A STUDY INTO THE BROADCASTING SURROUND PRODUCTION CHAIN

A. J. Horsburgh, Solent University, Southampton, UK
D. Turner, University of the West of Scotland, Paisley, UK
J. Size, University of the West of Scotland, Paisley, UK

1 ABSTRACT

In recent years, broadcasting companies have been adapting their delivery of audio content to utilize next generation technologies and enhance the viewer experience. This paper presents research which has been conducted through a partnership scheme between The University of the West of Scotland and the British Broadcasting Corporation Scotland on the quality of the listening experience by use of surround sound formats broadcast over HD channels.

A small scale study using survey based responses and listening tests examined the attitudes and experience of BBC Scotland production staff towards current and future audio-specific developments. The survey focused on audio workflow in production and post-production using foreseeable industry-wide formats. It considered the views of audio production staff on surround-sound compliance. The listening tests investigated the relative perceptual benefits and immersive quality of four established decoding schemes for surround sound.

2 AUDIENCE IMMERSION AND CURRENT GENERATION TECHNOLOGIES

The last of the six public purposes stated in the British Broadcasting Corporations' Royal Charter [1] expresses that the broadcaster should "… deliver to the public the benefit of emerging communications technologies and services …" through audience experience specific developments. The BBC must continue maintain excellence in research "… in broadcasting and other means for the electronic distribution of audio, visual and audiovisual material, and in related technologies …." and it must make an effective contribution to "… the preparation and maintenance of a highly-skilled media workforce across the audio-visual industry." Expanding audience experience through the development of immersive technologies is central to the future of broadcast technology. Examples of such improvements are; expansive visual qualities in 3D, Ultra High Definition, and the use of High Frame Rate. Developments of audio quality are enhancements such as '3D' sound, wider dynamic-range experiences, and the use of Higher Sample rates [2]. It is an aim of the presented research to further previous investigations in 'spatial' sound provided by seeking the critical listening experience of audio content producers in broadcasting.

2.1 CURRENT GENERATION SURROUND SOUND BROADCAST FORMATS

The delivery chain of broadcast audio is a complex route of a source capture, manipulation, real-time rendering, transmission, and delivery of content for consumption [Illustration 1]. Since the inception of monophonic radio in 1930s Britain – the BBC have been ensuring quality delivered across it's output. In the digital age however, the previously adopted 'one size fits all' model has to change. Many consumers are now accessing content on a range of devices: fixed location (TV / Radio), dynamic location (internet-based) with some content being consumed live or 'on demand'.

The variances of consumption now include audio systems, which are headphone, stereophonic internal loudspeaker systems, or external multichannel audio systems. Such variance in audio system compliance causes several legacy problems and expansion to a new format, either visual or audible, must include these.

Despite the complexity of potential consumption formats – there are relatively few methods of delivering high quality immersive audio in broadcast;

- Monophonic
- Stereophonic (amplitude panning, or binaural)
- Dolby E

Monophonic sound has limited spatial characteristics due to the single reproduction source, and stereophonic sources fares slightly better with the use of amplitude placement or techniques such as binaural. For multi-channel audio it has become common within broadcasting to use Dolby Laboratories ‘E’ encoding and decoding format.

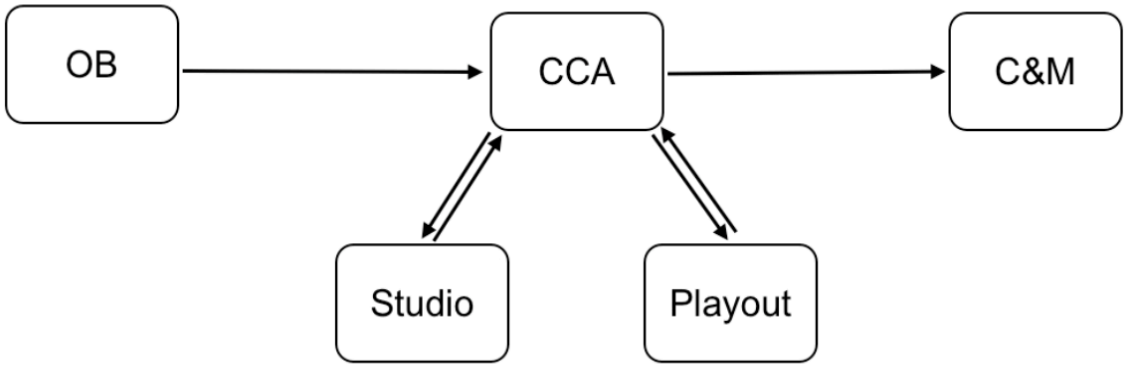

Illustration 1: Pomerai's [2] description of a HD OB signal path. Following Outside Broadcast (OB), through the Central Communications Area (CCA) to distribution via Coding & Multiplex (C&M).

The beneficial reasons behind the implementation of Dolby E being standardised were detailed extensively in a BBC review in 2008 [3], but are briefly covered in **2.1.1**.

2.1.1 DOLBY E

Dolby's audio compression codec within ‘E’ allows for the ease of distribution of multi-channel audio files, with a high degree of compatibility between companies and production houses, whilst keeping the required transmission bandwidth low. Key features of Dolby E are it's synchronous alignment with video frames, high quality en- and de-coding of multi-channel audio streams, inclusive metadata allows for crucial decoding and production information to retain with the digital file, and a known latency of en- and de-coding kept to one video sample. A production benefit is the transmission of a 5.1 soundtrack (6-channels, 16-bit) and a foreign language / specialised (2-track) version to be simultaneously distributed as a singular file. A version of Dolby E can accept up to 8 channels of broadbase PCM audio, and associated metadata, onto a 20-bit / 48kHz AES pair with an increased data rate of 1.92Mbps.

The synchronised principles that are the foundation of Dolby E provide support to audio supervisors and editors when integrated into the production environment.

2.1.2 Integration of Dolby E in the Production Environment

Integration of Dolby E into the audio broadcast production chain was adopted as the method of synchronisation between audio samples and video frames was defined clearly, with scope for future improvements. Pomerai's [3] description of audio production to be summarised as;

- Overlap of frames mean that simple audio edits include cross-fades
- Longer edits require decoding to PCM, editing, re-encoding to Dolby E
- Instability in consumer decoders when switching between 2.0 and 5.1 configurations
- Metadata attachment and separation during mixing

The application of metadata in loudness (*dialnorm*), dynamic range control (*DRC*), and mixing are also considerable factors to include throughout the production chain. Should metadata become corrupted or misaligned then significant errors may be present within the decoding of a 2-channel AES stream.

Transmission tools such as a Dolby E are however simply mechanisms of distributing audio in the wider broadcasting workflow. Audio monitoring of Dolby E productions are rigorously controlled via Dolby Incorporated's specification documentation. The specification features direction over acoustic environments and audio reproduction system benchmarks that as a result place considerable acoustical and equipment-related demands upon the content provider. The provision for a new production workflow should recognise such specific constraints. Compliance with other production houses, and internal BBC departments, restricts the type of software and formats of transmission and internally, broadcasters may differ significantly in a workflow but must ingest and transmit industry compliant media.

The qualitative listener experience is not ensured by the use of such en- and de-coding [2,3]. For Low Frequency Effects (.1, subwoofer) channel, it has been noted [4] that Dolby's E production recommendation is not compatible with LFE channel signals, due to both a gain difference and a filtering mismatch. Gain and filtering responses result in the engineer or supervisor being unsure of the end deliverable content to the audience at home [4]. Such unsure deliverables often yields in the LFE channel being largely unused or under represented in broadcast productions.

The compatibility of audio material for exchange between broadcasting houses features the associated cost of Dolby E in the form of licensing and hardware encoding systems for entire production chains adherence with Dolby's regulations. Any future formats (such as Dolby's ATMOS cinema sound) would require a new upgrade path for equipment expenditure; potentially affecting the viewer financially for the next version of immersion.

Improvements via choice of decoder for the listening experience are discussed in **2.1.3**.

2.1.3 Development of Future Technologies

The findings of research projects such as the FascinatE [5], Audio-Object [2] and other future audio formats [6] are continually being published. Challenges posed by the use of adaptive television, dual-presentations of visuals and adaptive audio content through 'object-based' formats, requires new systems to be designed and robustly investigated. The impact on production staff and their abilities has not been documented, but the findings of FascinatE demonstrated that it is possible to include audio-objects when combined with Higher Order Ambisonic signal encoding [5].

Development from one proprietary format (Dolby E) to another (Audio-Objects) may be interpreted as a side-step by the producers; and the difference in quality has yet to be clearly established promoting the presented research. The underlying problems with multiple encoding and decoding of files found within the traditional surround audio workflow could mitigate against this approach in the long term.

As the results of the presented research indicate - producers and consumers are currently satisfied with the quality of materials provided by broadcasters; but the changes to the production workflows can be informed by audio supervisors, requesting specific decoding methods for each programme.

2.2 AUDIO BROADCAST CHAIN WITHIN PACIFIC QUAY

Audio quality throughout the production chain, Illustration 1, has improved through continual development and the upgrading of facilities within the broadcasting centers culminating in the provision of full digital end-to-end High Definition systems. The upgrading of technology has led to the use of multi-channel delivery capabilities becoming a standard within the new Pacific Quay HD transmission chain. Increased bandwidth has now made it possible for 8 channels of audio to be transmitted with video, prompting a re-investigation to the use of matrixing audio signals [2, 5, 6].

The investigation has chosen specific genres of BBC Scotland productions to determine if a high quality of audience immersion can be delivered across the News, Documentaries, Sport and Light Entertainment productions using existing formats.

2.3 IMPROVEMENTS TO BROADCAST SIGNAL CHAIN

Current up- and down-mixing to proprietary formats is both complex and prone to error [3]. With multiple processing required between stages of production - the spatialisation of sound scenes are poorly matched to future developments in channel bandwidth and speaker requirements. Projects such as VAMBU [9] and FascinatE [5] have shown that other possibilities to deliverable content are available, but their implementations have been custom workflow setups and not had to traditionally interface with content pipelines.

The purpose of the investigation was to discover the cause of developmental mismatch between production and research. Within the literature [2, 3, 4, 6] there was a lack of data to show that the chosen decoding scheme was the best decision for the audience immersion. A small scale study was therefore commissioned using a survey and series of listening tests. Each of the data sets examined the attitudes and experience production staff towards current and future audio-specific developments.

3 EXPERIMENTAL CONSIDERATIONS

3.1 Survey

A building-wide survey of BBC Scotland Pacific Quay (PQ) was implemented to gather the largest possible response from a dynamic working environment. The use of an online package (Survey Monkey) allowed staff members to complete the questions quickly and submit anonymous responses. The survey results were then analysed and their conclusions referenced against the outcome of the listening tests.

3.2 Listening Tests

The listening tests were held in a small meeting room within the PQ building. The listening test procedure was a hybrid of the ITU-R B.S 1534 [7] (" Method for the subjective assessment of intermediate quality level of coding systems") and ITU BS.1116 [8] ("Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems"). The listening test was designed to be accessible to non-audio specialists. The test used samples of BBC sourced audio extracted from different programme genres, each modified by different decoding schemes.

Participants were provided with a 5-point scale (Excellent, Good, Fair, Poor and Bad) and asked to judge the 30 second clips on their initial impressions of space and immersion. Each clip was referenced to the 'mono' version of the stimuli. Writing space was provided on a response sheet so participants could include any notes or comments for each clip.

3.2.1 Materials

The extracts from existing broadcast programmes used as test materials were chosen for their categorised type;

- Documentary (John Lennon)
- Drama (River City)
- Music / Entertainment (T In The Park)
- Sport (BBC Scotland Sport)

The listening test was then presented to a small group of individuals with the following audio decoding schemes for each of the four categories;

- Ambisonic (2nd Order)
- Dolby 5.0 (No LFE channel)
- Mono
- Stereo

The reference format was monophonic. The sample size of the listening tests was 9. The small sample size number made it difficult to show a high value of statistical significance in the evidence of preference between the decoding types. We discuss this in the Results section of the report.

3.2.2 Equipment

The equipment used was provided by the UWS Spatial Audio Research Group. The system included 6 Genelec 8020B loudspeakers, Neutrik cabling, Mark of the Unicorn (MOTU) digital interfaces and Reaper Digital Audio Workstation (DAW) software. The loudspeaker array was calibrated at 78dBc to comply with EBU (technical document 3343) guidelines on multichannel listening environments. Placement of the loudspeakers conformed to a hexagon arrangement to include stereo, 5.0 and Ambisonic listening ideals.

3.2.3 Methodology

Critical perception and the evaluation of sound systems, especially in the measurement of multi-channel audio reproduction systems, can widely be covered by ITU-R BS1534-1 [6,7,8,9].

The performance of an audio system can be measured against specific metrics chosen by the investigators; in the work presented, the general source separation and spatial impression were two such areas. As the tests were supplementary to those already conducted by the BBC [6], it was decided that trained listeners, specifically, sound supervisors or audio-production staff, would be the main subject group. The materials chosen were also to be BBC-quality broadcast audio across five genres with which the production staff should be familiar in terms of the timbre and tonality of sources. The only difference between the stimuli clips was the decoding scheme used. No additional modifications to equalisation, dynamic processing, or tonal characteristics were made during the data preparation stage. Notes were made of any verbal comments by subjects on the positioning, size, depth, spatial-qualities or any other general perceptual aspect of the clips.

4 RESULTS

4.1 Survey

The survey was designed to be distributed across Pacific Quay's production chain therefore a series of questions were designed which related to broadcasting in a general manner, with increasing detail on audio and vision developments developed throughout the sections allowing respondents to comment according to their specialty. The respondents identified themselves via job function as members of four broad groups;

Production	(General)
Management	(Broadcast management, Management)
Audio	(Sound Supervisor, Radio)
Other	(Broadcast Resources, Programme Making, Acquisition)

All the respondents had been with the BBC for at least 5 years, and more than half had in excess of 20 years experience in broadcasting. All respondents also regarded their ability to critique audio quality as modest to excellent.

Despite a wide range of skills and areas of knowledge, the results demonstrated an opinion that screen resolution was a more significant factor in determining programme quality than audio transmission type. However, sound quality was seen as much more important than the provision of 3D images. Factors influencing this opinion range from cost, development time, implementation concerns and concerns about public perception.

With the recent production chain upgrade to now featuring end-to-end HD capabilities, most staff regarded the *aim* of implementing an end-to-end file based HD broadcasting facility to be 'highly important' or 'essential' to the company. However, they did not see this as a priority in consumers' eyes, reflecting the fact that only 20% of the respondents were equipped for 5.1 audio in their own homes. Respondents noted that limited awareness of HD audio would be present amongst general audiences. Progress in achieving this aim for advancements into 'HD audio' was seen as slow to moderate, with cost, lack of viewer demand and technical factors cited as reasons.

Only three respondents considered that the success and impact of HD visuals would accelerate the development of 'HD Audio' technologies, with the remainder of respondents being uncommitted. However 47% agreed that the BBC had an obligation under its charter to enhance the audio experience of its viewers.

The perceived cost/benefit trade off lead to a consensus that HD audio should only be considered for a subset of broadcast output based either on specific programme types or on the merits of individual programmes, Seven programme types were ranked according to anticipated benefit as: film, sport, drama, light entertainment, factual and documentary with little or no benefit expected for children's programming and news.

Whilst the majority of staff were moderately satisfied with the current production workflows, some issues were identified which indicated scope for improvement in general workflows and the majority of staff saw a need for new audio workflows and said these were important to the development of HD audio. It was generally recognized that standard audio and HD audio could not be covered by a single workflow for all programming.

Although members of each category identified different factors as a barrier to the implementation of HD audio as spatial sound, there was clear agreement on the benefits to programming of HD audio with the production and audio groups most enthusiastic and the management group expressing some reservations. Time, cost and compatibility factors were highlighted as potentially problematic, but there was some disagreement about the significance of the extra storage which HD audio would

require. This was a concern for management, but production technologists pointed out that the storage of audio is less demanding than the storage of pictures.

The survey identified training requirements since 70% of staff felt unprepared to work with HD audio.

The details of the survey have provided BBC Scotland with a useful insight into staff attitudes and opinions on HD audio. Despite a range of concern it is clear that there is a belief that the future of broadcasting will feature multi-channel audio in some form.

4.2 Listening Tests

The listening test implemented the 5-point rating system (Bad, Poor, Fair, Good, and Excellent) on a response system of general impression of spatial quality and immersion. The data presented follows two formats; direct values of ratings and the mean deviation from responses.

Format	Average of rating	Standard Deviation of rating	Average of deviation
Ambisonic	3.639	0.798	0.201
Dolby	3.639	0.762	0.201
Mono	3.056	0.893	-0.382
Stereo	3.417	1.079	-0.021

Table 1: Average test rating for audio formats across all categories

Results indicate that the Ambisonic and Dolby encoded surround files scored the highest preference of immersion. This is to be expected the presented files were providing multichannel information. Deviations of results show that there is a similar trend across format 1 and 2 (Ambisonic and Dolby, respectively). The largest preferential deviation was applied to monophonic material, and the lowest representing stereophonic material. By examining the standard deviation alone – it would be assumed that stereophonic material was consistently rated for all types of material.

Format	Material	Average of rating	Standard Deviation of rating	Average Deviation
Ambisonic	doc	3.7	0.71	0.2
	drama	3.9	0.78	0.3
	music	4.0	0.71	-0.5
	sport	3.0	0.71	0.7
Dolby	doc	3.8	0.83	0.5
	drama	3.7	0.87	0.2
	music	3.6	0.88	0.0
	sport	3.6	0.53	-0.8
Mono	doc	2.9	0.60	0.6
	drama	3.4	1.01	0.1
	music	2.8	1.09	-0.7
	sport	3.1	0.78	-0.7
Stereo	doc	4.1	0.78	-0.4
	drama	2.7	1.12	0.1
	music	2.8	0.83	-0.3
	sport	4.1	0.60	0.7

Table 2: Variation average figures for format and category

Expansion of rating across each of the four formats and category is provided in table 2. Variation between the average rating and material type indicates that preference for immersion may be linked to the type of content. An example is stereophonic material – scoring high in Documentary and Sport but a large deviation in drama and music. The ambisonic productions scored the highest mean, and lowest standard deviation, across all genres however this is not an indicator of a wider trend.

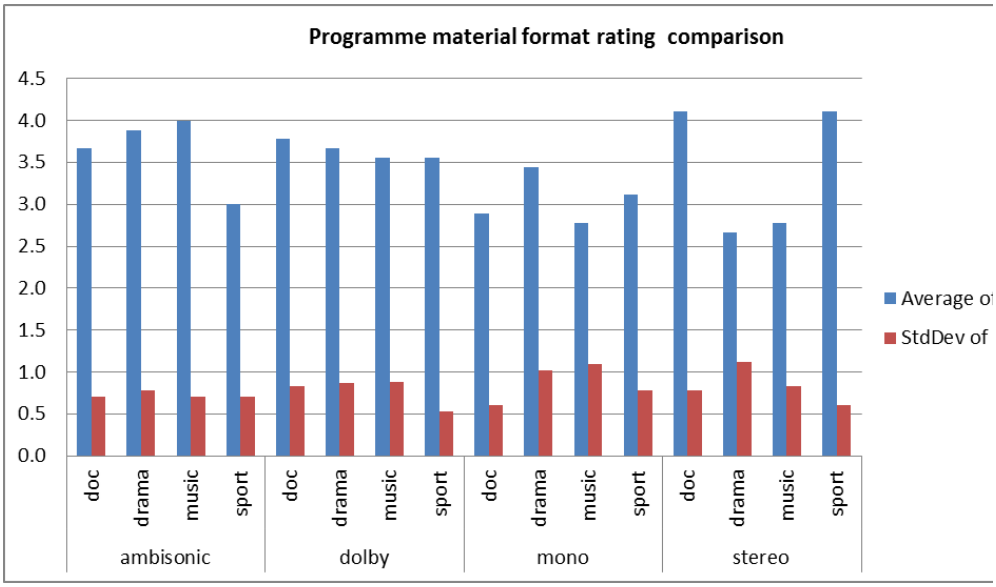

Illustration 2: Audio format test ratings for programme genres grouped by format

Visual representation of the results in table 2 is shown in Illustration 2.

4.2.1 General Result Analysis

These results point towards the suitability of surround formats for a range of programme materials including music and drama, but the standard stereo format is still preferred in some situations. It is also noted that stereo is rated most highly for sport which is a programming area which was identified in the staff survey as a genre likely to benefit from surround sound audio (Figure 8a). Surround sound formats were preferred for drama, which is consistent with the survey (Figure 8a).

The results demonstrate the value of the listening tests and also suggest the requirement for further research with a larger sample size and using more audio clips of each genre to make the results for the genres more representative. A number of questions are raised as to what motivates respondents' rating choices. It would also be useful to investigate whether the presence or absence of a sub-woofer has more influence on test rating than the format and also whether the preferences for formats remain the same in the presence of a subwoofer. Although there may be a low statistical significance in the results, they do represent the opinions of staff directly concerned with the issues under consideration.

Since the listening tests were restricted to audio clips it would be interesting in further research to investigate whether the audience respond to sound in a different way when it is coupled with a visual cue.

A more sophisticated selection of clips might be required to determine whether preferences were being made on the basis of localisation of: the main actor or performer, environmental ambience, sound effects or audience and crowd responses.

5 CONCLUSIONS

The essential findings of the survey can be summarised by the following key points:

- If staff attitudes are representative of those of general consumers, there is no established appetite for HD audio because they are not generally equipped with the technology to receive it in their homes. However, staff expect the consumer awareness of HD audio to increase as technology improves and more service is available.
- Staff considers HD audio development to be important to the BBC.
- Cost benefit considerations determine that HD audio should be applied to a subgroup of programming determined by genre or on individual merit.
- Films, Sport and Drama were regarded in the staff survey as the most likely genres to benefit from HD audio.
- New workflows are required to service HD audio production. It was not considered likely that HD and standard audio could be covered by one workflow process.
- Staff expressed specific concerns over details of the implementation of HD audio in terms of cost, time, storage and process.
- An urgent requirement for training was established if HD audio is to be implemented.
- Staff felt that HD audio had a positive future and staff awareness of formats was highest for 5.1 surround.
- Further research, informed by this study, would be beneficial to develop deeper understanding of audience response to different sound formats.
- Research on consumer electronics sales would be useful to provide trends on home AV equipment.

From the above we conclude that the future of audio productions within BBC Scotland is limited by cost-benefit considerations and a perceived low demand from consumers. The survey responses showed an open minded approach to the adoption of HD audio.

Listening tests have shown that the relative preferences expressed for different decoding types vary with the genre of programming material. For Drama and Music, Ambisonic and Dolby 5.0 surround sound were found to have improved the immersive experience. Drama was the only type of material that did not show any indication that stereophonic interleaving of channels immersed the audience.

The response to listening tests on different programming materials suggests that the decision on which decoding process to apply to a programme could be made on the basis of genre.
The staff survey supported this approach but with some further consideration to be made on a per programme basis.

Although viewers may benefit from the additional spatial-elements that surround sound can offer through the fuller use of available transmission channels, production staff remain to be convinced conclusively that the benefits of such techniques will outweigh the outlay in terms of time and money costs and also the technological challenges encountered in implementation.

There is a tension between the desire to keep pace with the continual development of technology and the requirement to produce programmes which meet audience expectations within a limited budget. This can only be resolved by the establishment of an efficient workflow and prioritised implementation on a segment of programming. Although the further research of immersive technologies and techniques is considered to be generally beneficial, producers are concerned with the impact on consumers and the potential for strain on budgets (both financial and time-based). The research conducted is viewed as being at an initial stage with no wider contextual thought to the workflows. The practical application is a real concern to staff with particular challenges in relation to the maintenance of quality and deadline achievement.

6 RECOMMENDATIONS

On the basis of the findings in this report, we make the following recommendations:

- Design and evaluate workflows for HD audio.
- Establish general staff training requirements in HD audio production.
- Consider the storage implication of archiving materials in formats which can be processed subsequently in current and future surround formats.
- Gather data on technical obstacles to HD audio implementation and suggest costed solutions.
- Perform a cost/benefit analysis of a small pilot HD audio-visual production.
- Confine the implementation of surround sound to programmes where the listening tests confirmed that the horizontal localisation and immersive aspects of surround may provide listeners with an enhanced experience.
- Further research in the area of listening tests:
    - Involvement of commissioning editors (which was a major omission from the initial sample).
    - Perform listening tests on genres informed by the report survey.
    - Refine the listening tests informed by the initial results reported and incorporating audio-visual content specifically prepared for test purposes over a larger sample set.

The study presented has successfully identified several key areas for development in the implementation and adoption of surround audio technologies within broadcasting. A further programme of research flowing from these recommendations has now commenced which it is hoped will inform the implementation of HD audio within PQ.

7 REFERENCES

1 BROADCASTING An Agreement Between Her Majesty's Secretary of State for Culture, Media and Sport and the British Broadcasting Corporation July 2006, ISBN 0-10-168722

2 Graham Thomas, O. Schreer, B. Shirley, J. Spille, 'Combining Panoramic Image and 3D Audio Capture with Conventional Coverage for Immersive and Interactive Content Production', BBC White Paper WHP 207, 2011

3 R. de Pomerai, 'Managing a Real World Dolby E Broadcast Workflow', BBC White Paper WHP175, 2009.

4 A. Mason, 'Use of the low frequency effects (LFE) channel in broadcasting', BBC White Paper WHP203, 2011.

5 http://www.fascinate-project.eu/, Accessed 07/10/13

6 C. Baume, T. Churnside, "White Paper WHP 221 (Upping the Auntie: A Broadcaster's Take on Ambisonics)", March 2012

7 International Telecommunication Union Recommendation – Broadcasting Standard 1534-1, 'Method for the subjective assessment of intermediate quality level of coding systems,', 2003

8 International Telecommunication Union Recommendation – Broadcasting Standard 1116-1, 'Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems', 1997

9 M. J. Morrell, C. Baume, J. D. Reiss , 'VAMBU sound: a mixed technique 4-d reproduction system with a heightened frontal localisation area', Audio Engineering Society's 25th Conference, 2012