Media Device Orchestration for Immersive Spatial Audio Reproduction

Jon Francombe
Russell Mason
Philip J.B. Jackson
Tim Brookes
University of Surrey
Guildford, Surrey GU2 7XH, UK
j.francombe@surrey.ac.uk

Richard Hughes
James Woodcock
University of Salford
Salford M5 4WT, UK

Andreas Franck
University of Southampton
Southampton SO17 1BJ, UK

Frank Melchior
Chris Pike
BBC Research & Development
MediaCityUK, Salford M50 2LH, UK

ABSTRACT

Whilst it is possible to create exciting, immersive listening experiences with current spatial audio technology, the required systems are generally difficult to install in a standard living room. However, in any living room there is likely to already be a range of loudspeakers (such as mobile phones, tablets, laptops, and so on). "Media device orchestration" (MDO) is the concept of utilising all available devices to augment the reproduction of a media experience. In this demonstration, MDO is used to augment low channel count renderings of various programme material, delivering immersive three-dimensional audio experiences.

AM '17, London, United Kingdom

© 2017 Copyright held by the owner/author(s). This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of AM '17, August 23–26, 2017*, <https://doi.org/10.1145/3123514.3123563>.

Figure 1: 22.2 system in the Institute of Sound Recording listening room

High-quality, immersive listening experiences can be produced by specialist reproduction systems (such as the 22.2 system in the Institute of Sound Recording listening room, pictured above). However, there are many restrictions that prevent such systems from being installed in realistic living rooms. For example, they might require a restrictively large number of loudspeakers (e.g. in the case of techniques such as wave field synthesis, or high channel count surround sound systems), or strictly accurate loudspeaker positioning (e.g. in higher-order ambisonics or channel-based surround sound). Furthermore, where multiple listeners are using the same space, it is not possible for them to be simultaneously in the sweet spot of a system, even if the loudspeakers are correctly positioned.

Sidebar 1: Operational constraints in domestic audio

CCS CONCEPTS

• **Applied computing** → **Sound and music computing**; • **Hardware** → *Sound-based input / output*;

KEYWORDS

Spatial audio; internet of things; audio perception; object-based audio; quality of listening experience

ACM Reference format:

Jon Francombe, Russell Mason, Philip J.B. Jackson, Tim Brookes, Andreas Franck, Richard Hughes, James Woodcock, Frank Melchior, and Chris Pike. 2017. Media Device Orchestration for Immersive Spatial Audio Reproduction. In *Proceedings of AM '17, London, United Kingdom, August 23–26, 2017*, 5 pages. DOI: 10.1145/3123514.3123563

INTRODUCTION

There are many spatial audio systems that can produce exciting, immersive listening experiences. However, such systems often have operational constraints that make them difficult or impossible to implement in realistic domestic situations (see the sidebar for examples). Because of such limitations, access to high-quality immersive listening experiences is generally limited to specialist facilities; bringing immersive audio experiences into the living room could benefit a huge number of consumers.

There has been much focus in the audio research community on developing systems that work well in laboratory settings but are not feasible in domestic reproduction. The drive towards higher channel count surround sound systems is an example of this; for example, it is possible to create excellent listening experiences with 22.2 reproduction (listeners have been shown to prefer replay systems with rear loudspeaker and height channels [2, 9]), although it is unlikely that this will ever see high uptake in living rooms. There is also a tendency to evaluate systems on their ability to produce stable and localisable virtual images; whilst this is certainly a desirable characteristic of a high-end system, preference studies have generally found that listeners—particularly untrained consumers—find the sensation of envelopment or surround sound fidelity more desirable than frontal spatial fidelity [2, 8], and that timbral fidelity is more important than spatial accuracy [7]. In a meta-analysis of attribute elicitation studies, Mason [4] found that higher-level attributes that might be peripherally related to localisation (such as envelopment) were more important than attributes specific to the location of a sound; where such location-specific sounds were important, distance was cited more often than azimuthal accuracy.

In order to create immersive spatial audio experiences that are widely accessible (i.e. that can be implemented in realistic living rooms), it is desirable to utilise equipment that consumers already own and are familiar with, rather than attempting to encourage uptake of reproduction methods that only work well in carefully controlled laboratory conditions. Whilst listeners may rarely be prepared

In January 2016, the S3A project organised a “hack week” event, hosted by the BBC Audio R&D group at MediaCityUK, Salford. The idea behind the week was to gather a group of researchers and a sound designer for an intensive week to work on investigating the creative possibilities that are opened up when mobile sound producing devices are used alongside an installed sound system. The hack week format was felt to be a good way of quickly prototyping ideas to determine if there was a benefit to this kind of technology and suggesting questions for further research. During the week, a number of pieces of demonstration content were produced, using a combination of a high quality 5.1 system (or stereo downmix) and a collection of loudspeakers of different qualities connected using Bluetooth or WiFi. Reactions to the demonstrations were positive, with many comments about the immersive and engaging listening experience, and excitement about the potential for widespread delivery (compared with similarly immersive listening experiences using more traditional high channel count loudspeaker setups).

Sidebar 2: S3A project “hack week”

Figure 2: Researchers at the hack week

to install high channel count surround sound systems in their living rooms, it is likely that there is already a considerable number of loudspeakers available; for example, loudspeakers built into televisions, games consoles, laptops, mobile phones, tablets, or other portable devices. As internet of things technologies continue to develop, it is increasingly likely that it will be possible to integrate such devices with existing reproduction systems. ‘Orchestration’ is an internet of things concept, defined as “multiple connected devices aware of each other working together” [1]. As technology supporting the internet of things develops, it becomes more likely that the *ad hoc* devices described above can be connected to one controller unit and used flexibly for spatial audio reproduction. In the S3A project (<http://www.s3a-spatialaudio.org>), preliminary investigations looking at how to use orchestration for optimal reproduction of immersive spatial audio have been conducted. The term “media device orchestration” (MDO) has been coined to describe the concept of utilizing any available devices for optimum reproduction of a media experience.

In this demonstration, we use MDO to enhance a listening experience by intelligently routing signals from an object-based audio mix to appropriate devices (for example, loudspeakers positioned at certain positions or near to particular items in the room, or loudspeakers of a certain quality). The demonstration was devised during a “hack week” conducted by the S3A project (described in the sidebar). Whilst this demonstration shows one application of MDO, there is vast potential for using such a network of connected devices for a variety of new and improved multimedia experiences.

MEDIA DEVICE ORCHESTRATION

As discussed above, few consumers will install high channel count systems in their living rooms. However, it is likely that there will already be a large number of loudspeakers in any given living room. Many people will have one or more personal portable devices with a loudspeaker (e.g. a smart phone or tablet); it is increasingly common for television viewers to access (and create) additional content on a “second screen” device [5]. Such devices could potentially be utilised as part of a spatial audio reproduction system, enabling access to multiple extra loudspeakers—and therefore being able to reproduce sound from multiple directions—without the requirement for the listener to purchase, install, position, and calibrate a specified number of devices.

Using an *ad hoc* array of devices in this fashion does have some limitations. For example, with a matched pair or trio of loudspeakers at known positions, it is possible to use amplitude panning to create ‘virtual sources’ at any position between the loudspeakers [6]. This is unlikely to be possible—or will at least present considerable challenges—with an array of different connected devices; for example, the frequency response and directionality of such devices is likely to be unmatched, and there may be different latencies that need to be accounted for. However, aside from these limitations, there are also various possibilities that are unlocked compared to a standard channel based system. For example, whilst installed loudspeakers are generally positioned equidistantly from the sweet spot,

Channel-based audio

For many years, audio content has been mixed for a specific loudspeaker setup, producing output channels prior to transmission. These audio channels are intended for reproduction by a set of loudspeakers at known ideal positions. Consequently, it is difficult to repurpose channel-based content for different systems.

Object-based audio

In object-based audio, an audio scene is transmitted as a set of objects (a single object might contain a certain aspect of the scene, e.g. the dialogue from a particular character) alongside accompanying metadata that describes how the objects should be reproduced. The loudspeaker signals can be created remotely (by a *renderer*), enabling adaptation to any loudspeaker setup (object-based audio is often described as being *platform agnostic*), as well as providing other benefits (such as personalisation).

Sidebar 3: Introduction to object-based audio

ad hoc devices are likely to be positioned at a range of distances; this means that distance cues can be enhanced or added into audio reproduction, benefitting an aspect of spatial audio reproduction that is often overlooked [4]. There may also be benefits of having devices with different frequency responses, particularly for emphasising the spatial offset of loudspeakers.

The number and position of the installed loudspeakers and additional devices utilised by MDO are likely to be different in each listening space. Current channel-based content delivery is sub-optimal for reproduction over such systems; therefore, the proposed MDO system would take advantage of the flexibility offered by object-based audio (see the sidebar for a brief introduction).

DEMONSTRATION

This demonstration features MDO reproduction using an object-based audio system; it is intended to highlight the improved immersion that can be achieved when a low channel count reproduction system is augmented with extra *ad hoc* devices and when object-based audio is used, allowing redistribution of the objects. Various object-based audio content is used; the audio objects are replayed from broadcast wave (BWF) files with the metadata stored in the BWF header as an XML chunk in the audio definition model (ADM) format [3] (augmented by an additional XML chunk of advanced metadata not currently defined in the ADM). The audio reproduction system is a two-channel ‘installed’ system with reasonable quality loudspeakers, which is augmented by a number of connected devices—in this case, small to mid-size portable loudspeakers connected by Bluetooth, WiFi, or cables. Loudspeaker feeds are created in real time by the *VISR* renderer (developed in the S3A project). In the enhanced MDO rendering, a number of the audio objects are routed to the extra devices. The top-level routing decisions are implemented using metadata adaptation in the *Metadapter*, a Python software package developed in the S3A project. A system diagram is shown in Figure 3.

DISCUSSION

The system being demonstrated has been found in informal listening tests to produce an exciting and immersive listening experience; the *ad hoc* loudspeakers augment the capabilities of the channel-based system even though they are fairly low quality and not positioned in controlled locations. The additional speakers give sounds a clear, distinct location in three-dimensional space (including enhanced distance cues). Further formal evaluation of the listening experience is planned.

The utilisation of extra devices for reproduction, coupled with the flexibility of object-based audio transmission, offers great potential for content creators and sound designers. The fact that MDO reproduction does not rely on a specific loudspeaker setup means that the producer’s intended emotional impact and immersion can be recreated regardless of the available devices. However, there are also potential pitfalls that will need further investigation. For example, taking an audio

Figure 3: MDO demo system diagram

object out of a surround sound mix and into a separate loudspeaker means that artefacts that were masked could potentially become audible.

In this demonstration, MDO was used for spatial audio reproduction; however, it could be extended to multimodal experiences. Some possibilities include: second screen video, virtual and augmented reality, connected lighting, haptics, interaction through wearable technology, computer gaming, and movies. Future work will focus on: developing the technology used to make best use of the *ad hoc* loudspeaker array; new methods of MDO reproduction; learning about the perception of MDO; working with content creators to find the best use for such technology; and extending MDO to multimodal experiences.

ACKNOWLEDGMENTS

This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1). The authors would like to acknowledge the input of many researchers on the S3A project who have contributed to developing the ideas behind media device orchestration as described above.

REFERENCES

- [1] Jennifer Elias. 2016. Google says "orchestration" is next big trend in the Internet of Things. (2016). <https://www.bizjournals.com/sanjose/news/2016/03/15/google-says-orchestration-is-next-big-trend-in-the.html>
- [2] Jon Francombe, Tim Brookes, Russell Mason, and James Woodcock. 2017. Evaluation of spatial audio reproduction methods (part 2): analysis of listener preference. *J. Audio Eng. Soc.* 65, 3 (March 2017), 212–225. DOI: <https://doi.org/10.17743/jaes.2016.0071>
- [3] ITU-R rec. BS.2076-0. 2015. *Audio definition model*. Technical Report. ITU-R Broadcasting Service Series.
- [4] Russell Mason. 2017. How important is accurate localisation in reproduced sound?. In *AES 142nd Conv.* Berlin, Germany.
- [5] Gianmarco De Francisci Morales and Ajay Shekhawat. 2013. The future of second screen experience. In *ACM Conference on Computer-Human Interaction*. Paris, France.
- [6] Ville Pulkki. 1997. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.* 45, 6 (June 1997), 456–466.
- [7] Francis Rumsey, Slawomir Zieliński, Rafael Kassier, and Søren Bech. 2005. On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *J. Acoust. Soc. Am.* 118, 2 (Aug. 2005), 968–976. DOI: <https://doi.org/10.1121/1.1945368>
- [8] Francis Rumsey, Slawomir Zieliński, Rafael Kassier, and Søren Bech. 2005. Relationships between experienced listener ratings of multichannel audio quality and naïve listener preferences. *J. Acoust. Soc. Am.* 117, 6 (June 2005), 3832–3840. DOI: <https://doi.org/10.1121/1.1904305>
- [9] Andreas Silzle, Sunish George, Emanuel A. P. Habets, and Thomas Bachmann. 2011. Investigation on the quality of 3D sound reproduction. In *Proceedings of the first International Conference on Spatial Audio, Detmold, Germany, 10–13 November 2011*. 334–341.