

Real-time Audio Coordination Framework for Immersive Sound Reproduction

Hyosu Kim, Jiyeon Lee, Hwidong Bae, Insik Shin

Dept. of Computer Science

KAIST, South Korea

{hskim, jylee, spilist}@cps.kaist.ac.kr, insik.shin@cs.kaist.ac.kr

Accurate sound reproduction - as close to the original sound as possible - creates a rich, engaging user experience. Music fans would like to listen to music on fine-tuned high-end audio systems, embracing the feeling of being at a live concert. The widespread use of mobile devices opens an opportunity to enable a new class of audio applications, namely *Mobile Multi-speaker Audio* (MMA) applications; they utilize loudspeakers on multiple mobile devices to deliver immersive in-situ sound reproduction even in poor acoustic setting (e.g., outdoors). As shown in Figure 1, imagine a group of friends on a camping trip. After dinner, one suggests watching an action movie together. They find no video/audio equipment at the camping site but the loudspeakers on their own smartphones. They have little expertise and experience for audio set-up. They would be happy with a click-away construction of a spectacular surround sound for movie through MMA applications.

Over several decades, many advanced acoustical systems such as multichannel surround sound systems have been developed for immersive sound reproduction [1], [2]. The systems typically create sounds by simulating two most important auditory cues in human sound perception: the difference in sound arrival time and the difference in sound pressure level [3]. For example, if multiple correlated sounds arrive having the arrival time difference below 1ms, listeners only perceive one sound and its direction is between the locations of the lead and lag sounds. Another acoustic effects appear when the correlated sounds arrive within the range of 1ms and 30ms after the first-arriving sound. An auditory event is then dominantly localized by the first-arriving sound and the delayed sounds are perceived as the reverberation of the first sound [4].

As such, audio systems provide high-quality acoustic features especially based on subtle difference in sound arrival time. Thus, for immersive sound reproduction, the systems require a high precision control over inter-speaker sound arrival time (up to 1ms). However, commodity mobile devices do not offer tight timing guarantees on the operation of playing audio signals, and thus systematically incur large and unpredictable time delays. Such a limited timing support generally causes an unintended sound arrival time difference of high magnitude and high variance between devices. For example, the difference is randomly distributed in [0.1, 35.7ms] for two phones running Android.

This work proposes RAC, a new *Real-time Audio Coordination* framework to make inter-speaker sound arrival time accurate and predictable. For the accuracy, RAC captures sound arrival time of devices using acoustic control signals, and adjusts the playback start time of each device adaptively based on the observation. It also introduces a new audio



Fig. 1: Usage examples of RAC

playback API which guarantees the predictability of both audio processing time and hardware preparation time. More specifically, RAC gives an enough preparation time interval for audio processing, and prevents needless device preparations so that a speaker immediately plays audio streams.

We implemented a prototype of RAC on commodity smartphones running Android 4.2.2, and achieved the highly-accurate sound arrival time synchronization in the order of tens of microsecond. In this demonstration, we show a MMA application, MobileTheater, implemented on top of the newly-introduced RAC API. MobileTheater basically runs with six devices: five audio players and one video player. It then playbacks a movie with 5.1 surround sound in coordination under the RAC framework.

REFERENCES

- [1] F. Toole, *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*. Taylor & Francis, 2009.
- [2] M. A. J. Baalman, "Spatial Composition Techniques and Sound Spatialisation Technologies," *Organised Sound*, vol. 15, no. 3, pp. 209–218, 2010.
- [3] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, revised edition, 1997.
- [4] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633–1654, 1999.