transaction on Signal and Information Processing: http://journals.cambridge.org/sip
- Paper Submission to APSIPA Transaction on Signal and Information Processing: http://mc.manuscriptcentral.com/apsipa
- APSIPA Industrial Activities: http://www.apsipa.org/industrial.htm
- APSIPA Friend's Lab: http://www.apsipa.org/friendlab/FriendLabs.htm.
- APSIPA Membership Registration/Renewal: http://www.apsipa.org/reg.asp

APSIPA Who's Who

President: Hitoshi Kiya, Tokyo Metropolitan University, Japan

President-Elect: Anthony Kuh, University of Hawaii at Manoa, USA

Past Presidents: Sadaoki Furui (2009-2012), C.C. Jay Kuo (2013-2014), Haizhou Li (2015-2016)

Immediate Past President: Wan-Chi Siu, The Hong Kong Polytechnic University, HK

VP - Conferences: Thomas Fang Zheng, Tsinghua University, China

VP - Industrial Relations and Development: Guan-Ming Su, Dolby Labs, Sunnyvale, CA, USA

VP - Institutional Relations and Education Program: Woon-Seng Gan, Nanyang Technological University, Singapore

VP - Member Relations and Development: Yoshinobu Kajikawa, Kansai University, Japan

VP - Publications: Kenneth Lam, The Hong Kong Polytechnic University, HK

VP - Technical Activities: Chung-Nan Lee, National Sun Yat-sen University, Taiwan

Members-at-Large:

Kosin Chamnongthai, King Mongkut's University of Technology Thonburi, Thailand

Nam Ik Cho, Seoul National University, Korea

Mingyi He, Northwestern Polytechnical University, China

Yo-Sung Ho, Gwangju Institute of Science and Technology, Korea

Tatsuya Kawahara, Kyoto University, Japan

Bonnie Law, The Hong Kong Polytechnic University, HK

Kai-Kuang Ma, Nanyang Technological University, Singapore

Susanto Rahardja, Northwestern Polytechnical University, China

Hiroshi Saruwatari, University of Tokyo, Japan

Toshihisa Tanaka, Tokyo University of Agriculture and Technology, Japan

Chung-Hsien Wu, National Cheng Kung University, Taiwan

Min Wu, University of Maryland, USA

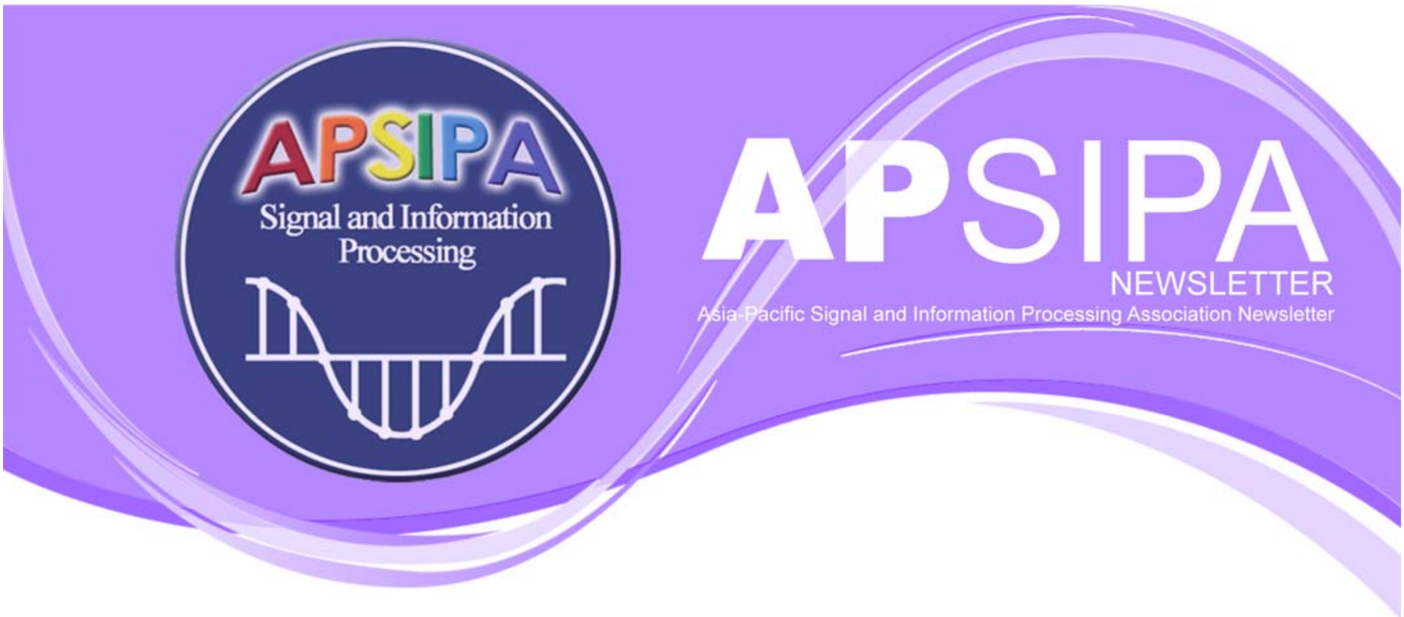
Headquarters

Address:

Asia Pacific Signal and Information Processing Association,
Centre for Signal Processing,
Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong.

Officers:

Director: Wan-Chi Siu, email: enwcsiu@polyu.edu.hk
Manager: Kin-Man Lam, Kenneth,
email: enkmlam@polyu.edu.hk
Secretary: Ngai-Fong Law, Bonnie,
email: ennflaw@polyu.edu.hk
Treasurer: Yuk-Hee Chan, Chris,
email: enyhchan@polyu.edu.hk



- **APSIPA Newsletter Editorial Board Members**

Are you an APSIPA member?
If not, then register online at
http://www.apsipa.org