A Library of Binaural Room Impulse Responses and Sound Scenes for Evaluation of Spatial Audio Systems

Frank Melchior[1], David Marston[1], Chris Pike[1], Darius Satongar[2] Yiu W. Lam[2]

[1] *BBC Research and Development, M50 2LH Salford, UK,* [2] *University of Salford, Salford, UK*

E-Mail: Frank.Melchior@bbc.co.uk

Introduction

Spatial audio systems, together with new concepts of content representation such as object-based audio, are likely to form the next generation of audio technology. To evaluate the wide variety of systems under consideration at the moment, two datasets have been created and are available for download. Together with publicly available rendering systems a test suite for spatial audio systems can be built. This consists of a set of binaural room impulse responses, a new format for representing audio scenes of various types, and example scenes for assessment of such systems. Each of these components is described in this paper.

Binaural Room Impulse Responses

A library of binaural room impulse responses (BRIR) was created using a B&K head and torso simulator (HATS). 15 listening positions have been measured in a 12 loudspeaker horizontal ring using the geometry given in Figure 1. The BRIR have been measured in 2° HATS-azimuth resolution on a regular grid. The room for the measurements was an ITU-R BS.1116 [1] compliant listening room at the University of Salford. The room has an average reverberation time of 0.27s and a background noise level of less then 6dBA. The room was equipped with Genelec 8030A loudspeakers. The 64,800 BRIR are available under [2] below using the SOFA format [3]. Measurements were made at sample rate 48kHz using sine sweeps of length 2^{18} samples, achieving a peak-to-tail signal-to-noise ratio of approximately 90 dB [4]. Room impulse responses made with an omnidirectional microphone are also available for each loudspeaker at each listening position.

Audio Definition Model

To represent audio content that can be rendered using various spatialisation techniques, a flexible content representation is required. The recently developed Audio Definition Model (ADM) is a formalised way to describe object-based, channel-based, and transformation-based audio content. It is part of the European Broadcasting Union Core XML schema [5] and is described in [6] and available for download. The initial supported file format is the Broadcast Wave Format (BWF) [7]. The model is for general use and not restricted to BWF files. Figure 2 presents an overview of the different elements of the format and their relationship.

The model is split into two parts, the format part, which is used to describe the technical aspects of the audio, and the content part, which is used to describe what the audio signal contains.

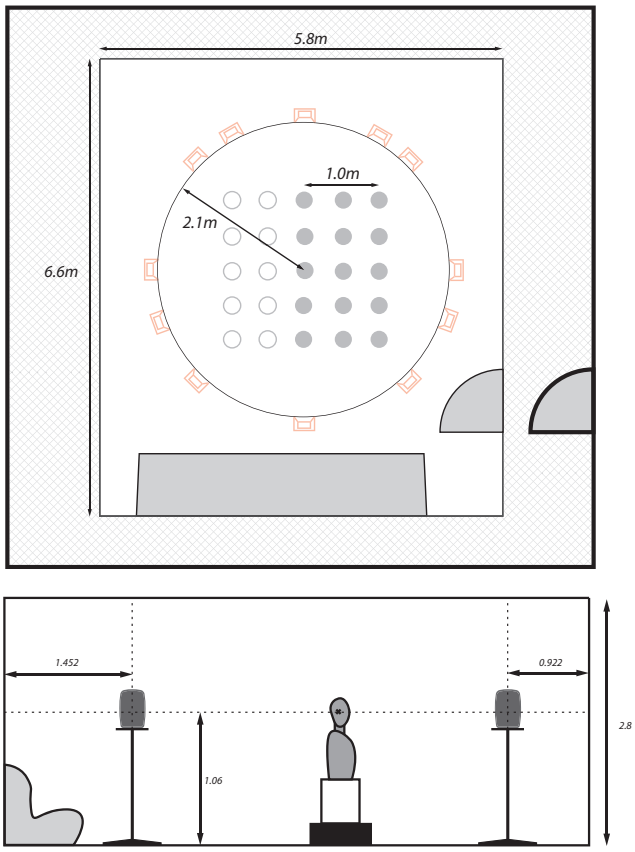

Figure 1: Geometry of the measurement setup to capture the BRIR database. Grey circles indicate the measurement position oft he dummy head.

The *TrackFormat* and *StreamFormat* elements in the format part are used to describe how the signal is formatted in the file (e.g. PCM). The *ChannelFormat* and *BlockFormat* elements are used to describe how the signal should be rendered (e.g. positional information or a preferred speaker). The *BlockFormat* element divides the *ChannelFormat* element along the time domain to allow dynamic metadata to be used (e.g. for moving objects).

The *PackFormat* element is used to combine multiple channels that belong together (e.g. two channels in a stereo pair).

In the content part the *Object* element links together the actual audio essence with its technical description from the format part. The *Content* and *Programme* elements are then used to describe the contents of the audio signals including loudness metadata.

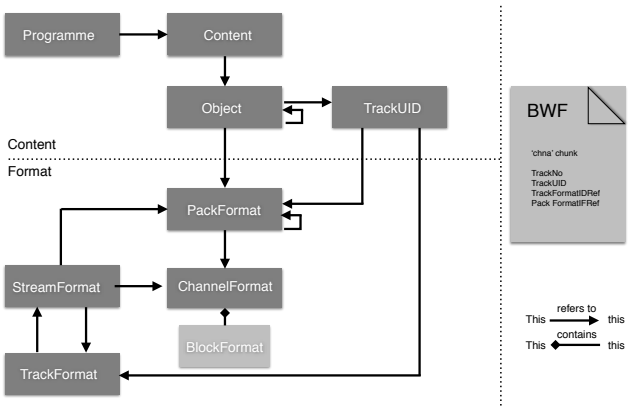

Figure 2: Relationship of different elements of the Audio Definition Model (ADM) used to describe audio content in a broadcast wave files (BWF).

The BWF file uses a look-up table that refers each of its audio tracks to a *TrackFormat* and *PackFormat* description. The ADM metadata can be stored as XML within a RIFF chunk in the BWF file.

Spatial Audio Quality Assessment Scenes

To evaluate different spatial audio systems a set of spatial audio scenes has been created. The aim is to develop these towards a complete library of test scenes to evaluate reproduction systems, codecs, and rendering systems. The scenes are stored using an ADM extension of the BWF as described in the previous section. The scenes have been created to demonstrate the concept and give a starting point for quality evaluation experiments. With this paper the following scenes have been published:

- *Speech Test*: Speech signals from multiple directions.
- *Panned Noise*: Amplitude-modulated noise circling the scene.
- *Applause*: Eight decorrelated applause signals.
- *Delibes - Flower Duet*: Trio of flute, cello, and viola.
- *Emma Heartbeat - The Machine (Aer Remix)* [Excerpt]: Electronic music.

This initial set of files can be found under [8]. These include the scenes in an object-based representation and their rendition into five channels.

Implementation Example

In order to playback and render the files a software library for the use with the SoundScape Renderer [9] is available for download. This includes reference software to read and write the Audio Definition Model extension, and to play it back and render it to the desired output format using the Soundscape Renderer.

In case the scenes are rendered for loudspeaker reproduction the BRIR provided can be used with a second instance of the SoundScape Renderer to virtualise the loudspeakers in a reference listening room. The rendered speaker feeds can then be reproduced using headphones for different listening positions in the room.

Conclusions

A library of binaural room impulse responses and spatial audio test scenes and software to enable their rendering/reproduction has been published. In future work the set of available scenes will be expanded. The library is open for further contributions of critical material. Some have already been used in binaural listening experiments [10].

References

[1] ITU, Methods For The Subjective Assessment Of Small Impairments In Audio Systems Including Multichannel Sound Systems (Rec. ITU-R BS.1116-1) 1997.

[2] http://usir.salford.ac.uk/30868.

[3] P. Majdak et al. “Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions”, presented at AES 134th convention, Rome, Italy, 2013.

[4] S. Müller and P. Massarani, “Transfer Function Measurement with Sweeps,” Journal of the Audio Engineering Society, vol. 49, no. 6, pp. 443–471, 2001.

[5] EBU Tech 3293, “EBU Core Metadata,” Geneva, January 2000.

[6] EBU Tech 3364, “Audio Definition Model,” Geneva, January 2014 (https://tech.ebu.ch/docs/tech/tech3364.pdf).

[7] EBU Tech 3285, “Specification of the Broadcast Wave Format,” Geneva, May 2011.

[8] data.bbcarp.org.uk/saqas.

[9] J. Ahrens, M. Geier, S. Spors. “The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods” presented at the 124th AES Convention, Amsterdam, The Netherlands, 2008.

[10] C. Pike and F. Melchior, “An assessment of virtual surround sound systems for headphone listening of 5.1 multichannel audio,” presented at AES 134th convention, Rome, Italy, 2013.

Acknowledgements

Parts of the research leading to the presented results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 610370.