Multichannel Audio for television

John Emmett
Broadcast Project Research

Television sound can no longer be considered as a single entity. We will soon have viewers (listeners?) demanding "5.1" cinema-quality sound from every programme, whilst others – possibly the elderly or hard-of-hearing – wanting just the programme dialogue to be clearly reproduced from a tiny portable television.

In this article, the author paints a picture of what he personally believes can be done with TV sound to cater for different user expectations – without making any fundamental changes to existing digital TV receivers, nor adding any significant costs at the production level.

One long-existing problem of any television sound system (my old boss said that he received the first complaints in 1960!) is that of obtaining an even, apparent, programme loudness when the viewer zaps between channels and between programme genres.

There is every hope however that, out of this apparent chaos, a genuine set of global recommendations can be established for the forthcoming multichannel television sound system. One part of this process has just commenced within EBU Project Group P/AGA – by "dusting off" the ten-year-old HD television sound recommendations created by the old G1 group, and commencing the process of aligning them wherever possible with current ATSC Television and DVD practice. We are helped in this not just by the nature of digital television – which enables the delivery of a range of suitable audio standards direct to the home – but also by economic advances in the overall sound production. Primary among these advances has been the incredibly fast adoption of the EBU-developed BWF computer-file format [1] within the global audio industry and community: one reason for this was that the BWF format built on existing hardware and software. We will see later – in the case of the DVD player – that this type of serendipity (or wise planning, depending on your point of view) is not unique and there exists a rich mine of consumer electronics now ready for us – as broadcasters – to dig into.

The eternal struggle for bit-rate
(Theatre is Art, Film is Life, Television is Furniture?)

An immersive experience

All new media seem to start as *immersive* experiences – we tend to stop all other activity around the house in order to watch or listen to the programmes. After a while (many years in the case of radio), some media may fade from interest in the typical home, or tend to become background entertainment. In some cases, it is the type of programmes that govern the form of interaction in the home. For example, feature films on television have over the years remained an immersive experience for all of the family, whilst daytime news has largely become background viewing.

Research into the psychological nature of the immersive experience is underway in many countries, but all the participants agree about the importance of the accompanying sound in creating an immersive atmosphere. You need only to watch children being totally immersed in the crashes and bangs that a DVD 5.1 soundtrack brings to our homes, in order to appreciate this.

There has been a viewpoint in the past that only HD pictures can justify the expense of 5.1 sound productions but, as we shall see, the real cost only comes in affording the frugal bit-rates needed for emission.

The general view that *surround sound* is unsuitable for television may simply have come about because of confusion over the term "surround sound". Going back to the 1970s, attempts to matrix two-channel audio – in order to fill a room with so-called "Surround Sound" – were hampered by the compromise that had been made in music recording standards some 30 years previously, which consigned any dialogue present in the recording to a virtual centre position between two "stereo" sources. This compromise made analogue 2-channel "stereo" recordings economically possible – where music was the main economic driver – but the value of a "real" centre dialogue channel – where the sound accompanied the pictures – had already been appreciated in cinema research as far back as the late 1930s.

The availability of such a centre dialogue channel in television today, brings additional advantages in both language choice and dialogue intelligibility [2].

The equipment in our homes
(The future is already with us?)

It may come as some surprise to find that many of the audio items that we have in our homes are already equipped for multichannel use. The reason behind this is that the DVD-player market has followed a completely different business model to conventional broadcasting (such as that followed for analogue radio, or even DAB). The conventional market approach leaves the reduction in receiver costs to commercial forces so, for instance, it took fifty years for the purchase price of radios to reduce to an almost trivial level. It has taken maybe three years for DVD technology to reach a similar level – and without the gift of any financial return from the media sales. The development process has created generic approaches to digital audio processing, and the interfaces used during this process are those that we use in our professional lives, out of simple pragmatism. After the largely software-based development process is over, these interfaces remain – ready to be used.

For instance, the consumer form of the AES/EBU audio interface [3] has been "borrowed" in the form of IEC 61937 [http://www.iec.ch/] in order to carry packets of multichannel audio data. The accompanying picture of the rear panel of a typical integrated "Hi-Fi" unit shows a bewildering range of connections, although the markings are quite correct in indicating that the unit will have the intelligence to auto-decode the input signal, be it in linear PCM, MPEG layer II, DTS or Dolby Digital (AC-3) form.
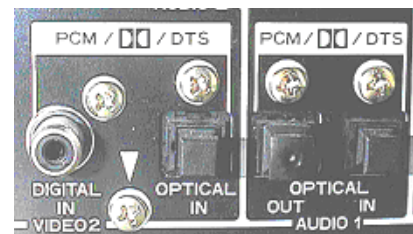
This range of *internal* capability – to which one recent addition has been that of MPEG "AAC" coding [1] – has been largely achieved without any equipment changes, as manufacturers have been introducing the software decoder "core" into the ever larger (and cheaper) processor chips used.

The multichannel audio-production process
(Just more of the same?)

The overall method of sound production for any television programme (or, for that matter, any sound programme accompanying pictures, whether for computer games or feature films), follows the same two stages illustrated in *Fig. 1*.

---

1. AAC (Advanced Audio Coding) is a non-backward compatible MPEG audio coding development that currently offers possibly the best performance at modest bit-rates for a 5.1 multichannel television service.

For most television programmes, the sound recorded at the production stage is used directly, whilst post-dubbed ADR (Automatic Dialogue Replacement) is used as the sole source of the sound, only in a few genres of film musical.

Part 2 of the mixdown process *(see Fig. 2)* exposes some of the economic problems facing television sound production in that, up to this stage, the costs of production are largely fixed whatever the output format is going to be. However, Part 2 shows many possible primary distribution mixes which will require operator attention, although there may be a few secondary mixes which may be obtained automatically – for instance at the emission site in order to feed legacy monophonic services.

One solution to this problem is to allow the viewer to produce his/her own preferred mix. This is done in a DVD programme by simply providing metadata and some parallel audio tracks from which the viewer can choose. To a large extent, this process will not be possible – nor desirable – in television production and it will be instructive to find out later just why this is.
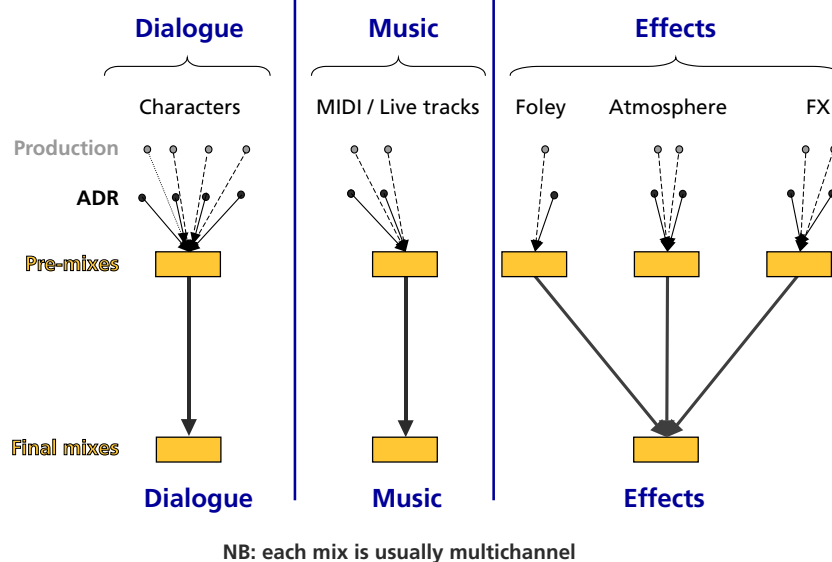
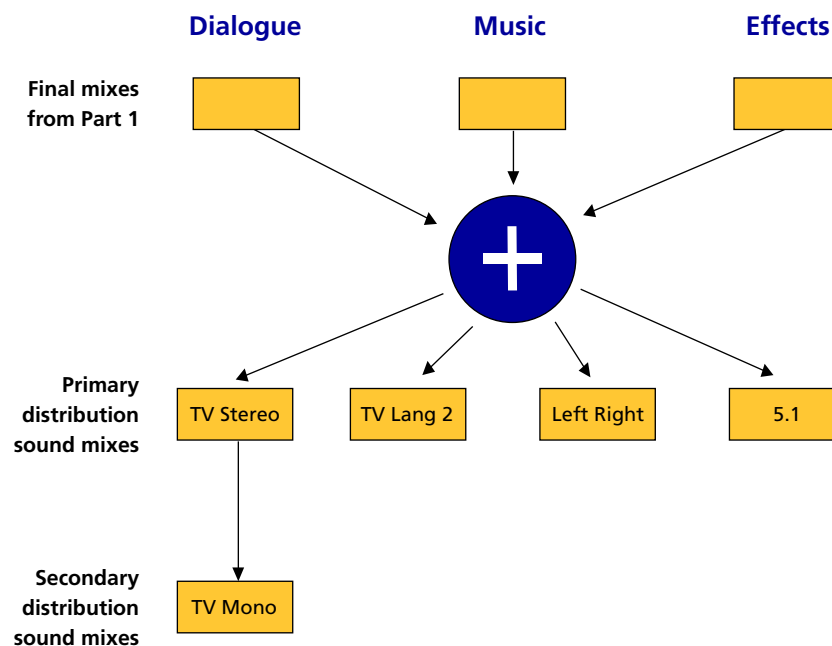## Giving viewers the choice

(Less may be more?)

In the cinema, the film director rules the playout process. He or she will have decided what you hear and how loud you hear it. There is no metadata in cinema practice, other than an agreement on the vital playback level setting, and once the playback system has been aligned to the agreed standard, the process is therefore complete [4].

NB: each mix is usually multichannel

Figure 1
A universal sound-mixdown route – part 1

Figure 2
A universal sound-mixdown route – part 2

DVD playback in the home follows this cinema process, but with the addition of recorded *metadata*. This metadata consists of extra information that may be engaged by the viewer for automatic corrections to the playback sound mix. Firstly, the *mean dialogue level* is usually recorded, so that the dialogue playback level may be normalized with respect to other audio sources in the home. Secondly, there may be recommended *mixdown* parameters included, in order to enable optimal 2-channel stereo or mono playback. Thirdly, *compression* parameters may be included in order to obtain a good mono playback with reduced dynamic range. This reduced dynamic range may be used for "intimate" low-level audio listening (it is sometimes described in DVD player literature as "midnight" mode), or it may be used in order to modulate an RF carrier for old-style television signal distribution.

So where does this leave us, in the case of multichannel television sound? Clearly, live television cannot produce the necessary metadata in advance in order to give programme directors total freedom in the sound mix and, anyway, would the viewers not prefer to be able to switch between channels and programme types without big differences in dialogue loudness?

The answer to this dilemma is simply to agree (i.e. create recommendations for) a single set of television metadata parameters in advance of any problems. Draft mixdown parameters are already present in [5][6], anyway, and loudness-metering standardization is underway in ITU-R WP 6P.

Of course this "look ahead" approach requires an educational and discussion period. Will all this restrict the artistic freedom of the programme makers? Are analogue "rule-based" programme-level controls (which we all accept for engineering reasons), any more restricting than a sensory-based loudness meter? These and many other questions will need patient study by the project group members over the coming months but I, for one, welcome the opportunities. After all, the introduction of NICAM stereo produced the extraordinary and unexpected result of actually *selling* television sets. Who knows what multichannel sound could do for digital television?

How do *we* get there?

(or… what if we just ignore it?)

Our viewers may well be "primed" ready to accept enhanced audio services, but what is the cost to us, as the broadcasters?

The good news is that, at the production level, the source material is all there for using, and thus 5.1 production will involve very little extra cost, maybe only in extending the mixing and monitoring facilities in many cases. Certainly, I have found when producing experimental 5.1 and 7.1 mixes that the "discrete track" nature of these productions actually seems to reduce the mixdown effort, compared to a two-channel stereo production. This is almost certainly because of the reduction of panning effort for the "virtual" dialogue images.

Perceived difficulties with 5.1 monitoring arrangements can almost certainly be dealt with in the same way that we succeeded in adding two-channel stereo monitoring to our largely "acoustically-unsuitable" television sound galleries. The major problem with the 5.1 loudspeaker arrangement in these galleries seems to lie in positioning the centre dialogue loudspeaker where the picture monitors and the mixing desk can least interfere with the sound, and vice-versa. The rear surround-channel reproduction may appear to be compromised by a lack of physical depth behind the operators. However, with small screen productions there should be far less surround activity than in a cinema type of production, and solutions such as using flat-panel loudspeakers of the NXT ™ type can work amazingly well in restricted rear-surround situations, as they often possess an exceptionally wide high-frequency distribution angle.

No, the major problems in the introduction of any multichannel sound for television may come in distributing and post-producing the extra audio tracks, particularly on "legacy" videotape formats [7][8]. On the horizon are the MXF- and proMPEG-based systems that are capable of carrying a flexible arrangement of BWF or AES/EBU audio data, but a legacy of fast-changing tape formats has left Europe with a huge infrastructure of DigiBeta-based systems with some years of economic life to run. In the US and Australian scenarios, the

Abbreviations

AAC	(MPEG-2/4) Advanced Audio Coding	**IEC**	International Electrotechnical Commission
ADR	Automatic Dialogue Replacement	**ITU**	International Telecommunication Union
AES	Audio Engineering Society	**ITU-R**	ITU - Radiocommunication Sector
ATSC	Advanced Television Systems Committee (USA)	**MPEG**	Moving Picture Experts Group
		MXF	Material eXchange Format
BWF	(EBU) Broadcast Wave Format	**NICAM**	Near-Instantaneous Companding And Multiplexing
DTS	Digital Theatre System		
HD	High-Definition	**PCM**	Pulse Code Modulation

Dr John Emmett joined the Engineering Department of Thames Television in London after gaining a Ph.D. at Durham University, and after starting the UK audio equipment manufacturer, EMO Systems. Whilst at Thames Television, he worked on subjects as diverse as film archive formats and psychoacoustics and, along the way, gained six international patents as sole inventor. Jointly with Lee Lighting, he received a Technical Oscar for developing the flicker-free Lighting Ballast.

John Emmett continued as R&D Manager with Pearson Television, and is currently Technical Director and Chief Executive of Broadcast Project Research, a new independent studio-based research group. He represents UK Independent Television at the EBU as Project Leader for the new P/AGA (Advisory Group on Audio) project, following on from leading the P/AFT project that created the BWF audio file format.

change of TV production to HD, at the same time as introducing multichannel audio, may actually assist them in a smooth changeover, whilst we in Europe need to look at the whole question of multichannel audio distribution on our current infrastructures. Even the file-based formats are not yet multichannel-ready. The BWF format is economically based on using existing WAV file arrangements, but there are few current IT-based applications for multichannel sound and, worse still, those applications that do exist (mainly games) are not compatible in channel allocation.

Let us not end on such a depressing note, however. EBU project groups such as P/AGA and P/FTP are well into these purely production-based problems: rarely in the history of television has so much benefit been possible without any fundamental changes, either in distribution or to the home reception system.

Bibliography

[1] R. Chalmers: **The Broadcast Wave Format – an introduction**
EBU Technical Review No. 274, Winter 1997
http://www.ebu.ch/trev_274-chalmers.pdf

[2] ITU Recommendation BS.775-1 (07/94): **Multichannel stereophonic sound system with and without accompanying picture**
http://www.itu.int/rec/recommendation.asp?type=items&lang=e&parent=R-REC-BS.775-1-199407-I

[3] EBU Tech Doc 3250: **Specification of the digital audio interface (2nd edition)**
http://www.ebu.ch/tech_t3250.html

[4] I. Allen: **Are movies too loud?**
SMPTE Film Conference, 22 March 22 1997.

[5] EBU Tech Doc 3276: **Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic**
http://www.ebu.ch/tech_t3276.html

[6] EBU Tech Doc 3276, Supplement 1: **Listening conditions for the assessment of sound programme material: Supplement 1: Multichannel audio**
http://www.ebu.ch/tech_t3276_s1.html

[7] EBU Technical Recommendation R96-1999: **Formats for production and delivery of multichannel audio programmes**
http://www.ebu.ch/tech_text_r96-1999.pdf

[8] EBU BPN 021: **Multichannel Audio: Different reproduction and delivery formats available on the market**
http://www.ebu.ch/ptech_bpn_021.pdf *(available to EBU Members only)*