Audio Engineering Society

Convention e-Brief 463

Presented at the 145th Convention
2018 October 17 – 20, New York, NY, USA

Creating Object-Based Stimuli to Explore Media Device Orchestration Reproduction Techniques

Craig Cieciura[1], Russell Mason[1], Philip Coleman[1], and Matthew Paradis[2]

[1]*Institute of Sound Recording, University of Surrey, Guildford, UK*
[2]*BBC Research & Development, Salford, UK*

Correspondence should be addressed to Craig Cieciura (c.cieciura@surrey.ac.uk)

ABSTRACT

Media Device Orchestration (MDO) makes use of interconnected devices to augment a reproduction system, and could be used to deliver more immersive audio experiences to domestic audiences. To investigate optimal rendering on an MDO-based system, stimuli were created via: 1) object-based audio (OBA) mixes undertaken in a reference listening room; and 2) up to 13 rendered versions of these employing a range of installed and ad-hoc loudspeakers with varying cost, quality and position. The programme items include audio-visual material (short film trailer and big band performance) and audio-only material (radio panel show, pop track, football match, and orchestral performance). The object-based programme items and alternate MDO configurations are made available for testing and demonstrating OBA systems.

1 Introduction

Creating immersive media experiences in the home is a topic popular in both academic and mainstream publications. The positive effects of spatialised audio on listener preference have been demonstrated [1, 2], and technologies for producing 3D spatial audio experiences are well established, with channel-based systems being most commonly used including standardised implementations incorporating up to 24 loudspeakers [3]. Whilst the effectiveness of these systems has been well demonstrated in laboratory and cinema-type environments, implementation in the domestic environment faces multiple obstacles including the seemingly poor tradeoff between the cost of additional loudspeakers and improvement in listening experience [4], a lack of knowledge and confidence in the setup of such a system, and a negative perceived effect on room aesthetics. The process of installing discrete channel-based systems has been reviewed in popular media, and soundbars were recommended as an alternative [5]. However, an experiment comparing two types of soundbars to both discrete surround and discrete two-channel stereo found that the soundbar systems performed less well than either format, based on a combination of timbral and spatial factors [6].

A new reproduction approach called "Media Device Orchestration" (MDO) has been described in a recent journal paper [7]. This approach looks to combine the flexibility of object-based audio (OBA) with, to make installation simpler, the increasing prevalence of interconnected devices in the home by utilising a combination of installed and ad-hoc audio capable devices controlled by an intelligent renderer.

In their paper, Francombe et al. [7] describe a system

which could be made up of any combination of ad-hoc and installed, wired and wireless loudspeakers, choosing to test a version consisting of an installed, high quality stereo pair in the front quadrant, one higher-quality ad-hoc loudspeaker on a centrally positioned table close to the audience position and three lower-quality ad-hoc loudspeakers on the left, rear and right of the audience position. Due to the variety of reproduction systems and flexibility of object-based audio, MDO could be used to enhance existing reproduction formats, such as built-in TV loudspeakers or discrete stereo, or to create novel listening experiences by re-rendering audio objects using rules based on object and loudspeaker quality, original position information and narrative importance of objects, as well as semantic information such as programme type and purpose, and listener customisation elements for accessibility and creative preference.

To explore these potential applications, beyond the current published scope of MDO evaluation, a two-part stimulus set was created. The first part of the set consists of mixed object-based programme items, with metadata captured in the Audio Description Model (ADM) format [8]. The second part consists of downmixes and alternate MDO rendering configurations of each of the items, from which loudspeaker feeds have been captured. The initial process of selecting and obtaining a broad range of programme items is described in Section 2, followed by a description of the process of creating the first part of the stimulus set in Section 3. The selection and creation of the alternate MDO versions is described in Sections 4 and 5, including the process of loudness matching.

Extracts from both parts of the stimulus set are made available for use to research and to demonstrate new reproduction technologies.

2 Selecting Suitable Programme Items

When examining the perceptual effects of varying spatial parameters, audio is often delivered in isolation (i.e., without accompanying pictures). However, a 2016 report by the UK communications regulator OFCOM stated that 38.7% of media consumption time was spent *watching* (the mean percentage calculated from figures given for 7 age groups) versus 18.7% of time *listening* [9]. To create a stimulus set for testing potential MDO applications, a system for reproducing content in the living room environment, both audio-only and audio-visual content is required, as the rules governing re-rendering of objects could be significantly effected by a visual component to the media. An existing dataset, consisting of three radio-dramas, has been produced by The S3A Project [10].

The BBC produces various genres of audio/audio-visual content[1] containing different combinations of production elements, and of sound objects which can be classified in the perceptual categories determined by Woodcock et al. [11]. To create a broad range of stimuli, the combinations of perceptual and production elements were considered and a desirable list of programme genres was created: *Drama, Factual, Music and Sport*, for which separate audio and audio-visual programmes were desired.

3 Creating the mixed object-based programme items

Six out of the eight desirable programme items were obtained, providing a reasonably representative spread of broadcast items, both audio-only and audio-visual. These consist of: (audio-visual) a short film trailer, a live big band performance, (audio-only) a radio panel show, a pop track, a football match recorded for radio, and an orchestral recording. Where possible, full project files and complete stems for each of the base programme items were obtained. For some items, limited stems were obtained, such as a stereo music mix and audience microphone feeds.

One of the principles of MDO is that the end reproduction format is unknown. As such, rather than mixing for any specific arrangement of loudspeakers, the production methodology was to initially mix for a system with the highest reproduction quality anticipated to be available for home listening, and then to create rendered configurations from this idealised mix. The OBA material was mixed in an ITU-R BS 1116 standard listening room equipped with a 22.2 system at the University of Surrey [12], with additional monitoring on a high-quality 2-channel stereo system.

A mixing engineer with extensive experience of mixing a variety of content for broadcast was employed to prepare the majority of the object-based programme items. Two production systems were used: the IOSONO Spatial Audio Workstation hosted in Steinberg's Nuendo

[1] https://www.bbc.co.uk/programmes/genres

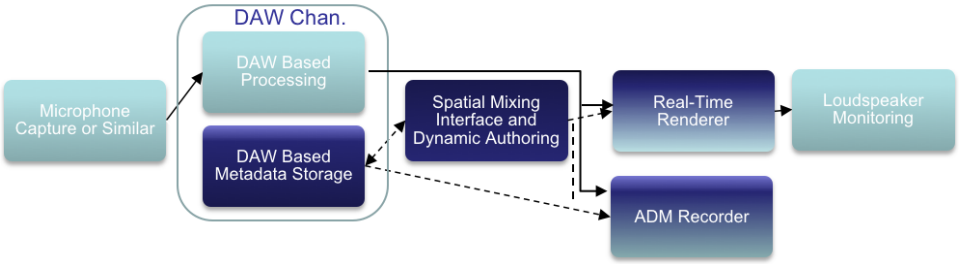

Fig. 1: Flow of audio and control data in both production systems. Light blue and solid lines represent the audio, and dark blue and dashed lines the metadata.

version 5.5, rendered using the S3A project renderer [13, 14]; and IRCAM's Panoramix software [15], a standalone system for mixing, panning and rendering spatial audio, with Avid Pro Tools HD v12 Native running IRCAM's ToscA plugin [16] to transmit control data in the Open Sound Control format between the two programmes. Metadata consisting of *azimuth* and *elevation* was stored for each item. Whilst both systems supported *distance*, this was not used as the distance-based sound effects in each programme were not standardised and the mixing engineer preferred to alter the perceived distance of each object using manual techniques including altering relative levels and direct to reverberant sound ratios. The Pro Tools system was predominantly used due to greater operator proficiency. Figure 1 displays an overview of the passage of audio and metadata within the two systems.

The final object-based programme items were captured using an IRCAM programme which receives the parallel audio and metadata streams and exports an ADM compatible broadcast WAV file.

4 Selecting the MDO Configurations

The next step was to determine alternative reproduction configurations to examine different potential rules for MDO. Reproduction configurations, in this context, refers to re-rendering the objects from the programme items described above and positioning them in specific positions or loudspeakers. The configurations were determined in groups, according to purpose, based on previous research in the S3A project[2], first hand discussions with various audio professionals both research and production based, and the first author's own production experience.

[2] http://s3a-spatialaudio.org/

The groups of reproduction configurations determined were:

- Channel-Based: configurations using the built-in TV loudspeakers, high-quality stereo loudspeakers and 5.0 system;
- Narrator-Feature: where the narrator or commentator object was positioned in either a low or medium quality ad-hoc loudspeaker;
- Spatialised-Dialogue: where all dialogue objects were positioned in different ad-hoc loudspeakers;
- All-Dialogue: in which the dialogue objects were reproduced from all loudspeakers simultaneously;
- Coarse-Quadrant: in which each quadrant of the room was treated as a zone and any object with metadata positioning them in each zone was reproduced from all loudspeakers in that zone (with appropriate level reduction);
- Vector-based amplitude panning (VBAP) [17]: VBAP without additional processing, incorporating the built-in TV and ad-hoc loudspeakers;
- Manual Rendering: two versions, one using the TV loudspeakers and one using the high-quality stereo loudspeakers, with manual positions designed to emulate how an intelligent renderer might operate, by combining some of the above configurations with object placement based on loudspeaker quality.

To examine the appropriateness of the selected configurations, a consultation with three industry experts and two spatial audio researchers was performed. The

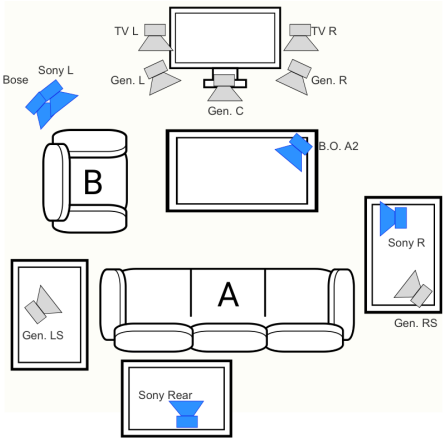

Fig. 2: *MDO Lab* room layout. Installed loudspeakers are coloured grey and ad-hoc loudspeakers in blue. The *Bose* and *B.O* labelled ad-hoc loudspeakers are of comparable and higher reproduction quality than the *Sony* labelled loudspeakers. The *Sony R* and *Gen. RS* loudspeakers are positioned on shelves on a bookcase.

aim was to discuss whether the reproduction configurations, applied to the OBA programme items, met three criteria:

1. Enhancing existing listening experiences;
2. Creating new types of listening experiences;
3. Informing about undesirable listening experiences.

For each reproduction configuration which was appropriate for each programme item, e.g. items without a commentator or narrator would not be re-rendered with a Narrator-Feature configuration, the majority of configurations were kept.

5 Creating the MDO Configurations

To create the stimuli, the reproduction configurations were then applied to the appropriate OBA items. The Pro Tools/Panoramix system was transferred to *The MDO Lab*, a 4.7m x 4.3m room outfitted to act as a living-room-type environment with the addition of various installed and ad-hoc loudspeakers. These included a stereo pair built into a TV; a 5.0 system positioned 'realistically' (loudspeakers were placed on available items of furniture rather than being positioned according to ITU-R BS.775-3 [18]), to reflect how such a system might be installed in a living room; and five ad-hoc loudspeakers of varying qualities and positions. A diagram of loudspeaker positions in *The MDO Lab* is displayed in Figure 2. Each stimulus was created in the Pro Tools/Panoramix system, from which the loudspeaker outputs were recorded.

To calibrate the system, pink noise was played from the calibration software at -28dBFS and measured at 68.5dBLeq, unweighted, in listening position A labelled in Figure 2. To match the levels between each stimulus based on the same programme item, the 5.0 version was set to a comfortable listening level, then the other versions were matched via an automated Matlab script incorporating the perceptual model for time-varying sounds by Zwicker and Fastl [19] using the implementation provided in the Genesis Loudness Toolbox [20].

6 Data Access

The second part of the stimulus set, specifically the alternate MDO configurations, has been used in two experiments at The University of Surrey, looking at potential applications for MDO, as well as multiple demonstrations to departmental visitors and the public. These excerpts are hosted at DOI: `10.5281/zenodo.1404797` along with license terms, such as referencing this E-Brief. Due to difficulties in obtaining the rights to distribute content produced by professional production companies and containing copyrighted material, only certain elements of the stimulus set can be made available to the public.

7 Acknowledgements

This work was supported as part of an EPSRC Industrial CASE PhD Studentship co-sponsored by BBC Research and Development. The first author would also like to thank Rupert Flindt for his efforts in mixing the majority of the OBA items and in obtaining several of them, including *Just Another Frame* by *The Hotel Whisky Foxtrot*. Thanks also to Jamie Gamache of LOWKEY Films for providing project files for short film: *Wander,* and permission to release derivatives created during the production of the stimuli described in this paper.

References

[1] S. Choisel and F. Wickelmaier, "Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference," The Journal of the Acoustical Society of America, vol. 121(1):388–400, (2007 Jan.). DOI: 10.1121/1.2385043.

[2] J. Francombe, T. Brookes, R. Mason and J. Woodcock, "Evaluation of Spatial Audio Reproduction Methods (Part 2): Analysis of Listener Preference," J. Audio Eng. Soc., vol. 65(3), pp. 212–225, (2003 Mar.).

[3] ITU-R rec. BS.2051, "Advanced Sound System for Programme Production," ITU-R Broadcasing Service (Sound) Series (2018 Jul.).

[4] T. Holman, "Surround Sound: Up and Running," (Taylor & Francis, 2008), ISBN: 978-0-240-80829-1, pp. 18

[5] E. A. Taub, "Reviewing Sound Bars: An Alternative to TV Home Theater Systems," The New York Times, (2017 Dec.).

[6] T. Walton, M. Evans, D. Kirk and F. Melchior, "A Subjective Comparison of Discrete Surround Sound and Soundbar Technology by Using Mixed Methods," presented at the 140th Convention of the Audio Engineering Society, convention paper 9592, (2016 May).

[7] J. Francombe, J. Woodcock, R. J. Hughes, R. Mason, A. Franck, C. Pike, T. Brookes, W. J. Davies, P. J. B. Jackson, T. J. Cox, F. M. Fazi and A. Hilton, "Qualitative Evaluation of Media Device Orchestration for Immersive Spatial Audio Reproduction," J. Audio Eng. Soc., vol. 66(6), pp. 414–429, (2018 Jun.).

[8] S. Füg, D. Marston and S. Norcross, "The Audio Definition Model—A Flexible Standardized Representation for Next Generation Audio Content in Broadcasting and Beyond," presented at the 141st Convention of the Audio Engineering Society, convention paper 9626, (2016 Sep.).

[9] S. Cape and J. Rees, "Communications market report 2016," The Office of Communications (2016 Aug.).

[10] J. Woodcock, C. Pike, F. Melchior, P. Coleman, A. Franck and A. Hilton, "Presenting the S3A Object- Based Audio Drama Dataset", presented at the 140th Convention of the Audio Engineering Society, EBrief 255, (2016 May).

[11] J. Woodcock, W. Davies, T. Cox and F.Melchior, "Categorization of Broadcast Audio Objects in Complex Auditory Scenes," J. Audio Eng. Soc., vol. 64(6), pp. 380–394, (2016 Jun.). ISSN: 15494950. DOI: 10.17743/jaes. 2016.0007.

[12] R. Mason, "Installation of a Flexible 3D Audio Reproduction System into a Standardized Listening Room," presented at the 140th Convention of the Audio Engineering Society, EBrief 256, (2016 May).

[13] A. Franck and F. M. Fazi, "VISR—A Versatile Open Software Framework for Audio Signal Processing," presented at the AES International Conference on Spatial Reproduction - Aesthetics and Science, conference paper P9-2, (2018, Jul.).

[14] P. Coleman, A. Franck, J. Francombe, Q. Liu, T. de Campos, R. J. Hughes, D. Menzies, M. F. S. Gálvez, Y. Tang, J. Woodcock, P. J. B. Jackson, F. Melchior, C. Pike, F. M. Fazi, T. J. Cox and A. Hilton, "An Audio-Visual System for Object-Based Audio: From Recording to Listening," IEEE Transactions on Multimedia, vol. 20(8), pp. 1919– 1931, (2018, Aug.). DOI: 10.1109/TMM.2018.2794780.

[15] T. Carpentier, "Panoramix: 3D mixing and post-production workstation," presented at the 42nd International Computer Music Conference (ICMC), (2016 Sep.).

[16] T. Carpentier, "ToscA: An OSC Communication Plugin for Object-Oriented Spatialization Authoring," presented at the 41st International Computer Music Conference (ICMC), (2015, Sep.).

[17] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," J. Audio Eng. Soc., vol. 45(6), pp. 456–466, (1997, Jun.).

[18] ITU-R rec. BS.775-3, "Multichannel Stereophonic Sound System with and without Accompanying Picture," ITU-R Broadcasing Service (Sound) Series (2012 Aug.).

[19] H. Fastl, "Psychoacoustics - Facts and Models", Springer Series in Information Sciences. 2nd edition, (1999).

[20] GENESIS, "Loudness online", `http://genesis-acoustics.com/en/loudness_online-32.html`, Accessed: 2018-08-24.