Audio Engineering Society

Convention e-Brief 624

Presented at the 149th Convention
2020 October 27–30, Online

"LDK-Cube" and "Time-Space-Energy-Equalization": A Case Study of Binaural 3D Recording and Production of an Orchestral Concert with a Sound Reinforcement System in Operation

Haiyan Xiang[1], Xiang Ji[2], Dakang Li[3], Jie Zhu[4], Peng Huang[5] and Shusen Wang[6]

[1] *Chengdu Radio and TV Station, 2# Gaopeng Dadao, Chengdu, China*

[2] *Shenzhen EZPro Sound & Light Technology Co., Ltd, No.1001, Zhongshan Yuan Rd, Nanshan District, Shenzhen, China*

[3] *Communication University of China, East Street of Dingfuzhuang No. 1, Beijing, China*

[4] *Beijing Dreamula Digital Technology Co., Ltd, No.39, Chunhe Road, Daxing District, Beijing, China*

[5] *Beijing Film Academy, Xitucheng Road No. 4, Beijing, China*

[6] *AIYISI (Beijing) Science & Technology Co., Ltd. Beijing, China*

Correspondence should be addressed to Haiyan Xiang (xianghaiyan36@gmail.com)

ABSTRACT

As for the real-time immersive sound recording and production for an orchestral concert amplified through a sound reinforcement system, how can a TV program be produced binaurally with real 3D ambience and audience capture, and potentially for the purpose of multi-platform distribution? This engineering brief describes innovative solutions applied at the 2020 Chengdu New Year Concert: a largely-spaced 3D microphone array for capturing hall ambience, named the LDK-Cube, with a process of High Order Ambisonics encoding and binaural rendering; and Time-Space-Energy-Equalization (TSEEQ), a novel concept which means optimizing time-space-energy composition, through an integrated process of spectral and spatial equalization/delay, timeline EQ/delay, and group delay, is first presented in this paper as an overarching principle for the solutions. The perception that sound reinforcement disarranges the recording, forcing broadcasting engineers to fall back on pre-recorded audience effects for years, has thus been subverted. Furthermore, binaural/transaural 3D encoded 2-channel recordings providing much better spatial perception, proved to be the future of the stereo, "bionic TSEEQ" is derivate thereof.

1 Introduction

The evolution from stereo to immersive audio production has enabled recording engineers to deliver an enhanced realistic experience for audiophiles, television viewers and online audience. The BBC Proms is an outstanding example of binaural 3D immersive audio production. Their workflow does not involve personalized HRTFs, but binaural production is integrated into the traditional stereo recording and production process [1]. Learning from this model, we decided to produce the TV program of the 2020 Chengdu New Year Concert with both stereo and binaural 3D versions, and to broadcast both on air and online. This concert was held in the 1500-seat Opera hall of Chengdu City Music Hall, with a sound reinforcement system in operation. For many years, broadcasting engineers have used pre-recorded audience effects because of the belief that PA systems disarrange the recording. For us, the greatest challenge was how to capture and mix the real 3D ambience and audience effects, as high SPL sound radiated simultaneously from the orchestra and the main line array speakers on both sides of the stage. A sound pressure level of 97 dB-C(average) was measured in the middle area of the audience seats during the performance. Ambience/audience microphones always play an important role in recordings and live broadcasts of classical music. Without them, the necessary information about the space and audience reactions could not be produced to match the HD/4k images.

A dummy head microphone was not our first choice as the main ambience microphone, due to its lack of flexibility and the visual dissonance it might cause while hung from the ceiling. Sounding tight is a common characteristic of the coincident microphone, so 1st order Ambisonics microphones were also not suitable for this task. Nonetheless, we found they both performed well as 3D audience spot microphones. See details below.

Prof Dakang Li, one of the authors of this paper, has for several years used an arrangement called the LDK-Cube to capture channel-based 3D ambience for audio research and commercial projects. This setup offers rich, loose and natural spatial information. In a standard 7.1.4 loudspeaker configuration, the recorded signals can be easily panned to front left/right, rear left/right channel at ear level, and 4 loudspeakers in the upper level. In this case, we exploited its potential for scene-based production: the perfect compatibility with HOA coding.

2 LDK-Cube

The LDK-Cube is a wide-spaced cuboid microphone array, named after its inventor, Professor Dakang Li, designed specifically for ambience recording. It is configured with four omni microphones on the bottom layer and four cardioid microphones on the top layer, with their axial directions facing the floor and the ceiling respectively, and edge lengths of around 4*8*7m (H*L*W) in fig. 1. In fact, the dimensions are flexible, and can be adapted to fit the venue. The rigging is simple and easy, particularly because orientation is not an issue. Visually, the LDK Cube is nearly invisible (also as shown in fig. 2).

There were several considerations that affected the choice of placement for the LDK-Cube:

1. The vertical faces of the Cube run parallel to the side walls of the hall and are proportional to the stage to achieve a relatively defined panorama of the hall and the orchestra. The edge lengths are usually less than 10m.
2. The cube was kept away from the balconies and air conditioner vents. The bottom layer was 7m above the ground, a distance that achieves a sense of community with the audience, and avoids possible noise.
3. Because of the flexibility of the LDK-Cube design, the cardioid microphones in the top layer could be located outside the coverage of the FOH speakers, and pointing upwards to capture more ceiling reflections.

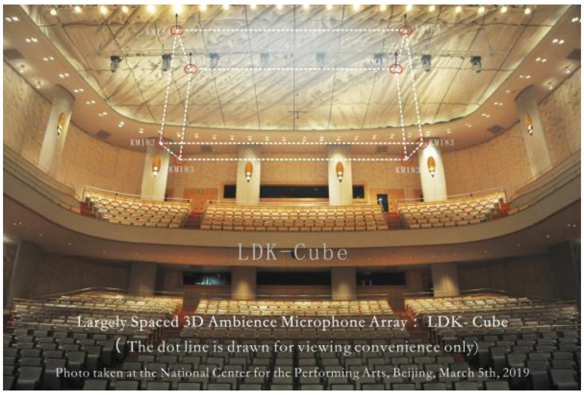

Fig 1. LDK-Cube at the National Centre for the Performing Arts, Beijing

Several subject listening tests indicated that the LDK-Cube outperformed coincident and/or small spacing arrays, such as 1st order Ambisonics and ORTF-3D microphones, for the capture of orchestral concert ambience while using HOA coding and binaural rendering. The LDK-Cube delivered a more natural ambience and fewer artifacts in higher frequencies. The result is consistent with the conclusions of Lee, Frank, and Zotter [2]. Of course, with HOA encoding, it also features in the compatibility to be decoded into various 3D loudspeaker configuration.

Fig 2. LDK-Cube on location at the 2020 Chengdu New Year Concert

3 Time Space Energy Equalization (TSEEQ) and Bionic TSEEQ

A sound reinforcement/recording system for a symphony orchestra included many close microphones positioned by music stands, and widely spaced ABC stereo miking. The LDK-Cube of ambience microphones was positioned in the middle of the space above the stalls. An Ambeo VR Mic and a KU100 and several MKH416s were used as spot mics for audience effects. The time-alignment rule between the stage spot microphones and the main ABC was applied to the stage capture with ambience/audience effects mics. Therefore, we took the midpoint between two bottom mics in LDK-Cube as reference zero, delayed the stage signal to make sure that there’s less than 30ms arrival discrepancy between these highly correlated signals in the same sound field. However, this created potential confusion with the delays in the sound production PGM, which are introduced to match the audio to the HD/4K images. To resolve the issue, H. Xiang conceived a new term: Time-Space-Energy Equalization (TSEEQ), means Optimizing Time-Space-Energy-Composition; this idea was developed into the Bionic TSEEQ, proposed as an integrated methodology for 3D sound processing. This approach incorporates spectral/spatial energy equalization/delay, timeline EQ/delay and group delay, and takes human anatomy, and psychoacoustics into consideration.

This process can be found in the extremely rigorous manipulation of the time alignment of the four capsules of the FOA microphones, which should theoretically be located at exactly the same point. These algorithms are common in the audio industry, but have been overlooked by broadcast engineers. This is why broadcasts have long used pre-recorded audience effects.

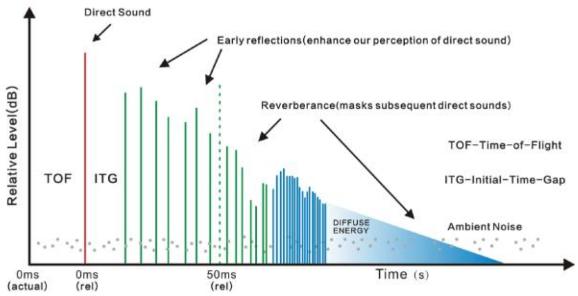

Fig 3. The Energy-Time Curve [3]

Figure 3 illustrates the relationship of time and energy to psychoacoustic perception within a cross section of space. However, new research concludes that for reproduction of a binaural source through headphones, the threshold value for echo perception is only 15ms, instead of 50ms as commonly assumed [4]. Thus, time-space-energy relationship is worth close examination for binaural production.

This prompted the concept of Time-Space-Energy Equalization (TSEEQ), which can work as a hybrid tool that makes use of time-space conversion relationships to align signals in space either precisely or roughly. Each type of alignment has its own scope of application, and can work as effective equalization before any spectral energy EQ operation.

In this case, we also found that even in a stereo mix, the ambience with height information sounded more vivid and immersive than simply downmixing the capture of the LDK-Cube to stereo. Height information was introduced by binauralizing the

HOA signal encoded from LDK-Cube. The reasons may be (1) HOA encoding describes a more detailed spherical space with 16-64 spatial components; (2) Binaural rendering applies time-space-energy equalization by applying filters/filter sets sequentially according to the measured HRTF/HRIR data in the timeline. This means that the time-space-energy relationships of the original signal can be adapted to the psychological perception of the human ear. However, there was no way to design signal processes to achieve this effect in traditional EQ + fader + panpot + reverb stereo production. The success and popularity of ORTF microphones remind us of the importance of bionic processes, and of the difficulty of going beyond the dummy head microphone to deliver natural and subtle spatial information. Bionic TSEEQ was therefore developed based on the TSEEQ concept.

Applying TSEEQ and Bionic TSEEQ to the ambient signals allowed us to maximize the value of the LDK-Cube for binaural 3D production. It allowed the ambient signal to be mixed with the stage signal as early reflections to enhance the distance and depth of a realistic 3D space [5]. It also enabled reproduction familiar to our perceptions in a realistic 3D sound field.

But we didn't align the stage signal to the centre of the LDK-Cube. The stage signal should be leading in time, e.g. 10ms ahead of the front edge of LDK-Cube. A 'perfect' alignment is impossible; and it could degrade the clarity and timbre of the stage signal as illustrated in fig.3. For example, we delayed the signal from the stage mic on the suona (a Chinese brass-like folk instrument) relative to the reference zero until no echo/timbre degradation was perceived.

On the other hand, with a sound reinforcement system in operation, the loudspeakers and stage/room acoustics worked together as a system, exhibiting unique transfer functions which introduced magnitude and phase distortions to the sound of orchestra. Optimising the sound reinforcement system is a must before recording. TSEEQ processes was undoubtedly put into use. Playing a sine sweep tone on either L or R main speakers as the reference signal, the magnitude and phase response were calculated after compensating for propagation delay from the loudspeaker to the test microphone. Converting the phase response into group delay provided a clearer picture of the time delays for different frequency ranges. An all-pass filter and custom generated FIR filters were used as handy tools for time alignment across the whole bandwidth of the sound reinforcement system. Here, delay was utilized again as spatial/spectral energy equalization. Other operations were as below:

1. Reducing excessive high frequencies in the lower LDK-Cube mics created a more natural and warm sound. Keeping a certain amount mid/high frequency content from the upper array was helpful for height information perception.
2. After equalization, the gain of all 8 microphones of the LDK-Cube should be adjusted to make sure proper level balance is kept. This prevents their inherent time relation being disrupted because of distinct level difference.

4 HOA Coding and Binauralization/Transauralization

For the stage capture, the mixing of main ABC microphone array and spot microphones was basically done in stereo. Adding ambience/audience microphones made a huge difference to listeners' spatial and immersive experiences. We encoded the LDK-Cube to 3rd order Ambisonics, and then binauralised the 3OA signal. This created a better sound than binauralizing the Cube directly, because more spatial components are generated through 3OA encoding. The audience effects spot mic, an Ambeo VR Mic was up-mixed from FOA to 3OA, then binauralized/transauralized. This immediately produced more precise spatial details and an amazingly vivid sound. It performed much better than mixing several MKH416 microphones together. The KU100 was used as the reference, and produced natural sounding 3D audience effects that could be mixed without any processing.

With the use of a number of spatial audio tools, both the ambient signals and the stage capture could be mixed binaurally/transaurally, to produce both stereo and binaural/transaural PGM. We preferred the binaural/transaural versions, because they delivered a distinctive 3D expansion of the sound field, clearer lines for each part, and had no perceptible coloration of the music timbre. The signal process flowchart is shown as fig 4.

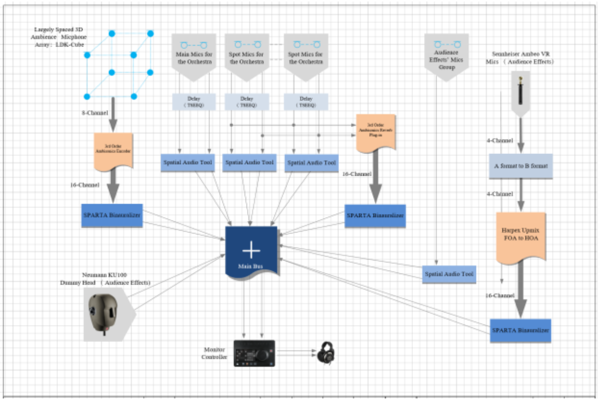

Fig 4. The signal process flowchart

5 Conclusions

The LDK-Cube is a largely-spaced 3D ambience microphone array for immersive audio recording and production, channel-based or scene-based. This layout perfectly suits the recording of orchestral music no matter whether a sound reinforcement system is in operation or not.

A new term of Time-Space-Energy-Equalization (TSEEQ) which means Optimizing Time-Space-Energy-Composition is conceived by H. Xiang as an integrated processing methodology in the recording and production of 3D sound. It is a combination of spectral and spatial energy equalization/delay, timeline EQ/delay, and group delay. This obviates the longstanding practice of using pre-recorded audience effects for live production.

In the case described here, the application of the TSEEQ process proved that no trace of disarrangement from the PA system could be perceived in the binaural 3D recording. We were able to obtain more vivid, immersive ambience with the real acoustic signature of the venue, as well as real 3D audience effects.

Bionic TSEEQ is a derivative of TSEEQ that improves the spatial feel in 2-channel recordings through binaural/transaural 3D encoding. It represents the future of traditional stereo. Of course, the results depend on the spatial audio tools chosen. Extensive experimentation and careful selection are important. We recommend binaural/transaural mixing as the most economic and promising approach to 2-channel 3D sound.

Acknowledgements

The first author would like to thank Prof Cathy Lane, Prof John Wynne and Ms Lisa Hall at CRiSAP of the London College of Communication, University of the Arts London; Dr Chris Pike at the BBC Research & Development; China Scholarship Council (File No. 201808510055).

References

[1] T. Parnell and C. Pike, “An efficient method for producing binaural mixes of classical music from a primary stereo mix”, *presented at the 144th Convention* (2018 May), Convention e-Brief 430.

[2] H. Lee, M. Frank, and Franz Zotter, “Spatial and Timbral Fidelities of Binaural Ambisonics Decoders for Main Microphone Array Recordings,” *presented at the International Conference on Immersive and Interactive Audio* (2019 March), conference Paper 75.

[3] X. Chen, “Sound Reality Technology: The Exploration of “Real Sound” In the Headphones”, *In China Excellent Master thesis Full-text Database* (2017).

[4] P. Brown, “Sound System Design”, Synergetic Audio Concepts, p13, (2016).

[5] G. Theile, H. Wittek, “3D Audio Natural Recording”, *presented at the 27th VDT international convention*, pp 6-7 (2012 November).