

SCALING NEW HEIGHTS IN BROADCASTING USING AMBISONICS

Chris Baume

BBC Research and Development,
Centre House, 56 Wood Lane,
London, W12 7SB
chris.baume@bbc.co.uk

Anthony Churnside

BBC Research and Development,
New Broadcasting House,
Oxford Road, Manchester, M60 1SJ
anthony.churnside@bbc.co.uk

ABSTRACT

As the world's biggest broadcaster, the BBC transmits over 400 hours of audio content every day – the vast majority of which is in stereo. This paper will look at why the BBC is interested in Ambisonics, and describe recent experiences in trying out the technology in its first-order format. Two subjective listening tests are described, which attempt to discover how Ambisonics compares to current technology, and how much the height dimension contributes towards the listening experience. Finally, some suggestions are made on how to make Ambisonics more accessible, in the hope that more Ambisonic content would be created as a result.

1. INTRODUCTION

One of the six public purposes of the BBC is to 'deliver to the public the benefit of emerging communications technologies and services' [1]. For this reason, BBC R&D thoroughly explore what technology is available, and advise on which can deliver the best experience and value for the audience. For nearly 40 years, the BBC has been broadcasting the vast majority of its audio content in stereo [2]. The only change from this has come from the BBC HD television service, which since 2006 has been broadcasting most of its output using 5.1 Dolby Digital [3].

With modern technology allowing broadcasters to transmit content in new and interesting ways, the BBC is looking at what improvements can be made to how audio is created and delivered. 5-channel surround is considered to be the easiest option, due to its wide-scale adoption by film studios, television and some radio stations. However, before an investment is made in a particular format, it would be wise to see if an alternative approach could provide a longer-term, more flexible solution, whilst still being able to handle existing formats.

2. PROBLEMS

There are many problems with current audio formats which either make it difficult for the audience to listen to content appropriately, or demand extra cost and effort from the broadcaster. In this paper 'current formats' are considered to be stereo and 5-channel surround, as they are the two formats currently used by the BBC. 5-channel surround is defined as the ITU 5.1 (3 front / 2 side) speaker layout, but ignoring the '.1' LFE channel in this instance. Some of the problems that broadcasters face with current audio formats are explored in this section. However, most of the problems are not exclusive to broadcasting, and are more industry-wide.

2.1. Trend

Looking at where the future of 'surround sound' is heading, it can be seen that 7-channel surround (or 7.1) is already an established format within the film industry, and there is already talk of using 9.1 and 11.1 formats. In addition, NHK research labs in Japan are proposing 22.2 as a future audio format [4]. There is a clear trend of simply adding more discrete speaker channels, which is simply not sustainable. The problem for broadcasters is predicting where this trend will end, and knowing to which format to commit, and when.

2.2. Compatibility

The trend of increasing 'X.1' formats is not necessarily a problem in itself, but rather the problem lies in their incompatibility with each other. Changing between formats requires processing using up- or down-mixing algorithms. This problem has already raised its head in the BBC HD channel, where programmes are sometimes created with a stereo soundtrack, then up-mixed to 5.1 for broadcast. This processing can compromise the audio quality and, as there are no standards governing their use, the end result can be unpredictable.

The problems resulting from incompatibility extend to the production end, where two separate mixes need to be made. For some BBC programmes, three separate audio mixes are made using three separate broadcast vehicles: a stereo mix for radio, a stereo mix for SD television and a 5-channel mix for HD television. Being able to create these mixes simultaneously would bring a significant cost benefit.

An increasing number of formats also brings problems when it comes to archiving material. In archiving, it is beneficial to reduce the number of formats to as few as possible, to ensure that they can be replayed in the far future.

2.3. Speaker positions

A problem with all current audio formats is that they are based on discrete speaker feeds, requiring that they are replayed using a specific speaker layout. While it is usually not a problem placing a pair of stereo speakers, it is much more difficult to arrange a 5-channel surround layout correctly. This is particularly a problem with the centre speaker, which often needs to share its space with a television screen. Although this technique means that there is no processing to be done between the source and amplifier, it places restrictions on where people can position speakers. In many (if not most) instances, this will cause listeners to position speakers in the wrong place, and result in a compromised listening experience.

Similarly, many people listen to stereo content over headphones. This situation is far from ideal, as the sound ends up coming from ‘inside’ the listener’s head. Binaural technology can help by filtering the content using head-related transfer function data, but commercial solutions for this are currently limited.

3. CASE STUDY: THE LAST NIGHT OF THE PROMS

The BBC Proms is an eight-week season of classical music concerts, held primarily in the Royal Albert Hall in London. The event is famous for the final concert of each season – the ‘last night’ – in which popular classical pieces are played, ending in a sequence of very British music, including ‘Rule Britannia’ and the national anthem.



Figure 1: Last Night of the Proms 2009, with the sound field microphone circled. Image credit: Chris Christodoulou

3.1. Setup

The Proms is recorded for both radio and television using upwards of 120 microphones, the layout of which varies very little from year-to-year. Previously, sound field microphones have been used either for crowd noise, or for 5-channel surround. In 2009, BBC R&D added a Soundfield DSF-2 microphone for the purposes of recording for B-format. The microphone was placed in a central position, approximately 3 metres behind, and 5 metres above the conductor’s position. It was connected to a microphone controller backstage, whose B-format digital outputs were recorded using a pair of synchronised sound cards and a laptop computer. The end of the microphone was pointed towards the centre of the strings section of the orchestra, and the controller was set to ‘end fire’ mode. This places the sound image of the orchestra mainly in front, and the audience behind and below.

3.2. Result

The resulting recording, when replayed using a suitable periphasic speaker array, is an excellent example of what can be achieved using only a single microphone. The acoustic of the space is clearly captured, and there is a good balance of sound from the orchestra. There is a significant amount of audience noise, partly due to there being a noisy crowd, but also due to the fact that the microphone captures sound directly below it, which picks up coughs very well. Those experienced in performing at the Royal Albert Hall, who have listened to the recording, have commented on instantly being able to recognise the acoustic of

the space – something that would be much harder to notice with a stereo mix.

Critics of the recording would say that the orchestra lacks definition in instrument positions, and that most of the audience comes from below, which sounds unnatural.

3.3. Spot mics

In addition to the B-format recording of the concert, the multi-track of all the raw microphone outputs was recorded. Mixing the spots-mics in with the B-format recording will allow for a much more balanced sound, and will help address the problem of audience noise. In addition, this could be done at a much higher order, providing a better spatial resolution.

There has been interest in using the combination of B-format and spot microphone recordings to develop an algorithm for automatically panning and setting the levels for each microphone output in the sound field. This could potentially allow engineers to arrange the spot mics, place a sound field microphone in the ideal listening position, and for the microphone signals to be panned and mixed automatically.

4. CASE STUDY: THE WIZARD OF OZ

In November 2009, a radio drama entitled ‘The Wonderful Wizard of Oz’ was created for BBC Radio 4. The hour-long show is an interpretation of the famous Wizard of Oz story, told through voice acting, music and sound effects. BBC R&D joined the production team in an attempt to gather material suitable for putting together a periphasic demonstration piece of radio drama. In return, we provided them with material that was used to create a 5.1 mix of the programme.

4.1. Setup

Recording for the programme was done over three days in BBC New Broadcasting House in Manchester, using one of the radio drama studios. The studio contains two rooms, with moveable partitions to allow the creation of various acoustic spaces. Many different props are on-hand, and there is a small team of foley artists available. One of the rooms is a dead-space, isolated from the other, and with padded walls to reduce reverberation.

Various combinations of microphones are used, often arranged as a Blumlein pair, and a couple of spot-mics. They are arranged to optimise the stereo image, and the techniques have been developed over many decades.

BBC R&D joined the production team for the recording, and brought a Soundfield ST250 microphone. The microphone was arranged alongside the usual setup to capture the voice acting in B-format. For many situations, the microphone was just placed in front of the actor in the same way as a spot mic. For other situations, actors worked together in a circle around the sound field microphone, which took advantage of its periphasic nature.

The final mix was a combination of dialogue – using both B-format and panned mono sources – and panned stereo sound effects, using both custom foley recordings and sound effect library content. The mix was made using Steinberg Nuendo and a VST Ambisonic encoder.

4.2. Results

The Wizard of Oz was chosen as a suitable programme due to the number of scenes that could exploit the added height dimension.

For instance, in the first scene, Dorothy's house is sucked into a tornado with her inside, creating an excellent opportunity to have wind and objects spinning around the listener.

It was very encouraging to see the sound engineer and producers getting excited about how height could be used in the programme. They would often bounce ideas off each other and be visibly enthused over where to place tornados and flying monkeys in the sound field. At one point, this led to the producers asking one of the actors to lie on the floor under the Sound-field microphone, to capture the mother's scream from below as Dorothy was sucked into the tornado.

The resulting sound provided a convincing atmosphere and environment, with the tornado effect being particularly pronounced. However, the localisation of the dialogue was less than satisfactory.

5. LISTENING TEST: EFFECT OF HEIGHT

Although Ambisonics is much more than a method of recording and replaying 'with-height' audio, it is often cited as a '3D' audio format. There is no doubt that the ability to include height is one of the major draws of Ambisonics, and it gets people interested and excited about the technology. However, there are very few people who have experienced – never mind heard of – 'with-height' audio. For this reason, the effect that height information has on the listening experience is not well understood. In deciding whether it is something worth investing in, the effect needs to be investigated to find what works well and what doesn't.

In order to do this, a subjective test was designed in which a number of audio items were replayed using a variety of speaker layouts, some of which included speakers above and below the listener. Although Ambisonics was used to record and replay the test items, the intent was not to test the performance of Ambisonics itself.

5.1. Setup

The test was conducted using the MUSHRA test method [5]. Four different speaker layouts (or configurations) were considered, plus a hidden reference and an 'anchor'. Participants were required to give each of the six configurations a score, based on how it compared to a given reference configuration. Five audio test items were used for the test, each one being played in a 30-second loop while the participant rated each configuration.

A listening room in BBC R&D's former base in Kingswood Warren in Surrey was used to conduct the listening test. Twelve PMC DB1-SA active monitors were used in the layout shown in Table 1 and Figure 2. Six of the speakers were arranged in a hexagon layout in the horizontal plane, with three arranged in a triangle layout above the listener, and three in a triangle below (rotated 180°). A hexagon was chosen as it is something akin to a 5-channel surround setup. Triangles were chosen because using more speakers would mean they couldn't be placed with a great enough elevation.

The twelve speakers were used in five different configurations for the purposes of the test:

- **Hex** - Hexagon of speakers in the horizontal plane
- **HexTri** - All of the speakers, consisting of the hexagon in the horizontal plane, and the triangles above and below the listener.

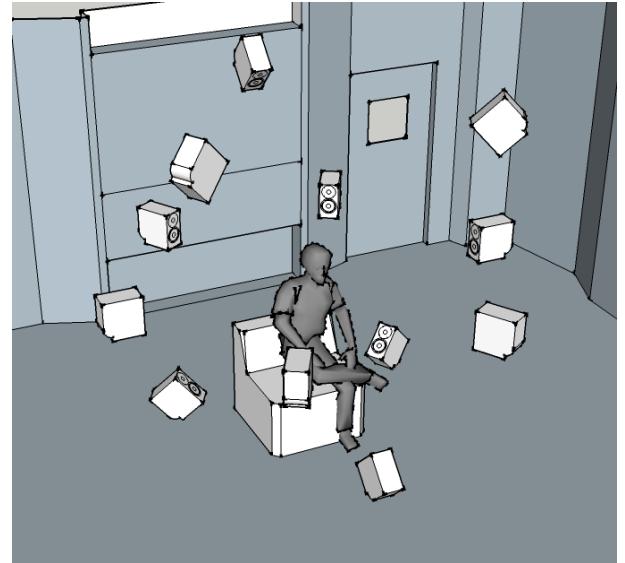


Figure 2: 3D model of the speaker layout used for the height listening test

Speaker	X	Y	Z	Azi	Ele
1	Front	-	Down	0°	-45°
2	Back	Left	Down	120°	-45°
3	Back	Right	Down	240°	-45°
4	Front	Left	-	30°	0°
5	-	Left	-	90°	0°
6	Back	Left	-	150°	0°
7	Back	Right	-	210°	0°
8	-	Right	-	270°	0°
9	Front	Right	-	330°	0°
10	Front	Left	Up	60°	+45°
11	Back	-	Up	180°	+45°
12	Front	Right	Up	300°	+45°

Table 1: Speaker positions for the height listening test

- **HexUp** - Hexagon in horizontal plane, and the triangle above the listener
- **HexDown** - Hexagon in horizontal plane, and the triangle below the listener
- **Tri** - Triangles above and below the listener

Decoding matrices were generated for each configuration, which ensured that the overall sound level would be fairly consistent across each one. Maximum velocity decoding was used for frequencies below 400Hz, and maximum energy decoding was used for frequencies above. The decoder used was Fons Adriaensen's 'AmbDec' [6], and the decoding matrices used are listed in Tables 5 to 9.

As part of the MUSHRA recommendation, an 'anchor' must be included in the stimuli. A 3.5kHz low-pass filtered version of the reference is recommended, however for the purposes of this test it was considered unsuitable. Instead, a version of 'Hex-Tri' was used where the Z channel was ignored, which has the effect of removing the height information. This configuration was named 'HexTriNoZ'. Interestingly, the speakers above and below the listener will still be used, outputting the horizontal

components of incoming waves. The results of this configuration in relation to ‘HexTri’, should show whether the sensation of height is due to the inclusion of height information, or just because speakers are placed above and below the horizontal plane.

To speed up the test process, a user interface was designed to let the test participants dynamically control the speaker configuration in use. The GUI was modelled on examples in the MUSHRA recommendation, and was implemented using Java and Swing. The interface could be controlled using a keyboard and/or a mouse, and allowed users to control the speaker configuration, as well as give a score for each one. The user’s actions and final results were saved locally in spreadsheets for later analysis. The software worked by sending MIDI messages to another PC running Steinberg Nuendo, which was used to play the audio.

The descriptive anchors used for the test were ‘Much better’, ‘Slightly better’, ‘About the same’, ‘Slightly worse’ and ‘Much worse’. To line up with these, a numeric scale of +20 to -20 was used.

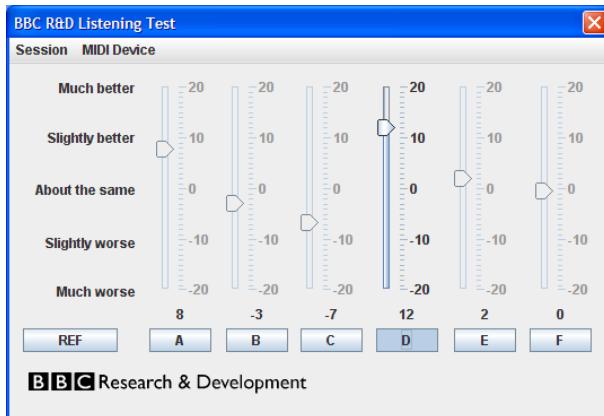


Figure 3: GUI used for the height listening test

5.2. Items

Five separate audio test items were used in the listening test. Selecting suitable items was one of the most difficult parts of the experiment, as it can greatly influence the results. A variety of music and atmospheric items were chosen, with only two of the items containing explicit audio sources above the listener, the rest relying on reverberation for height content. Each of them is described below:

• Proms music

The raw B-format output of the Soundfield microphone system was used, containing a clip of classical music followed by a bit of applause. No point sources were mixed into the content, as that was not available at the time. As mentioned previously, the orchestra appears in the front of the sound field, with the audience below and behind.

• Thunder

This item was recorded using a Soundfield ST-250 microphone at Kingswood Warren during a thunder storm. The microphone is held under an umbrella, so there is significant height content from the rain hitting the umbrella directly above. In addition to the rain, there are two large cracks of thunder in the distance, which can be heard echoing around the listener.

• Wizard of Oz

A clip taken from the B-format version of the Wizard of Oz drama. It contains both dialogue and atmospheric sound effects, followed by a loud, swirling tornado effect.

• Classical music

This clip was taken from Ambisonia.com, and was made by Aaron J Heller. The recording is of an orchestra playing Beethoven’s Symphony No. 4 in B-flat major, and was made using a Calrec Soundfield MKIV No. 99. Although there is noticeable reverberation, the orchestra sounds very close and is much ‘drier’ than the Proms recording. There is also more bass content, and unlike the Proms clip, it contains no applause.

• Proms atmosphere

Taken from the same recording as the ‘Proms music’ clip, this does not contain music, but rather only audience noise. This includes clapping, laughing and horns from around the venue. The reverberation of the space is very noticeable in this clip.

5.3. Results

18 people took part in the listening test, 6 of whom had heard periphonic audio previously, and 9 of whom had experience of critical listening. Each of the speaker configurations will be considered in turn, looking at any interesting results and comments that occurred.

• Reference

Every participant was able to identify the hidden reference the vast majority of the time. Only in 5 out of 95 cases was it scored outside of the range -5 to +5. See Figure 4.

• HexUp

As the speakers below the listener were not used for this layout, the speakers above had to output sounds from both above and below. This configuration was expected to perform badly because of its irregularity, however some participants found the sound preferable. Whilst many found this configuration to sound much like the reference, some participants spoke of the sound image being ‘squashed’ or becoming ‘narrower’ – particularly in the music items. In the more diffuse items, some enjoyed the sensation of height that the speakers above provided, but many commented on tonal differences. See Figure 5.

• HexDown

Again, being an irregular layout, the speakers below the listener outputted sound from both above and below. Similarly to HexUp, some participants found the sound preferable in some situations. Many – if not most – participants commented on the louder bass. This was attributed to having speakers on the floor, which raises an interesting issue that affects the type of speaker that should be used. Some participants commented that the configuration was more atmospheric and had better localisation, whilst others made opposing comments. See Figure 6.

• Tri

Despite using none of the speakers on the horizontal plane, the Tri configuration was quite popular in a number of situations. These tended to be diffuse, atmospheric soundscapes where there were few identifiable sources, such as the ‘thunder’ test item. The height was noticeable by most,

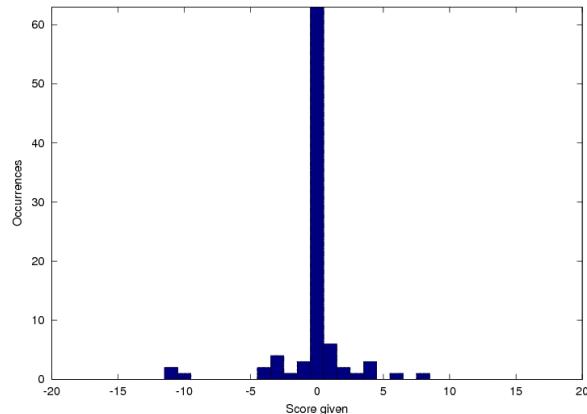


Figure 4: Frequency distribution of all scores for Hex

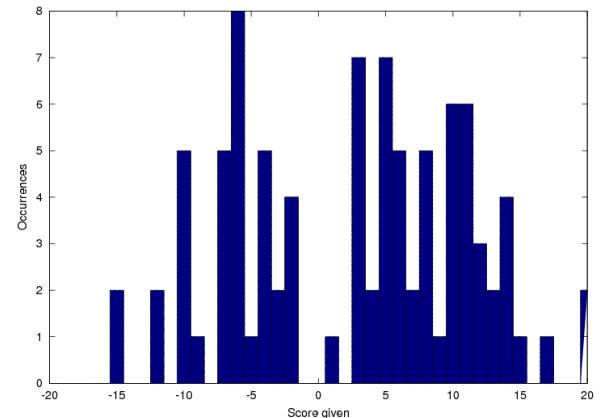


Figure 7: Frequency distribution of all scores for Tri

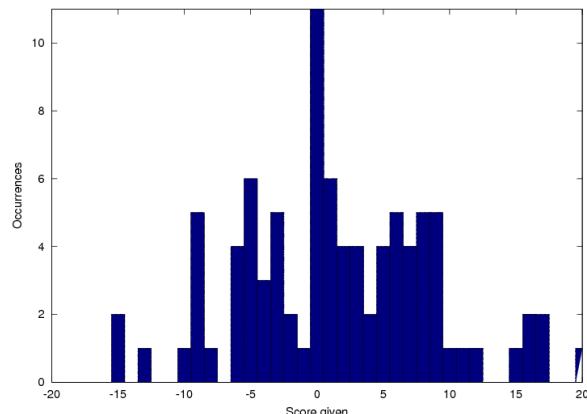


Figure 5: Frequency distribution of all scores for HexUp

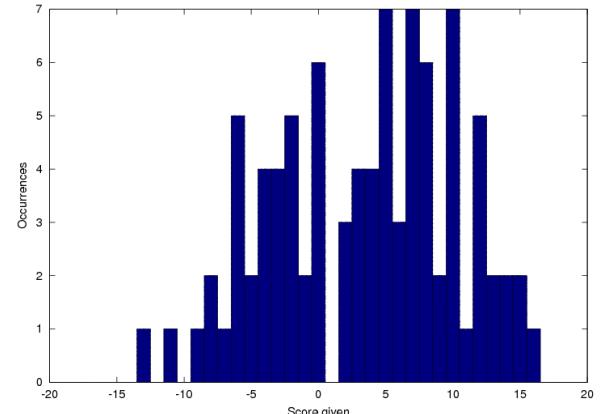


Figure 8: Frequency distribution of all scores for HexTri

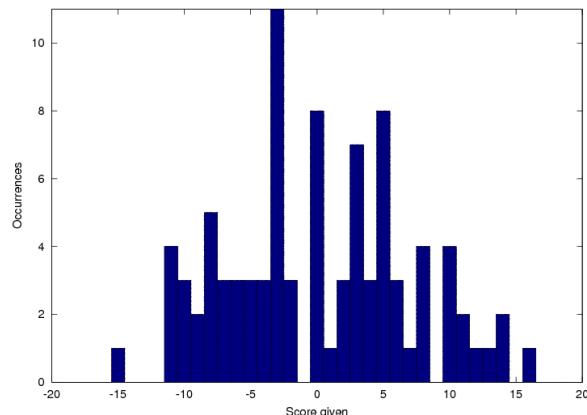


Figure 6: Frequency distribution of all scores for HexDown

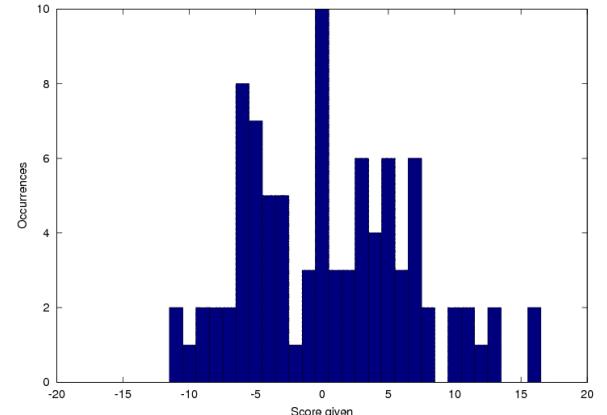


Figure 9: Frequency distribution of all scores for HexTriNoZ

and comments included words such as ‘immersive’ and ‘enveloping’. As with HexDown, many noticed the increased bass due to having speakers on the floor. Negative comments included sound noticeably coming from the front-down speaker, and having large gaps in the sound field. See Figure 7.

• **HexTri**

This configuration, being the most regular and covering the entire 3D soundfield, was expected to perform much better than anything else. Although it did emerge with the highest overall score, the gap was not as large as expected. Some commented that it didn’t sound very different from the reference, but many more commented on the space, distance and atmosphere that the configuration brought. See Figure 8.

• **HexTriNoZ**

Having lost its height information, this configuration was expected to sound and perform much like the reference. The overall score turned out to be very close to zero, but this is mainly due to an equal amount of positive and negative scores, rather than a cluster of scores close to zero. Predictably, many commented on its similarity to the reference, or that it sounded ‘flatter’ or ‘duller’. However, some commented on its ‘good height’ or better ‘sense of space’.

5.4. Conclusions from height listening test

It is hard to draw firm conclusions from the results shown in Figures 4 to 9 and Table 2. In the end, the periphonic configuration HexTri emerged with a narrow lead, but also drew criticism from many participants. What did become clear is that ‘with-height’ audio works much better in some situations than others. There was a clear preference for speakers above/below when using atmospheric, non-directional content. However, for music – where most sources are in-front and in the horizontal plane – there was no clear preference for the use of high/low speakers.

An issue which arose as part of the test is placement of speakers on the floor. For rear-ported speakers, this brings an undesirable bass boost, so front-ported speakers should be used where available. The effect could also be filtered, but the phase-shift incurred by this may cause problems with the resulting sound.

Overall, it can be said that for some people and situations, the effect of height is noticeable and desirable. However, it is unclear why opinion often differs when considering the same content, using the same configuration. The test described here was very general, and did not look at any one property of the listening experience. Further work is warranted into trying to find specific reasons into how height can improve the listening experience, and what needs to be done to achieve that.

Config	Average
Hex	-0.11
HexUp	1.24
HexDown	0.74
Tri	0.88
HexTri	2.66
HexTriNoZ	0.08

Table 2: Mean scores for each configuration, for all items (range of -20 to +20), in the height listening test

6. LISTENING TEST: STEREO/5.0 COMPARISON

In attempting to assess the value of Ambisonics, it is important to directly compare its performance against that of current technologies. The two most used formats in the BBC are stereo and 5-channel surround, so these were used as a benchmark. A listening test was conducted to directly compare stereo, 5.0 and first-order Ambisonics in terms of listening experience. A total of 15 people took part in the test.

6.1. Items

The nature of the test requires that the material used needs to exist in all three formats. As this is quite rare, most of the items were created for the test. Five items were used, and each item was mixed to make best use of the format being mixed for. This meant that the 5.0 mixes used the rear speakers where appropriate, and the Ambisonics mixes used the height dimension. Each of the items were limited to 30 seconds to reduce listener fatigue. A description of each is written below:

1. Classical Music: BBC Proms 2009.
A clip of a classical piece, made by mixing the sound field and point source microphone signals.
2. Radio Drama: The Wonderful Wizard of Oz.
A mixed piece from the radio drama, where Dorothy’s house is sucked into a tornado. Contains both dialogue and sound effects.
3. Popular Music: The Get Out Clause.
A contemporary band recording, with a simple acoustic sound.
4. Jazz Music: BBC Proms 2009.
Using the same technique as the previous recording, but with a jazz piece, featuring a singer and trumpeter.
5. Radio Drama: The Wonderful Wizard of Oz.
As above, but for the opening of the drama, where the sound effects are much more subtle and spacious.

6.2. Set Up

The listening test was conducted in an old radio studio in BBC New Broadcasting House in Manchester, with low reverberation. 16 PMC DB1-SA active monitors were used – 14 for Ambisonic playback (arranged as shown in Table 3), with another two for stereo. The 5.0 layout used the stereo speakers, and three from the Ambisonic layout, following the ITU-R BS.1116-1 [7] recommendation. The Ambisonic layout consisted of a hexagon in the horizontal plane, and a cube where the speakers had an elevation of $\pm 45^\circ$ relative to the listener’s head.

6.3. Method

The MUSHRA test method was used for this test. Using the 5.0 signal as a reference, stereo, Ambisonics, a hidden reference and two hidden anchors were tested against it. The hidden anchors were corrupted 5.0 (rear L/R replaced with front L/R at -6dB), and corrupted Ambisonics (Z channel ignored). Participants were asked to score each configuration relative to the reference using the ± 3 scale recommended in Miyasaka [8] (shown in Table 4). Each 30-second test item was played on a loop, until the subjects had finalised their score for each configuration. The

Speaker	X	Y	Z	Azi	Ele
1	Front	Left	Down	-45°	-45°
2	Front	Right	Down	45°	-45°
3	Back	Right	Down	135°	-45°
4	Back	Left	Down	-135°	-45°
5	Front	Centre	-	0°	0°
6	Front	Right	-	60°	0°
7	Back	Left	-	120°	0°
8	Back	Centre	-	180°	0°
9	Back	Right	-	-120°	0°
10	Front	Left	-	-60°	0°
11	Front	Left	Up	-45°	+45°
12	Front	Right	Up	45°	+45°
13	Back	Right	Up	135°	+45°
14	Back	Left	Up	-135°	+45°

Table 3: Speaker positions used for the playback of the Ambisonic material in the stereo/5.0 comparison listening test

participants could dynamically switch between configuration using the custom test software described in Section 5.1. In addition to the scores, participants were asked for any verbal comments on the sound of each configuration.

3	Much better
2	Better
1	Slightly better
0	The same
-1	Slightly worse
-2	Worse
-3	Much worse

Table 4: Scoring system used for the stereo/5.0 comparison listening test

6.4. Quantitative Results

Figure 10 shows average scores, and 95% confidence intervals for stereo and Ambisonics, when compared to 5.0. The results are separated into all test items, musical test items and the drama test items to highlight the different results given for each style. Although Ambisonics is the favourite in most cases, it is far from conclusive. With the musical items, stereo is clearly not satisfactory, but the difference between 5.0 and Ambisonics is more subtle. In most other cases, stereo is slightly worse and Ambisonics slightly better.

6.5. Qualitative Results

Although the quantitative results don't display a clear preference, the comments of the participants shed a little more light on the situation. Almost no positive comments were given for the stereo item, with the majority of negative comments made against the classical and jazz music from the Proms. With the drama pieces, more than half of the participants commented that they struggled to score the piece because they preferred the sound effects when played using Ambisonics, but preferred the dialogue when using stereo. A typical comment would be "the sound effects are really good but vocal is not so good, she sounds

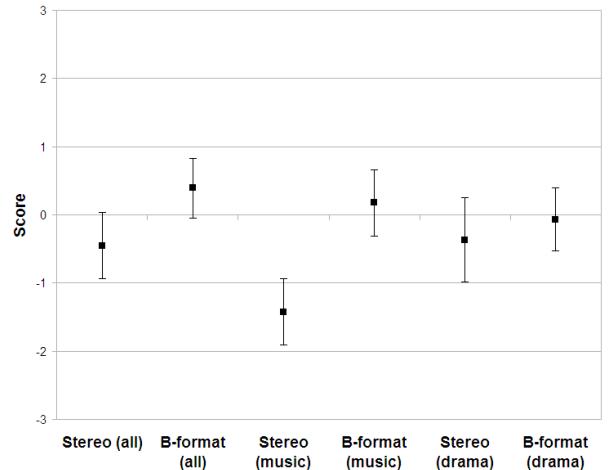


Figure 10: Mean scores, shown with 95% confidence intervals, for the stereo/5.0 comparison listening test

muted, the sound effects are lovely". More than half of participants commented on how they felt like they were in the performance, when using Ambisonics. However, this did not always correlate with the highest score. Typical comments included "I feel like I'm too in it" and "You feel like you're in it, but not listening to it".

6.6. Conclusions from stereo/5.0 listening test

The quantitative results of the test are somewhat inconclusive. However, by looking at the comments as well, some tentative conclusions can be drawn. Musical material appears to make the best use of speakers around the listener. A preference for the Ambisonic playback of the Proms classical piece was shown, as it contains a large amount of sound surrounding the listener. 5.0 was preferred for pieces that featured more obvious point sources, such as dialogue, while less directional sound sources worked better with Ambisonic playback.

Further investigation into how higher-order Ambisonics could further improve the listening experience is of interest.

7. CHALLENGES

Despite being almost 40 years old, the majority of interest in Ambisonics still remains in academia and the living rooms of enthusiasts. The past decade has seen a large rise in its use outside of these environments – namely by the video games industry and artists looking to play with 3D audio – but there appears to be little or no interest from the film, television or radio industries. This section aims to analyse the reasons which prevent Ambisonics from being used in the 'mainstream', based on the experiences detailed in this paper, and the perspective of a broadcasting context. It hopes to spark conversation into how the technology can be promoted, and made available to a wider community.

7.1. File format

It is no secret that there is a desperate need for a standardised Ambisonics file format. Although the '.amb' format is widely

used and accepted, it can only support B-format and does not address the future needs of the technology. Metadata should be at the heart of the standard, where as much information as possible is included. Such a format should be scalable, and include crucial information such as the order (including support for mixed orders), channel order and which normalisation function was used to encode.

7.2. Tools

There are a large number of tools available for the creation, manipulation and decoding of Ambisonic signals. The vast majority are created by academia and enthusiasts for their own purposes, and in many cases are made available publicly using the internet. However, most of these tools are unsuitable for a broadcast environment.

Rarely is there documentation about how the tools are put together, and it is therefore difficult to know exactly what is happening to the audio without reverse engineering it. The tools are usually limited to one interface such as the VST plugin standard, and compiled for only one platform. This causes problems when many broadcast environments use Pro Tools or Logic on Mac OS X. In addition, most are only capable of handling first-order Ambisonics, which makes it difficult to experiment with higher orders.

One particular area that lacks suitable tools is decoding. As it manages the playback of content, a decoder could be considered to be the most important part of the ‘chain’. Speaker setups and listening rooms can vary wildly, so decoding Ambisonic signals in a suitable way can be complex. A number of decoders are publicly available, but are limited to either a choice of preset layouts, or require a decoding matrix to be supplied. Similarly to other tools, the interface is usually limited to VST or JACK. The ideal decoder would support a range of interfaces, and allow the user to specify their speaker setup using an easy interface. Features would include support for higher orders, near-field compensation, distance compensation, shelf-filtering, a choice of decoding ‘flavours’, and handling of irregular layouts.

It is often said in the BBC that ‘content is king’, so tools should be designed to work around content producers, rather than those working on the technical aspects. One of the benefits of Ambisonics is that there are many interesting things that can be done to manipulate the sound field. Exploring what creative things can be done with the sound field, and making more tools for manipulation available, may encourage some content producers to try it out.

7.3. Production methods

The ability to place sounds anywhere in space brings with it questions surrounding how to make best use of such freedom. Early examples of quadrophonic audio and stereoscopic film – to take two examples – show that producers enjoy creating ‘gimmicky’ content in the beginning, such as waving a stick in the viewer’s face, or placing instruments behind the listener. As the technology matures, the novelty tends to wear off, and producers gain a better understanding of how to use the technology effectively. As periphonic audio is a new concept to most, techniques in using the technology effectively are likely to be in the early stages.

Another thing to think about is periphonic audio’s compatibility with stereo and 5-channel surround. Current surround technology uses a technique of ‘folding’ the audio from the

side/rear speakers into the front. With this in mind, methods of folding audio with a strong vertical component into a horizontal-only setup need to be explored.

7.4. Microphones

Sound field microphones have been in existence for a long time, with new ones still being developed. Higher-order microphones, such as the MH acoustics ‘Eigenmike’, are in existence, but are expensive and not readily available. Although multiple-capsule microphones are still in active development, it would be desirable to have a low-cost, readily available higher-order microphone available for the wider community to use.

8. CONCLUSIONS

An audio format is nothing without content, and when competing with other formats, it is often the one with the most content that is adopted. If widespread use of Ambisonics is to become a reality, it is important to look at the technology from the content producer’s point-of-view. Although many tools are available, there is still a need for an easy-to-use package that covers everything from encoding to decoding. The code to realise this already exists in various free software tools, so it only needs to be brought together and packaged correctly.

In creating Ambisonic content, there are many questions surrounding creative use of height, and compatibility with stereo/surround. Techniques for mixing ‘with-height’ audio are still to be developed fully, but should come naturally when more content producers have the opportunity to use the technology. Some issues to be looked at include how to mix with consideration for how it would sound on various common speaker arrangements, and what to do with vertical sound content when played over horizontally-placed speakers.

Initial investigations into the effect of height showed that its inclusion did not always improve the listening experience. Some situations benefited from height more than others, particularly non-directional atmospheric audio, but generally opinion was divided over whether the inclusion of height brought a benefit to the listening experience.

Similarly, when comparing first-order Ambisonics to stereo and 5.0, non-directional content performed well when using Ambisonics, while more directional content worked better with 5.0.

9. FURTHER WORK

This investigation has only looked at first-order Ambisonics. To fully consider the potential performance of the technology, higher-order content should be looked at in greater detail. This should include judging the sound of recordings made with a higher-order microphone, and of mono sources panned into a higher-order format. It would also be beneficial to compare HOA to 5- and 7-channel surround, to see how it compares when using the same bandwidth.

10. ACKNOWLEDGMENTS

The authors would like to extend their thanks to Simon Tuff, Rupert Brun, Steve Brooke, Nadia Molinari, SIS LIVE, Richard Furse, Simon Goodwin, Bruce Wiggins and Peter Lennox for their help and input. Special thanks go to Andrew Mason and David Marston for their invaluable knowledge and assistance.

Speaker	W	X	Y	Z
1	0.117851	0.157135	0.000000	-0.235702
2	0.117851	-0.078567	0.136083	-0.235702
3	0.117851	-0.078567	-0.136083	-0.235702
4	0.117851	0.192450	0.111111	0.000000
5	0.117851	-0.000000	0.222222	0.000000
6	0.117851	-0.192450	0.111111	-0.000000
7	0.117851	-0.192450	-0.111111	-0.000000
8	0.117851	-0.000000	-0.222222	-0.000000
9	0.117851	0.192450	-0.111111	0.000000
10	0.117851	0.078567	0.136083	0.235702
11	0.117851	-0.157135	0.000000	0.235702
12	0.117851	0.078567	-0.136083	0.235702

Table 5: Decoding matrix for ‘HexTri’

Speaker	W	X	Y	Z
4	0.235702	0.288675	0.166667	0.000000
5	0.235702	0.000000	0.333333	0.000000
6	0.235702	-0.288675	0.166667	0.000000
7	0.235702	-0.288675	-0.166667	0.000000
8	0.235702	-0.000000	-0.333333	0.000000
9	0.235702	0.288675	-0.166667	0.000000

Table 6: Decoding matrix for ‘Hex’

Speaker	W	X	Y	Z
4	0.235702	0.230940	0.133333	-0.235702
5	0.235702	0.000000	0.266667	-0.235702
6	0.235702	-0.230940	0.133333	-0.235702
7	0.235702	-0.230940	-0.133333	-0.235702
8	0.235702	0.000000	-0.266667	-0.235702
9	0.235702	0.230940	-0.133333	-0.235702
10	0.000000	0.094281	0.163299	0.471405
11	0.000000	-0.188562	-0.000000	0.471405
12	0.000000	0.094281	-0.163299	0.471405

Table 7: Decoding matrix for ‘HexUp’

Speaker	W	X	Y	Z
1	0.000000	0.188562	-0.000000	-0.471405
2	0.000000	-0.094281	0.163299	-0.471405
3	0.000000	-0.094281	-0.163299	-0.471405
4	0.235702	0.230940	0.133333	0.235702
5	0.235702	0.000000	0.266667	0.235702
6	0.235702	-0.230940	0.133333	0.235702
7	0.235702	-0.230940	-0.133333	0.235702
8	0.235702	0.000000	-0.266667	0.235702
9	0.235702	0.230940	-0.133333	0.235702

Table 8: Decoding matrix for ‘HexDown’

Speaker	W	X	Y	Z
1	0.235702	0.471405	0.000000	-0.235702
2	0.235702	-0.235702	0.408248	-0.235702
3	0.235702	-0.235702	-0.408248	-0.235702
10	0.235702	0.235702	0.408248	0.235702
11	0.235702	-0.471405	-0.000000	0.235702
12	0.235702	0.235702	-0.408248	0.235702

Table 9: Decoding matrix for ‘Tri’

11. REFERENCES

- [1] BBC Website: Public Purposes, http://www.bbc.co.uk/aboutthebbc/purpose/public_purposes/
- [2] Wikipedia: FM Broadcasting: Stereo, http://en.wikipedia.org/wiki/FM_broadcasting#FM_stereo
- [3] BBC Website: BBC HD: What is HD?, http://www.bbc.co.uk/bbchd/what_is_hd.shtml
- [4] K. Hamasaki, S. Komiya, K. Hiyama and H. Okubo, “5.1 and 22.2 Multichannel Sound Productions Using an Integrated Surround Sound Panning System”, NAB Broadcast Engineering Conference, April 18-21, 2005, Las Vegas.
- [5] ITU Radiocommunication Assembly, “Recommendation ITU-R BS.1534-1: Method for the subjective assessment of intermediate quality levels of coding systems”, January, 2003.
- [6] Linux Audio projects at Kokkini Zita, <http://www.kokkinizita.net/linuxaudio/>
- [7] ITU Radiocommunication Assembly, “ITU-R recommendation BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multi-channel sound systems”, October, 1997.
- [8] E. Miyasaka, “Methods of Quality Assessment of Multi-channel Sound Systems”, 12th International AES Conference: The Perception of Reproduced Sound, June 28–30, 1993, Copenhagen, Denmark.