Advantages of Object-Based Audio and Possible Future Optimizations

Nianze Liu
New York University

This article firstly goes through the advantages of object-based audio compared with traditional channel based audio as well as sound field representation, leading to the discussion of possible optimizations to current object-based audio production workflow. In order to make the pipeline more productive, the concept of Internet of Things is introduced with some real world demonstration, ending with a brief discussion about the future challenges in combining Internet of Things into the field of immersive audio.

Keywords: Immersive Audio, Internet of Things, Object-Based Audio

Introduction

With its flexibility to dynamically render the output audio according to the context requirement at runtime, object-based audio becomes more and more popular nowadays, especially in the field of virtual reality and augmented reality.

Advantages of Object-Based Audio

Traditionally, in order to get best playback quality, producer dealing with channel-based audio must have the knowledge of the target reproduction system: for headphone setup, binaural recordings will provide better experience; for loudspeakers, stereo microphones are preferred during the recording phase. In the end, for different reproduction devices setup, upmixing and downmixing, as well as signal processing accordingly, are required to optimize the playback experience, which is inflexible, if not unrealistic, in the context of immersive audio since we might want the audio to be dynamically updated according to the listener's head direction. Object-based audio, as showed in Susal, Krauss, Tsingos, and Altman (2016), has the advantage of being format-agnostic, which makes it more efficient to produce high quality audio playback.

There is another audio representation that could be used in immersive audio: sound field representation. However, sound field representation has a big disadvantage: the PCM channels required to encode more detailed spacial information grows very quickly: in Higher-Order Ambisonics, 1st order requires 4 channels, 2nd order requires 9 channels, and 3rd order grows up to 16 channels. For practical reasons, soud field representation might be more appropriate with lower order in the context of 3 degrees of freedom (DOF) playback scenarios. For more immersive 6 DOF environment, object-based audio representation is preferred. What's more, with the associated audio metadata, it is easier to modify object-audio representation to provide better interactivity and achieve spacial fidelity compared to the sound field representation.

Optimizing Object-Based Audio

Pipeline Optimization

With all the advantages above, object-based audio is a promissing immersive audio representation. However, current workflows in object based audio production have not reached its full potential.

First, in the capture phase, the object-based audio is usually created manually by producer. This could be optimized by adopting new recording facilities able to directly generate metadata and create object-based audio. For example, in Coleman et al. (2018), a kinect2 sensor is used to get position data, which will be directly encoded into the audio object metadata.

Second, in the production phase, there are multiple different formats to represent the metadata, such as Audio Scene Discription Format(ASDF) and SpatDIF, where ASDF contains the complete scene information in a static file, inlucding the whole temporal progress, while SpatDIF has limited extensibility. According to Coleman et al. (2018), it is suggested to adopt the more flexible JSON format, which is widely used in the internet as a versatile means to represent object-oriented information. The proposed format is a streaming representation, so the temporal change of audio scene could be transmitted repeatedly via a network connection.

Last but not least, in the rendering phase, it is preferred to follow the component-based design pattern with predefined application programming interfaces(API) shared by the whole industry, so that software frameworks can be built in small modules and extensibility and portability might be achieved. Modules might focus on the signal pocessing part, as well as the signal's interaction part. With the common API, the modules could be easily exchanged and switched,

making the rendering procedule more flexible.

Assisted with Internet of Things

To make the object-based audio pipeline more productive, techniques from Internet of Things might be helpful.

Internet of Things, according to Turchet, Fischione, Essl, Keller, and Barthet (2018), refer to "embedded systems that are connected to the internet, which are able to interact with each other and cooperate to reach common goals." Following this direction, in Lee (2017), speakers are combined with beacon and Raspberry Pi to form a group of smart speakers that are aware of their relative positions in respect to other speakers. Following this thought, by embedding sensors in recording devices, it is possible to build next generation of "smart microphones" that are smart enough to generate object-based audio metadata. Since the optimized object-based audio pipeline uses standardized JSON format for the audio object metadata as well as component-based common API, it is even possible for the capture devices to communicate with the devices in the later production phases.

Future Challenges

It is exciting to apply Internet of Things are exciting techniques to the field of immersive audio, especially in the context of coming 5G network, where faster and more reliable inter-device communication becomes possible. However, there are still engineering challenges before their complete merge. Firstly.

References

Coleman, P., Franck, A., Francombe, J., Liu, Q., de Campos, T., Hughes, R. J., ... Woodcock, J., et al. (2018). An audio-visual system for object-based audio: From recording to listening. *IEEE Transactions on Multimedia*, *20*(8), 1919–1931.

Lee, C. H. (2017). Location-aware speakers for the virtual reality environments. *IEEE Access*, *5*, 2636–2640.

Susal, J., Krauss, K., Tsingos, N., & Altman, M. (2016). Immersive audio for vr. In *Audio engineering society conference: 2016 aes international conference on audio for virtual and augmented reality*. Audio Engineering Society.

Turchet, L., Fischione, C., Essl, G., Keller, D., & Barthet, M. (2018). Internet of musical things: Vision and challenges. *IEEE Access*, *6*, 61994–62017.