Presenting the S3A Object-Based Audio Drama dataset

James Woodcock[1], Chris Pike[2], Frank Melchior[2], Philip Coleman[3], Andreas Franck[4] and Adrian Hilton[3]

[1]Acoustics Research Centre, University of Salford, Salford, UK*

[2]BBC R&D, Dock House, MediaCityUK, Salford, UK

[3]Centre for Vision, Speech, and Signal Processing, University of Surrey, Guildford, UK

[4]Institute of Sound and Vibration Research, University of Southampton, Southampton, UK

June 20, 2016

Abstract

This engineering brief reports on the production of 3 object-based audio drama scenes, commissioned as part of the S3A project. 3D reproduction and an object-based workflow were considered and implemented from the initial script commissioning through to the final mix of the scenes. The scenes are being made available as Broadcast Wave Format files containing all objects as separate tracks and all metadata necessary to render the scenes as an XML chunk in the header conforming to the Audio Definition Model specification (Recommendation ITU-R BS.2076 [1]). It is hoped that these scenes will find use in perceptual experiments and in the testing of 3D audio systems. The scenes are available via the following link: `http://dx.doi.org/10.17866/rd.salford.3043921`.

0 Introduction

As object-based audio becomes an increasingly important approach to broadcasting, there is a need for programme material that can be used to evaluate object-based reproduction systems and to effectively demonstrate the features of 3D audio. Object-based audio describes an approach to audio production, storage, and transmission whereby audio signals are stored

*j.s.woodcock@salford.ac.uk

and transmitted as separate audio objects along with metadata describing properties such as the spatial and temporal position of each object [2]. These data are then used to render the sound scene at the listener end, potentially removing the need to produce separate mixes for different loudspeaker layouts (i.e. a separate mono, stereo and surround mix). This is in contrast to traditional channel based broadcast where individual loudspeaker signals are stored and transmitted.

Object-based platforms are currently being developed both commercially [3, 4] and via standardisation [5]. Despite the growing interest in object based broadcast, there have been very few productions to date that have taken an object-based approach throughout the entire production process. BBC R&D have produced some radio content with aspects of object-based production and there has been some investigation into using object-based broadcast to develop responsive content.

Armstrong et al. [6] demonstrated the use of object-based audio to provide three alternate soundtracks to the same TV programme. Churnside et al. [7] used an object-based approach to produce a responsive audio drama in which the narrative adapted to the geographic location of the listener. In another production, Churnside [8] used object based audio to generate 2 broadcast quality mixes of a radio drama production of Pinnochio. Armstrong et al. [9] demonstrated the use of object-based broadcasting in the production of a variable length radio documentary. Each of these productions has demonstrated various benefits and challenges associated with object-based production; however, in each case the object-based aspects of the production have been additional to a traditional stereo or surround production meaning that, necessarily, compromises were made.

This engineering brief describes the production of 3 audio drama scenes. In each case, object based production and 3D reproduction were considered from the outset. These considerations therefore informed the generation of the initial scene ideas, scriptwriting, recording, and production. The aim of this exercise was threefold: 1) to observe and investigate how the workflow of an object based production differs from a traditional stereo production, 2) to identify the technical and creative challenges arising from the production, and 3) to develop a toolbox of object based productions that can be used to evaluate 3D audio systems and can be used as stimuli for perceptual testing.

1 The scenes

3 scenes were commissioned, each of which were designed to demonstrate different features of 3D audio. The first scene, Family, is a domestic drama in which a family are discussing attending a protest. The scene involves dialogue taking place at different levels of a family

home (the ground floor, the first floor, and the attic). The main technical challenges related to this scene included creating a convincing acoustic for the family home and creating convincing movement for the characters.

The second scene, Protest, depicts the protest which was being discussed in the Family scene. The protest is being staged outside a bank. The scene begins inside the bank and evolves from a front dominant image to full 3D sound as the action moves from indoors to the protest outside. The scene demonstrates immersive crowd atmos, individually identifiable voices popping out of the atmos, and moving localizable sources at different heights. The main technical challenges related to this scene involved recreating the acoustic of the bank lobby, and creating an immersive crowd atmosphere.

The third scene, The Turning Forest, is a fantasy scenario set in a forest. The scene opens with two children playing in an autumnal forest. After one of the children runs away, the remaining child encounters a large friendly monster. The child walks with the creature through the forest and rides on the creature's back as it swims across a river. When the creature and the child reach the other side of the river, the season has changed from autumn to winter. The scene includes a narrator throughout and non-diegetic music. The main technical challenges related to this scene involved creating a convincing 3D atmosphere for the forest and creating a creature that gave the impression of having large physical dimensions.

2 Recording sessions

Material for the 3 scenes was recorded over 3 days; 2 days in studio (MPAS studio at Dock10, MediaCityUK) and 1 day on location recording atmos and wildtrack for The Turning Forest scene. Throughout the process, there was a general aim to record all material dry and well separated from other sources. This was done to ensure that clean recordings of each source were available as audio objects for the mix.

2.1 Studio sessions

The main objectives of the studio sessions were to capture dialogue and to record material that could be used to build the crowd for the Protest scene. Dialogue for the Protest and Family scenes, along with the narration for The Turning Forest, was recorded in a semi-anechoic environment using Neumann 191 microphones. Each actor's dialogue was recorded on a separate channel and the actors were asked not to overlap their lines wherever possible.

An array of 12 Sennheiser MKH40 cardioid microphones facing into the room was used in the semi-anechoic space to capture material for the crowd in the Protest scene. These

microphones were distributed around the edge of the space at different heights. The crowd sounds were created by 15 people chanting and moving around the space. This was captured multiple times to allow the material to be layered in the mix.

2.2 Location recording session

A small wooded area located at Rode Hall, Cheshire, UK was used for the location recordings for The Turning Forest scene. The main objectives of this session was to capture atmos tracks, wildtrack, and some additional dialogue. 3 child actors were used for the additional dialogue and wildtrack. Atmos was captured using a double layered circular array of radius 2.5m. Each layer of the array consisted of 8 regularly spaced super-cardioid MKH8050 microphones. The microphones in the lower layer were 1.35m above the ground and pointed outwards. The microphones in the upper layer were 2.45m above the ground and pointed upwards and outwards. 4 additional MKH20 omnidirectional microphones were used in the lower layer to improve the low frequency response of the array. The array was designed to provide decorrelated signals but still give some sense of movement (i.e. wind moving through trees). A diagram and photograph of the array setup are shown in Figure 1 and Figure 2 respectively.

The child actors were equipped with lapel mics to capture dialogue and breathing. Wildtrack such as footsteps, rustling of leaves, and twigs snapping were captured using a Neumann 191 on a boom operated by the sound designer.

3 Mix sessions

The mix sessions took place in the listening room at BBC R&D in MediaCity. This listening environment meets the specifications set out in ITU BS.1116, and is equipped with 32 Genelec 8030B loudspeakers and 2 Genelec 7040 subwoofers. A detailed description of the production environment is given by Nixon et al. [10].

The main DAW used in the mix session was Nuendo v5.5 and the scene composition software used was the IOSONO Spatial Audio Workstation plugin (SAW) (v2.3.0). Audio was rendered using the VBAP based algorithm that is being developed as part of the S3A project. Reverb for the family scene and the protest scene was generated using the Max/MSP implementation of SPAT from IRCAM, which is feedback delay network based and takes into account individual object positions.

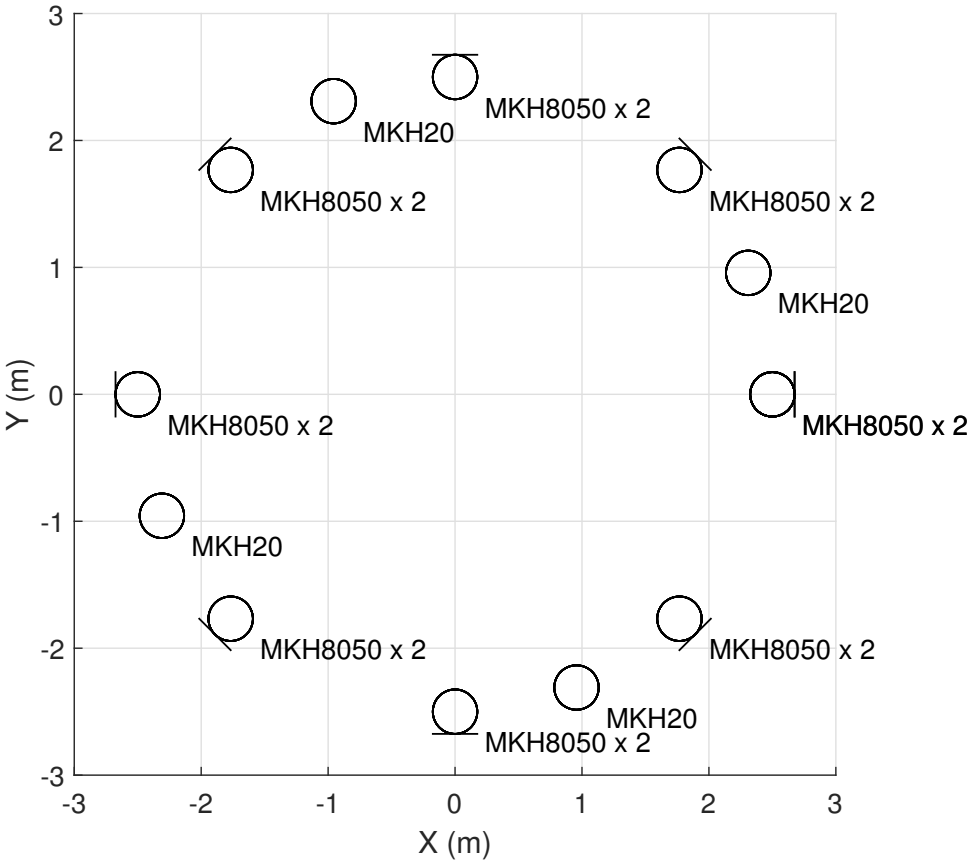

Figure 1: Plan view of the arrangement of microphones used in The Turning Forest atmos recoding session. Note, each of the MKH8050 positions includes 2 microphones, one at 1.35m elevation and one at 2.45m elevation.

4 The dataset

The data arising from these sessions are made available in object-based format as BWF files. The meta-data required to render the scenes is stored in the BWF header as an XML chunk using the Audio Definition Model [1]. The Audio Definition Model provides a formal description of the audio format within the BWF file, whether the signals are channel-based, higher-order ambisonics, or object-based. Figure 3 presents a simplified view of the model structure, the *¡chna¿* chunk within the BWF header uniquely identifies each track within the file. The audio format description then identifies what the signals within those tracks represent and relationships between them. For these drama scenes, the format description includes the dynamic 3D positioning data for the audio objects. The content description provides high-level descriptive meta-data, but also allows for grouping of related components of the content (e.g. dialogue and music) and representing alternate versions of the programme from subsets of the content (e.g. multiple languages).

A suite of C++ libraries for creating and editing ADM data and reading from and writ-

Figure 2: Array setup for The Turning Forest atmos recordings.

ing to BWF files is available from BBC R&D `http://www.bbc.co.uk/rd/publications/audio-definition-model-software`. These libraries were used in a software application to convert the DAW sessions to ADM BWF files. The scenes can be accessed from the following link: `http://dx.doi.org/10.17866/rd.salford.3043921`.

5 Summary

3 object-based audio drama scenes have been produced as part of the S3A project. The scenes are being made available as BWF files containing the meta-data as an XML chunk in the header. These data will be useful in perceptual experiments and in the testing of 3D audio systems.

6 Acknowledgements

This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and the BBC as part of the BBC Audio Research Partnership. The authors would like to thank Eloise Whitmore, the sound designer who worked on the production, and Shelley Silas who wrote the scripts.

The dataset is available from `http://dx.doi.org/10.17866/rd.salford.3043921`.

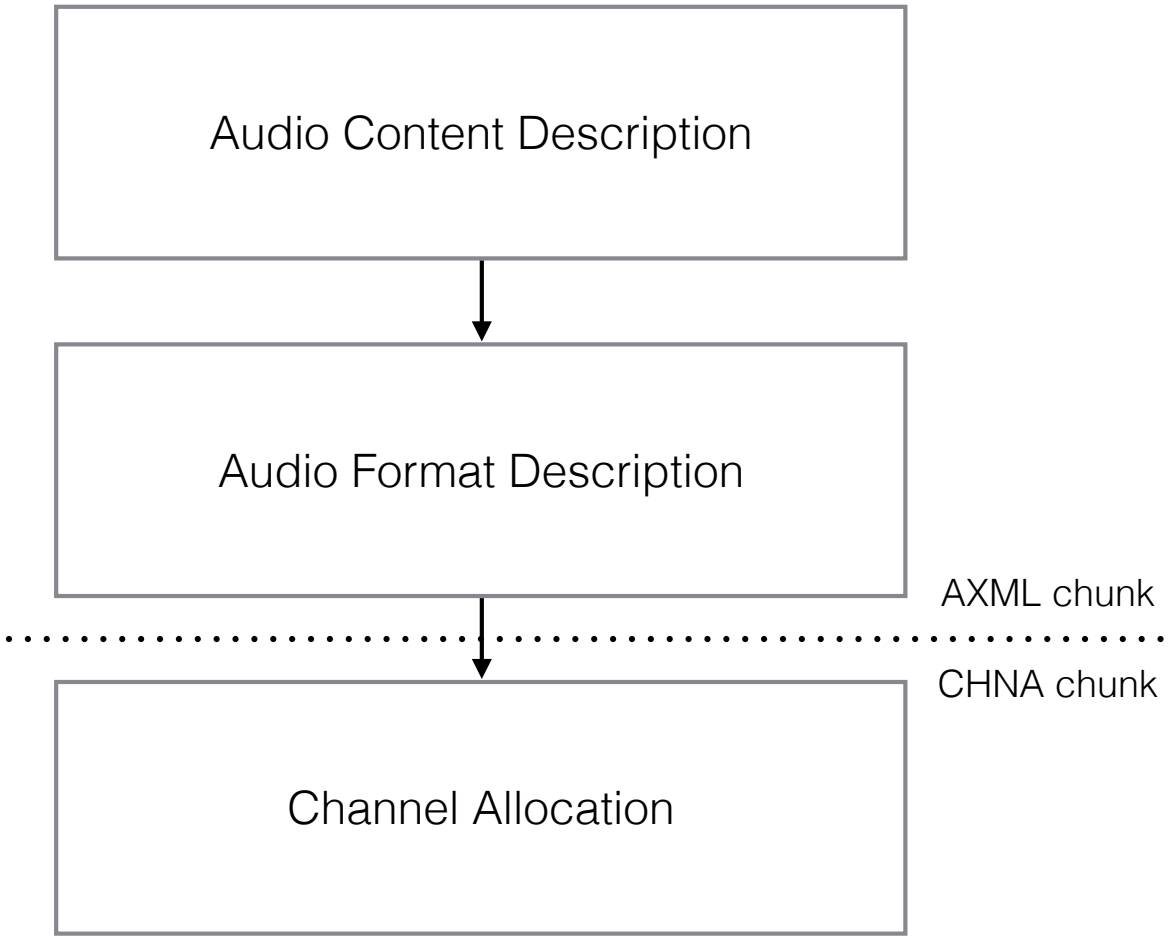

Figure 3: Simplified structure of the Audio Definition Model [1].

References

[1] ITU, "ITU BS.2076 : Audio Definition Model," 2015.

[2] Shirley, B., Oldfield, R., Melchior, F., and Batke, J.-M., "Platform independent audio," *Media Production, Delivery and Interaction for Platform Independent Systems: Format-Agnostic Media*, pp. 130–165, 2013.

[3] Mehta, S., Onders, T., and Riedmiller, J., "Recipes for Creating and Delivering Next-Generation Broadcast Audio," in *Annual Technical Conference and Exhibition, SMPTE 2015*, pp. 1–12, SMPTE, 2015.

[4] Jot, J.-M., Smith, B., and Thompson, J., "Dialog Control and Enhancement in Object-Based Audio Systems," in *Audio Engineering Society Convention 139*, Audio Engineering Society, 2015.

[5] Herre, J., Hilpert, J., Kuntz, A., and Plogsties, J., "MPEG-H 3D Audio—The New

Standard for Coding of Immersive Spatial Audio," *Selected Topics in Signal Processing, IEEE Journal of*, 9(5), pp. 770–779, 2015.

[6] Armstrong, M., Barrett, J., and Evans, M., "Enabling and enriching broadcast services by combining IP and broadcast delivery," *BBC Research & Development white paper*, 185, 2010.

[7] Churnside, A. and Forrester, I., "The Creation of a Perceptive Audio Drama," *NEM Summit, Istanbul, Turkey*, pp. 16–18, 2012.

[8] Churnside, A., "Pinocchio," `http://www.bbc.co.uk/rd/blog/2012/12/pinocchio`, 2012, [Online; accessed 2016-03-30].

[9] Armstrong, M., Brooks, M., Churnside, A., Evans, M., Melchior, F., and Shotton, M., "Object-based broadcasting-curation, responsiveness and user experience," in *IBC2014 Conference*, 2014.

[10] Nixon, T., Bonny, A., and Melchior, F., "A Reference Listening Room for 3D Audio Research," in *ICSA*, 2015.