

Multi-channel and binaural spatial audio: an overview and possibilities of a unified system

Brian Carty

1. Overview

Multi-channel reproduction of spatial audio can be approached in several ways. Typically, several loudspeakers are placed in a listening space and a discrete signal is sent to each. Many loudspeaker configurations and signal capturing/generation techniques exist. Systems vary from the omni-prevalent two channel stereo to large Wave Field Synthesis loudspeaker arrays. Stereo and quadraphonic spatialization algorithms are typically based on amplitude panning, which has recently been extended using vector based techniques. Other approaches include Ambisonics, which employs spherical harmonics to decompose the sound field, and Wave Field Synthesis, which is based on the Huygens Principle of Wave Propagation. Respective systems can be critically assessed based on attributes such as spatial accuracy, complexity and practicality.

Spatial audio can also be artificially simulated using binaural techniques. Binaural systems involve modelling/measuring the transformations made on sound from a particular location from source to tympanic membrane. These transformations can then be imposed on a non-localized sound, thus making it appear to originate from the measured location. An interesting fusion of the two techniques involves using binaural techniques to artificially localize loudspeakers as point sources. Any of the multi-channel loudspeaker setups above can thus be artificially recreated as headphone signals. Therefore, a listener can theoretically audition a desired loudspeaker setup in a desired listening space in headphones using binaural processing.

This chapter presents a brief introduction to spatial audio, followed by a discussion of the main approaches to multi-channel audio. Binaural techniques are then considered, from the point of view of possibilities of a unified system. Finally, benefits, limitations and applications of such a system are presented. A non technical approach is taken to the often complex algorithms and phenomena involved, in the

hope that an audience who may not be familiar with the topics covered can appreciate the final application.

2. Spatial audio

Spatial audio, in the context of sound reproduction, refers to sound existing in a three-dimensional space around a listener, and is primarily concerned with the location specific qualities of a sound/sound field. Sound spatialization refers to how sound is distributed in a specific environment. Sound localization, conversely, deals with how listeners perceive location specific parameters of the sound field in their listening environment.

Binaural hearing is the main factor involved in sound localization. Essentially, listening with two ears (rather than one) affords the brain two independent signals from the left and right ear respectively, which can be compared from a spatial point of view. Interaural Time Difference (ITD) is the name given to the time it takes a sound to reach one ear after it has first reached the other, and is one of the two main binaural phenomena used in sound localization. If a sound source is further from one ear than the other, a delay will occur in the time it takes the sound to reach the further ear. The further the sound source is from the lateral centre of a listener's environment, the greater this delay will be. ITD works best for lower frequencies, whose wavelengths are long with respect to the distance between the ears. Higher frequencies with shorter wavelengths can be ambiguous, thus causing the breakdown of ITD.

The other main binaural localization cue is Interaural Intensity Difference (IID). IID is based principally on the head acting as a barrier to sound and uses varying respective intensities of a signal at each ear to locate source sounds. Conversely to ITD, IID works best for high frequencies, as low frequencies tend to diffract around a listener, enveloping the head.

These binaural cues rely on differences between signals arriving at each ear to derive information about where sound sources lie in a sonic environment. However, when a source is directly in front of, behind, or above a listener (or indeed anywhere in the median plane), there are essentially no interaural differences. Monaural information can provide important localization cues in these cases. Monaural refers to independent information attainable from one ear. The physiology of the outer ear is complex and causes incoming sounds to be altered as

the sound waves interact with the various folds of the pinna. These interactions vary with source location. Thus the pinna's interaction with the sound is the main factor involved in localization in the median plane, particularly in determining if the source is in front of or behind a listener, where the back of the pinna will filter out higher frequency components of the sound.

Another phenomenon which becomes important, particularly when considering multiple loudspeakers is the Precedence Effect. If a similar sound arrives at a listener's ears more than once in quick succession, from apparently different locations, the sound is localized according to the first arriving wavefront.¹ This can become particularly pertinent when a listener's location relative to the loudspeakers in a multi-channel setup is non-ideal/variable.

3. Multi-channel audio

Multi-channel audio refers to sound reproduction using a number of loudspeakers. In a real world scenario, sound arrives at our ears from everywhere in our three-dimensional environment, implying the need for an infinite number of loudspeakers. Typically, however, a discrete number of loudspeakers are used in a multi-channel situation. An ongoing audio research challenge is how to best represent complex spatial environments with a discrete number of loudspeakers. A number of approaches will be discussed.

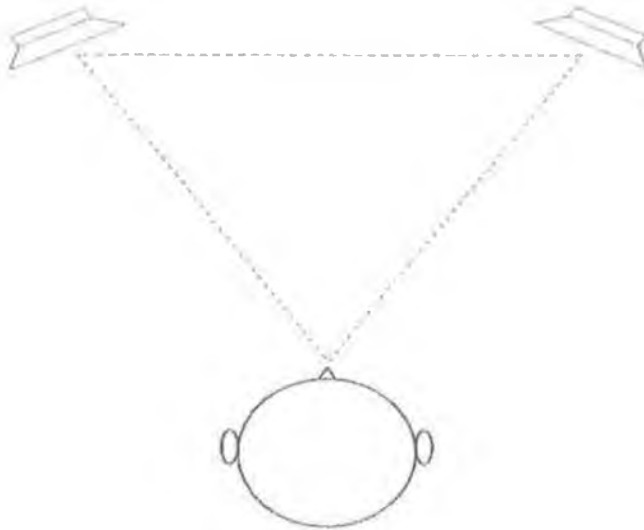
3.1 Two channel stereo

Strictly speaking, stereophony refers to any three-dimensional sound system; however the omnipresent two channel/loudspeaker stereophonic approach has commandeered the term and is typically referred to as 'stereo'. An equilateral triangle describes optimal listening conditions, with the listener just behind the base point/in the 'sweet spot', as shown in figure 1, below. Wider loudspeaker angles distort the spatial image leading to a 'hole in the middle' of the sound scene as the source collapses to loudspeaker locations. Amplitude/intensity panning is typically used in stereo systems, allowing a source to be artificially spatialized between the loudspeakers (by simply taking a mono signal

¹ B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th edn (London: Elsevier Academic Press, 1977), pp. 253–56.

and sending relatively more of it to the loudspeaker nearest the desired source).

Figure 1: Optimal stereo setup



Alternatively, stereo recording methods exist, using specific microphone techniques. Microphones can have different polar patterns/directional responses. For example, a figure of eight response will pick up sound from in front of and behind the microphone, whereas a Cardioid microphone will pick up sound predominantly from the front. Two figure of eight microphones, one facing left and the other right, in the same location can thus be used as a stereo microphone technique. The microphone facing left will pick up more sounds located to the left of the setup and vice versa. Upon playback on loudspeakers, a source that presented more energy to the left facing microphone will appear to come from the left. This technique is based predominantly on intensity differences. Alternatively, two microphones can be placed in a spaced pair configuration, perhaps to the left and right of a performing group. This setup will introduce time delays between the left and right microphone of the pair. If, for example, a source is nearer to the left microphone, it will arrive at this microphone first. Thus time differences are introduced to the reproduction of a spaced pair recording.

The stereo reproduction system is strictly limited and exhibits several drawbacks. The source can only exist between loudspeakers, severely limiting the spatial image, not allowing sources behind a listener or at the extremes of the horizontal plane, and not considering source height at all. As the user moves away from the sweet spot, spatialization gets progressively worse, as binaural cues become compromised. Also, amplitude panning does not provide exact spatial characteristics over the audible spectrum. Another relevant point to mention when considering any multi-channel loudspeaker setup is that a non-reverberant reproduction room is recommended to avoid reverberation, which can alter spatial images (unless a sound designer wishes to use the listening room's reverberant characteristics as part of the reproduction).

3.2 Vector Base Amplitude Panning

Vector Base Amplitude Panning (VBAP) can be described essentially as an extension of stereo. It was recently suggested by Pulkki and reformulates amplitude panning to vector bases for simplicity and efficiency.² VBAP can extend two channel stereo to any number of channels, even incorporating height information, leading to the possibility of three-dimensional sound. In VBAP, a source will use only the nearest loudspeakers to its desired location for spatialization using amplitude panning techniques. Therefore, if a source is at a loudspeaker location, only that loudspeaker will be used. If a source is between loudspeaker locations, only the two/three nearest loudspeakers will be used. In horizontal-plane-only reproduction, loudspeaker pairs will be used. In such a setup, sources above/below the loudspeaker array are impossible. However, in full three-dimensional reproduction, speaker triplets will be used, as the nearest three loudspeakers to the left and right of the source as well as above and below it will be used. Like stereo, a listener should be equidistant to loudspeaker pairs/triplets, ideally in the sweet spot, outside of which, the spatial image will be degraded.

² V. Pulkki, 'Virtual Sound Source Positioning Using Vector Base Amplitude Panning', *Journal of the Audio Engineering Society*, 45/6 (1997), 456–66.

3.3 Ambisonics

Ambisonics is a complete system of sound capture and reproduction, aiming to provide full three-dimensional spatialization with accuracy depending on the number of reproduction channels. Specific methods exist for sound field capture, storage and reproduction. The technique stores signals in B-format, which describes a sound field's overall pressure level and directional velocity levels. This can be done using a specific kind of microphone (a sound field microphone) to capture a sound field, or artificially using mathematical formulae to encode a particular source location to B-format. B-format signals store details on how much sound energy is coming from the front/back, left/right and up/down directions, as well as the overall sound pressure level.

Ambisonics is flexible in the number/location of loudspeakers employed for playback, although a regular layout is best. Typically, all loudspeakers work together to create the spatial image. B-format signals can then be decoded for the specific loudspeaker setup desired. This decoding can use psychoacoustically motivated formulae, splitting the task into a low frequency and high frequency process. The low frequency process uses ITD sensitive formulae, and the high IID based computations. Ambisonics constitutes a more physically-based approach to sound spatialization, and can be increased to a higher order, which gives increased spatial precision. Again, listeners should ideally be in the sweet spot.

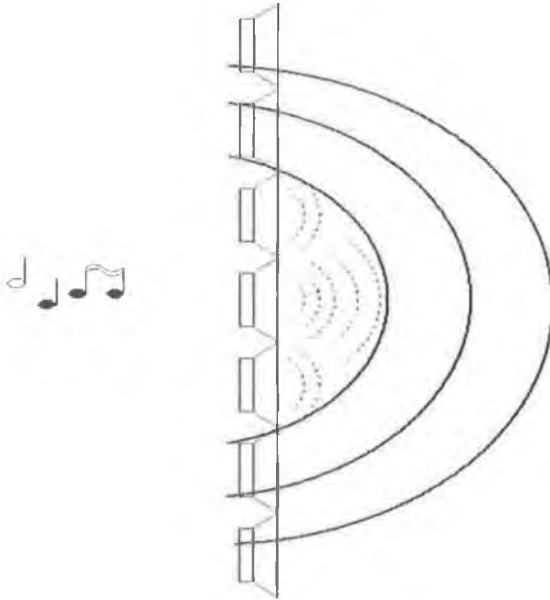
3.4 Wave Field Synthesis

Wave Field Synthesis (WFS) can be thought of as a 'Wall of Loudspeakers' approach. It is based on the Huygens Principal, which states that a wavefront can be represented by an infinite number of point sources, whose wavefronts add to recreate the original. In WFS, this principal is extended to a discrete array of loudspeakers, as shown in figure 2, below.

The main benefit of WFS is that there is no 'sweet spot'; the wavefront is recreated for the reproduction space. The wavefield is physically recreated, no longer trying to 'trick' a listener, as with previously mentioned approaches. Virtual sources will be located at the same point anywhere in the listening space, allowing a user to move around and gain an impression of the whole sound scene. Virtual sound sources can be placed on/outside the array of loudspeakers, infinitely

far away (with direction) or even between a listener and the loudspeakers.

Figure 2: A Wave Field Synthesis loudspeaker array. The outputs of the loudspeakers (dashed lines) add to represent the output of the source (full lines).



Practical difficulties include the sheer amount of loudspeakers needed, for example the WFS setup at TU Berlin consists of 840 loudspeakers in a lecture hall.³ As above, reverberation issues can be problematic. Accurate spatial reverberation for WFS is a complex and extremely computationally costly task. It also raises the issue of listening room interference. Mathematical issues also play a part, related to not using the ideally infinite number of loudspeakers (for example, spatial aliasing and diffraction at the edge of the array). Practically, a limitation to the horizontal plane is also logical (which not only limits the source sound to the horizontal plane, but also prohibits the reproduction of

³ M. A. J. Baalman, T. Hohn, S. Schampijer and T. Koch, 'Renewed architecture of the sWONDER software for Wave Field Synthesis on large scale systems', *Proc. 5th Linux Audio Conference*, Berlin, Germany, March 2007, 76–83.

reverberant reflections from the roof and ceiling, which are important for realistic environmental processing).

3.5 Surround sound (5.1)

The pervasive ‘surround sound’ that is swiftly becoming a consumer standard and is typically used in cinemas is referred to as 5.1, referring to five standard audio channels and one low frequency/bass channel. Three channels are located in close proximity in front of the listener and two behind. From the outset, the technique was always meant as more of an ‘experience enhancer’ than a true spatialization tool, and does not aim for full three-dimensional spatial accuracy. It is a front centric system, designed with a visual screen in mind. Part of the 5.1 protocol is to be compatible with stereo, as above, so a narrow frontal region, where spatial images are sharpest, is somewhat inevitable. Problems are similar to those of stereo systems, as amplitude based panning is typically employed (although ambisonic decoding for 5.1 has been suggested).⁴ Surround and front left/right channels subtend large angles, leading to phantom source problems. Listener location also poses a difficult challenge. In a scenario designed for a large audience, listeners may be significantly far from the ideal sweet spot, thus potentially ruining the spatial image.

4. Binaural

As multi-channel refers to multiple loudspeakers, binaural refers to headphone reproduction. Binaural techniques aim to accurately model how a source sound from a particular location is perceived at our ears, and will include all the localization cues mentioned above (for example ITD). Binaural techniques use headphones, so do not suffer from multi-channel loudspeaker listener location drawbacks (for example, there is no sweet spot).

4.1 Head Related Transfer Functions

Head Related Transfer Functions (HRTFs) are functions that describe how a sound from a specific location is altered from source to eardrum, and are typically used in binaural systems. For any particular source sound, a pair of transfer functions exist (for the left and right ear) for

⁴ F. Rumsey and T. McCormick, *Sound and Recording: An Introduction*, 4th edn (Oxford: Focal, 1992), p. 407.

any location relative to a listener. By imposing the HRTFs for a particular location on a mono source sound (which inherently includes no direct spatial information), it can be artificially spatialized to that location. Essentially, the process involves boosting or attenuating and delaying the frequencies contained in the source sound in accordance with how the ear treats the appropriate frequencies. ‘Real world’ sounds are made up of combinations of simple periodic sounds, with different frequencies, amplitudes/magnitudes and phases. HRTFs alter these component frequencies depending on the direction they come from. To reiterate, artificial binaural spatialization can be summarized thus: find out how the ears treat sound from a particular location and treat a source sound in the same way.

4.2 Moving sources

HRTFs are typically measured at discrete points around a listener/dummy head. Interpolation is needed if sources are required to be artificially spatialized at non measured points or are required to move. For example, if the HRTFs for a source directly in front of a listener and ten degrees to a listener’s right are known, a source can be placed at both locations. If the source is required to move from one location to the other, and the HRTFs are simply switched, the source will jump, and there will be a click/some noise in the output due to the complex processes involved in applying the HRTF to the source. In a real world scenario, the source sound will move smoothly between the two points, rather than jump between them. So, for a full three-dimensional system, HRTF measurements are needed for all locations. As this is not practical, a HRTF for ‘in between’ points is needed. This is a very complex task, with several difficulties inherent to the nature of HRTFs and the discipline of Digital Signal Processing (DSP).⁵ Briefly, the complexities and fine degree of detail involved in the HRTFs make it a non trivial task to derive an accurate intermediate HRTF between measured points.

If the HRTF is considered in the frequency domain, it is possible to see how the ear treats individual frequencies that make up a source sound, with regard to magnitude boosts/attenuations and

⁵ B. Carty, ‘Artificial Simulation of Audio Spatialization: Developing a Binaural System’, *Maynooth Musicology*, 1 (2008), 271–96.

delay/phase shifts. It is reasonable to use a simple interpolation algorithm to derive how an intermediate HRTF will treat magnitudes of component frequencies of a source sound, by looking at how it treats magnitudes of nearest measured points, and interpolating/deriving a relative 'in between' value. However, this technique is not successful for phase information (which contains time delay details, so is important for ITD), as phase is a cyclical property.

Various solutions to this challenge, including system updates and novel algorithm development, have been recently presented by the author.⁶ To summarize, the novel algorithms use the above mentioned magnitude interpolation and two novel approaches to phase interpolation. The first simply truncates phase values to the nearest known values. When a source is moving, a jump from one set of nearest phase values to another may cause an interruption in the audio output. This is dealt with using a user definable cross fade, fading out the old phase information and in the new. Thus any discontinuity in the output is perceptually removed. The second assumes the head is a sphere and uses geometry to calculate phase values. This method also looks more closely at the phase values in the available locations, in the low frequency range where phase delays/time differences are more important. This data is closely analysed and compared to the geometrically derived data. Inconsistencies are then corrected, leading to a psychoacoustically motivated solution, with accurate phase information, and thus accurate ITD in the important low frequency range.

5. Multi-channel binaural

HRTF processing can place a source sound anywhere in a listener's environment, as above. A simple, yet high potential link between multi-channel and binaural systems can thus be drawn. Virtual multi-channel systems can be simulated by placing a HRTF source at each loudspeaker location for a specific multi-channel setup. Multi-channel binaural systems can thus spatialize a source sound at each loudspeaker point for any of the mentioned multi-channel setups, using the process outlined above, with interpolation if necessary. The source sound at each

⁶ V. Lazzarini and B. Carty, 'New Csound Opcodes for Binaural Processing', *Proc. 6th Linux Audio Conference*, Cologne, Germany, February 2008, 28–35.

location will be the sound at the respective loudspeaker for the multi-channel setup in question. For example, to virtualize a stereo system, process the left and right stereo output feeds to appear to be located at the corners of the equilateral triangle described above. So, take the left channel of a stereo output and place it 30 degrees to the left of a listener using HRTF processing. Then do the same with the right output, placing it at 30 degrees to the right of the listener. Sending these processed signals to the listener's headphones creates a virtual optimal stereo listening configuration. Therefore, if a source is panned left, the virtual left loudspeaker will get more amplitude, as the amplitude panning law dictates. Crucially, the listener is always in the 'sweet spot' if these loudspeakers are kept static. Equally, using dynamic HRTF processing, a user can move around within a virtual listening space. Therefore a listener may, for example, virtually move to the left of the sweet spot and observe how the spatial image is distorted. Similarly, the signals derived from ambisonic (or indeed any) panning algorithms can be sent to multiple 'virtual loudspeakers'. A binaural stereo system is shown in figure 3, below.

Figure 3: Virtual stereo using multi-channel binaural techniques



5.1 Source distance from listener

Sound source distance from listener should be considered in a multi-channel binaural setup, as sound takes time to travel from source to listener. Also, the farther a sound travels to reach a listener, the more amplitude it loses. HRTFs are typically measured at fixed distances from a listener. Sources can therefore be delayed/attenuated to simulate distance.

5.2 Reverberation

When a sound source exists in an enclosed space, environmental processing/reverberation is also critical. Typically, if sounds exist in a room, the room interacts with the source. Sound will be reflected off walls/objects. Binaural systems do not interact with the listening room, as they are ideally reproduced using headphones. Therefore, reverb can be added artificially, or the multi-channel signal can be reproduced directly. Adding reverberation is a much more complex scenario than simply processing direct sounds using HRTFs, as it introduces the necessity to consider lots of reflections.

Multi-channel signals may have artificial reverberation added to them with vastly varying degrees of spatial accuracy. Alternatively, multi-channel recordings may include the natural reverb of the recording location (for example sound field microphone ambisonic recordings), or may contain no reverb at all (for example a synthetically created sound), allowing the listening room interaction to constitute all the reverberation. In the move to headphones two approaches can be taken. The system can artificially recreate how a specified multi-channel system/source sound would react in a user defined room, with user defined characteristics. Alternatively, the system can just play the multi-channel source, in a theoretically anechoic room. In the first scenario, the reverb is added virtually as part of the multi-channel binaural process (various approaches can be taken). In the second, the reverb is assumed to be encoded into the multi-channel source (for example an ambisonic recording in a reverberant environment, or a purposefully anechoic reproduction). Note that using HRTF based reverb may provide more accurate spatial reverb resolution than a limited multi-channel system.

5.3 Artificial reverb

As discussed above, a multi-channel binaural system requires an artificial reverb module for more accurate source spatialization. Artificial reverb involves modelling how a specific environment will affect sound. In an enclosed environment, the direct sound will reach a listener, followed closely by reflections off walls and other obstacles. The nature of these obstacles will define how they affect the sound. For example, curtains will absorb more sound than painted plaster walls, therefore sound reflected from plaster walls will be louder. Also, different surfaces will affect different frequencies in a non linear fashion, for example a particular surface may absorb high frequencies very well, but not lower frequencies.

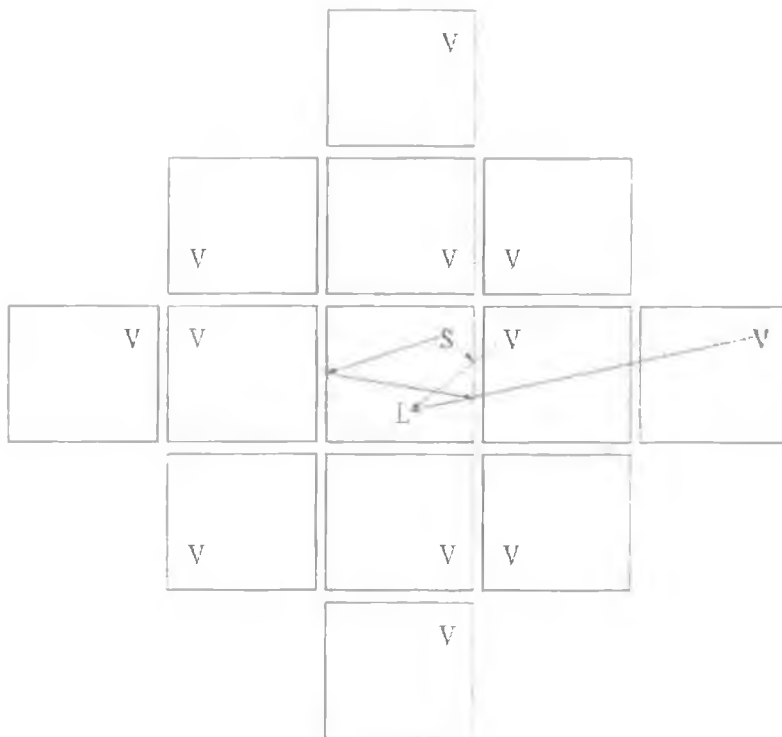
Reverberation in a room can be broken down into two significant parts. After the direct sound reaches a listener, a number of distinct early reflections promptly arrive. Early reflections help to inform the listener on the nature of the space, as well as the location of the source.

After the distinct early reflections, reflections begin to arrive at the listener from all locations, as the direct sound has had time to travel around the room in several trajectories. Sound is thus reflected off several walls before it reaches the listener, and is affected by each of these walls, as above. Thus this later reverb begins to fade out; as each reflection is reduced in amplitude as it travels further to reach the listener and is absorbed by several surfaces. As there are very many possible reflections reaching the listener at any one moment, the reverb swiftly becomes dense and no longer contains individual, localized reflections. This later reverb gives the listener further insight into the characteristics of the listening space and the distance of the source.

Several approaches can be taken to modelling reverberation artificially. Highly complex modelling is possible, but for a reliable result offering real time processing, the geometric image model can be effective.⁷ This model works by assuming there are virtual sources in rooms adjacent to the listening room. Figure 4 illustrates this model.

⁷ J. Allen and D. Berkley, 'Image Model for Efficiently Simulating Small Room Acoustics', *Journal of the Acoustical Society of America*, 65/4 (1979), 943-50.

Figure 4: The image model for artificial reverberation



The source is at point S, the listener at L, and virtual/image sources at each V. The source and listener are in the actual listening room. Other rooms are virtual/image rooms, containing virtual sources. Each virtual source can then be dealt with individually, with respect to distance from listener, effect of reflective surfaces and location. For example, the source in the virtual room to the immediate right of the actual room is further from the listener than the direct source, so will take slightly longer to arrive. It also originates more to the listener's right than the direct source and will be spatially perceived accordingly. The actual trajectory of the source is shown by the trajectory line in the actual room, showing the source hitting the right hand wall. Therefore this source will be filtered/attenuated according to the characteristics of this wall. Each virtual source can be treated in this way: reflections are traced back to virtual sources. Trajectories get more complex as sound is reflected off two or more walls before it reaches the listener. This can

be seen in the next virtual room to the right of the listener, which will be further again from the listener, at a more severe angle, and be affected by both the left and right walls.

The image model is desirable as early reflections can be accurately modelled using the first group of virtual sources. If the source or listener is moving, interpolated HRTF algorithms can be used to provide smooth movement of both the direct source and the early reflections. Thus the spatial cues that early reflections give regarding source location and room characteristics can be maintained. The later reverberation can be similarly modelled. As the arrival of the reflections becomes diffuse, the location of specific reflections is no longer necessary, so a generalized function can be derived.

This suggested model for artificial reverberation thus constitutes a useful application of the developed HRTF processes. The direct sound and early reflections are modelled accurately and a binaural generalized later reverb is employed. Crucially, the listener and sources can move smoothly in this setup, which maintains spatially accurate dynamic early reflections. The split of early and later reverb greatly increases computational efficiency, allowing real time processing. By design, the image model simplifies the nature of reflections and the reverberation process. However, a natural, efficient environmental processing tool is possible using the technique.

5.4 State of the art

Multi-channel binaural systems have been implemented. Binaural Ambisonics are used in auralization software, typically commercial products which go to incredible lengths to accurately model a room. 5.1 for headphones has also been implemented, essentially constituting a virtual cinema in headphones (used, for example, on long distance flight film presentations). An interesting new approach would be to allow a user to choose the multi-channel system algorithms/amount of loudspeakers/room size etc. Typically, systems are setup for sweet spot listening.⁸ It would be interesting to allow the user to move around within the loudspeaker system, i.e. out of the 'sweet spot', thus testing

⁸ M. Noisternig, T. Musil, A. Sontacchi, and R. Holdrich, '3D Binaural Sound Reproduction using a Virtual Ambisonic Approach', *IEEE Symposium on Virtual Environments*, Lugano, Switzerland, July 2003, 174-78.

the multi-channel system. In summary, the novel tool being suggested is a binaural *loudspeaker setup simulation*, as opposed to a room simulator/auralization tool.

5.5 System benefits

With HRTF only processing, each source has to be processed separately. In a multi-channel binaural setup, the number of HRTF processes is always the same as the number of loudspeakers. Therefore a large number of sources can be represented using a fixed amount of HRTF processes. For example a film will have many layers of sound effects, dialog, atmospheric effects, music etc., each of which may be spatialized to a different location. In a multi-channel binaural scenario, the number of HRTF processes will at most be the same as the number of loudspeakers, irrespective of the number of sources.

The practicality of not having to set up loudspeaker displays is also clearly advantageous. Furthermore, the often complex algorithms used to derive loudspeaker signals can be incorporated in the software, allowing immediate results.

Output can also be reproduced in a virtual anechoic room, a specialized and expensive facility to procure. WFS, for example, is designed for reproduction in an anechoic room.

5.6 Limitations

HRTFs are individual specific (everyone's ears are different), although generic sets perform quite well.⁹ Headphone monitoring is not always desirable; head tracking is needed for a more virtual reality based approach. The proposed tool also suffers from the limitations of the multi-channel system itself. For example, stereo will only allow frontal images. However, this is also an advantage from a system analysis point of view.

6. Conclusions and applications

Spatial audio has been introduced and discussed from psychoacoustical, multi-channel and binaural points of view. The possibilities of a flexible multi-channel binaural system have been considered. Such a system has

⁹ E. M. Wenzel, M. Arruda, D. J. Kistler and F. L. Wightman, 'Localization using Non-individualized Head Related Transfer Functions', *Journal of the Acoustical Society of America*, 94/1 (1993), 111–23.

many applications, some of which follow. A composer can audition a new multi-channel output using only headphones, providing a very practical solution to an otherwise difficult and perhaps daunting task. Films can be mixed in a 'virtual cinema'. Loudspeaker setups in specific rooms can be tested (with accuracy proportional to the complexity of algorithms employed). Of particular interest, multi-channel algorithms can be tested, allowing users to walk out of the sweet spot, for example. Collaboration with research into sound source localization (finding out where a sound lies in space) could allow the up-mix of binaural signals to any multi-channel format. Significant development of the room model and a collaboration with instrument modelling research could even allow antiquated instruments to be heard in models of historic auditoria (although this moves away from the loudspeaker approach).

Acknowledgements

This work is supported by the Irish Research Council for Science, Engineering and Technology: funded by the National Development Plan and NUI Maynooth.

Select bibliography

Primary bibliography

- Allen J. and Berkley, D., 'Image Model for Efficiently Simulating Small Room Acoustics', *Journal of the Acoustical Society of America*, 65/4 (1979), 943–50
- Baalman, M. A. J., Hohn, T., Schampijer, S. and Koch, T., 'Renewed architecture of the sWONDER software for Wave Field Synthesis on large scale systems', *Proc. 5th Linux Audio Conference*, Berlin, Germany, March 2007, 76–83
- Carty, B., 'Artificial Simulation of Audio Spatialisation: Developing a Binaural System', *Maynooth Musicology*, 1 (2008), 271–96
- Lazzarini, V. and Carty, B., 'New Csound Opcodes for Binaural Processing', *Proc. 6th Linux Audio Conference*, Cologne, Germany, February 2008, 28–35
- Moore, B. C. J., *An Introduction to the Psychology of Hearing*, 5th edn (London: Elsevier Academic Press, 1977)
- Noisternig, M., Musil, T., Sontacchi, A. and Holdrich, R., '3D Binaural Sound Reproduction using a Virtual Ambisonic Approach',

IEEE Symposium on Virtual Environments, Lugano, Switzerland, July 2003, 174–78

Pulkki, V., 'Virtual Sound Source Positioning Using Vector Base Amplitude Panning', *Journal of the Audio Engineering Society*, 45/6 (1997), 456–66

Rumsey, F. and McCormick, T., *Sound and Recording: An Introduction*, 4th edn (Oxford: Focal, 1992)

Wenzel, E. M., Arruda, M., Kistler, D. J. and Wightman, F. L., 'Localization using Non-individualized Head Related Transfer Functions', *Journal of the Acoustical Society of America*, 94/1 (1993), 111–23

Secondary bibliography

Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization* (London: MIT Press, 1997)

Howard, D. M. and Angus, J., *Acoustics and Psychoacoustics*, 3rd edn (Oxford: Focal Press, 1996)

Malham, D. G. and Myatt, A., '3-D Sound Spatialization using Ambisonic Techniques', *Computer Music Journal*, 19/4 (1995), 58–70

Theile, G., 'Wave Field Synthesis — A Promising Spatial Audio Rendering Concept', *Proceedings of the 7th International Conference on Digital Audio Effects (DAFX '04)*, Naples, Italy, October 2004, 125-32