



This Convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library, <http://www.aes.org/e-lib>. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

A microphone array for recording music in surround-sound with height channels

David v.R. Bowles

Swineshead Productions, LLC, Berkeley, CA 94708 USA
bowles@swineshead.com

ABSTRACT

In the past few years, sound recordings with spatial audio have moved from the realm of theoretical research to the actuality of physical and digital releases in the market. At present, three Blu-Ray disc formats utilize a traditional 5.1 surround-sound recording, with an added 4-channel layer of height channels. The topic of this paper is how to capture vertical localization effectively within this release format, utilizing existing research on hearing localization and techniques learned in the field. The proposed microphone array has time-of-arrival differences between all microphones, yet mixes down to 5.1 and stereo without excessive comb-filtering or other artifacts.

1. INTRODUCTION

As more recordings include height channel information we need to develop recording techniques based on an understanding of our own perceptual abilities, minus the ways we have learned to compensate with our “golden ears”. This includes not only understanding our listening environment, how the recordings will be released, how we hear in a three-dimensional plane, and how different microphone arrays represent a height channel layer. Our perception of sound sources is increasingly unreliable the further that source is located in a vertical arc. Therefore, a height channel layer needs information specific to “what is coming from above”, even if it sounds deficient when listened to alone, minus the context of the main layer.

1.1. Previous research

Multichannel recording and reproduction in three dimensions challenges us to regard ceiling reflections as important. Klepko points out that the ceiling is indeed the largest of an acoustic space's reflective surfaces, and presents challenges in surround-sound recordings [1]. Merimaa [2] and Williams [3] propose 12-channel arrays using the same type of microphone throughout. With the potential for cinematic and consumer releases containing 3-D audio, an extension of 5.1 surround with a 4-channel height layer was generally agreed upon. Theile and Wittek outlined some basic principles in considering 3-D microphone arrays [4]. Different microphone arrays are proposed within this framework by Lee and Gribben [5].

1.2. Listening environment

My critical listening to, and evaluating of height channel recordings was done in several different listening environments. This provided opportunities for me to hear recordings in differently-sized rooms, and with different speaker sizes and configurations as well. These were done at:

- 1) Dolan Studios, Steinhardt School of Music, New York University (three rooms: a control room large enough to hold twenty people; a small listening room with fifteen small matched speakers arranged in height, main and depth layers, and a studio space adapted for 9.1 playback), and
- 2) Dolby Laboratories, San Francisco (two rooms: a newly-constructed control room designed specifically for final mix and Atmos® encoding, and the theatre in their 100 Brannan Street headquarters)

1.3. Parameters of testing different mic arrays

I did some preliminary work when recording live classical music concerts for radio broadcast and/or for the clients' sound archives. I initially worked with just front height channels, as some of these situations allowed for mic stands to be placed only onstage (and where hanging microphones was not possible). This gave me some valuable initial insights: in some cases I was able to incorporate some of the height channel layer for additional ambience. Later on, I was able to start including rear height channel microphones in some of my projects for commercial release. In all cases, the principal consideration was that the main layer microphones could not be affected adversely by shadowing or reflections from additional microphones. A secondary consideration, equally important, was that the height layer needed to be incorporated into 5.1 surround and stereo mixes without introducing unwanted artefacts.

However, during test recording sessions, and in teaching at New York University, I was able to set up two concurrent height channel arrays without the above consideration; these experiments were invaluable for A/B comparison. I was able to compare different types of mic arrays, as well as microphones of different polar patterns placed together in coincidence.

2. LISTENING ENVIRONMENT AND RELEASE FORMAT(S)

Also important was to consider the targeted listening environment, how listeners are situated, and eventual release format.

2.1. Differences between studio, home theatre and cinema setup

For mastering studios, the mix position is set up so that tweeters of the main layer monitors are at 0°. The height layer is at maximum of 45°; those monitors are above respective L/R/LS/RS monitors, with time-alignment applied. Based on the ITU-R BS.775-2 specification [6], the surround layer is in a circular configuration. The mix engineer is sitting straight up with a head angle of 0°.

For cinemas, the room is more of a shoebox shape, with raked rows of seats. Front speakers are in columns or line arrays for proper dispersion throughout the space. Transitions between side and rear surround channels are blended to avoid sound effects "snapping" between speakers. Both the main and height layers are mounted considerably higher because of the aforementioned raking (which is more extreme in larger cinemas). While no additional dynamic compression is used, the X-curve EQ is applied on playback. These speakers are often mounted in the ceiling. The listener can be reclined at up to 30°.

For home theatres, the main layer is assumed at 20-30° from the horizontal plane, and the height layer at 45-55°. Room shapes, types of speakers and their placement vary widely. Speaker systems for height channels are evolving, with some models offering upwards-facing cones for outputting main and corresponding height channels simultaneously. In general, ceiling heights are low in this listening environment. In comparison to a cinematic release, home releases have varying amounts of compression applied: Blu-Ray has the least amount, followed by downloadable files and finally streamed content with the most compression. The listener is usually reclined at around 15°.

In general, height layer speakers will be more compact, and of lighter weight than the main layer speakers for ease in mounting and budgetary reasons alike. There is a resultant rolloff in low frequencies, which is compensated for by the subwoofer(s)!

2.2. Release format

The Blu-Ray standard not only introduced a disc of higher capacity but with a higher data transfer rate as well [7]. Multichannel audio-only Blu-Ray recordings were possible in 5.1 and 7.1 formats, with or without the Pure-Audio AES Standard [8]. The latter also introduced 9.1 playback: 5.1 surround plus a height channel layer. Blu-Ray releases in Auro3D [9] and Dolby® Atmos® [10] formats followed, with decoders/receivers in the consumer market in 2014. In 2015 DTS:X was announced as well. All of these systems utilize “lossy” encoding to ensure their discs will play normal surround-sound in earlier model players. Music-only Blu-Ray releases are rare, though the number of portals offering multichannel digital downloads is increasing. Digital downloads can offer up to full-bandwidth audio files (at the expense of download time and hard disc space); streaming has the highest amount of compression.

3. HEARING LOCALIZATION

The next step is to consider our hearing in the context of a three-dimensional field. Localization of sound sources therefore needs to be considered with the vertical axis in addition to the axes of width and depth.

3.1. Front vs. rear

It is well documented that our hearing is “forward-oriented”, and that head-shadowing plays a role, not only in defining left from right, but also in defining sound sources which originate from behind [11]. When moving beyond 90° left and right in the horizontal plane, specific localization is less as less possible, particularly at 180°. Pinnae also reduce perception of sound pressure levels for those sound sources which emanate from the rear [12]. As a result, surround channel levels need to be increased in order to be perceived at a similar loudness to front channels. Note that in the ITU-R BS.775-2 specification, angular placement of rear speakers varies from 100° to 120° - a total of 40° variance overall! Is this leeway a result of our loss of rear localization?

3.2. The vertical axis

Our lower body shadowing also serves to identify height cues as different from those coming from in front or below. Timing differences in floor reflections also identify whether those sounds are coming from above or

below. Localization accuracy starts to decrease at 45° in the vertical plane, and is particularly poor directly overhead [13]. We have adapted by turning our heads and bodies in reaction to sound cues to help us fix their position in space, particularly when our “fight or flight” reflex comes into play. While moving sound sources are easier to localize, music recording contains mostly stationary sound sources (with the exception of opera or musical theatre productions). Lastly, our traditional understanding of how musical instruments radiate sound is still based on a paradigm of a two-dimensional plane; re-learning sound radiation in a three-dimensional paradigm will help understand height channel recording better.

3.3. The listener’s position(s)

Keeping our hearing localization in mind, we need keep our heads vertical when listening critically in the mastering studio. We also need to be aware that listeners at home and in the cinema might not hear optimally if they lean too far back, and/or if the main and height speakers aren’t placed optimally.

4. A SUMMARY OF CHALLENGES IN HEIGHT CHANNEL RECORDINGS

In surround-sound recording, it is permissible for the rear channels to contain information which is general rather than specific, as the majority of surround-sound recordings use rear channels to convey ambience and reflections coming back from rear walls. Though we are less able to localize sounds coming from the rear, the surround layer is nevertheless in a horizontal plane with corresponds with good localization overall. The main challenge in identifying a height channel layer is to gather and represent the vertical axis in a way which compensates for our reduced localization at increased vertical angles. The height layer therefore needs to contain information specific to sound coming from above, whether it be ceiling reflections, or in some cases the sound sources themselves (see next paragraph below). Finally, the height layer should be able to be introduced into 5.1 surround and stereo mixes without excessive comb-filtering or other undesirable artefacts.

A secondary challenge is to determine whether the height channel layer will contain ambience alone, or whether it will also contain direct signals from members of the ensemble itself. This will depend on the size of ensemble, type of music being recorded and the focus desired by the producer: is the listener drawn to the

performers, or to the acoustic space? For groups on several levels of risers, the listener can similarly hear the sweep of the ensemble from floor to choir loft as well as from front to back and side to side. For the ultimate sound source of great height and width – the pipe organ – the listener can finally hear high ranks of pipes placed in proper vertical context of the acoustic space in which it is located!

5. MICROPHONE ARRAYS UNDER CONSIDERATION

Based on earlier experiments, my first step was to consider extending traditional microphone arrays using similar capsules throughout:

5.1. Single-point array

Height channels can be derived from a single-point array; it was always part of the Ambisonic specification to capture height information. No time-of-arrival differences exist between any of the channels. When deriving surround-sound channels from a single-point array, 1st-order polar patterns overlap excessively, as compared to a surround-sound array with some degree of spacing between microphones. Furthermore, when deriving height channels from a single-point microphone, resulting polar patterns will overlap significantly between the two layers. While 2nd-order polar patterns can theoretically be derived from a tetrahedral array, those patterns could be too narrow to provide proper imaging in the front channels; I suspect the same would hold true for the height channels. There are few choices of single-point microphone arrays currently on the market; it is my observation that many of these are not of high enough quality for music recording.

It is also possible to create a quasi-coincident array with three small-element Braunmühl-Weber microphones: two placed in the traditional M/S configuration with the third one pointed upwards (aiming the null axis at the horizontal plane). Geluso applied this principle as a single unit, and in combination with a spaced array in the main layer [14]. In general, a single-point microphone is an excellent choice for where portability and complete mixdown compatibility are desired, such as recording outdoor soundscapes or background ambience for film.

5.2. Semi-coincident array

These provide some time-of-arrival differences between main and height channel layers. In order to maintain the -3dB crossover point between the three front channel polar patterns, I found it necessary to space the left and right microphones further apart than the 17 cm. ORTF specification. This was confirmed by Williams' proposed multichannel cardioid array [15]. When using this semi-coincident surround array, imaging in front channels is excellent, but rear channels can sound "spotty" if they are spaced too far apart. When adding a height channel layer, with the height microphones above their corresponding channels, imaging was also good. Due to the wide polar pattern of cardioid microphones, this layer needed to be spaced well above the main layer if one desires clear separation. Semi-coincident arrays can be fashioned using high-quality microphones, can be made fairly compact and can be mounted on a single microphone stand. These are also good choices in excessively reverberant acoustic spaces, as the low-frequency rolloff tames excessive boominess. However, when recording large ensembles, extreme left and right sides aren't always well represented unless supporting "flank" or "outrigger" microphones are used. Those additional microphones should include their own height microphones to pick up height information. The final consideration is the many acoustic reflections between these nine closely-spaced microphones.

5.3. Spaced pairs with M/S height channels

Using Braunmühl-Weber dual element microphones, and recording each element discreetly, the engineer has an infinite possibility of polar patterns in post-production. This array can be spaced at any distance between left and right, and front and rear. At each left and right channel, an upwards-facing directional microphone is placed. Using M/S processing, the upwards-facing element can be narrowed to a figure-of-8 lobe (therefore eliminating any signal from the horizontal plane) and isolated as a separate signal. When recording single or just a few musicians, this height channel layer worked well, and could be mixed easily into 5.1 surround or stereo. However, there are no time-of-arrival differences between the main and height channel layers, making the differences between those two layers more difficult to discern. This array is more flexible with regard to ensembles of different sizes – as well as having the choice of any polar pattern for main and height layers alike!

5.4. Omnidirectional arrays

These provide full-frequency response in all channels and work well with large ensembles or instruments such as the pipe organ. As with all spaced arrays, the distance between each of the microphones depends on the size of ensemble being recorded, the acoustic space and the desired imaging in the final mix. Since the polar pattern IS omnidirectional, a height layer spaced too closely to the main layer won't sound all that much different. Moving the height layer further away improves time-of-arrival difference between the two layers, making them sound distinct from each other. Further improvements result from pointing the microphones of the main layer at a downwards angle of 30-45°, and the microphones of the height layer at a corresponding upwards angle. [This is due to the behavior of most omnidirectional microphones: as the frequency rises past 1 kHz, the response is increasingly directional.] However, too much distance between the two layers creates excess time-of-arrival difference and a "hole in the middle" between the main and height layers! Another problem of this microphone array can be excessive low-frequency presence in the height layer, making those sound sources sound poorly-localized (as if they came from more than one area). My observation is that while main and height layers sound good on their own, when played together (or mixed down to surround or stereo) it is sometimes possible to discern comb-filtering artefacts.

6. ARRAYS WITH DIFFERENT TYPES OF MICROPHONES IN THE HEIGHT CHANNEL LAYER

6.1. Spaced omnidirectional array with centre hypercardioid pair

There are plenty of situations where I record live performances in dry acoustical spaces, and can't use more than one microphone stand. To an array of spaced omnidirectional microphones, I added an upwards-facing semi-coincident pair of hypercardioid microphones (spaced with the ORTF distance of 17 cm. but at an angle of 90°). The two polar patterns would then overlap at their respective -3dB points. At times, I also used a directional microphone in the main array to pick up a centre signal. This height channel information isolated sound reflections coming from above, and could also be blended into a stereo mix as ambience. However, I was aware of the differences in time-of-arrival between the two pairs. This meant that a similar hypercardioid pair could not be used with surround

channels (which I space further apart than the main pair).

6.2. Spaced omnidirectional array with coincident figures-of-8

The next step was to experiment with upwards-facing directional microphones semi-coincident to the spaced omnidirectional microphones. This was a similar approach to 5.3 above, but with microphones from different manufacturers. First was to place a figure-of-8 microphone with the null axis pointing horizontally and the positive lobe facing upwards. Using M/S decoding, the resultant "upward-facing" signal was used for a height channel, and the "downward-facing" signal discarded. Floor reflections were minimized and information in the height channels came from directly above the corresponding main layer microphone. There were some difficulties using this configuration: 1) placing the two microphones closely enough without the figure-of-8 blocking the omnidirectional microphone, 2) a rather narrow horizontal imaging of the height channel layer (with a potential "hole in the middle" in rear height channels), and 3) an excess low-frequency build-up resulting from specific two microphones chosen.¹ In post-production considerable DSP resources needed to be used in order hear the main layer and four height channels derived from multiple mid-side chains. In theory this meant perfect coincidence between a channel in the main layer and its corresponding height channel. In practice the lack of time-of-arrival differences between the main and height layers reduced the perceived differences between these two layers.

6.3. Spaced omnidirectional array with spaced hypercardioid layer

I now acknowledged the need to separate the main and height layers entirely, and to keep the height channel information specific to sound arriving from above. I started out with the surround array I use for most of my

¹ With most M/S recordings, pressure-gradient microphones are used; these have a low-frequency rolloff. As described, I was using a true pressure omnidirectional microphone along with a figure-of-8 which had been equalized internally to have a flat response. Whether the signals were played in 9.1, 5.1 or stereo, the phase-coherence meant that the low frequencies were emphasized, particularly when recorded in an already reverberant acoustic space.

recordings². For the height channel layer, I used four hypercardioid microphones located directly above their corresponding omnidirectional microphones. This layer was now placed above the main layer; angling the hypercardioids at 60° upward so their null axis pointed directly below. I then angled them outward to provide better separation between left and right height channels (with resultant wider imaging in the height layer). These angles helped to define the height layer as distinct from the main layer. Additionally, since hypercardioids have a low-frequency rolloff there is less chance of undesirable low-frequency build-up when listening to the full array.

7. LISTENING AND EVALUATION

7.1. Mixing down vs. 9.1 reproduction

My facility has separate rooms for mixing stereo and surround-sound recordings, therefore I was able to evaluate mixing down to these formats immediately. Mixing some of the height layer signal into these mixes introduced ambience without losing too much focus on the main microphones. When playing these mixes to different clients, they commented on how “real” or “spacious” they sounded.

When played back on a 9.1 system, listeners commented upon the presence of the height layer, but that the effect was not exaggerated.

7.2. Enhancements and improvements

For large ensembles, more “upward-facing” hypercardioid microphones can be used for gathering height information from sound sources. This will help in gathering height information from the centre and edges of larger ensembles, as well as from the audience noise in the case of live pop concert recordings.

² My system for recording 5.1 surround-sound consists of a main layer of omnidirectional microphones for left and right front and rear channels. Depending on size of musical group and/or desired imaging, a directional microphone can be used for the centre channel. Surround microphones are spread further apart than the left and right pair, pointing backwards at angles approximating speaker spread and angle in the ITU-R BS.775-2 specification. Taking into account head shadowing, this increased distance between the surround microphones helps to establish “rear imaging”.

Further improvements to the current Bowles array: 1) using additional microphones coincidentally-mounted with the hypercardioids, then using M/S processing to split the front and rear lobes of each height microphone (finally discarding the rear lobe signal entirely); or, 2) use of DSP-enabled microphones which carry out the same function internally. For the second option one danger is that one model on the market is a shotgun microphone; the resulting polar pattern might be too narrow for proper imaging in the height channel layer. Either of these two options will 1) enable the height channel layer to be isolated even better from the main layer, and 2) enable similarly-isolated recording of a top “voice of god” layer for Auro-3D encoded Blu-Ray releases.

8. CONCLUSIONS

When considering microphone arrays, it is important to visualize sound sources, and microphone polar patterns (and their frequency response) in three dimensions. This is even more vital when recording in surround-sound and with height channels. Recording the vertical axis also forces us to consider ceilings just as important as floors, side walls and rear walls in acoustic spaces. Representing this vertical axis accurately is important because of our diminished ability to localize sounds coming from vertical angles greater than 45° - similar to our diminished ability to localize sounds accurately coming from behind.

This “Bowles array” proves to be a good solution for a surround-sound recording with a height channel layer: it contains information specific to ceiling reflections, with sufficient isolation from floor reflections. This height layer can also be mixed into a traditional surround-sound or stereo recording as ambience, without excessive comb-filtering. It is in a true 9.1 playback environment that these recordings really stand out, presenting the listener with a height layer which is clearly defined yet integrates with the surround layer.

9. ACKNOWLEDGEMENTS

This work was supported by Professor Paul Geluso and Dr. Agnieszka Roginska, New York University

10. REFERENCES

- [1] Klepko, John: 5-Channel Microphone Array with Binaural-Head for Multichannel Reproduction (Faculty

of Music, McGill University, Montreal); doctoral thesis, September 1999

http://www.collectionscanada.gc.ca/obj/s4/f2/dsk1/tape9/PQDD_0016/NQ55349.pdf

[2] Merimaa, Juha: Applications of a 3-D Microphone Array (Helsinki University of Technology, Lab. of Acoustics and Audio Signal Processing, HUT, Finland); Paper no. 112 presented at AES112, 1 April 2002

<http://www.aes.org/e-lib/browse.cfm?elib=11337>

[3] Williams, Michael: The Psychoacoustic Testing of the 3D Multiformat Microphone Array Design, and the Basic Isosceles Triangle Structure of the Array and the Loudspeaker Reproduction Configuration (Sounds of Scotland, Le Perreux sur Marne, France); Paper no. 8839 presented at AES134, 4 May 2013

<http://www.aes.org/e-lib/browse.cfm?elib=16740>

[4] Theile, Günther; Wittek, Helmut: Principles in Surround Recordings with Height (VDT, Geretsried, Germany; Schoeps Mikrofone, Karlsruhe, Germany); Paper no. 8403 presented at AES130, 13 May 2011

<http://www.aes.org/e-lib/browse.cfm?elib=15870>

[5] Lee, Hyunkook; Gribben, Christopher: On the Optimum Microphone Array Configuration for Height Channels (University of Huddersfield, Huddersfield, UK); eBrief no. 93 presented at AES134, 3 May 2013

<http://www.aes.org/e-lib/browse.cfm?elib=16694>

[6] ITU-R BS.775-2:

https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.775-2-200607-S!!PDF-E.pdf

[7] Blu-Ray capacity and data transfer rate: http://www.blu-raydisc.com/Assets/Downloadablefile/White_Paper_General_3rd_Dec%202012_20121210.pdf

[8] AES-21id, Screen-less navigation for high-resolution audio on Blu-ray Disc <http://www.aes.org/standards/blog/2011/4/aes-21id-screenless-nav-for-blu-ray>

[9] van Daele, Bert; van Baelen, Wilfried (R&D Engineer Auro Technologies NV; CEO Auro Technologies NV)

<http://www.auro-3d.com/wp-content/uploads/documents/Professional-Workflow-White-Paper-v0-6-20120228.pdf>

[10] Dolby® Atmos® Specifications

<http://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-specifications.pdf>

[11] D. Wesley Grantham, Spatial Hearing and Related Phenomena (Chapter 9, "Hearing", ed. Brian C.J. Moore; Academic Press, London 1995):

<https://books.google.com/books?id=OywDx9pxCMYC>

[12] Computational Auditory Scene Analysis (Chapter 5, "Binaural Sound Localization", R.M. Stern, De L. Wang and G. Brown; IEEE, 2006):

<http://www.cs.cmu.edu/~robust/Papers/SternWangBrownChapter.pdf>

[13] Parseihian1, Gaëtan; Jouffrais, Christophe and Katz1Brian F. G.: Reaching nearby sources: comparison between real and virtual sound and visual targets (Frontiers in Neuroscience, 2 September 2014)

<http://journal.frontiersin.org/article/10.3389/fnins.2014.00269/full>

[14] Geluso, Paul: Capturing Height: The Addition of Z Microphones to Stereo and Surround Microphone Arrays:

<http://www.aes.org/e-lib/browse.cfm?elib=16233>

[15] Williams, Michael: Magic Arrays — Multichannel Microphone Array Design Applied to Microphone Arrays Generating Interformat Compatibility (Sounds of Scotland, Le Perreux sur Marne, France); Paper no. 7057 presented at AES122, 1 May 2007

<http://www.aes.org/e-lib/browse.cfm?elib=14042>