Issue 12
July 2016

APSIPA NEWSLETTER

Asia-Pacific Signal and Information Processing Association Newsletter

Message from Editor-in-Chief

Greetings!

There are six Technical Committees (TCs) in APSIPA. They are,

- Signal Processing Systems: Design and Implementation (SPS)
- Signal and Information Processing Theory and Methods (SIPTM)
- Speech, Language, and Audio (SLA)
- Biomedical Signal Processing and Systems (BioSiPS)
- Image, Video, and Multimedia (IVM)
- Wireless Communications and Networking (WCN)

To allow members to know more about the TCs, we have invited the TC chairs to give some information about their TCs in the newsletters. In this issue, we are happy to have contributions from Professor Gwo Giun Lee who is the TC chair of SPS (Signal Processing Systems: Design and Implementation). We hope that members can know more about SPS technical committee and its activities. In the new few issues, other TC chairs will be invited to share with you their activities.

Besides technical committees, APSIPA has various activities for its members, such as Friend Labs, APSIPA Annual Submit and Conference (ASC), Distinguished Lecturer Program and Industrial Distinguished Leader Program. A summary of activities about the Friend Labs for the past three months is provided in this issue.

Contributions from members are always highly appreciated. In this issue, we are happy to have contributions from Dr Jianjun He and Professor Woon-Seng Gan about their work on "Spatial Audio Reproduction using Primary Ambient Extraction". All members are invited to send us your contributions to publish in APSIPA Newsletter. Don't forget to give us thoughts about the further development of APSIPA Newsletter. Enjoy reading this issue!

Bonnie Law

APSIPA Newsletter EiC

In this issue

Message from TC Chair of SPS (Signal Processing Systems: Design and Implementation)

Professor Gwo Giun Lee
National Cheng-Kung University, Taiwan
Email: clee@mail.ncku.edu.tw

In the 1960s, Marshall McLuhan published the book entitled "The Extensions of Man", focusing primarily on television, an electronic media as being the outward extension of human nervous system, which from contemporary interpretation marks the previous stage of Big Data. Based upon mathematical fundamentals as foundations for complexity-aware algorithms, intelligent, flexible, and efficient, analytics architectures including both software and hardware for GPU, multicore, high performance computing, and distributed systems, the Signal Processing Systems Technical Committee envisions yet even further inward extension of human information perception experiences into genomic, proteomic, and other physiological and biomedical phenomena and exchange of information, in expediting the field of signal and information processing systems into futuristic new era of the Internet-of-Things over Cloud and Fog.

About the TC Chair

Gwo Giun (Chris) Lee (S'91- M'97-SM'07) received his M.S. and Ph.D. degrees in Electrical Engineering from University of Massachusetts. Dr. Lee has held several technical and managerial positions in the industry including System Architect in former Philips Semiconductors, USA; DSP Architect in Micrel Semiconductors, USA; and Director of Quanta Research Institute, Taiwan before joining the faculty team of the Department of Electrical Engineering in National Cheng Kung University (NCKU) in Tainan, Taiwan. Dr. Lee has authored more than 100 technical documents. Dr. Lee also serves as the Associate Editor for IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) and Journal of Signal Processing Systems. He received the Best Paper Award for the BioCAS track in IEEE ISCAS 2012. His research interests include Algorithm/ Architecture Co-design for System-on-a-Chip, Visual Signal Processing and Communication, Internet of Things, Biomedical Image Processing, Bioinformatics.

Information about Signal Processing Systems: Design and Implementation (SPS) Technical Committee

Chair:	Prof Gwo Giun Lee, National Cheng-Kung University, Taiwan
Purpose:	To promote advancement and exchange of the research fields of design and implementation related to signal processing systems in the Asia-Pacific region
Field of Interest:	The fields of interest of the SPS TC shall be the following:

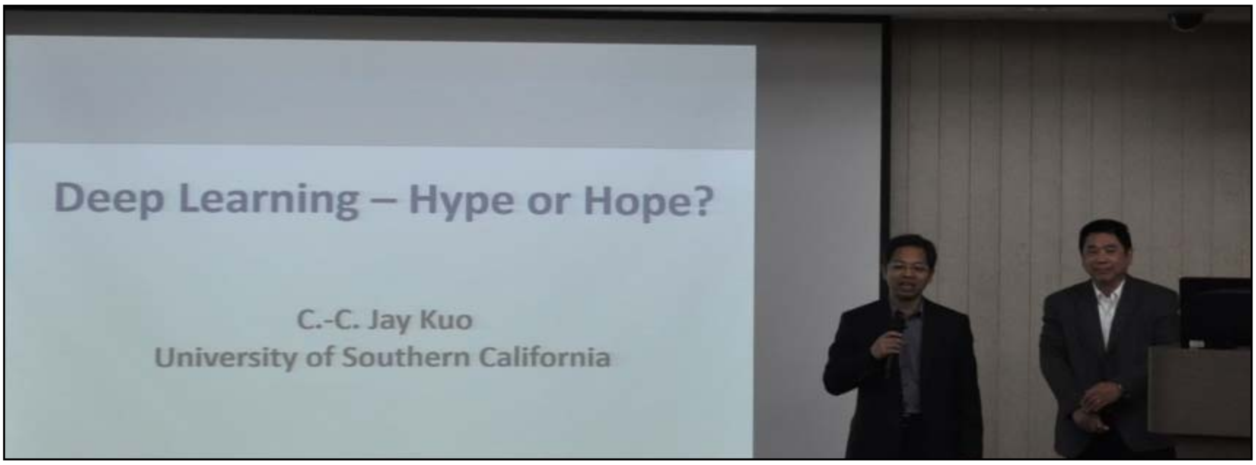
- Analog/Digital LSI and Systems
- Hardware/Software Systems
- Image/Video Processing and Coding Systems
- Multimedia Systems
- Computer Vision/Graphics Systems
- Speech/Audio Processing Systems
- VLSI communication Systems
- Design Methodologies and CAD Tools for Signal Processing Systems

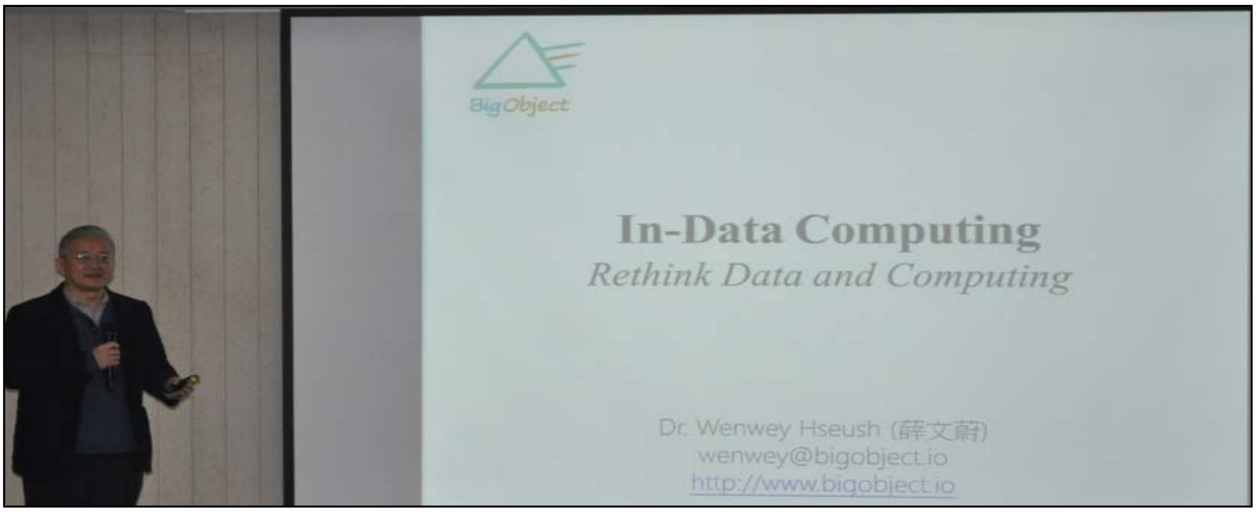
Recent Activities: “Big Data and Deep Learning”

On Mar. 14 2016, Dr. Wen-Huang Cheng from Research Center for Information Technology Innovation (CITI) of Academia Sinica Taiwan organized an activity for APSIPA members in Taiwan, “Workshop for Big Data Analysis and Reunion for APSIPA Friend Labs.“ The invited speakers are Prof. C. –C. Jay Kuo from University of Southern California (USC) and Dr. Wenway Hseush from BigObjectR.

The topic of the first talk is “**Deep learning: Hype and Hope**” by Prof. Kuo. Recent achievements in speech recognition and computer vision through deep learning algorithms are addressed. The pros, cons and future perspectives of the deep learning methodology are discussed. The topic of the second talk is “In-Data Computing, Rethink Data and Computing”. Dr. Hseush introduced a new technology, called “In-Data computing”, which is a set of principle for storing and computing big data. The workshop, held right after the breaking news that AlphaGo won the first three games against legendary Go player, thus attracts much attention and shows its importance.

In the afternoon, a Q&A session is arranged for all the attendees and two invited speakers. The attendees from academia, industry and commercial company raised several questions from different points of views regarding the big data and deep learning. As suggested by the speakers: a bridge or a cooperation is required that closes the gap between the senior’s professional experiences and the junior’s creative imagination, between the industry who owns the source of big data and the academia who exploits the possibilities.





Friend Labs Activities

Friend Labs is one of the recent activities of APSIPA to provide education, research and development exchange platforms for both academia and industry. All APSIPA Friend Labs are listed in the APSIPA website. Each lab has one page to post lab information, photos and a link to the lab home page. An academia or industrial lab is qualified to become an APSIPA Friend Lab if it has at least 10 current or former lab members who are full or associate members of APSIPA. For more information, please visit http://www.apsipa.org/friendlab/FriendLabs.htm.

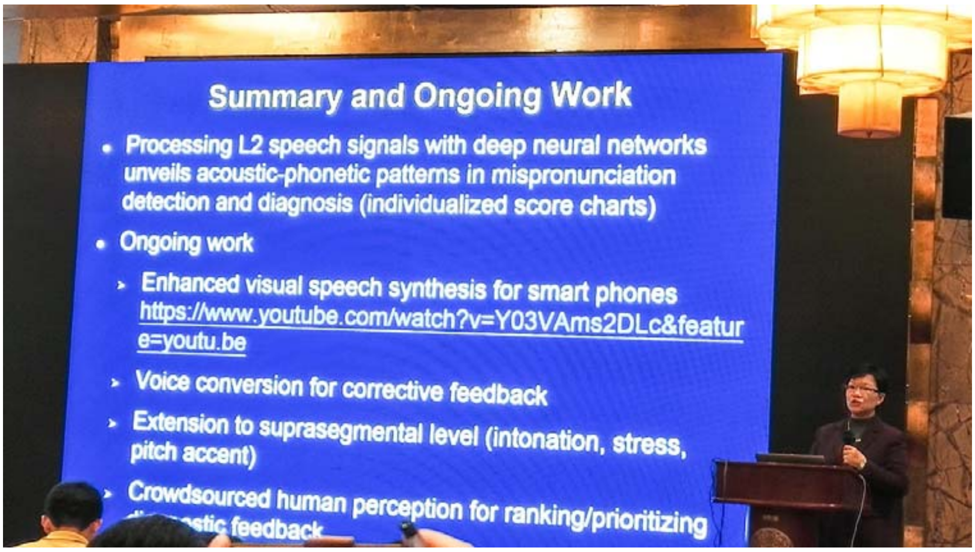
- **2016 Reunion of APSIPA Friend Labs in Taiwan**

On March 14, 2016, a reunion of APSIPA Friend Labs was held at Academia Sinica in Taiwan. About 50 participants met and exchanged experiences. It was a very successful event. In the future, there will be more exchange activities held among the APSIPA Friend Labs.

- **SIDAS 2016**

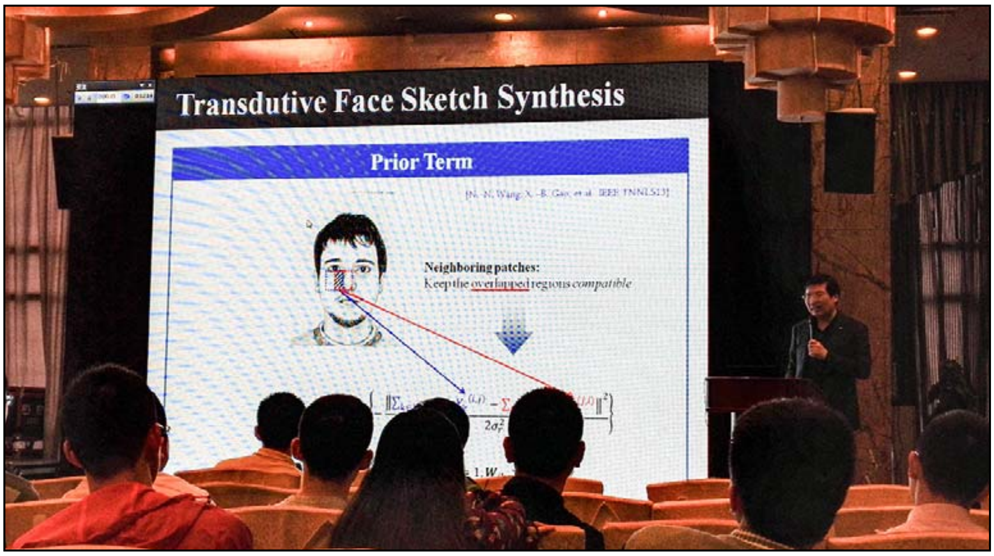
IEEE Signal and Data Science Forum (SIDAS 2016) was held on 24-26 April 2016 in Wuhan, China under the theme of "Signal and Data Science - Powering Our Digital Life". SIDAS 2016 is co-located with VALSE and co-sponsored by IEEE Signal Processing Society, APSIPA, VALSE and other institutions.

SIDAS 2016 provided a platform for the exchange of the latest developments in the field of signal processing. The Forum included two thematic reports, three forums as well as 29 invited presentations. More details can be found at: http://mclab.eic.hust.edu.cn/sidas2016/program.html.



Professor Helen Meng from the Chinese University Hong Kong

Dr Tao Mei from the Microsoft Research – Asia



Professor Gao Xinbo from Xidian University

- **APSIPA BioSiPS Workshop 2016: July 14, 2016**

The Fourth APSIPA Workshop on the Frontier in Biomedical Signal Processing and Systems (BioSiPS 2016) in conjunction with APSIPA Friend Lab Meeting in Tokyo and TUAT Institute for Global Innovation Research Open Seminar will be held on July 14, 2016 in Tokyo, Japan. This joint workshop is organized in order to promote and exchange recent advances of biomedical engineering and signal processing research among technical committee members and members in APSIPA Friend Labs, and provide opportunities of meeting and getting together among young researchers and students. The first BioSiPS workshop was held at Mahidol University in Bangkok, Thailand in March 2013, the second was held at University of Rajshahi in Bangladesh in March 2014, and the third was held in Shanghai, China in June 2014.

Co-sponsors

- APSIPA Technical Committee on Biomedical Signal Processing and Systems
- APSIPA Friend Labs
- KAKENHI Multidisciplinary Computational Anatomy Project
- TUAT Institute of Global Innovation Research (Prof. Ortega's lecture)

Workshop Agenda

- Date: July 14, 2016 (afternoon)
- Venue: Tokyo University of Agriculture and Technology (Koganei Campus)
- Tentative program:

  1. Poster presentation by young researchers and students
  2. Friend Lab session: Short technical talks by Friend Lab PIs and members
  3. Plenary talks and tutorials (tentative speakers and titles)
     - Prof. Akinobu Shimizu
       Tokyo University of Agriculture and Technology
       "Computational Anatomy"
     - [GIR Open Seminar] Prof. Antonio Ortega
       University of Southern California, USA)
       "Graph Signal Processing"
  4. Reception

Committee Members

- General Co-Chairs
  Tomasz M. Rutkowski (University of Tokyo, Japan)
  Jianting Cao (Saitama Institute of Technology, Japan)
- Local Arrangement Co-Chairs
  Toshihisa Tanaka (Tokyo University of Agriculture and Technology, Japan)
  Yuichi Tanaka (Tokyo University of Agriculture and Technology, Japan)

For enquires, please contact

Toshihisa Tanaka (Tokyo University of Agriculture and Technology)
tanakat@cc.tuat.ac.jp

Contribution from Members

The following six pages are about a recent work on "Spatial Audio Reproduction using Primary Ambient Extraction". It is contributed by our members who want to share their works with other APSIPA members.

Spatial Audio Reproduction using Primary Ambient Extraction

Jianjun He and Woon-Seng Gan

Digital Signal Processing Lab (an APSIPA Friends Lab), School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore
(jhe007@e.ntu.edu.sg, ewsgan@ntu.edu.sg)

Abstract: Recreating a natural listening experience is the aim of spatial audio reproduction system. Despite the increasing popularity of object-based audio, majority of the legacy audio contents are in channel-based format, which is dependent on the desired playback system. Considering the diversity of today's playback systems, the quality of reproduced sound scenes degrades significantly when mismatches between the audio content and the playback system occur. To this end, this article introduces an overview of the spatial audio reproduction techniques referred to as primary ambient extraction (PAE) that facilitates the development of an efficient, flexible, and immersive spatial audio reproduction.

1 Background

Sound is an inherent part of our everyday lives for information, communication and interaction. Sound improves the situational awareness by providing feedback for actions and situations that are out of the view of the listener. An advantage of sound is that multiple sound sources can be perceived from any location around the head in the three dimensional (3D) space [1]. The role of natural 3D sound, or spatial sound, is very essential in high stress applications like flight navigation systems [2], in route guidance for visually impaired people [3], and in medical therapy for patients [4]. Last but not least, the ever growing market of consumer electronics calls for spatial audio reproduction for digital media and virtual/augmented reality applications [5]. The reproduction methods generally differ in the formats of audio content. Despite the growing interest in object-based audio formats [5], such as Dolby ATMOS [6], DTS X [7], most existing digital media content is still in channel-based formats (such as stereo and multichannel signals). The channel-based audio is usually specific in its playback configuration, and it does not support flexible playback configurations in domestic or personal listening circumstances [5]. Considering the wide diversity of today's playback systems [8], it becomes necessary to process audio signals to achieve the best quality (especially spatial quality [9]) with the actual playback system. This is also in line with the objective of the new MPEG-H standard for 3D audio [8].

Depending on the actual playback system, the challenges in spatial audio reproduction can be broadly categorized into two main types: loudspeaker playback and headphone playback. The challenge in loudspeaker playback mainly arises from the mismatch of loudspeaker playback systems in home theater applications, where the number of loudspeakers [10] or even the type of loudspeakers [11] between the intended loudspeaker system (based on the audio content) and the actual loudspeaker system is different. Conventional techniques to solve this challenge are often referred to as audio remixing [10], [12], which basically compute the loudspeaker signals as the weighted sums of the input signals. For headphone playback, the challenge arises when the audio content is not tailored for headphone playback. Virtualization is often regarded as the technique to solve this challenge [1], which is achieved by binaural rendering, i.e., convolving the channel-based signals with head-related impulse responses (HRIRs) of the corresponding loudspeaker positions. These conventional techniques are capable of solving the compatibility issue, but limit the spatial quality [13]-[15]. To improve the spatial quality, MPEG Surround and related techniques were proposed to synthesize the multichannel signals using one-channel down-mixed signal and the spatial cues, which better suit the reproduction of the directional signals as compared to the diffuse signals [14], [16].

To further improve the quality of the reproduced sound scene, the perception of the sound scenes can be understood as a combination of the foreground sound and background sound, which are often referred to as primary (or direct) and ambient (or diffuse) components, respectively [17]-[20]. The primary components consist of point-like directional sound sources, whereas the ambient components are made up of diffuse environmental sound, such as the reverberation, applause, or nature sound like waterfall [16], [21]. Due to their perceptual differences, different rendering schemes should be applied to the primary and ambient components for optimal reproduction [16], [22]. However, the channel-based audio provides only the mixed signals, which necessitate the process of primary ambient extraction (PAE).

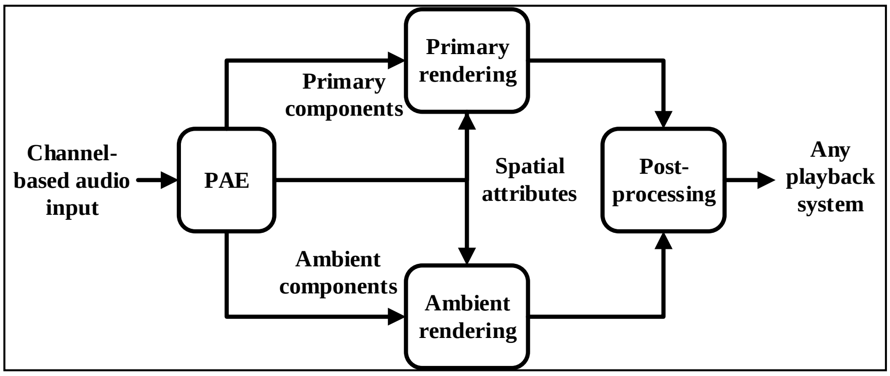


Figure 1 Block diagram of PAE based spatial audio reproduction

As a spatial audio processing tool [10], [13], [16], [14], [19], [22], PAE can also be incorporated into spatial audio coding systems, such as spatial audio scene coding [17], and directional audio coding [23]. Essentially, PAE serves as a front-end to facilitate flexible, efficient, and immersive spatial audio reproduction. First, by decomposing the primary and ambient components of the sound scene, PAE enables the sound reproduction format to be independent of the input format, hence increasing the flexibility of spatial audio reproduction [23]. Second, PAE based reproduction of sound scenes does not require the individual sound objects as in object-based format (which is the most flexible), but is able to recreate perceptually similar sound scenes, hence maintaining the efficiency of spatial audio reproduction [18]. Last but not least, PAE extracts the two key components of the sound scenes, namely, directional and diffuse sound components. These components are highly useful in recreating an immersive listening experience of the sound scene [17], [24]-[27].

Figure 1 illustrates the PAE based spatial audio reproduction system, where the primary and ambient components undergo different rendering schemes [28]. The rendering schemes differ for loudspeaker or headphone playback [21], [24], [29]. For loudspeaker playback, the primary components are reproduced using vector base amplitude panning (VBAP) [30] or vector base intensity panning [31], [32] to reproduce the accurate direction of the sound sources. The ambient components, on the other hand, are further decorrelated and distributed to all the loudspeaker channels to create an envelopment effect of the sound environment [17], [33]. For headphone playback, the conventional virtualization that simply applies binaural rendering to the mixed channel-based signals suffers from virtual phantom effect as discussed in [13], [16]. PAE based virtualization resolves this problem by applying binaural rendering to the extracted primary components, creating accurate virtual sound sources in the desired directions [16] for headphone playback [19]. Similar to the loudspeaker playback case, the ambient components are decorrelated using artificial reverberation [14], [17], [21], [22] to create a more natural sound environment. A natural sound rendering framework that incorporates PAE with several other signal processing techniques to tackle headphone reproduction issues is introduced and evaluated in [19].

2 Primary ambient extraction approaches

Given that PAE could be employed to improve the performance of spatial audio rendering, the sound decomposition process of PAE itself is not necessarily easy. In this section, we will briefly present an overview of various approaches from the PAE literature [34]. As discussed in previous section, we consider the audio scene as a sum of the primary components and ambient components. The primary components are usually composed of directional point-like sources, whereas the ambient components are diffuse sound determined by the sound environment. The target audio format of PAE is channel-based signals. Therefore, we classify the PAE approaches based on the number of channels in the input signals: single channel (or mono), stereo, and multichannel. From another perspective, the complexity of the audio scenes affects the performance of PAE greatly. Based on the existing PAE work, the complexity of audio scenes can generally be classified into three levels, namely, basic, medium, and complex. Besides sharing the common spatial characteristics that differentiate primary components from ambient components, other conditions in the three complexity levels are different. The basic complexity level refers to the audio scene where there is usually one dominant source in the primary components, with its direction created using only amplitude panning techniques. The medium complexity level only requires the condition of one dominant sources, without restricting how its direction (using amplitude panning, delay, etc.) can be created. In the complex audio scene level, we consider multiple dominant sources in the primary components. Based on these two perspectives, we shall classify the PAE approaches into different categories, as summarized in Table 1. It is observed that most of the PAE works mainly focus on the stereo signals, thanks to the large amount of stereo content. There are some works for multichannel signals, whereas very limited works on single channel signals.

Table 1: An overview of recent work in PAE

No. of channels	Complexity of audio scenes: Basic (single source, only amplitude panning)	Medium (single source)	Complex (multiple sources)
Stereo	**Time frequency masking:** [35], [21], [36], [23] **PCA:** [37], [36], [16], [38], [40] **Least-squares:** [33], [31], [24], [29], [41] **Linear estimation**: [18] **Ambient spectrum estimation**: [42], [43] **Others:** [13], [22], [44]	**LMS:** [25] **Shifted PCA:** [45] **Time-shifting:** [46]	**PCA:** [28], [47]
Multichannel	**PCA:** [16] **Others:** [48]	**ICA and masking:** [49] **Pairwise correlations:** [50] **Pairing**: [52]	**ICA:** [51]
Single	**NMF**: [53] **Neural network**: [54]		

2.1 Stereo signals

One of the earliest works in primary or ambient extraction was from Avendano and Jot in 2002 [35]. In this work, a time-frequency masking approaches was proposed to extract ambient components from stereo signals by deriving the ambient mask using a nonlinear function of the inter-channel coherence [21]. Following works derives the ambient mask based on the characteristic that ambient components have equal level in the two channels of the stereo signal [36] or using diffuseness measured from B-format microphone recordings [23].

Principal component analysis (PCA) has been the most widely studied PAE approach [16], [36]-[40]. The key idea is to extract the principal component with the largest variance as the primary components. Variants of PCA include the modified PCA that ensures uncorrelated ambience extraction [38], enhanced post-scaling to restore the correct primary -to-ambient energy ratio [39] and correct power of primary and ambient components [40]. Least-squares is another type of commonly used PAE approaches [33], [26], [24], [29], [41]. Based on the basic stereo signal model, least-squares algorithm derives the estimated primary and ambient components by minimizing the mean-square-error (MSE) of the estimation of these components [33]. Furthermore, other least-squares variants were introduced to improve the spatial quality of the extracted primary and ambient components [24], [41].

These existing PAE approaches could be generalized into a linear estimation framework [18]. Under this framework, two groups of measures are introduced to yield a more complete performance evaluation of the timbre and spatial quality of the PAE approaches. For the timbre quality, a series of performance measures are proposed to identify the components that contribute to the extraction error, including distortion, interference and leakage. Therefore, the PAE problem can be solved based on specific objectives, for example, minimizing the primary ambient components correlation in PCA, minimizing MSE in Least-squares, and three variants of least-squares based on more specific extraction objectives. The linear estimation framework allows us to obtain a comprehensive study on the performance of these PAE approaches.

The performance of linear estimation based PAE approaches is inferior in strong ambient power cases, due to the limitations to cancel the uncorrelated ambient components without distorting the primary components. To solve this problem, a new framework based on ambient spectrum estimation was introduced [42], [43]. Based on the observation that uncorrelated ambient component tend to preserve a close or equal magnitude between the two channels, we can reformulate the PAE problem as the problem of estimating ambient spectrum (i.e., ambient phase or magnitude). Solutions to ambient spectrum estimation are obtained by exploiting the sparsity of the primary components in the time-frequency domain. The objective and subjective evaluations reveal that the ambient spectrum estimation based approaches could reduce the extraction significantly, especially when the ambient power is relatively higher.

Other PAE approaches in this category include [13] that derives an out-of-phase signal as ambient components; [22] that considers ambient components as the sum of a common component and independent component; and [44] that classifies various models for extraction.

Practical signals are usually more complex than what is assumed in the basic signal model for PAE. One common case considers that the primary components are partially correlated, mainly due to the inclusion of time/phase differences (i.e., medium complexity). Using conventional PAE approaches for these complex signals degrades PAE performance. To handle these stereo signals, Usher and Benesty proposed an adaptive approach using normalized least-mean-squares (NLMS) to extract reverberation from stereo microphone recordings [25]. However, this adaptive approach cannot always yield a good performance in a short time. In contrast, we proposed shifted PCA [45] and extended it to a general time-shifting technique [46], which could increase the primary correlation to its maximum. Thus, the input signal is closest to the basic signal model, making conventional PAE approaches re-usable.

Though one dominant source is found to be quite common, it is still possible to encounter the cases with multiple dominant sources in some movies and games. Two methods were proposed to deal with primary components with multiple sources. The first technique considers subband decomposition before performing PAE in each subband [28]. The partitioning of the frequency bins into subbands is found to be critical, where the adaptive top-down partitioning method outperforms other methods. The other multi-shift technique [47] involves multiple instances of time-shifting, performing extraction for each shifted signals, and combining the extracted components from all shifting versions. The weighting method based on inter-channel cross-correlation is found to yield the best performance.

2.2 Multichannel signals

Besides the extensive study on PAE for stereo signals, PAE on multichannel signals is less well studied. PCA was originally proposed to work for multichannel signals with only one dominant amplitude-panned source in [16]. There are several works [48] that only briefly mention the idea for multichannel PAE without in-depth studies. For other multichannel signals with one dominant source, independent component analysis (ICA) can be combined with time-frequency masking to extract the dominant sources [49]. Another approach that was extended from [24], achieves primary ambient extraction using a system of pairwise correlation [50]. In the case of multiple sources in multichannel signals, blind source separation techniques can be employed for the purpose of primary ambient extraction. When the number of dominant sources is equal to or less than the number of channels (as it is the case for PAE), ICA is a common technique [51]. Compared to stereo signals, PAE with multichannel signals is easier to solve since there are more information available.

Moreover, PAE approaches based on stereo signals can be extended to multichannel signals [52]. The most straightforward way is via down-mixing the multichannel signal into the stereo signal and then apply PAE on the down -mixed stereo signal. A more general way to solve multichannel PAE problem is to apply pairwise stereo PAE. Considering the large number of combinations, one critical problem in this approach is how to select the pairs from multichannel signals. Considering the pairwise amplitude panning nature of multichannel signals, pairing every two neighboring channels seems more desirable.

2.3 Single channel signals

In contrast to stereo and multichannel signals, PAE with single channel signals is quite challenging due to the limited amount of information available. A critical problem in the single channel case is that how primary and ambient components can be defined and characterized since there are no inter-channel cues. Nevertheless, two works from Uhle shed some light on solving such a problem. In [53], it is considered that ambient components exhibit a less repetitive and constructive spectra structure than primary components. Therefore, when applying non-negative matrix factorization (NMF), the residue can be considered as ambient components. To reduce the computational complexity and latency in NMF, Uhle and Paul trained a neural network to obtain ambient spectra mask. Subjective tests in [54] validated the improved perceptual quality of the up-mix systems employing these PAE approaches.

3 Concluding remarks

Spatial audio reproduction is essential in creating immersive and authentic listening experience, as per the increasing need from the consumer market. Primary ambient extraction can be applied in spatial audio reproduction to alleviate the rigorous requirements of the channel-based audio format on the audio reproduction system configuration. Thereby, PAE facilitates flexible, efficient, and immersive spatial audio reproduction. This article presents an overview of the recent advancements on PAE. These works greatly improve the performance of PAE in dealing with various types of stereo signals. Nevertheless, more work still needs to be carried out for PAE with complex signals. In the future, we shall see more PAE techniques applied in spatial audio reproduction applications, including headphones, loudspeaker systems with height channels, and also stereo sound from mobile phones and tablets. With these advanced PAE approaches readily applied, the listeners can thus immerse him/her-self in the reproduced sound scenes, without the limitation on the audio contents or playback systems.

References

[1] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*. Cambridge, MA: AP Professional, 2000.

[2] D. R. Begault, E. M. Wenzel, M. Godfroy, J. D. Miller, and M. R. Anderson, "Applying spatial audio to human interfaces: 25 years of NASA experience," in *Proc. 40th AES Intl. Conf. Spatial Audio*, Tokyo, Oct. 2010.

[3] J. M. Loomis, J. R. Marston, R. G. Golledge, and R. L. Klatzky, "Personal guidance system for people with visual impairment: a comparison of spatial displays for route guidance," *J. Vis. Impair blind*, vol. 99, no. 4, pp. 219-232, Jan. 2005.

[4] Y. C. Arai, S. Sakakibara, A. Ito, K. Ohshima, T. Sakakibara, T. Nishi, et al., "Intra-operative natural sound decreases salivary amylase activity of patients undergoing inguinal hernia repair under epidural anesthesia," *Acta Anaesthesiologica Scandinavica*, vol. 52, no. 7, pp. 987-990, May 2008.

[5] ITU, "Report ITU-R BS.2159-4: Multichannel sound technology in home and broadcasting applications," 2012.

[6] Dolby Atmos-Next Generation Audio for Cinema (White Paper). 2013. Available online: http://www.dolby.com/uploadedFiles/Assets/US/Doc/Professional/Dolby-Atmos-Next-Generation-Audio-for-Cinema.pdf

[7] J. M. Jot, and Z. Fejzo, "Beyond surround sound - creation, coding and reproduction of 3-D audio soundtracks," in *Proc. 131st AES Conv.*, New York, NY, Oct. 2011.

[8] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H audio – the new standard for universal spatial/3D audio coding," *J. Audio Eng. Soc.*, vol. 62, no. 12, pp. 821–830, Dec. 2014.

[9] F. Rumsey, "Spatial quality evaluation for reproduced sound: terminology, meaning, and a scene-based paradigm," *J. Audio Eng. Soc.*, vol. 50, no. 9, pp. 651–666, Sept. 2002.

[10] F. Rumsey, *Spatial Audio*. Oxford, UK: Focal Press, 2001.

[11] W. S. Gan, E. L. Tan, and S. M. Kuo, "Audio projection: directional sound and its application in immersive communication," *IEEE Sig. Process. Mag.*, vol. 28, no. 1, pp. 43-57, Jan. 2011.

[12] M. A. Gerzon, "Optimal reproduction matricies for multispeaker stereo," *J. Audio Eng. Soc.*, vol. 40, no. 7/8, pp. 571–589, Jul./Aug. 1992.

[13] J. Breebaart and E. Schuijers, "Phantom materialization: a novel method to enhance stereo audio reproduction on headphones," *IEEE Trans. Audio, Speech, Lang. Process.*, vol.16, no. 8, pp. 1503-1511, Nov. 2008.

[14] J. Breebaart and C. Faller, *Spatial Audio Processing: MPEG Surround and other Applications.* Hoboken, NJ: Wiley, 2007.

[15] S. K. Zielinski, and F. Rumsey, "Effects of down-mix algorithms on quality of surround sound," *J. Audio Eng. Soc.*, vol. 51, no. 9, pp. 780–798, Sept. 2003.

[16] M. Goodwin and J. M. Jot, "Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement," in *Proc. ICASSP*, Hawaii, 2007, pp. 9-12.

[17] M. Goodwin and J. M. Jot, "Spatial audio scene coding," in *Proc. 125th AES Conv.*, San Francisco, 2008.

[18] J. He, E. L. Tan, and W. S. Gan, "Linear estimation based primary-ambient extraction for stereo audio signals," *IEEE/ACM Trans. Audio, Speech, Lang. Process*., vol. 22, no. 2, pp. 505-517, Feb. 2014.

[19] K. Sunder, J. He, E. L. Tan, and W. S. Gan, "Natural sound rendering for headphones: integration of signal processing techniques," *IEEE Sig. Process. Mag.*, vol. 32. no. 2, pp. 100-113, Mar. 2015.

[20] K. Kowalczyk, O. Thiergart, M. Taseska, G. Del Galdo, V. Pulkki, and E. A. P. Habets, "Parametric spatial sound processing," *IEEE Sig. Process. Mag.*, vol. 32, no. 2, Mar 2015, pp. 31- 42.

[21] C. Avendano and J. M. Jot, "A frequency-domain approach to multichannel upmix," *J. Audio Eng. Soc*., vol. 52, no. 7/8, pp. 700-709, Jul./Aug. 2004.

[22] F. Menzer and C. Faller, "Stereo-to-binaural conversion using interaural coherence matching," in *Proc. 128th AES Conv.*, London, UK, 2010.

[23] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503-516, Jun. 2007.

[24] S. W. Jeon, Y. C. Park, S. Lee, and D. Youn, "Robust representation of spatial sound in stereo-to-multichannel upmix," in *Proc. 128th AES Conv.*, London, UK, 2010.

[25] J. Usher and J. Benesty, “Enhancement of spatial sound quality: a new reverberation-extraction audio upmixer,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 7, pp. 2141-2150, Sept. 2007.

[26] C. Faller, “Matrix surround revisited,” in *Proc. 30th AES Int. Conf.*, Saariselka, Finland, Mar. 2007.

[27] P. Kleczkowski, A. Krol, and P. Malecki, “Multichannel sound reproduction quality improves with angular separation of direct and reflected sounds,” *J. Audio Eng. Soc.*, Vol. 63, No. 6, pp. 427-442, Jun. 2015.

[28] J. He, W. S. Gan, and E. L. Tan, “A study on the frequency-domain primary-ambient extraction for stereo audio signals,” in *Proc. ICASSP*, Florence, Italy, 2014, pp. 2892-2896.

[29] C. Faller and J. Breebaart, “Binaural reproduction of stereo signals using upmixing and diffuse rendering,” in *Proc. 131th Audio Eng. Soc. Conv.*, New York, 2011.

[30] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, Jun. 1997.

[31] M. Goodwin and J.-M. Jot. “A frequency-domain framework for spatial audio coding based on universal spatial cues,” in *Proc. 120th AES Conv.*, May 2006.

[32] J. M. Jot, V. Larcher, and J. M. Pernaux, “A Comparative Study of 3-D Audio Encoding and Rendering Techniques,” in *Proc. 16th AES Int. Conf.*, Rovaniemi, Finland, 1999.

[33] C. Faller, “Multiple-loudspeaker playback of stereo signals,” *J. Audio Eng. Soc.*, vol. 54, no. 11, pp. 1051-1064, Nov. 2006.

[34] J. He, “Spatial audio reproduction with primary ambient extraction,” PhD thesis, School of Electrical and Electronic Engineering, Nanyang Technological University, 2016.

[35] C. Avendano, and J. M. Jot, “Ambience extraction and synthesis from stereo signals for multi-channel audio up-mix,” In *Proc. ICASSP*, pp. 13-17, May 2002.

[36] J. Merimaa, M. Goodwin, J. M. Jot, “Correlation-based ambience extraction from stereo recordings,” in *Proc. 123rd AES Conv.*, New York, Oct. 2007.

[37] R. Irwan and R. M. Aarts, “Two-to-five channel sound processing,” *J. Audio Eng. Soc.*, vol. 50, no. 11, pp. 914-926, Nov. 2002.

[38] M. Goodwin, “Geometric signal decompositions for spatial audio enhancement,” in *Proc. ICASSP*, Las Vegas, 2008, pp. 409-412.

[39] S. W. Jeon, D. Hyun, J. Seo, Y. C. Park, and D. H. Youn, “Enhancement of principal to ambient energy ratio for PCA-based parametric audio coding,” in *Proc. ICASSP*, Dallas, 2010, pp. 385-388.

[40] Y. H. Baek, S. W. Jeon, Y. C. Park, and S. Lee, “Efficient primary-ambient decomposition algorithm for audio upmix,” in *Proc. 133rd Audio Eng. Soc. Conv.*, San Francisco, 2012.

[41] C. Uhle, and E. A. P. Habets, “Direct-ambient decomposition using parametric wiener filtering with spatial cue control,” in *Proc. ICASSP*, Brisbane, Australia, 2015, pp. 36-40.

[42] J. He, W. S. Gan, and E. L. Tan, “Primary-ambient extraction using ambient phase estimation with a sparsity constraint,” *IEEE Signal Process. Letters*, vol. 22, no. 8, pp. 1127-1131, Aug. 2015.

[43] J. He, W. S. Gan, and E. L. Tan, “Primary-ambient extraction using ambient spectrum estimation for immersive spatial audio reproduction,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 9, pp. 1430-1443, Sept. 2015.

[44] A. Härmä, “Classification of time-frequency regions in stereo audio,” *J. Audio Eng. Soc.*, vol. 59, no. 10, pp. 707-720, Oct. 2011.

[45] J. He, E. L. Tan, and W. S. Gan, “Time-shifted principal component analysis based cue extraction for stereo audio signals,” in *Proc. ICASSP*, Vancouver, Canada, 2013, pp. 266-270.

[46] J. He, W. S. Gan, and E. L. Tan, “Time-shifting based primary-ambient extraction for spatial audio reproduction,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 10, pp. 1576-1588, Oct. 2015.

[47] J. He, and W. S. Gan, “Multi-shift principal component analysis based primary component extraction for spatial audio reproduction,” in *Proc. ICASSP*, Brisbane, Australia, 2015, pp. 350-354.

[48] M. Goodwin and J. M. Jot, “Binaural 3-D audio rendering based on spatial audio scene coding,” in *Proc. 123rd AES Conv.*, New York, 2007.

[49] H. Sawada, S. Araki, R. Mukai, and S. Makino, “Blind extraction of dominant target sources using ICA and time–frequency masking,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 6, pp. 2165–2173, Nov. 2006.

[50] J. Thompson, B. Smith, A. Warner, and J. M. Jot, “Direct-diffuse decomposition of multichannel signals using a system of pair-wise correlations,” in *Proc. 133rd AES Conv.*, San Francisco, 2012.

[51] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ: Wiley, 2004.

[52] J. He, and W. S. Gan, “Applying primary ambient extraction for immersive spatial audio reproduction,” in *Proc. APSIPA ASC*, Hong Kong, Dec. 2015.

[53] C. Uhle, A. Walther, O. Hellmuth, and J. Herre, “Ambience separation from mono recordings using non-negative matrix factorization,” in *Proc. 30th AES Int. Conf.*, Saariselka, Finland, 2007.

[54] C. Uhle, and C. Paul, “A Supervised Learning Approach to Ambience Extraction,” in *Proc. DAFx*, Espoo, Finland, 2008.

Summary of Links

- APSIPA ASC 2016: http://www.apsipa2016.org
- Videos for APSIPA ASC 2016 **Keynotes**:
  - **Whither Speech Recognition? - Deep Learning to Deep Thinking**

    Professor Sadaoki Furui, Toyota Technological Institute at Chicago, USA

    [PDF] http://www.apsipa.org/doc/APSIPA%20ASC%202015%20Keynote%20-%20Whither%20speech%20recognition.pdf

    [Video] https://youtu.be/mP9QMJKtJyU

  - **Graph Signal Processing: Filterbanks, Sampling and Applications to Machine Learning and Video Coding**

    Professor Antonio Ortega, Signal and Image Processing Institute, Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA

    [PDF] http://www.apsipa.org/doc/GSP-Intro-Learning-Dec-2015.pdf

    [Video] https://youtu.be/AKxwZaEavRw

  - **Can I Trust this Photo?**

    Professor Alex Kot, School of EEE Associate Dean, College of Engineering Director, Rapid-Rich Object Search (ROSE) Lab School of EEE, Nanyang Technological University, Singapore

    [Video] https://youtu.be/QXdu8oPXawY

- APSIPA Friend's Lab: http://www.apsipa.org/friendlab/FriendLabs.htm.

APSIPA Who's Who

President: Haizhou Li, Institute for Infocomm Research, A*STAR, Singapore

President-Elect: Wan-Chi Siu, The Hong Kong Polytechnic University, Hong Kong

Past President: C.-C. Jay Kuo, University of Southern California, USA

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

VP - Industrial Relations and Development: Shan Liu, MediaTek, USA

VP - Institutional Relations and Education Program:

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

VP - Member Relations and Development: Kenneth Lam, The Hong Kong Polytechnic University, Hong Kong

VP - Publications: Tatsuya Kawahara, Kyoto University, Japan

VP - Technical Activities: Anthony Kuh, University of Hawaii at Manoa, USA

Members-at-Large:

Waleed H. Abdulla, The University of Auckland, New Zealand

Mrityunjoy Chakraborty, Indian Institute of Technology, India

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

Homer Chen, National Taiwan University, Taiwan

Woon-Seng Gan, Nanyang Technological University, Singapore

Wen Gao, Peking University, China

Hsueh-Ming Hang, National Chiao-Tung University, Taiwan

Yoshinobu Kajikawa, Kansai University, Japan

K.J. Ray Liu, University of Maryland, USA

Ming-Ting Sun, University of Washington, USA

Kazyua Takeda, Nagoya University, Japan

Toshihisa Tanaka, Tokyo University of Agriculture and Technology, Japan

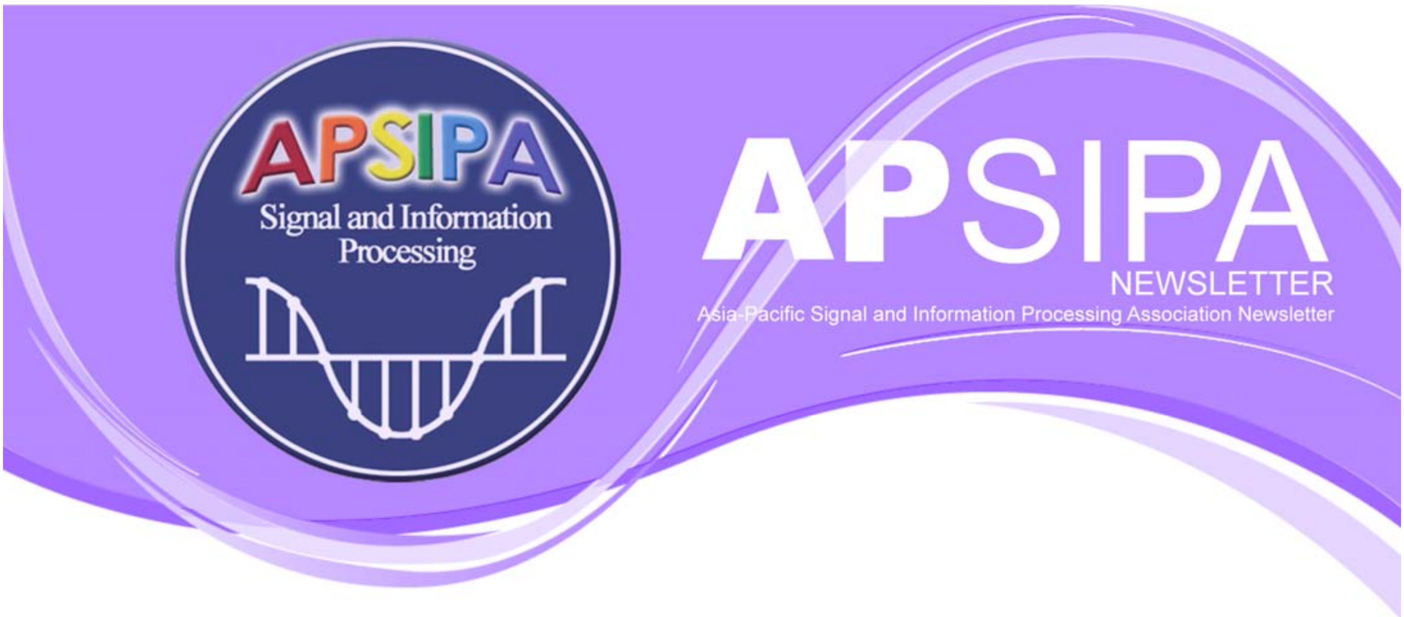
Headquarters

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**

More members to come!

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