On object-based audio with reverberation

Philip Coleman[1], Andreas Franck[2], Philip Jackson[1], Richard Hughes[3], Luca Remaggi[1], and Frank Melchior[4]

[1]*Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey, GU2 7XH, UK*

[2]*Institute for Sound and Vibration Research, University of Southampton, Southampton, Hampshire, SO17 1BJ, UK*

[3]*Acoustics Research Centre, University of Salford, Salford, M5 4WT, UK*

[4]*BBC Research and Development, Dock House, MediaCityUK, Salford, M50 2LH, UK*

Correspondence should be addressed to Philip Coleman (p.d.coleman@surrey.ac.uk)

ABSTRACT

Object-based audio is gaining momentum as a means for future audio productions to be format-agnostic and interactive. Recent standardization developments make recommendations for object formats, however the capture, production and reproduction of reverberation is an open issue. In this paper, we review approaches for recording, transmitting and rendering reverberation over a 3D spatial audio system. Techniques include channel-based approaches where room signals intended for a specific reproduction layout are transmitted, and synthetic reverberators where the room effect is constructed at the renderer. We consider how each approach translates into an object-based context considering the end-to-end production chain of capture, representation, editing, and rendering. We discuss some application examples to highlight the implications of the various approaches.

1. INTRODUCTION

Reproduction of a room effect is a critical part of audio production, whether the intention is to convey the sense of being in a specific real room or to carry the listener into a new world imagined by an artist. Technology for audio capture, production and reproduction must support these applications and any in between. Ideally, this requires the room effect to be recordable, intuitively editable, efficiently represented in transmission, and reproducible on a wide range of reproduction systems.

A spatial audio scene, or components of a scene, can be represented by *channel-based*, *transform-based*, or *object-based* approaches [1]. For channel-based approaches, such as stereo or 5.1 [2], the engineer must provide a mix for each target reproduction, and the actual loudspeaker feeds are transmitted. Similarly, recordings of spatial scenes can be made with microphone techniques designed for a specific reproduction layout [3].

Transform-based (or *scene-based* [4]) approaches map the scene onto a set of orthogonal basis functions which can be decoded at the receiver and mapped to the available loudspeakers. Reverberant scenes can be captured and represented directly onto basis functions, e.g. with a B-Format recording or using Ambisonics [5].

A scene in object-based audio is instead composed of a number of objects, each described by audio accompanied by metadata. The metadata are interpreted by a renderer which derives loudspeaker feeds. This approach allows audio content to be *format-agnostic*, i.e. produced once and replayed on many different kinds of devices [6].

If reverberation were rendered in an object-based way, the benefits of format-agnostic audio, such as greater immersion, personalization, and intelligibility, would also apply. In terms of reverberation, an object-based representation could give greater immersion by allowing the renderer to reproduce early reflections independently and precisely, taking into account the reproduction layout [7]. Opportunities for personalization or interaction can be envisaged both for producers (e.g. editing the room acoustics) and consumers (e.g. modifying the reverberation based on the listening room acoustics). Ideally, object-based reverberation would allow for independent control of the room effect due to each source, allowing objects to behave intuitively if, for instance, the level is adjusted. Finally, speech intelligibility may be enhanced in some cases by allowing listeners to increase the direct to reverberant ratio (DRR).

However, there is little in the literature to suggest that an

object-based room representation is supported by contemporary object standards. Rather, one common approach is to use a set of objects in conjunction with channel-based or Ambisonic-encoded ambience [8], [9]. Alternatively, one could synthesize reverberation at the renderer based on some physical or perceptual parameters such as those specified in MPEG-4 [10]. In this paper, we discuss the advantages and disadvantages of these kinds of approaches in content recording, production, editing, and rendering. We also apply the discussions to three application examples: sonic art, live recording, and rendering on mobile devices.

In Sec. 2, we outline the background to audio object descriptions and room acoustics. In Sec. 3 we discuss approaches to reverberation in the context of an end-to-end object-based production chain, and in Sec. 4 we consider the implications of the signal representation for rendering and personalization. In Sec. 5 we discuss application examples, and finally we summarize in Sec. 6.

2. BACKGROUND

As preliminary topics for the central discussion on object-based reverberation, in this section we first review previous proposals of object formats and current standards. Then, we briefly review the literature relating to the physical properties and perception of room acoustics.

2.1. Object parameters

The capabilities of object-based audio are directly linked to the metadata describing audio streams, and how these metadata are interpreted by the renderer. Alongside the position of the objects in the scene and their level, recent standardization activities consider objects with varying size or diffuseness. For instance the European Broadcasting Union's audio definition model (ADM) [11] allows an object to be diffuse, or to have non-infinitesimal dimensions, and MPEG-H [4], [12] includes a spread parameter. The signal-processing blocks necessary in the renderer to interpret these parameters may give opportunities to extend the metadata and reproduce reverberant signals, such as the parametric approach of [7]. However, there is no standard set of parameters for such an approach to reverberation.

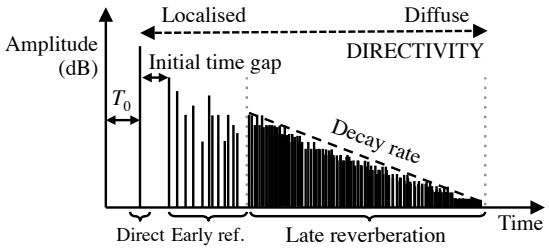
There have previously been proposals to include rooms in object schemes. In MPEG-4 [10], the reverberation could be synthesized at the renderer based on a physical model or a set of perceptual parameters (see Sec. 3.2). A detailed set of metadata for rendering a virtual scene including room parameters was detailed in [13], based on MPEG-4. In the proposed spatial sound description interchange format (SpatDIF) [14], reverb is considered to belong to the *spatial coding layer* of the scene, rather than to each individual source. The process encodes signals containing spatial information while remaining format-agnostic. One example of the use of this layer is to add surround effects by Ambisonic B-Format convolution. These approaches have not been widely adopted.

2.2. Room acoustics

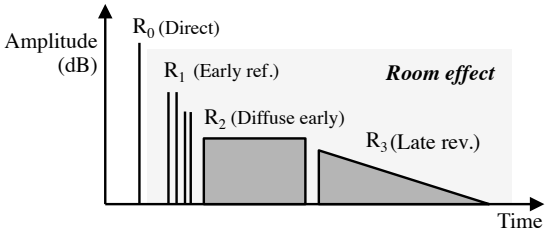
To represent the effects of a room in an object-based manner, it is useful to consider both the physical and perceptual effects of the space on the soundfield. The influence of the room alters perception of a sound source and the environment it is in.

At low frequencies, when the wavelength is comparable to the dimensions of the room, modal behaviour dominates. This occurs below a transition frequency typically estimated using the Schroeder frequency [15]. The perception of modal behaviour in typical enclosed listening environments is often considered to be a monaural timbral effect. When the wavelength becomes smaller relative to the room and surface dimensions, reflections are often thought of (and modelled) as sound rays following the principles of geometric acoustics [16]. After the direct sound, early reflections are initially sparse in time, appearing as distinct contributions arriving from specific directions determined by the room geometry. Reflections arriving within the first 5–10 ms affect localisation, and typically (when not sufficiently strong to break the precedence effect) are associated with a perceived image shift and broadening of the primary sound source [17]. Early reflections can also lead to a change in perceived timbral quality of the direct sound, producing colouration through comb-filtering [18]. The *initial time gap* separating the direct sound and the first reflection is thought to affect perception of the presence or intimacy of the room and its apparent size [19].

As time progresses, the soundfield becomes a mix of diffuse reflections and specular reflections of decaying level and increasing temporal density and spatial diffuseness, displaying behaviour more statistically random in nature [20]. Later reflections and the reverberant decay affect predominantly spatial attributes such as perceived envelopment and spaciousness [19]. The later soundfield also provides cues to source distance, with the DRR playing a significant role [21], as well as providing further cues to room size.



(a) Generic RIR model (Based on [20], Fig. 1)



(b) Spat RIR model (Based on [22], Fig. 5)

Fig. 1: RIR models; (a) direct sound arriving time T_0, early reflections beginning after the initial time gap, and late reverberation, showing sound becoming increasingly diffuse with time; (b) direct sound (R_0), discrete early reflections (R_1), diffuse early reflections (R_2), and late reverberation (R_3).

Figure 1 shows two generic reverberation models representing the development of the room impulse response (RIR) over time. The first model (Fig. 1(a)), comprises the direct sound arriving after time T_0, a number of early reflections, and late reverberation characterised by an exponential decay curve. The second model (Fig. 1(b)) is similar, having direct sound (R_0) and early reflections (R_1), but also specifically including diffuse early reflections (R_2) before the late reverberation (R_3).

It is clear from the above that the influence of a room acoustic not only affects how individual sound sources are perceived, but also provides important auditory cues used by the listener to gain a sense of the space they are in. Including the most perceptually relevant aspects of the reverberant room response can hence deliver a more realistic impression of being transported to an alternative space.

3. CREATING REVERBERATION

Current approaches for recording and representing reverberation fall into two main approaches: recording in reverberant spaces; and synthesising reverberation in post-production. These approaches are described below. Throughout the discussion, we will refer to Fig. 2, which illustrates the steps of capture and parameterization, production, representation, and reproduction. The object-based approach, together with parametric reverb, is found towards the top, and signal paths for scene-based and channel-based reverb recordings are illustrated towards the bottom. Italicized terms in the following description relate to blocks depicted in the figure.

3.1. Recording in reverberant spaces

Spatial microphone techniques are often used to make recordings with a spatial impression. Channel-based and scene-based approaches exist, with different characteristics and limitations. Channel-based spatial microphone techniques are intended to be reproduced over a specific reproduction layout. *Main microphone* techniques have been developed, first from stereo techniques [3], and with increasing numbers of microphones added first for 5.1 surround, e.g. [23], and more recently for with-height systems, e.g. [24], [25]. Where there is an opportunity to include a separate *room microphone* array such as a Hamasaki square [26], [27], diffuse room sound may also be captured. Once captured, the room effect is moderately editable (by combining microphone signals with different mixing gains [28]), illustrated by the *mix to production layout* block in Fig. 2. However, the spatial aspects of the recording are fixed and only properly reproduced over the target loudspeaker arrangement.

Alternatively, reverberant signals can be captured under a scene-based approach. The most common of these is the low order Ambisonic Soundfield microphone, which gives 3D information encoded onto orthogonal basis functions (usually B-Format) and is sometimes used for broadcast [29]. Higher-order spherical microphone arrays may also be used for recording reverberation or ambience with increased spatial resolution [8]. Ambisonic signal representations give a compact description of a whole scene, and loudspeaker feeds can be derived at the renderer based on the target loudspeaker arrangement. The representation also allows for rotation, scaling, and spatial filtering to enhance or attenuate certain directions [30], [31]. The process of recording using circular, spherical or B-Format arrays, mapping to basis functions, and editing, is shown as *rotate/re-emphasize scene* in Fig. 2.

These approaches to content recording are familiar to the engineers, allow for their creative intuition to influence the recording, and achieve high fidelity recordings when

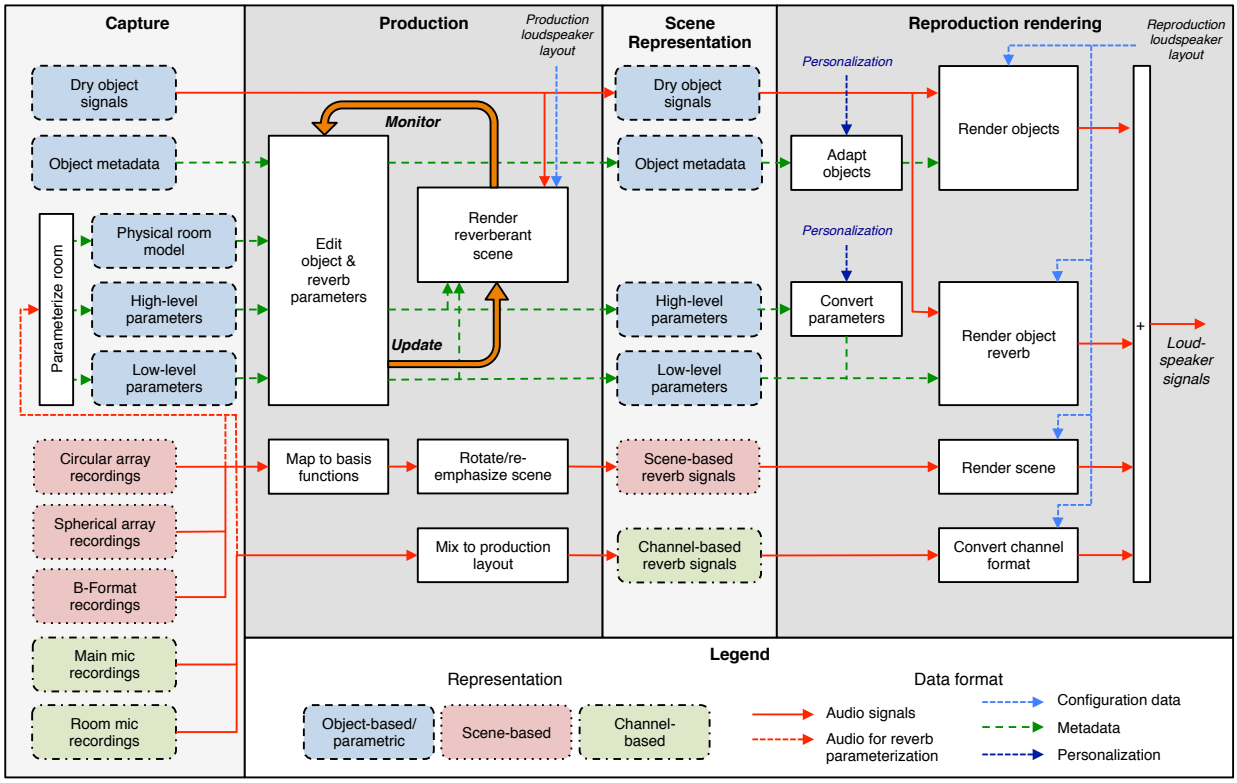


Fig. 2: Signal and metadata flows for capturing, editing, representing/transmitting and rendering reverberation.

equipment of sufficient quality is used. However, in the context of object-based audio, the signals may be difficult to edit and represent in a format-agnostic way. One approach could be to render room signals as wide or diffuse sources. The challenge with this approach would be to maintain the spatial properties achieved by various channel-based techniques, which are designed to be reproduced over specific channel layouts.

3.2. Synthetic reverberation

Digital synthesis of reverberation is a topic that has attracted much research over many decades [20], [32]. Synthesis of reverberation can generally be split into three categories: delay networks, convolutional, and computational acoustic [20]. Delay networks (based on the same principle as digital waveguides) synthesize reverberation by feeding the input signal through a series of feedforward and feedback delays to achieve the desired room impression. Convolution reverbs take a measured or synthetic RIR and create reverberation by convolving this with dry audio content. Computational acoustic techniques may be directly applied, or compute RIRs offline for later use via convolution.

The signal flow in Fig. 2 accommodates each approach. A set of parameters describing the reverberation are defined in the *parameterization* process, optionally using recordings of various formats from real rooms. Then in production, the producer *edits object and reverb parameters*, using a local version of the renderer to *render the reverberant scene* and *monitors* the production. The reverberant signals are finally represented as audio and metadata streams.

In this section we provide an overview of synthetic reverberation. We briefly mention convolution reverbs based on recorded RIRs in Sec. 3.2.1. In Sec. 3.2.2 we discuss the parameterization and synthesis of rooms based on *low-level* parameters directly available to the renderer, and in Sec. 3.2.3 we discuss approaches based on *high-level* parameters. Table 1 summarizes the discussion of the parametric approaches. The reader is referred to [20] for a detailed discussion of other synthetic reverberators.

3.2.1. Convolution reverbs

Convolution reverbs are commonly used in audio and film post-production. The underlying assumption for a convolution reverb is that the RIR of a reverberant space can be applied as a finite impulse response (FIR) filter, e.g. [33], [34]. Therefore, any dry signal can be made reverberant in post-production by convolution with a pre-recorded set of RIRs. There are many commercial prod-

ucts on the market that use this technology based on libraries of measured RIRs, giving various degrees of control to the producer by allowing them to search and select various room types and sizes, or to apply signal processing operations (e.g. time stretching) for creative modification [35].

For the present discussion, we simply note that the application of a convolution reverb is analogous to making a recording in the space where the RIR was recorded (or computationally generated). This carries the limitation of the channel-based and transform-based approaches outlined above, in that the spatial content of the reverberator is tied to the arrangement of real or virtual microphones used to record the RIR. On the other hand, many convolution reverbs have distinct characteristics that make them popular with producers.

3.2.2. Room parameterization and synthesis

For parametric reverbs, some kind of analysis of the sound field at the listening position is first required. Approaches have been proposed to analyse the sound field from physical and perceptual perspectives. The analysed sound field is then encoded as a set of low-level parameters from which the room effect may be synthesized.

One method of efficiently parameterising a RIR is the spatial decomposition method (SDM) [36]. The SDM is based on the assumption that the RIR is composed of image sources in the far field. In each time segment, a microphone array is used to determine the direction of arrival (DOA) of the most prominent image source. This information is combined with the actual RIR recorded at a real or virtual omnidirectional microphone at the centre of the array, to give a way of spatialising the omnidirectional signal. In the context of object-based audio, the omnidirectional microphone RIR would likely be broadcast over an audio channel, and would require a convolution engine to be implemented at the renderer to be combined with the dry object audio.

Spatial impulse response rendering (SIRR) [37] (which underpins much of the analysis and synthesis in directional audio coding (DirAC) [38]) is an alternative framework for analysing, encoding and synthesing a spatial RIR. The analysis part is based on a B-Format RIR. The principle of the analysis part is that, for a particular time-frequency window, the spatial response can be represented by three parameters: the DOA (azimuth and elevation), and a diffuseness coefficient. Editing of DirAC metadata in production was considered in [39] to achieve effects including rotation, zoom, compression and spatial filtering. However, the encoding of the reverberant sound source is such that it is not clear how to edit the parameters for creative or interactive adjustment of the room, for instance to move the sound source or listener with respect to the room. For synthesis, the direct sound component is panned via vector-base amplitude panning (VBAP) [40], while the diffuse portion is decorrelated and sent to all loudspeakers [41]. The parametric nature of SIRR means that it can be synthesized flexibly, for instance over headphones with binaural processing, or using a sound field synthesis approach such as wave field synthesis (WFS). Recent improvements to DirAC been achieved by recording with a higher-order microphone, leading to greater spatial resolution [42].

A system for capturing, editing and rendering room effects based on a plane wave description of the sound field was proposed in [43] with WFS as the target rendering approach. We refer to this as reverberant WFS (R-WFS). First, the wave field is analysed based on measurements from a circular microphone array [44], [45]. Then, RIRs in the plane wave domain are divided into an early part and late part. In addition, strong early reflections are extracted by spatio-temporal windowing. This leads to a representation of the room comprising discrete early reflections, and the early part (reflections and building diffuseness) and late part (reverberation tail) of the room response. The discrete early reflections may be modified based on the position and directivity of the direct sound, whereas the early part and late part of the reverberation are fixed for each room [46]. Representation of the sound field as a sum of physical point sources and plane waves gives a high resolution and inherently object-based approach. The point sources can be used for discrete early reflections (R_1,cf. Fig. 1(b)) and diffuse early reflections (R_2), and at least ten plane wave sources [47] distributed evenly around the listener in 2D are used to render the late reverberation (R_3). The same approach has been used for surround sound and binaural reproduction [48].

Alternatively, the reverberant spatial audio object (RSAO) [7] describes a compact set of parameters for parametric reverberation based on RIRs measured using a circular array. The RIR was characterized by the direct sound, L early reflections, and late reverberation, similar to the model in Fig. 1(a). Early reflections are considered as peaks in the time domain RIR, generated by first-order specular reflections arriving from a specific direction. The late reverberation is modelled as Gaussian

noise having an exponential decay, generated by the superposition of all high-order specular and diffuse reflections [49]. For these reasons, times of arrival (TOAs) and DOAs with respect to the array are the main parameters chosen for the early reflections, with the mixing time and exponential decays estimated in octave subbands used for the late reverberation. In addition to this the spectral envelope is extracted from the early reflections in order to provide the proper coloration. The filtered direct sound and directional reflections are treated as image sources and rendered by applying the appropriate delay and panning. The late reverberation is rendered by delaying the signal based on the mixing time estimate, convolving it with a filter built by applying the subband decay constants to white noise, and finally rendering as a diffuse source. The set of parameters is flexible and editable in production.

3.2.3. High-level parameters and synthesis

There is provision in MPEG-4 [10] for parametric reverberation based on high-level parameters. The parameters may be used to build a virtual-reality scene (i.e. one designed to mimic the real world), or an abstract effect (i.e. providing freedom to a sound designer to have complete control over the environment created). Instructions for interpreting the effect parameters are transmitted via the structured audio orchestra language (SAOL).

The physical and perceptual parameters for room modelling are fully described in MPEG-4 v2 [10], [51], [52]. The physical parameters are specified in terms of the transmission paths in the environment and frequency-dependent directivity models for sound sources. The perceptual parameters include six parameters specific to each source in the scene, and an additional three parameters describing the late reverberation (and applied to all sources) [53]. The (high-level) perceptual parameters map to low-level delay network coefficients to control various portions of the RIR. The mapping in MPEG-4 is based on the model described in [22]. In the synthesis, the R_0 portion (cf. Fig. 1(b)) is panned directly, the R_1 portion is created by panning delayed versions of the direct sound, and the R_2 and R_3 portions are created by passing the source through feedback delay networks which contain a matrix governing the mixing time and decay time. The mapping from high-level to low-level parameters is represented by the *convert parameters* stage in Fig. 2.

4. REPRESENTING REVERBERATION

The chosen approach to reverberation impacts the features of the object-based reproduction systems as well as the architecture of the renderer. The *reproduction rendering* block in Fig. 2 depicts a versatile signal flow for an object-based renderer that includes several of the approaches outlined in the previous section.

In channel-based approaches, the renderer needs to *convert the channel format* to adapt the received loudspeaker signals to the actual reproduction layout [12]. As trivial matrixing approaches can lead to processing artifacts due to inter-channel correlation and comb-filtering, sophisticated algorithms are required for good quality [54]–[56]. At the same time, user interactivity of this approach is limited to a control of the total reverberation level, because the channel signals' mixture does not allow for adaptation of individual parts of the reverberant sound scene.

Scene-based approaches avoid the need for a format conversion. Instead, a scene renderer, for instance a higher-order Ambisonic decoder, produces the channel signals for the reproduction system, which might also account for the actual loudspeaker positions. The user interaction capabilities are limited in the same way as for channel-based approaches.

In contrast, object-based reverberation rendering enables comprehensive user interactivity and control, or *personalization*. Modifications to individual objects or object groups are applied in the *adapt objects* stage. Examples include the selection of different commentaries, the attenuation or emphasis of specific parts of the audio scene, or the movement of individual audio objects, which could, for instance, affect the discrete reflection pattern. This approach also enables the user to control room parameters of the reverberation such as level or reverberation time. In Fig. 2, this form of control is represented by the incorporation of personalization data into the *convert parameters* facility which transforms high-level parameters into a low-level representation for the *render object reverb* stage. Thus, the viability of user personalization depends on the abstraction level of the transmitted parameters. Low-level parameters, such as those describing recorded RIRs, reduce the level of flexibility. Moreover, the admissible changes to the room parameters will often be limited in order to preserve the artistic intent of the production. These limits need to be transmitted as part of the metadata, similar to interac-

	Capture	Parameters	Editing	Rendering
Spat [22]	Estimate physical room properties and perceptual correlates from mono RIR	Perceptual parameters (linked to delay network coefficients)	Modify source/room in perceptual parameters domain	Apply feedback delay network coefficients
MPEG-4 [10], [50]	Estimate room dimensions/surface filtering properties	Source directivity and FIR surface filter or SAOL specification	Edit room description	Computational acoustics room rendering/convolution
RSAO [7]	Circular array RIR	TOA/DOA for direct sound & discrete early reflections; octave-band decay for late reverb	Modify image sources (early) and octave-band reverb time (late)	Split signal: pan non-diffuse and decorrelate diffuse
SIRR [37]	B-Format RIR	Time-frequency-wise azimuth, elevation and diffuseness	Edit TF-cell parameters	Split signal: pan non-diffuse and decorrelate diffuse
R-WFS [43]	Circular array RIR	High resolution plane wave RIR	Edit response in plane wave domain	Render source types by WFS, convolve with audio
SDM [36]	≥ 4 microphone RIR in a 3D layout	Omni RIR plus DOA for each time segment	Low-level editing of omni RIR or DOA for an image source	Convolve and render to target DOA

Table 1: Summary of parametric reverberation in the context of an object-based production pipeline

tivity limits for general objects, e.g., [4]. The *render object reverb* processor might produce different output formats. One possibility is to directly create channel signals for the reproduction layout. Alternatively, the object renderer might generate a new set of audio objects, e.g., point sources for discrete reflections and plane waves or diffuse sound objects for the late reverberation. Finally, the reverberation renderer can also output a scene-based representation of the sound field, e.g., [57], [58].

5. APPLICATION EXAMPLES

In this section we consider some application examples and discuss the implications for producing reverberation, referring to the methods in Table 1.

5.1. Sonic art

In sonic art, for instance production of popular music or a radio drama, reverberation is used predominantly as a creative effect. There is not necessarily a real-world reference point with which the reproduced reverberation is compared, and different components of the acoustic scene may have differing reverberation that would not be physically possible in practice. The key aspects of this application are the flexibility of editing or tuning the reverberant effect, and the preservation of producer intent through to reproduction. Currently, reverb tools would be used in post-production, mixed to a specific reproduction format. For existing object-based productions, the reverb would be likely to be mixed down to a channel-based bed transmitted alongside the objects. On the other hand, parametric approaches could provide significant flexibility. As representing the sense of being in a specific physical space is not a priority, the Spat approach of defining and editing high-level parameters would likely be appropriate. However, with a suitable library of parameterised rooms, a producer might be able to more precisely tune the room effect by modifying image sources and reverb time, for instance with the RSAO or with a suitable production tool for editing SIRR. One risk with a parametric approach is that the producer devolves some amount of creative control, relying on the renderer to faithfully reproduce the content. However this might be outweighed by the opportunity for intelligent rendering whereby reverberant content could be rendered over many loudspeaker layouts, or headphones, while preserving the sense of immersion.

5.2. Live recording

For live recording such as broadcasting a classical music concert, the acoustic space is an inherent part of the performance, affecting the conductor and musicians [59].

Thus, the recording process is likely to begin by faithfully reproducing the room acoustics, while allowing the engineer to modify the room impression if desired. For instance, the Royal Albert Hall in London has a strong echo that may be undesirable [59]. Channel-based room techniques might be appropriate and would supply good fidelity, but editing the reverberant content of the room depends on the skill of the recording engineer. Furthermore, the resulting mix would be inflexible over different reproduction systems. An alternative approach would therefore be to combine close-microphone recordings of the orchestra sections with a parametric description of the acoustic space. The low-level parametric approaches based on recorded RIRs might be the most appropriate. Using the previous example of a problematic strong reflection, the R-WFS and RSAO approaches would both allow the reflector to be edited in the parameter domain. On the other hand, if minimal editing of the space were required then the SIRR or SDM approaches might be most appropriate. These methods are straightforward to capture and reproduce, but are not as intuitive to edit. Using a parametric representation, the impression of the concert hall could also be conveyed flexibly over a range of reproduction systems.

5.3. Rendering on mobile devices

Another aspect of object-based reverberation is to consider the required computation, power and bandwidth for rendering. While it could be assumed that professional studio environments, and probably home systems, are able to supply sufficient resources to render parametric reverb, the same cannot be assumed for mobile devices. Indeed, the resources on mobile devices for 3D audio may already be stretched by any binaural processing. Artificial reverberators based on delay networks are likely to be efficient enough to operate on mobile devices, but those requiring convolution in the renderer may not be appropriate. On the other hand, parametric approaches could be used for capture and production, which would allow flexible editing. Eventually, a channel-based or scene-based representation might be the best approach, with content either produced directly or by rendering parametric reverberation in post-production. This would limit the opportunities for listener personalization, although a scene-based approach would still allow some flexibility for rendering over headphones or various loudspeaker arrangements.

6. SUMMARY

The topic of reverberation for object-based audio was discussed, comparing approaches creating reverberant signals in recording and post-production with those that synthesise reverberation at the renderer. Traditional approaches to capturing and editing reverberant content or RIRs do not permit the intelligent rendering and interactivity that object-based audio promises, but they do achieve good results and limit the complexity at the renderer. On the other hand, parametric approaches can be efficiently captured, edited, transmitted, and modified at the renderer to suit the actual loudspeaker locations and listener preferences. While this allows the benefits of object-based audio to be maintained, producers' control over their content may need to be reconceived. Future work might include encapsulating channel-based or scene-based signals as objects, and evaluating the parametric approaches with listening tests.

7. ACKNOWLEDGEMENTS

This work was supported by EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1).

8. REFERENCES

[1] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, *et al.*, "Spatial sound with loudspeakers and its perception: a review of the current state," *Proc. IEEE*, vol. 101, no. 9, pp. 1920–1938, 2013.

[2] ITU, "Recommendation ITU-R BS.775-3, Multichannel stereophonic sound system with and without accompanying picture," International Telecommunication Union (ITU), August 2012.

[3] F. Rumsey, *Spatial audio*. Oxford, UK: Focal Press, 2001.

[4] S. Füg, A. Hölzer, C. Borß, C. Ertel, M. Kratschmer, *et al.*, "Design, coding and processing of metadata for object-based interactive audio," in *137 Conv. Audio Eng. Soc.*, Los Angeles, CA, USA, 2014.

[5] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *J. Audio Eng. Soc.*, vol. 33, no. 11, pp. 859–871, 1985.

[6] B. Shirley, R. Oldfield, F. Melchior, and J.-M. Batke, "Platform independent audio," in *Media Production, Delivery and Interaction for Platform Independent Systems*, John Wiley & Sons, Ltd, 2013, pp. 130–165.

[7] L. Remaggi, P. J. B. Jackson, and P. Coleman, "Estimation of room reflection parameters for a reverberant spatial audio object," in *138 Conv. Audio Eng. Soc.*, Warsaw, Poland, 2015.

[8] H. Stenzel and U. Scuda, "Producing interactive immersive sound for MPEG-H: a field test for sports broadcasting," in *137 Conv. Audio Eng. Soc.*, Los Angeles, CA, USA, 2014.

[9] R. Oldfield, B. Shirley, and J. Spille, "An object-based audio system for interactive broadcasting," in *137 Conv. Audio Eng. Soc.*, Los Angeles, CA, USA, 2014.

[10] E. D. Scheirer, R. Väänänen, and J. Huopaniemi, "AudioBIFS: describing audio scenes with the MPEG-4 multimedia standard," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 237–250, 1999.

[11] EBU, "Tech 3364, Audio Definition Model," European Broadcasting Union (EBU), January, 2014.

[12] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H 3D audio — The new standard for coding of immersive spatial audio," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 770–779, 2015.

[13] G. Potard and I. Burnett, "An XML-based 3D audio scene metadata scheme," in *Proc. 25th AES Int. Conf.*, London, UK, 2004, pp. 17–19.

[14] N. Peters, T. Lossius, and J. C. Schacher, "The spatial sound description interchange format: principles, specification, and examples," *Computer Music Journal*, vol. 37, no. 1, pp. 11–22, 2013.

[15] M. R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *J. Audio Eng. Soc.*, vol. 35, no. 5, pp. 299–306, 1987.

[16] M. Vorländer, "Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm," *J. Acoust. Soc. Am.*, vol. 86, no. 1, pp. 172–178, 1989.

[17] S. E. Olive and F. E. Toole, "The detection of reflections in typical rooms," *J. Audio Eng. Soc.*, vol. 37, no. 7/8, pp. 539–553, 1989.

[18] S. Bech, "Timbral aspects of reproduced sound in small rooms II," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3539–3550, 1996.

[19] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, "Perception of reverberation in small rooms: a literature study," in *Proc. 55th AES Int. Conf., Helsinki*, 2014.

[20] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Trans. Audio Speech Lang. Proc.*, vol. 20, no. 5, pp. 1421–1448, 2012.

[21] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: a summary of past and present research," *Acta. Acust. united Ac.*, vol. 91, no. 3, pp. 409–420, 2005.

[22] J.-M. Jot, "Efficient models for reverberation and distance rendering in computer music and virtual audio reality," in *Proc. Int. Computer Music Conference*, Thessaloniki, Greece, 1997.

[23] G. Theile, "Multichannel natural recording based on psychoacoustic principles," in *108 Conv. Audio Eng. Soc.*, Paris, France, 2000.

[24] G. Theile and H. Wittek, "Principles in surround recordings with height," in *130 Conv. Audio Eng. Soc.*, London, UK, 2011.

[25] H. Lee and C. Gribben, "On the optimum microphone array configuration for height channels," in *134 Conv. Audio Eng. Soc.*, Rome, Italy, 2013.

[26] K. Hamasaki, T. Shinmura, S. Akita, and K. Hiyama, "Approach and mixing technique for natural sound recording of multichannel audio," in *Proc. 19th AES Int. Conf.*, Schloss Elmau, Germany, 2001.

[27] K. Hamasaki, "Multichannel recording techniques for reproducing adequate spatial impression," in *Proc. 24th AES Int. Conf.*, Banff, Canada, 2003.

[28] J. Francombe, T. Brookes, R. Mason, R. Flindt, P. Coleman, *et al.*, "Production and reproduction of program material for a variety of spatial audio formats," in *138 Conv. Audio Eng. Soc.*, Warsaw, Poland, 2015.

[29] G. Thomas, A. Engström, J.-F. Macq, O. A. Aziz Niamut, B. Shirley, *et al.*, "State of the art and challenges in media production, broadcast and delivery," in *Media Production, Delivery and Interaction for Platform Independent Systems*, John Wiley & Sons, Ltd, 2013, pp. 5–73.

[30] D. G. Malham and A. Myatt, "3D sound spatialization using ambisonic techniques," *Computer Music Journal*, vol. 19, no. 4, pp. 58–70, 1995.

[31] M. Frank, F. Zotter, and A. Sontacchi, "Producing 3d audio in ambisonics," in *Proc. 57th AES Int. Conf.*, Los Angeles, CA, USA, 2015.

[32] B. A. Blesser, "An interdisciplinary synthesis of reverberation viewpoints," *J. Audio Eng. Soc.*, vol. 49, no. 10, pp. 867–903, 2001.

[33] W. G. Gardner, "Efficient convolution without input-output delay," *J. Audio Eng. Soc.*, vol. 43, no. 3, pp. 127–136, 1995.

[34] A. Reilly and D. McGrath, "Convolution processing for realistic reverberation," in *98 Conv. Audio Eng. Soc.*, Paris, France, 1995.

[35] E. Deruty, *Creative convolution: new sounds from impulse responses*, https://www.soundonsound.com/sos/sep10/articles/convolution.htm, September 2010, accessed 9th July 2015.

[36] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc.*, vol. 61, no. 1/2, pp. 17–28, 2013.

[37] J. Merimaa and V. Pulkki, "Spatial impulse response rendering I: analysis and synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 12, pp. 1115–1127, 2005.

[38] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516, 2007.

[39] A. Politis, T. Pihlajamäki, and V. Pulkki, "Parametric spatial audio effects," in *15th Int. Conf. Digital Audio Effects (DAFx-12)*, York, UK, 2012.

[40] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, 1997.

[41] V. Pulkki and J. Merimaa, "Spatial impulse response rendering II: reproduction of diffuse sound and listening tests," *J. Audio Eng. Soc.*, vol. 54, no. 1/2, pp. 3–20, 2006.

[42] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-based parametric sound field reproduction in the spherical harmonic domain," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 852–866, 2015.

[43] F. Melchior, C. Sladeczek, A. Partzsch, and S. Brix, "Design and implementation of an interactive room simulation for wave field synthesis," in *Proc. 40th AES Int. Conf.*, Tokyo, Japan, 2010.

[44] E. M. Hulsebos and D. de Vries, "Parameterization and reproduction of concert hall acoustics measured with a circular microphone array," in *112 Conv. Audio Eng. Soc.*, Munich, Germany, 2002.

[45] E. M. Hulsebos, "Auralization using wave field synthesis," PhD thesis, Delft University of Technology, 2004.

[46] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 675–705, 1999.

[47] J. Nowak, J. Liebetrau, and T. Sporer, "On the perception of apparent source width and listener envelopment in wave field synthesis," in *5th Workshop on Quality of Multimedia Experience (QoMEX)*, IEEE, Klagenfurt, Austria, 2013, pp. 82–87.

[48] F. Melchior, "Investigations on spatial sound design based on measured room impulse responses," PhD thesis, Delft University of Technology, 2011.

[49] J.-M. Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *Proc. ICASSP'92*, IEEE, San Francisco, CA, USA, 1992, pp. 221–224.

[50] R. Väänänen and J. Huopaniemi, "Advanced AudioBIFS: virtual acoustics modeling in MPEG-4 scene description," *IEEE Trans. Multimedia*, vol. 6, no. 5, pp. 661–675, 2004.

[51] M. Honkala, *Acoustics modeling in MPEG-4*, http://www.tml.tkk.fi/Opinnot/Tik-111.590/2002s/Paperit/honkala_MPEG4_OK.pdf, accessed 28th May 2015.

[52] J. Schmidt and E. F. Schroeder, "New and advanced features for audio presentation in the MPEG-4 standard," in *116 Conv. Audio Eng. Soc.*, Berlin, Germany, 2004.

[53] J.-M. Trivi and J.-M. Jot, "Rendering MPEG-4 AABIFS content through a low-level cross-platform 3D audio API," in *Proc. ICME'02*, IEEE, vol. 1, Lausanne, Switzerland, 2002, pp. 513–516.

[54] S. K. Zieliński, F. Rumsey, and S. Bech, "Effects of down-mix algorithms on quality of surround sound," *J. Audio Eng. Soc*, vol. 51, no. 9, pp. 780–798, Sep. 2003.

[55] J. Vilkamo, A. Kuntz, and S. Füg, "Reduction of spectral artifacts in multichannel downmixing with adaptive phase alignment," *J. Audio Eng. Soc*, vol. 62, no. 7/8, pp. 516–526, 2014.

[56] A. Adami, E. Habets, and J. Herre, "Downmixing using coherence suppression," in *Proc. ICASSP2014*, IEEE, Florence, Italy, 2014, pp. 2878–2882.

[57] J. Anderson and S. Costello, "Adapting artificial reverberation architectures for B-format signal processing," in *Ambisonics Symposium 2009*, Graz, Austria, 2009.

[58] F. Lopez-Lezcano, "An architecture for reverberation in high order ambisonics," in *137 Conv. Audio Eng. Soc.*, Los Angeles, CA, USA, 2014.

[59] L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, 2nd ed. New York: Springer-Verlag, 2004.