

Spatial Audio Reproduction with Primary Ambient Extraction

JianJun He

Springer, New York, (2017),
132 pp., 54.99 USD, Softcover,
ISBN 978-981-10-1550-2

This publication appears to be a popularized dissertation of JianJun He's PhD work at Nanyang Technological University in Singapore. Primary ambient extraction (PAE) is the name given to any set of digital signal processing (DSP) techniques, performed during the playback of an audio file, and having the goal of optimizing the accuracy and immersiveness of the perceived spatial audio reproduction relative to the intended soundscape. The PAE separates the primary sources (those having coherent directivity) from the ambient sounds in the audio recording in preparation for subsequent DSP and signal routing to loudspeakers or headphones, with paths tailored for each type of signal.

Typically, PAE is not required for audio files having formats matched to the playback system, such as stereo files played over stereo speakers or surround-encoded files played over a matching set of surround speakers. In these cases, one expects the mix to be optimized for the designated playback arrangement. However, PAE will be required when attempting to translate audio recorded in one format for optimal spatial reproduction in a mismatched playback arrangement, such as stereo files played over a surround speaker arrangement or surround-encoded files played over a 3D speaker arrangement.

The book is premised on the assumption that the performance of existing PAEs on existing recordings (having a variety of mixing and recording formats) may be quantitatively measured. With this measurement tool in hand, improved extraction techniques can then be quantitatively judged.

Throughout the book, various methods are reviewed for extracting primary and ambient signals from single-, two-, and multi-channel recordings. Performance measures are chosen, improved processes are suggested and developed, and the old and new methods are graded on their abilities

to extract primary and ambient signals from several benchmark recordings. Each of these benchmark recordings is manufactured for the purpose at hand, so everything is known about their primary and ambient sources, microphones, mixes and recording formats a priori.

Although the author contends that all existing audio recordings can be described and judged by a finite set of recording and mixing processes, what is missing is some sort of statistical sampling of how well those real recordings fit the finite assumptions used to create the benchmarks with which the book gages process performance. JianJun He touches on this issue in the very last paragraph of the book, noting that "PAE is a blind process . . . performance relies heavily on how effective the signal model is . . . not one signal model could satisfy any audio content", and proposes machine learning as a possible solution.

With that proviso out of the way, I found the book to be an excellent primer and reference on the title subject. Each chapter ends with a generous bibliography of additional references for specific PAE subjects. The author's selection of PAE type groupings and methods of grading performance are clear and logical. The development of improved PAE is straightforward and the benchmarking experiments are defensible. Reading may be complicated by an excessive use of two- and three-letter abbreviations throughout the text, with a couple abbreviations missing their initial definitions; however, those working within the discipline should have few problems.

The final chapter reviews the current state of PAE, the book's contributions in improving those processes, and recommendations for further development. I recommend *Spatial Audio Reproduction with Primary Ambient Extraction* as a good reference for every engineer working behind the scenes to improve the user audio experience.

Jon W. Mooney
IMEG Corp.

Anaheim, CA, 928905, USA
jon.w.mooney@imegcorp.com