Demo: A Spatial Audio System for the Internet-of-Things

Frank Liu, Robert LiKamWa
Arizona State University
Tempe, Arizona
fwliu1@asu.edu, likamwa@asu.edu

ABSTRACT

Current speakers on IoT hubs are limited, as they do not provide spatial audio experiences. Current spatial audio techniques do not scale to IoT needs, due to high cost, maintenance, and latency. To provide spatial audio for existing IoT environments, we propose and demonstrate a distributed spatial audio system that uses time-domain crosstalk cancellation among a multi-speaker IoT system to position virtual audio throughout a room.

1 The need for spatial audio in IoT

Spatial audio – the ability to discern directionality of sound – would add significant richness to the Internet of Things (IoT) application space. Virtual multidirectional sounds could bring to life an audio-based home guidance system or search engine. Smart speakers on IoT hubs, e.g. Amazon Echo or Google Home, do not currently provide spatial audio to the users; audio sounds like it emanates directly from the device. We propose to develop a system that leverages existing IoT infrastructure to create sounds that seem to emanate from arbitrary physical locations.

1.1 The state-of-the-art of spatial audio

Binaural synthesis processing frameworks, e.g., Resonance Audio, translate spatial virtual sound environments into two-channel audio signals delivered through headphones. However, delivering spatial audio through loudspeakers creates *crosstalk*, where unwanted sound from each speaker reaches the opposite ear. *Wave field synthesis* creates the effect by carefully synchronizing speaker arrays, which is both expensive and non-scalable. *Ambisonic* [3] and *amplitude panning* [2] systems create the effect by controlling the amplitude of surrounding loudspeakers, limiting the virtual sounds to a line segment between loudspeakers. *Dynamic crosstalk cancellers* use pairs of loudspeakers to create spatial audio for a tracked user through frequency-domain audio processing but have issues with high latency and inaccuracy [4] or require significant calibration to model a room and user's transfer function [1].

Hotmobile'19, Feb 27-28, 2019, Santa Cruz, CA, USA

ACM ISBN 978-1-4503-6273-3/19/02
https://doi.org/10.1145/3301293.3309567

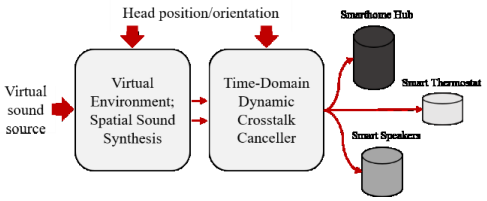

Figure 1: Diagram of the dynamic spatial audio system

2 Our demo: a proposed spatial audio system

We need spatial audio systems built from *existing* IoT infrastructure, and that adapt to the user and environment. We propose a distributed spatial audio system (Fig. 1), revolving around a time-domain dynamic crosstalk cancellation technique which uses amplitude gain and time delay modifications for real-time low-latency spatial audio.

In this demonstration we integrate: (i) The Unity Game Engine to place virtual sounds, (ii) Google Resonance Audio to generate spatial audio, (iii) HTC Vive tracking to provide user head location, (iv) our time-domain crosstalk cancellation algorithm, (v) multiple speakers, representing IoT devices. We compare our system against a headphones-based system to assess efficacy against the state-of-the-art.

REFERENCES

[1] M. Song, C. Zhang, D. Florencio, H. Kang. 2011. An Interactive 3-D Audio System with Loudspeakers. In IEEE Transactions on Multimedia (Volume: 13, Issue: 5, Oct. 2011)
[2] V. Pullki. 1997. Virtual sound source positioning using vector base amplitude panning. J. Audio Eng. Soc., vol. 45, pp. 456–466.
[3] D. Artega. 2018. Introduction to Ambisonics.
[4] H. Kurabayashi, M. Otani, K. Itoh, M. Hashimoto, M. Kayama. 2013. Development of dynamic transaural reproduction system using non-contact head tracking. In Proc. of IEEE Global Conference on Consumer Electronics (GCCE), (Tokyo, Japan).