Reproduction of Room Sound-fields for Subjective Assessment

Densil Cabrera

Faculty of Architecture, University of Sydney, NSW 2006, Australia

ABSTRACT

This paper reviews techniques for the recording and reproduction of room sound fields used by the author in a variety of room acoustics research projects. The paper approaches this first through discussion of various criteria for assessing the usefulness of sound-field reproduction/simulation systems. The paper highlights issues in the recording of impulse responses for audio reproduction (through convolution) including in the design of the sound source, the design of microphone arrays, measurement signals, treatment of impulse responses and larger problems with this process. The greatest focus of this paper is on audio reproduction methods, with an emphasis on non-individualized two-channel techniques.

INTRODUCTION

In recent years, techniques and technologies for spatial audio recording and reproduction have expanded considerably. In consumer audio, multichannel sound reproduction has become main-stream. In research and high-end applications, advances have included refinements of binaural technology, wave-field synthesis (WFS), ambisonics, as well as a host of other techniques and hybrid solutions for high resolution spatial audio. Such advances can be exploited in architectural acoustics research involving subjective assessments. Many architectural acoustics studies are concerned with the qualities of multiple acoustic environments – which may be scattered around the globe – and it is unfeasible to transport subjects to each. Spatial audio recording and reproduction techniques allow such acoustic environments to be reproduced in a laboratory, with instant random access to any one environment.

The selection of an appropriate technique for soundfield recording and reproduction depends on many factors, including the availability of resources and convenience. Perfection is elusive in this area, so decisions on technique selection and implementation should be informed by the limitations of various possibilities. Even given the inherent limitations of a technique, its effective application can be undermined by any of a multitude of oversights and errors. This paper gives an overview of some issues that might be considered in the selection and application of techniques for soundfield recording and reproduction, with an emphasis on the simple techniques (eg using two audio channels), and especially those with which the author has practical experience.

CRITERIA FOR ACCURATE SOUND-FIELD REPRODUCTION

Various approaches could be taken to assess he accuracy of reproduced sound. The most obvious of these is to compare physical measurements of the sound-field in the original space to physical measurements in the reproduction. Measurements that might be made include room acoustical parameters (reverberation time, early decay time, clarity index, lateral energy measures), sound pressure level as a function of frequency, the details of an impulse response and spatial variation of the soundfield. Given that some distortion in the reproduction is, in most cases, inevitable, a sensible appreciation of the importance of various distortions is required. Furthermore, in some contexts, there are not many measurements of the original space available for detailed comparisons.

It is often assumed that distortions that are less than one just noticeable difference (JND) can be tolerated in a sound reproduction, with no effect on the perceived sound quality. This assumes that interaction effects do not occur when there are small changes of multiple acoustic parameters. JNDs for commonly used room acoustical parameters are given in ISO3382, and are typically 1 dB for values in decibels (eg strength factor or clarity index), or 5% for ratio scale values such as reverberation time or early decay time.

A subjective judgment such as perceived distance or localisation angle can be related to a simply measurable physical counterpart (physical distance or angle of the sound source). Hence, veridicality in such judgments might be taken as a measure of reproduction fidelity. However, caution is required here, since there may be systematic discrepancies between physical and perceived source positions in real situations. For example, auditorily perceived source-receiver distance can be influenced by several factors, such as visual cues, direct to reverberant sound ratio, and many others (Zahorik 2002). Localisation angle is also affected by factors apart from the source angle (eg Cabrera *et al.* 2005). Therefore a better subjective test of reproduction accuracy is to compare perceptions in the reproduced environment with those in the actual environmnent.

Finally, the subjective rating of realism can be a useful approach to assessing reproduction quality. If the reproduced soundfield does not seem to be realistic, then it has almost certainly failed. This approach is most applicable when multiple audio systems are being compared. One difficulty with this is that audio in entertainment often aims to enhance the reality represented – and listeners might confuse that approach with the notion of realism. Audio in films, for example, emphasises the sounds most important for the narrative, excluding others, and is normally much louder than the equivalent sounds in reality. Nevertheless, the experiences available in the spatial audio reproduction techniques discussed in this paper are strikingly different in character to those commonly experienced in entertainment, and can offer the listener an unexpected degree of realism.

TYPES OF DISTORTION

The effectiveness of an audio system can be assessed in terms of various types of distortion. The systems that we are considering often include a measurement loudspeaker and microphone array in the room being measured. An anechoic recording of speech or music may be convolved with the measured room impulse response. The result is conveyed to a listener via a playback system – which includes the room in which sound is reproduced in addition to the audio components. In this type of multi stage system – where various stages may be implemented at quite different times – there are many possible sources of distortion from the reality represented. Understanding and managing such distortion at every stage of the process is key to a system's success.

Gain Distortion

Loudness is certainly one of the most salient auditory dimensions, and yet its role in sound-field assessment is easily overlooked, given the focus in research literature on temporal and spatial attributes of the sound-field. Sound level calibration of a sound-field reproduction can be difficult, but without it, any reproduced sound conveys an impression of mediation through audio technology rather than of simply the sound of an environment. Even without absolute calibration, if the *relative* gain between stimuli is not maintained, then the gain changes may have a futher undue effect on subjective impressions. In room acoustics, strength factor (G) is a measure of the added sound pressure level given by a room (and also from source-receiver distance), compared with the same (omnidirectional) source in an anechoic space at 10 m distance (ISO3382). In the author's experience, this is the most difficult standard room acoustical parameter to measure – because of the calibration requirements of the sound source. Similar demands exist even if G is not being measured, and great care is required if measurements made in multiple rooms, months or years apart, are to be compatible for simulation experiments.

While pistonphones are available for generating calibration signals for measurement microphones, there is often good reason to use other microphones (especially multi-channel) for room acoustics measurements. Calibration signals for such microphones may require improvised devices for efficient on-site calibration. Without this, there is a risk of gain distortion of individual channels (resulting also in spatial distortion) as well as gain distortion to all channels together.

In calibrating loudspeakers, the author favors noise signals (eg octave band noise centred on 1 kHz). Pure tone calibration, on its own, can be less reliable because loudspeaker frequency response may have fine peaks and dips that might not be entirely cancelled in inverse filtering. These fine variations in frequency response may not be audible in listening to normal program material, but have significant effects on pure tone reproduction.

There are several published sources of anechoic recordings of speech and music, which could be used for room acoustics simulation when convolved with room impulse reponses. However, generally these do not come with information that could be used for gain calibration, meaning that the playback level when such sources are used is either arbitrary or derived through educated guesswork. In some instances, the author has made and used anechoic recordings that have included a microphone calibration tone – yet even in these instances, the microphone only represents the sound in one direction from the source.

In some contexts, constant calibrated gain may not be desirable. For example, if loudness has a strong known effect on subjective impressions, yet a research project aims to investigate more subtle effects, then loudness-matched stimuli may be advantageous. While the exact means for doing this is not well defined, simply 'normalising' the stimuli (i.e. scaling the sound file to maximum) is not advisable, even if it is convenient. Other approaches exist, such as A-weighted or B-weighted L_{eq} matching, or applying a psychoacoustical loudness model for stimulus matching. The assumption behind such approaches is that there are not interaction effects between loudness and other parameters studied.

Spectral Distortion

The intelligent application of inverse filtering algorithms makes it possible to remove spectral distortion of a recording and playback system in many situations (accounting for both phase and amplitude distortions). However, there are limits to inverse filtering – and large spectral distortions may not be possible to compensate for (eg, it is not possible to meaningfully compensate for a frequency component that has no amplitude transferred in the system). Hence, audio system componenents with relatively little spectral distortion in the frequency range under study are desirable in both the measurement and playback system, since these are most amenable to the application of inverse filters. For measurement sound sources, inverse filtering is largely applicable to directional sources – for which there is a favoured direction. The complexity of current omnidirectional sources makes them poor candidates for inverse filtering, and simply flattening the power magnitude spectrum may be done instead.

Following standards such as ISO3382, a dodecahedral loudspeaker is frequently used in room acoustics measurements. Such devices are far from perfect for audio recording and reproduction in many respects, even though they are very useful for measuring room acoustical parameters. Dodecahedral loudspeakers introduce substantial spectral, temporal and spatial distortion (and, potentially, non-linear distortion). Usually these devices are designed for measurement frequencies in octave bands between 125 Hz and 4 kHz, and provide substantially less power outside of this range. Low frequencies may be conveyed by the addition of a subwoofer. Some compensation at high frequencies can be made through treatment of the measurement signal. Nevertheless, above about 1 kHz (depending on the size of the loudspeaker), the loudspeaker has a complex directivity pattern, due to beaming from each loudspeaker driver, as well as interference between drivers. This means that the frequency response in any one direction will not be flat, even if the spatially averaged power spectrum is flat. For room acoustical measurements, this means that the direct sound may have significant high frequency colouration, but that in general terms the reverberation is not coloured.

Temporal Distortion

Temporal distortion can be the result of errors in impulse response measurement – for example, temporal aliasing which occurs when the tail of an impulse response is wrapped around to the head due to an insufficiently long but repeated measurement signal, or false peaks in an impulse response which occur due to non-linear distortion in a maximum length sequence measurement signal. Such errors are easily avoided through an informed selection and application of the measurement technique.

A dodecahedral measurement loudspeaker has an impulse response length determined by the loudspeaker size. Considering the spectral and spatial distortion associated with dodecahedral loudspeakers, this temporal distortion may be a second order effect.

In room acoustics modelling, large simplifications are routinely made to room impulse responses. The first part of the impulse response is modelled with greatest precision, while the reverberant decay is modelled through statistical procedures. Such simplifications, which could be thought of as temporal distortion, are thought to have little or no perceptual effect when implemented correctly.

Spatial Distortion

This is, of course, the key question addressed in various ways by competing spatial audio room recording and reproduction techniques. However, spatial distortion also occurs prior to these through the measurement sound source (eg, when an impulse response from an omnidirectional loudspeaker is used to represent a directional musical instrument, or indeed a musical ensemble). As was implied above, an omnidirectional loudspeaker might be better described as multi-directional at high frequencies. Anechoic recordings, used in impulse response convolutions, tend to be made from a single microphone, which only represents the soundfield of the source in one position (the frequency-dependent directivity pattern of the source generally is neither recorded nor reproduced). Given the amount of research effort directed towards microphone array design for room acoustical measurements, greater effort is warranted in the design of measurement sources, and the representation of complex sound sources such as musical ensembles. The multiway dodecahedral loudspeaker of Witew and Behler (2005) is one instance of an improved omnidirectional source.

Apart from the dodecahedral source, other readily available options for measurement sources include other omnidirectional source designs (none of which comes close to ideal over the audible frequency range) and more conventional loudspeakers (such as studio monitors). Studio monitors may be effective for simulating the directivity of some sound sources, and have the advantage of a quite flat on-axis frequency response. For human speech and perhaps singing, a mouth simulator or head and torso simulator (HATS) may be appropriate. However, there is often a trade-off between available source sound power and its spatial fidelity – and a dodecahedral source has the advantage of substantial available sound power within its designed-for frequency range.

An option currently under development involves the recording of a sound source in an anechoic environment in multiple directions, and then reproducing the sound using a multichannel loudspeaker capable of approximately recreating the radiation pattern of the original sound source.

A well-known source of spatial distortion occurs with binaural recording using a dummy head microphone. Auditory localisation uses pinna-related spectral cues for identification of source angles around the aural axis, and differences between the pinnae of a dummy head and those of the listener results in vague and/or false localisation angles – although left-right discrimination is preserved. It is, perhaps, unfortunate that the most used source angle is directly ahead of the dummy head, an angle for which this type of spatial distortion is maximal. Auditory images tend to be localised above or behind a listener for this source angle, using binaural headphone reproduction (Cabrera and Gilfillan 2002).

Multichannel microphone and loudspeaker techniques, such are used for wave field sysnthesis, have a high frequency limit based on their spatial sampling interval (Start *et al.* 1995). Critical sampling occurs when the sample positions are half a wavelength apart, and spatial aliasing occurs above this frequency. This limitation is easier to solve for microphone arrays than loudspeaker arrays, because of the larger size of loudspeakers. For some aspects of spatial hearing, spatial accuracy in the high frequency range is important, posing a design challenge for arrays of this type.

Every type of spatial audio reproduction is limited in its effective reproduction space. Examples include the head-locked reproduction space of headphones (without head-tracking), the specific head positioning required for binaural cross-talk cancellation systems (using loudspeakers), and the limited 'sweet spot' of most multi-loudspeaker systems. These restrictions on the listener can be thought of as a type of spatial distortion.

Non-linear Distortion

The term 'distortion' is used most to refer to this type of distortion, which often occurs with excessive signal levels in loudspeakers and other analog and digital parts of an audio system. Non-linear distortion can have other causes (eg low bit depth in digital signal representation), and can even occur with very high sound pressure levels in the air (eg in the throat of a horn loudspeaker, or with sound-induced rattles of elements in a room). Since a static room acoustical environment is generally considered to be a linear system, distortion is neither desirable in the measurement system, nor a focus of measurement and analysis.

Non-linear distortion becomes an issue in measurement of rooms because signal-to-noise ratio concerns can make a high measureement signal sound power level desirable. If so, then the selection of a method with high imunity to non-linear distortion is required. Non-linear distortion can also be a problem in playback systems, where a large dynamic range is required, including in the low frequency range. Cross-talk cancelling systems may 'waste' low frequency energy, and so can require greater low frequency sound reproduction power than conventional loudspeaker systems.

Noise

Noise can be divided into the noise of the measurement system and the ambient noise in the situation being represented. The first type of noise is always undesirable, and fortunately current IR measurement techniques can achieve very high levels of noise immunity, at least when used intelligently in favourable circumstances. Common measurement noise artefacts include:

- The acoustic noise of the measurement equipment (eg cooling fans, hard disc drives);
- Noise in the analogue part of the electro-acoustical system;
- Digital audio noise (eg in signal processing);
- Temporal aliasing artefacts in impulse responses;
- Non-linear distortion artefacts in impulse responses;
- Time variance artefacts in impulse responses;

The second type of noise is usually undesirable, but may be desirable when it is considered to be part of the environment to be represented. Nevertheless, recording a noise-free impulse response, and recording the noise separately allows control of the noise in the reproduction, and may be preferable to simply recording noisy impulse reponses. Considering that background noise in many situations varies in time, this separate recording of noise also allows a much more detailed recreation of the background noise environment than a noisy impulse response.

TECHNIQUES AND THEIR LIMITATIONS

Impulse response techniques

While there are many techniques for measuring impulse responses, this section considers the two most popular ones for room acoustics: maximum length sequence and swept-sine.

The maximum length sequence (MLS) technique uses a pseudo-random signal which has approximately a white noise spectrum (Borish and Angell 1983). The signal could be thought of as a frequency-modulated square wave, as it alternates between two values. Cross-correlating the original signal with the signal returned through the system yields an impulse response.This is done efficiently through the fast Hadamard transform.

One limitation of the MLS technique is that the test signal has a white power spectrum, whereas it can be advantageous to use a pink power spectrum in room acoustical situations (because loudspeakers tend not to be designed for high power white noise, so the tweeter of a multi-way loudspeaker can be damaged; and because the background noise in rooms tends to be predominantly low frequency, so a signal-to-noise ratio advantage is obtained using a signal with greater low frequency content). Pre- and post-filtering the measurement signal is often done to create a pink spectrum.

MLS measurements are very sensitive to time variance, and the result of time variance is high frequency noise in the impulse response (increasing at 3 dB per octave). This limitation can also be thought of as an advantage, because the noise-free impulse responses from MLS measurements are characterised by precise temporal features (eg a well defined start to the direct sound).

The swept-sine measurement signal (also known as ‘chirp’ and ‘time-stretched pulse’) has become increasingly popular as computing power has increased. Convolving a swept-sine signal (swept at a constant rate per unit frequency) with itself time-reversed yields an impulse, and when the swept sine is played through a system it yields an impulse response. There are many variants of this technique, one of which is to sweep at a constant rate per octave, providing the same advantage as pink noise (versus white noise) does for MLS measurements. A further advantage of the constant rate per octave sweep is that is separates non-linear distortion components (especially harmonic distortion) from the undistorted system impulse response (Farina 2000).

Although a time-varying system cannot be fully represented by a single impulse response, the swept sine technique can be useful in acquiring workable impulse responses in such situations. Rather than being expressed as high frequency noise, time variance during the measurement period is primarily expressed as a temporal smearing of the spectrum, which allows room acoustical parameters to be measured effectively, and probably provides workable impulse responses for auralisation. Experiments in a stadium on a windy day (in which the loudspeakers were swinging in the wind) found that useable impulse responses are obtained in this way, while MLS signals yielded little more than the direct signal plus noise (Willasallen and Cabrera 2004).

Binaural systems

Binaural systems are attractive for room acoustics simulations because, in their simplest form, they are quite undemanding with respect to audio components, and yet they have the potential to recreate at a listener’s ears the sound-field at the measurement microphone’s artificial ears. Møller (1992) provides a summary of fundamental issues in binaural techniques.

Making a binaural recording involves more than just reproducing the sound recorded by dummy head microphones over headphones. If that is all that is done, then effectively the sound has a double effect of the pinna – first through the dummy head’s ears, and then through the listener’s ears. The transfer function between the headphones (on the dummy head) and the dummy head microphones should be compensated for by equalisation or inverse filtering for effective binaural reproduction via headphones.

Comment has already been made on the spatial distortions encountered in non-individualised binaural systems. To some extent, a loudspeaker-based binaural reproduction system known as stereo dipole overcomes some lof these imitations. Stereo dipole is a cross-talk cancelling system, with the loudspeaker pair close together (eg at 10 degrees) (Kirkeby *et al.* 1998). This arrangement provides a reproduced sound-field which is robust to minor head movements (Takeuchi *et al.* 1997). Furthermore, having a physical sound source directly in front of the listener appears to assist in creating a frontally located auditory image – which is very difficult to achieve with non-individualised headphone reproduction.

Substantial improvements in binaural reproduction occur with individualisation of head related transfer functions (Minaar *et al.* 2001) and head tracking (Spikofski and Fruhmann 2001).

Conventional stereophony

Conventional stereophony does not aim to recreate a sound field with the precision of other techniques outlined here. Therefore it is not widely used in room acoustics simulations for empirical work, despite being very widely used in entertainment audio. There are many different stereophonic techniques, but one that deserves mention is the ORTF technique, which uses two cardioid microphones, 110 degrees apart at a distance of 17 cm. Since the microphones are separated by the inter-aural distance, this technique is effective for both headphone and loudspeaker reproduction.

In a comparison of various stereophonic microphone arrays, Hugonnet and Jouhaneau (1987) find that coincident techniques (such as XY and MS) yield the most accurate lateral localization, while closely spaced techniques (including O.R.T.F.) yield the finest distance discrimination. In another comparison, Ceoen (1972) found a subjective preference for recordings made using the O.R.T.F. system (these were recordings of an orchestra in an auditorium), and this preference appears to be due to the configuration’s ability to convey the spatial impression of the auditorium.

Ambisonics

First order ambisonics has been a popular technique for spatial audio recording and reproduction, following a proposal by Gerzon (1975), and due to the commercial availability of the ‘soundfield’ microphone. First order ambisonics has also been supported by Lake Technology’s Huron and related spatial audio processors, and by Catt-Acoustic, which is one of the leading room acoustics modelling and auralisation computer programs. Ambisonics represents the sound-field using spherical harmonics, in a manner analogous to the representation of an audio waveform through Fourier analysis. First order ambisonics uses just four harmonics, in the form of an omnidirectional component and three mutually orthogonal cosine components (with a 'figure of eight' function). In the author's experience, this is a case of oversimplification of the sound-field – the results give a spacious auditory impression, but fail to give any precise sound localization. It is very difficult to achieve a frontally located phantom sound image using this system, and clearly audible phase disturbance occurs around the sweet spot in diffuse sound fields (so that in many cases the sound quality may be preferable outside the sweet spot). Much better results can be achieved with higher order ambisonics, but there are no production microphone systems for this. Farina and Ugolotti (1999) have also found such limitations to first order ambisonics for car cabin simulations.

Hybrid binaural + ambisonic systems (known as ambiophony) have been proposed, taking advantage of the best features of each technique (Farina *et al.* 2001). However, a more elegant approach is desirable for empirical applications.

Wave field synthesis (WFS)

WFS is an approach to soundfield reproduction proposed by Berkhout *et al.* (1993), which records and reconstructs waveforms using large arrays of spaced microphones and loudspeakers. In typical application, WFS is restricted to the horizontal plane. Microphone arrays may form a line, cross, circle, or other form (Hulsebos *et al.* 2002). The loudspeaker array may almost completely enclcose the listening space, and in some instances consists of hundreds of loudspeakers.

Current WFS systems can produce impressive results, but have practical limits of spatial sampling (based on the distance between loudspeakers, which imposes a high frequency or reproduction angle limit on accurate spatial reproduction) and the restriction to horizontal (which is merely due to the considerably greater expense in covering entire surfaces with independently controlled loudspeakers). Developments in loudspeaker design may help to solve these problems.

Ambisonics and WFS are two styles of sound field repressentation, both part of a more general concept called holophony (Jessel 1973). These techniques show great promise for high resolution spatial recording and reproduction, especially with advances in computing power and transducer design.

EXPERIENCE WITH AUDIO SYSTEMS

Members of the University of Sydney's acoustics laboratory have conducted many studies involving the simulation of acoustic spaces. This has included making room acoustical measurements using binaural and ambisonic microphone systems, and reproducing these for subjective tests (eg Bassett *et al.* 2003, Cabrera and Gilfillan 2002, Cabrera *et al.* 2004, Cabrera *et al.* 2005, Choi and Cabrera 2005, Osman *et al.* 2003). Some studies have employed the re-recording of program material in the room being measured, while others have involved impulse response measurement and convolution. In one currently unpublished study, a musical ensemble was recorded during several rehearsals, for comparison with impulse response measurements from the stage.

A majority of the subjective studies have involved non-individualised binaural reproduction using headphones. The studies have been concerned with acoustical quality, auditory spatial impression, auditory distance perception and auditory room size perception. The authors have been aware of the limitations of using this convenient technique, and so have recently been considering possible improvements.

In 2004-5, the author was involved in a study comparing four non-individualised two-channel techniques: ORTF stereophony, binaural headphones, stereo dipole and double stereo dipole (Martignon *et al.* 2005). The soundfields reproduced were derived from stereophonic and binaural impulse responses in five concert auditoria – the large, medium and small auditoria of Rome's *Parco della Musica*, Parma's *Auditorium Paganini*, and Kirishima's *Miyama Conseru*. The impulse responses in these auditoria had been measured using exactly the same equipment and technique (Farina and Ayalon 2003), with system gain settings fully recorded. A dodecahedral loudspeaker had been used as the sound source, and the measurement signal was a logarithmically swept sine wave adjusted for constant power in 1/3-octave bands.

The spectral and spatial distortion of the dodecahedral loudspeaker, which is especially evident in the direct component of a room impulse response, was compensated for by substituting an ideal direct sound for that recorded. For the ORTF impulse responses, a single sample impulse was used. For the dummy head impulse responses, an anechoic impulse response for the same dummy head (using a loudspeaker with a short impulse response) was used. In both cases the ideal versions replaced the originals at the same energy in the 500 Hz octave band. This frequency was chosen because it is within the omnidirectional band of the dodecahedral loudspeaker used. Some directionality was also desired for these impulse responses, so the remainder of each was attenuated by 3 dB so as to crudely simulate a weakly directional source facing the listener.

Gain distortion was minimised by making a calibrated anechoic recording of the musical instrument to be convolved with the auditorium impulse responses (piano accordion); ensuring that all gain adjustments were accounted for in impulse response preparation; and verifying that the direct component of the impulse responses followed the inverse square law (at least, in the vicinity of 500 Hz - where various high frequency artefacts are absent). A 500 Hz octave band calibration signal, with a known sound level relative to the calibrated anechoic recording, was convolved with the direct component of the impulse responses, and reproduced by the audio systems, allowing their gain to be calibrated. Then the anechoic music recording, convolved with the full auditorium impulse responses, could be played at about the same sound pressure level as would have occurred in the actual auditoria.

The response of each audio reproduction was corrected by inverse filtering to the listening position. Sound absorption was introduced into the reproduction room. A subjective experiment was conducted, in which subjects rated the room size, estimated the source distance, and rated the realism of the sound.

The subjective experiment found that the double stereo dipole system and binaural headphone system were rated as significantly less realistic than the ORTF stereo and stereo dipole systems. The result for double stereo dipole is easily explained by the fact that the subject's head was not fixed - and the front and rear stereo dipoles produce a standing wave pattern which has a very small sweet spot. Hence, when a listener moves their head, the high frequency content changes substantially due to interference between front and rear sources, defeating one of the primary advantages of single stereo dipole. This technique is inappropriate to use without either head clamping or precise head tracking.

The result for headphones was, to some extent, predictable, but also somewhat disturbing, since non-individualised binaural headphone reproduction is widely used for room acoustics auralisations and simulations. The low realism ratings for headphones seem likely to be due to the vague spatial imagery in non-individualised binaural headphone reproduction, as well as the head-locking of the sound field. The result for auditory distance perception stands in contrast to that of Minaar *et al.* (2001), who found that in a single moderately sized room, this reproduction technique can be effective for conveying auditory distance. For the auditorium situations tested, auditory distance perception using headphones bore little resemblance to actual distance, whereas a greater resemblance was found for other audio systems.

It seems perverse that conventional stereophony should receive high realism ratings compared to binaural headphones, because binaural techniques reproduce the sound field at the ears with greater precision. Perhaps this result can be partly attributed to familiarity with the sound of stereophony – considering that the listeners would have been familiar with the sound of commercially released recordings. However, the ORTF reproduction results showed almost no distinction between room size ratings and distance perception, whereas these were distinguished in the three binaural systems, apparently through the influence of inter-aural cross correlation.

The single stereo dipole system received similarly high realism ratings, as well as plausible distance estimates that were distinct from the room size ratings. Based on these results, as well as informal listening, the author favours the stereo dipole technique for two-channel non-individualised spatial audio rendering of room acoustic situations. Since that experiment, the author has worked futher with non-individualised stereo dipole for auditorium acoustical quality assessment (Jeon *et al.* 2005), and the resulting simulations in that experiment appeared to be much more effective than the headphone alternative.

Audio researchers at the University of Sydney are in the process of setting up a relatively high resolution spatial audio recording and reproduction system, using a 64-channel microphone system, and a hemi-anechoic reproduction space (carpeted floor, anechoic lining on walls and ceiling) with 76 loudspeakers. Participants include Craig Jin, Andre van Schaik, Simon Carlile, David Alais, the author, and others. The reproduction system will have eight full range loudspeakers (Tannoy DMT-15), 64 smaller loudspeakers (Tannoy V6) and four subwoofers (Whise 319A). There are many possible applications of this facility, including the simulation of acoustic environments such as has been discussed in this paper. One interest of the author is to investigate the usefulness of much simpler techniques (such as stereo dipole) through comparison with the comparatively high resolution of the new facility.

CONCLUSIONS

This paper outlines some of the issues encountered in attempting to record and reproduce room sound fields. The multitude of issues makes this a challenging task requiring great attention to detail. While there is still large room for improvement, effective results can be achieved using current techniques.

REFERENCES

Bassett, J, Osman, R, Choi, Y-J and Cabrera, D 2003 'A means of documenting the acoustic quality of spaces', *Australian and New Zealand Architectural Science Association Conference*, Sydney, Australia, pp. 104-116

Berkhout, AJ, de Vries, D and Vogel, P 1993, 'Acoustic control by wave field synthesis', *Journal of the Acoustical Society of America*, vol. 93, no 5, pp. 2764-2778

Borish, J and Angell, JB 1983, 'An efficient algorithm for measuring the impulse response using pseudorandom noise', *Journal of the Audio Engineering Society*, vol. 31, no. 7/8, pp. 478-488

Cabrera, D and Gilfillan, D 2002, 'Auditory distance perception of speech in the presence of noise', *International Conference on Auditory Display*, Kyoto, Japan (http://www.icad.org)

Cabrera, D, Ferguson, S, Tilley, S and Morimoto, M 2005, 'Recent studies on the effect of signal frequency on auditory vertical Localization', *International Conference on Auditory Display*, Limerick, Ireland (http://www.icad.org)

Cabrera, D, Jeong, D, Kwak, HJ and Kim, J-Y 2005, 'Auditory room size perception for modeled and measured rooms', *Internoise*, Rio de Janiero, Brazil

Cabrera, D, Nguyen A and Choi, Y-J 2004, 'Auditory versus visual spatial impression: A study of two auditoria', *International Conference on Auditory Display*, Sydney, Australia (http://www.icad.org)

Ceoen C 1972, 'Comparative stereophonic listening tests', *Journal of the Audio Engineering Society*, vol. 20, no. 1, pp. 19-27

Choi, Y-J and Cabrera, D 2005, 'Some current issues in computer modeling for room acoustic design,' *Acoustics Australia*, vol. 33, no. 1, pp. 19-24

Farina, A 2000, 'Simultaneous measurement of impulse response and distortion with a swept-sine technique,' *108th Audio Engineering Society Convention*, Paris, France

Farina A and Ayalon R 2003, 'Recording concert hall acoustics for posterity', *24th International Audio Engineering Society Conference on Multichannel Audio*, Banff, Canada

Farina, A, Glasgal, G, Armelloni, E and Torger, A 2001, 'Ambiophonic principles for the recording and reproduction of surround sound for music', Audio Engineering Society 19th International Conference.

Farina, A and Ugolotti, E 1999, 'Subjective comparison between stereo dipole and 3D ambisonic surround sound systems for automotive applications', *16th International Audio Engineering Society Conference*, Rovaniemi, Finland

Gerzon, MA 1975, 'Recording concert hall acoustics for posterity', *Journal of the Audio Engineering Society*, vol. 23, no. 7, pp. 569-571

Hugonnet C and Jouhaneau J 1987, 'Comparative spatial transfer function of six different stereophonic systems', *82nd Audio Engineering Society Convention.*, London, Great Britain

Hulsebos, E, de Vries, D and Bourdillat, E 2002, 'Improved microphone array configurations for auralization of sound fields by wave-field synthesis', *Journal of the Audio Engineering Society*, vol. 50, no. 10, pp. 779-790

Jeon, JY, Kim, YH, Kim, SY, Cabrera, D and Bassett, J 2005, 'The effects of visual input on the evaluation of the acoustics of an opera house', *Forum Acusticum*, Budapest, Hungary.

Jessel, M 1973, *Acoustique Théorique : Propagation et Holophonie*, Masson et cie, Paris, France

Kirkeby, O, Nelson, PA and Hamada, H 1998, 'The "stereo dipole" – a virtual source imaging system using two closely spaced loudspeakers', *Journal of the Audio Engineering Society*, vol. 46, no. 5, pp. 387-395

Martignon, P, Azzali, A, Cabrera D, Capra, A and Farina, A 2005, 'Reproduction of auditorium spatial impression with binaural and stereophonic sound systems', *118th Audio Engineering Society Convention*, Barcelona, Spain

Minaar, P, Olesen, SK, Christensen, F and Møller, H 2001, 'Localization with binaural recordings from artificial and human heads', *Journal of the Audio Engineering Society*, vol. 49, no. 5, pp. 323-336

Møller, H 1992, 'Fundamentals of binaural technology', *Applied Acoustics*, vol. 36, issue 3-4, pp. 171-218

Osman, R, Cabrera, D and Fricke, F 2003, 'Acoustic design of small music practice rooms', *Australian and New Zealand Architectural Science Association Conference*, Sydney, Australia, pp. 476-486

Spikofski, G and Fruhmann, M 2001, 'Optimization of Binaural Room Scanning (BRS): Considering inter-individual HRTF-characteristics', *19th International Audio Engineering Society Conference*, Bavaria, Germany

Start, EW, Valstar, VG and de Vries, D 1995, 'Application of spatial bandwidth reduction in wave field synthesis', *98th Audio Engineering Society Convention*, Paris, France

Takeuchi, T, Nelson, PA, Kirkeby, O and Hamada, H 1997, 'Robustness of the performance of the "stereo dipole" to misalignment of head position', *102nd Audio Engineering Society Convention*, Munich, Germany

Willsallen, S and Cabrera, D 2004, 'Assessment of music audio quality in a sports stadium', *117th Audio Engineering Society Convention*, San Francisco, USA.

Witew, I and Behler, GK 2005, 'Low-mid frequency measurement of single number parameters in room acoustics with multi-way dodecahedron loudspeakers', *Internoise*, Rio de Janiero, Brazil

Zahorik, P 2002, 'Assessing auditory distance perception using virtual acoustics', *Journal of the Acoustical Society of America*, vol. 111, pp. 1832-1846