


# Guest Editors' Note—Special Issue on Spatial Audio

Woon-Seng Gan <sup>1,\*</sup> and Jung-Woo Choi <sup>2</sup> 

<sup>1</sup> Digital Signal Processing Lab, School of Electrical and Electronic Engineering,  
Nanyang Technological University, Singapore 639798, Singapore

<sup>2</sup> School of Electrical Engineering, Korea Advanced Institute of Science and Technology,  
Daejeon 34141, Korea; jwoo@kaist.ac.kr

\* Correspondence: ewsgan@ntu.edu.sg; Tel.: +65-6790-4538

Received: 1 August 2017; Accepted: 2 August 2017; Published: 3 August 2017

## 1. Introduction

Three-dimensional (or spatial) audio is a growing research field that plays a key role in realizing immersive communication in many of today's applications for teleconferencing, entertainment, gaming, navigation guidance, and virtual reality (VR)/augmented reality (AR). Technologies in spatial sound capture and binaural recording are becoming an add-on module to our mobile devices to capture the surrounding soundscape, pickup directional and ambient cues, and create an immersive 3D audio media for playback. In total, eight research papers and two review papers are published in this special issue. The research papers reported on new research techniques that resulted in higher quality and more immersive spatial audio reproduction; while the two review papers account for the state-of-the-art spatial audio recording and reproduction for playback in VR and AR headsets. A detailed accounts on these papers are summarized as follows.

## 2. Advanced Signal Processing Technologies for Spatial Audio

The current state-of-the-art technologies involved with spatial audio are summarized in two review papers [1,2]. The review paper written by Zhang, Samarasinghe, Chen, and Abhayapala [1] delivers a broad overview of existing and emerging spatial audio technologies involved with spatial audio. The paper begins with a summary of binaural technologies based on the head-related transfer function and covers sound field recording/reproduction techniques utilizing multichannel microphones and loudspeakers. The ending of the paper is devoted to the multi-zone sound reproduction problem, which aims to deliver multiple audio programs over multiple spatial regions.

The paper by Hong, He, Lam, Gupta, and Gan [2] puts a strong emphasis on the use of signal processing tools for the design of soundscapes. The review paper discusses the sound recording and reproduction technologies to render auditory sceneries resulting from the interaction of sound objects and surrounding environments. Beyond the simple reproduction of existing auditory scenery, soundscape design problems to improve the existing poor acoustic conditions are presented. The augmented reality in audio is especially highlighted as a means to provide an improved listening experience.

Proper localization of a sound source has been an important issue in the spatial audio for a long time, and it has been realized in various ways for stereo, discrete multichannel, and sound field control systems. Some new aspects of the localization problem are dealt with in this special issue, especially for proper recording, reproduction, perception, and evaluation.

In the paper written by Gößwein, Grosse, and van de Par [3], the authors propose a stereoscopic recording technique for enhancing the direct sound field in a reverberant environment. They employ two crossed linear microphone arrays combined with a super-directive endfire beam pattern. It is shown that the array recording can reduce the reverberation, while keeping compatibility with the amplitude panning technique.

The perception of sound localization has been studied over a decade, but the localization of elevated sound sources still remains a challenging problem. Wallis and Lee [4] study the influence of the interchannel time and level differences between two loudspeaker layers of different heights on the localization threshold. The results show the dependence of the localization threshold on the interchannel time difference. The required directivity of recording microphone in height direction is also discussed based on the identified localization threshold.

Objective evaluation of the reproduced sound field is another important issue. Mean squared error (MSE) has been popularly employed as a measure of similarity between target and reproduced sound fields. However, in their work [5], Chang and Jeong propose to use beamforming powers derived from given sound fields as a new measure. The primary reasoning behind the proposed measure is the weakness of MSE against room reflections and for 2.5-D reproduction techniques, such as wave field synthesis, that inevitably exhibit amplitude bias along the distance. The beam-power measure is expected to provide a more robust means to evaluate the directional cue or the direct component of a sound field.

The study by Mieth and Zölzer [6] deals with the objective evaluation problem for the pairwise panning-based upmix algorithm. To access the sound quality of upmix algorithms without subjective evaluations, they propose detailed procedures and measures regarding the direction of a virtual sound source, the amount of residual direct sound, loudness, and correlations in the frontal and surround channels.

The localization of sound is not only involved with the directional cue. Wendt, Zotter, Frank, and Höldrich [7] investigated the way to control the perceived distance through the variation of source directivities. The influence of the auralized room, source-listener distance, signal, and single-channel reverberation are considered together to build a model predicting the perceived distance. They tested various third-order beam patterns in a real room, which demonstrate that the distance perception caused by the source directivity is coupled with the sense of apparent source width.

For spatial audio, there are many auditory impressions to be carefully controlled along with the localization cues. Auditory sceneries deliver various spatial impressions such as stage width and ambience. The synthesis of late reverberation is studied in the paper written by Välimäki, Holm-Rasmussen, Alary, and Lehtonen [8]. They segmented the late part of a room impulse response and approximated the segments as filtered velvet noises that are very sparse in time but sound smoother. It is demonstrated that filtering with velvet noises greatly reduces the computational cost, only resulting in minor subjective differences for transient sounds.

The paper authored by Bai, Chung, Wu, Chiang, and Yang [9] proposes a general strategy to tackle the inverse problem that is often encountered in solving the source identification and separation problems for the spatial audio signal processing. Various inverse problem solvers, for both underdetermined and overdetermined problems, are investigated and compared in terms of PESQ and segSNR. Guidelines for choosing the right algorithm and regularization parameter are provided, with detailed examples of sound field analysis and synthesis problems.

Another inverse problem discussed by Gómez, Astley, and Fazi [10], is for the interactive auralization of sound fields in a low-frequency region. They utilized the finite element method to simulate a sound field in a room and then transformed the result using a plane wave expansion technique. Plane wave expansion has been popularly used for its simplicity in realizing the interactive sound rendering system that requires translation and rotation of sound fields. The transform of a sound field using plane wave expansion is a typical inverse problem, in which the determination of a regularization parameter is important to prevent singularity problems. The effect of regularization on the sound field representation is discussed, in view of plane waves' energy density and the size of sweet spot.

### 3. Summary

The trends reflected in the above papers stress that the signal processing technologies for spatial audio are heading towards a more natural reproduction of auditory impressions. The whole signal processing chain from recording to reproduction and evaluation is being revisited to cope with emerging applications, such as VR/AR and to render new auditory impressions. Although there is still a long way to go for the complete understanding of human listening and perfect control of auditory sceneries, the directions presented in this special issue demonstrate that a great deal of improvement can be made through the combination of perceptual and physical sides of spatial audio.

**Acknowledgments:** The editors would like to thank the strong administration support rendered by the MDPI editorial team, which include Managing Editor, Xiaoyan Chen; and Assistant Editors, Alice Zhang, Daria Shi, Candice Zhuo, Sydney Sun, and Jennifer Li. Their effort and quick responses in handling paper reviews and responding to our many questions has resulted in the fast appearance of this Special Issue on Spatial Audio.

**Conflicts of Interest:** The authors declare no conflict of interest.

### References

1. Zhang, W.; Samarasinghe, P.N.; Chen, H.; Abhayapala, T.D. Surround by Sound: A Review of Spatial Audio Recording and Reproduction. *Appl. Sci.* **2017**, *7*, 532. [[CrossRef](#)]
2. Hong, J.Y.; He, J.; Lam, B.; Gupta, R.; Gan, W.-S. Spatial Audio for Soundscape Design: Recording and Reproduction. *Appl. Sci.* **2017**, *7*, 627. [[CrossRef](#)]
3. Gößwein, J.A.; Grosse, J.; van de Par, S. Stereophonic Microphone Array for the Recording of the Direct Sound Field in a Reverberant Environment. *Appl. Sci.* **2017**, *7*, 541.
4. Wallis, R.; Lee, H. The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localisation Thresholds for Natural Sound Sources. *Appl. Sci.* **2017**, *7*, 278. [[CrossRef](#)]
5. Chang, J.-H.; Jeong, C.-H. A Measure Based on Beamforming Power for Evaluation of Sound Field Reproduction Performance. *Appl. Sci.* **2017**, *7*, 249. [[CrossRef](#)]
6. Mieth, M.; Zölzer, U. Objective Evaluation Techniques for Pairwise Panning-Based Stereo Upmix Algorithms for Spatial Audio. *Appl. Sci.* **2017**, *7*, 374. [[CrossRef](#)]
7. Wendt, F.; Zotter, F.; Frank, M.; Höldrich, R. Auditory Distance Control Using a Variable-Directivity Loudspeaker. *Appl. Sci.* **2017**, *7*, 666. [[CrossRef](#)]
8. Välimäki, V.; Holm-Rasmussen, B.; Alary, B.; Lehtonen, H.-M. Late Reverberation Synthesis Using Filtered Velvet Noise. *Appl. Sci.* **2017**, *7*, 483.
9. Bai, M.R.; Chung, C.; Wu, P.-C.; Chiang, Y.-H.; Yang, C.-M. Solution Strategies for Linear Inverse Problems in Spatial Audio Signal Processing. *Appl. Sci.* **2017**, *7*, 582. [[CrossRef](#)]
10. Gómez, D.M.M.; Astley, J.; Fazi, F.M. Low Frequency Interactive Auralization Based on a Plane Wave Expansion. *Appl. Sci.* **2017**, *7*, 558.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).