THE QUEEN'S AWARDS
FOR ENTERPRISE
2001

BBC

R&D White Paper

WHP 051

October 2002 (revised July 2004)

Audio Description:
what it is and how it works

N.E. Tanton, T. Ware *and* M. Armstrong

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP051

Audio Description: what it is and how it works

Nick Tanton, Trevor Ware & Mike Armstrong

Abstract

For understandable reasons many television programmes rely on visual content and composition to help to tell their story. The provision of additional description of the scene or action can therefore provide a considerable benefit to understanding, especially to the visually impaired.

Audio description (AD) is an ancillary component associated with a TV service which delivers a verbal description of the visual scene as an aid to understanding and enjoyment particularly (but not exclusively) for viewers who have visual impairments.

This White Paper describes the user requirements for audio description, how audio description can be implemented for digital television and how it works for digital terrestrial television (DTT). It also briefly describes two particular practical implementations.

Additional key words: Access services, accessibility, DTV,

Document revision history

v1.0	2002	original
v2.0	2004	added explicit specification for practical pan characteristic, updated implementation section and minor additions

White Papers are distributed freely on request.

Authorisation of the Chief Scientist is required for publication.

What is Audio Description ?

For understandable reasons many television programmes rely on visual content and pictorial composition to help to tell their story. The provision of additional description of the scene or action can therefore provide a considerable benefit to understanding, especially to the visually impaired.

Audio description (AD) is an ancillary component associated with a TV service which delivers a verbal description of the visual scene as an aid to understanding and enjoyment particularly (but not exclusively) for viewers who have visual impairments [1 & 2].

The description is typically confined to gaps in the normal programme narrative. The opportunities to describe a scene are therefore dependent on the programme genre and on the editing of the main programme sound. Some programmes are naturally more suited to description than others.

For example news programmes provide little opportunity for description and anyway tend to be self-documenting. Similarly the presenters of some (but not all) cookery programmes provide a seamless uninterrupted description of what they are doing on screen which precludes significant opportunities for description.

Science and informative programmes, on the other hand, tend to include relatively long gaps in the narrative which are associated with significant visual content; in such cases the gaps provide ample opportunity to fully describe the concurrent visual images.

At the other extreme, action drama and “soaps” are edited more tightly and typically provide predominantly brief windows in the dialogue which allow only concise description events. Because drama and soaps make significant use of purely visual imagery for dramatic purposes AD is particularly beneficial despite the briefness of the description windows.

In any programme there may be long periods during which there is no suitable opportunity for description. Any means of delivering AD must therefore accommodate extended periods of description silence (representing near-continuous narrative in the main programme sound) and so should provide the user with means of confirming that, under these conditions, description silence does not necessarily imply failure in delivery of the service or in the receiving equipment.

---

[1] Readers of this note who have yet to be persuaded of the benefits of AD are invited to try fully comprehending an episode of a soap or drama or of a science and features programme (such as “Horizon”) with their eyes closed.

[2] Audio description became available with some US analogue terrestrial channels more than a decade ago. A brief and small-scale UK experiment (called Audetel) using CELP-coded audio conveyed as teletext data was led by the ITC in the early 1990s. This further demonstrated the value of such a service and, en passant, exposed several practical issues and user requirements.

Audio Description – determining the requirements

An ancillary service such as AD will not suit all viewers so the first requirement is that it is provided as a "closed" or "elective" system where the user elects whether or not to hear the description. This is directly analogous to "closed captioning" or subtitles.

Those gaps in programme narrative which present opportunities for description will nevertheless often include sound effects or music which can make an added description hard to discern. Furthermore the loudness of programme sound during these description opportunities will vary from one gap to the next. An important requirement of AD is therefore to be able to adjust the relative level of programme sound and description in the mix which the AD user hears and, significantly, to be able to adjust this relative level on a passage-by-passage basis. Determining the appropriate depth of this fade is best be done by the programme maker under controlled conditions typically using the final transmission copy of the programme at the stage when the AD component is being authored.

Individual AD users will have different aural acuity, different describers will have their own style of vocal delivery (voice pitch and timbre), several voices may be used to describe a single programme and there are, in practice, differences in audio signal level for different home receivers. This makes it very desirable for the AD user to be able to make minor adjustments to the volume of the description signal to suit his or her condition[3].

Description content is voice only. It thus makes sense to assume that the description signal need not have the same audio bandwidth as programme sound and that, where appropriate, a separate description signal can be conveyed as mono rather than stereo – this saves bandwidth or bit-rate. The combined programme sound and description signal as heard by the AD user will however need the normal bandwidth appropriate accurately to convey the programme sound.

As noted above there will sometimes be considerable intervals between successive description passages. The AD user who has selected a described channel during one of these gaps will therefore find it hard to determine from the audio itself whether the temporary absence of description is intentional or is the result of a fault in scheduling, transmission or decoding. There is a strong requirement for the user to be able to confirm that AD is being transmitted as scheduled, preferably using an audible indication since many AD users will have visual impairments.

In summary the user requirements for AD are

- a closed system,
- ability to adjust relative volume of description and
- ability to promptly determine that a programme is currently being described.

From the service provider/multiplex operator's viewpoint the requirements should include

- bandwidth or bit-rate frugal delivery of the service,
- a delivery mechanism that uses existing & open standards (e.g. ISO/IEC 13818-x, DVB etc.),

Additional desirable features for AD decoder implementations include having separate hi-fi and VCR outputs [4] and providing an output for headphones should the AD user wish to listen in the company of others who do not wish to hear the description.

[3] A similar clear requirement for user control of description level was been identified in a recent (2003) usability study of "spoken subtitling" conducted by SVT in Sweden.

[4] Many set-top boxes for example already have separate SCART connections for the TV and VCR.

Audio Description on Digital Terrestrial Television

Digital television offers considerable flexibility for the delivery of new types of service or of additional service components; it therefore provides an excellent opportunity for adding AD to appropriate television services.

The deliverable bit-rate for each platform (satellite, terrestrial or cable) is not, however, limitless and bit-rate efficient methods of delivery are important. This is as true for the vision component of a digital television service as for subtitles or AD; thus service providers & multiplex operators all take steps (such as statistical multiplexing and opportunistic data insertion) to ensure that bit-rate is used effectively. Of the three platforms DTT has the least deliverable bit-rate per multiplex and the following text in this note describes the mechanism by which AD is coded, signalled and decoded for DTT [5]. The technique [6] could equally be applied to DSat and to DCable.

One of the practical conclusions of the "Audetel" project was that means should be provided of controlling the level of the main programme sound during a description passage and restoring it when the description was over. The level of the AD and overall level of main mixed with description would preferably be under the control of the listener. The principles are shown diagrammatically below in figure 1.

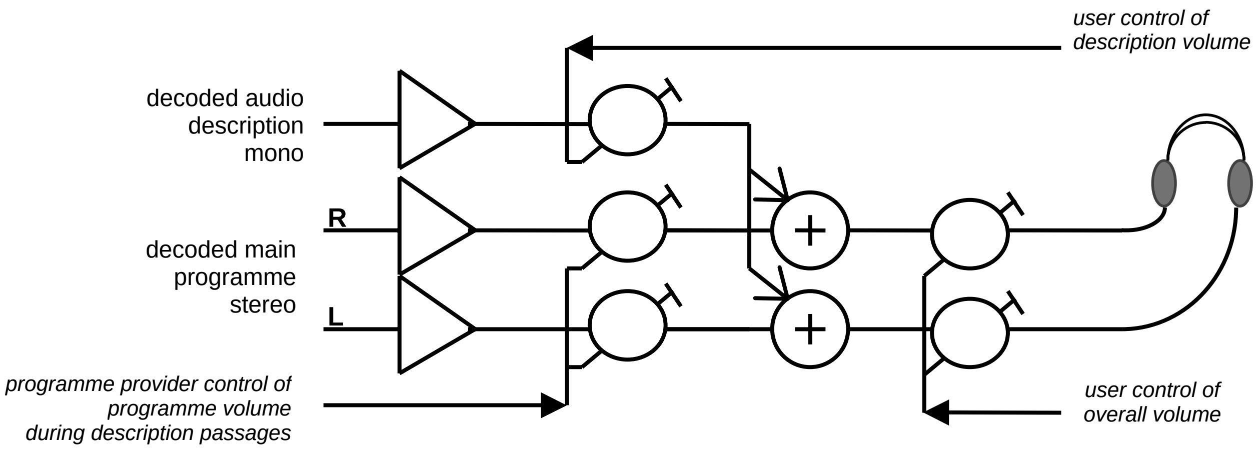


fig 1 : functionality of AD processing

To support this functionality there is thus a need for the programme provider to signal to the receiver/decoder so as to fade the main programme sound to a suitable level when the accompanying AD is active. For practical reasons this signalling is best done by embedding the information in the AD stream, thereby ensuring appropriate timing and conveniently binding the control signal to the component itself. As the level of fade depends on the level of the programme sound during each particular description passage the fade value needs to be adjustable rather than simply "on" or "off".

Fading the programme sound

In UK DTT the transmitted **fade** instruction is an unsigned byte value, 0x00 representing 0 dB, each increment representing a nominal 0.3 dB, 0xFE therefore representing approximately –77 dB whilst the fade value 0xFF represents completely mute programme sound. The programme provider is normally able to ensure standard signal levels in the description authoring process and throughout studio infrastructure and, where gaps in the narrative permit, subtle stepped fades are possible. This obviates having to have very conspicuous "crashes" in and out of a description

[5] The provision of audio description with digital terrestrial television services was mandated in the UK Broadcasting Act 1996. The UK Communications Act 2003 sets out a framework for providing AD on digital satellite and cable platforms as well as DTT.

[6] This is called "receiver-mix AD" and is documented in ETSI document TR101 154 v1.5.0 (2004-01). AD broadcast as a pre-mixed combination of programme sound and description is called "broadcast-mix AD".

passage when the corresponding programme sound is very loud. If the opportunity to describe is only short the fade and consequent recovery may however need to be very prompt.

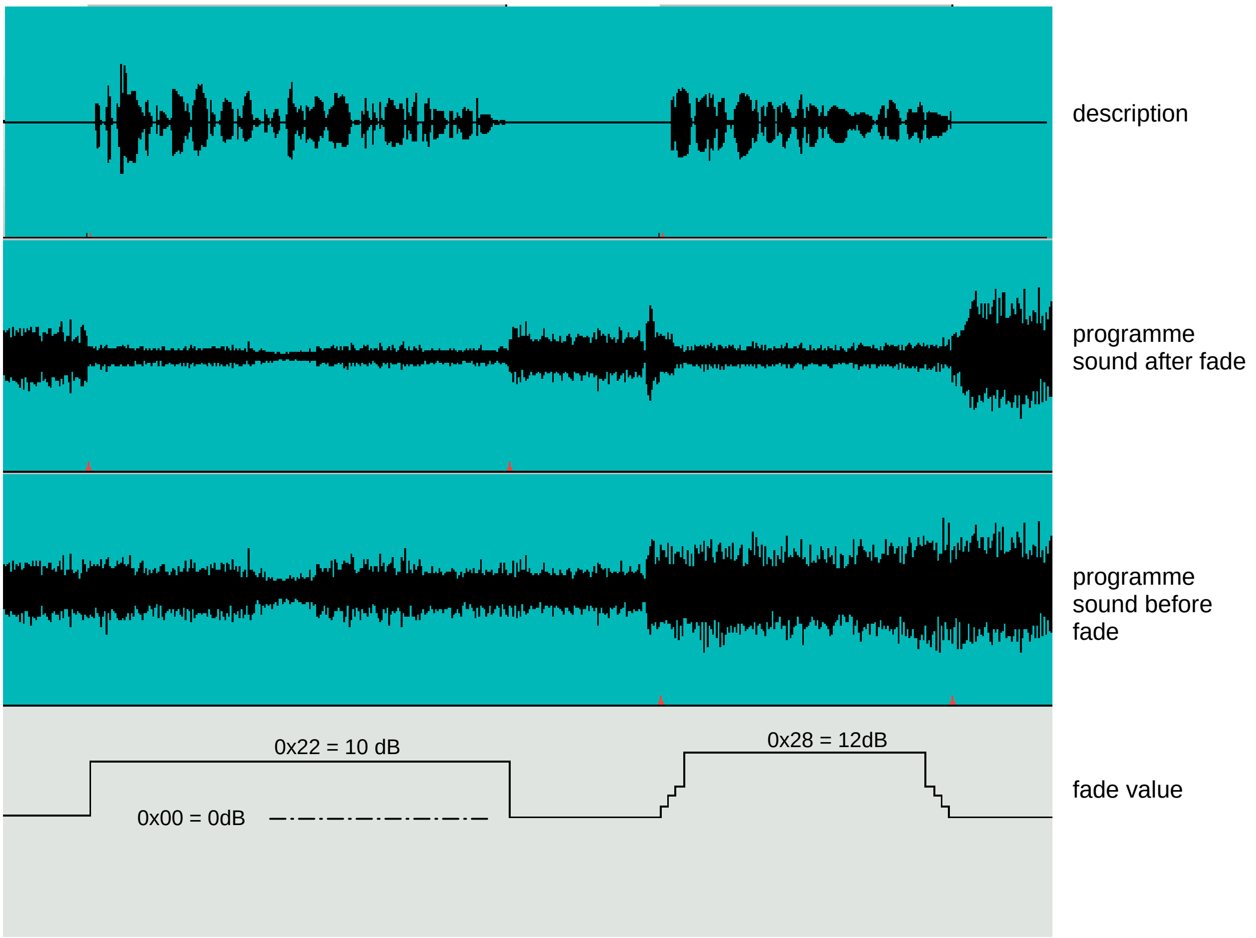


figure 2 – principles of fading programme sound during description passages

Figure 2 above illustrates these principles using real waveforms from a described BBC programme.

During the first description passage (leftmost) the programme sound requires only 10 dB of fade to avoid obscuring the description but the amount of description required to explain the scene means that the fade must be abrupt.

During the second description passage (right) more attenuation of the programme sound is required but there is less to describe; a more gradual fade in and out of the description passage is therefore possible.

Panning the description

As the means of embedding this control data in the compressed audio stream leaves some small bit-rate resource unused, a second control value (**pan**) is also transmitted. This allows the decoded AD signal to be panned within the sound stage of the main programme sound. Pan control enables the programme maker to place the "describer" at any preferred horizontal position within the sound field (in the same way that speech from out-of-vision commentators is sometimes placed to one side of the stereo image).

As with the fade, transmitted pan is a byte value, 0x00 representing centre front, each increment representing about 1.4° clockwise looking down on the listener (see figure 3 below).

For stereo the pan value will be restricted to ±30° of the centre front (i.e. to the range 0xEB..0xFF & 0x00..0x15) but the syntax of the signalling allows for any future use in which an AD component might be provided with a surround-sound main programme audio.

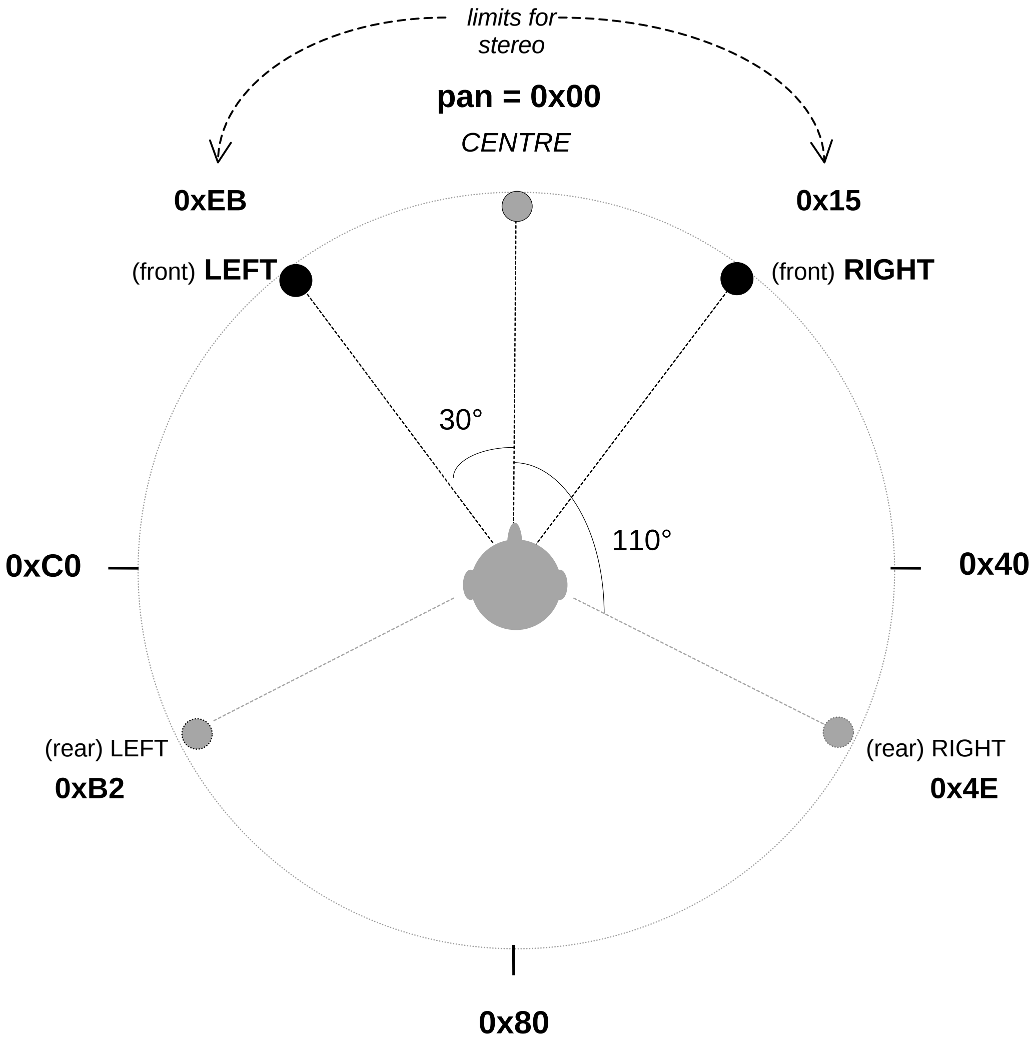


figure 3 : interpretation of audio description pan value

(seen from above the listener; includes mapping onto multi-channel sound presentation)

AD pan and line up levels

UK television broadcasters use line up levels of -18 dBfs for all audio signals, whether mono or stereo, with peak level at -10 dBfs (otherwise known as **M6** line up where mono signal is derived by the formula M = (A + B) - 6 dB). To avoid incompatibility and clipping, AD signals need to be recorded with the same line up as the main audio channels.

In the main audio central speech usually peaks up to -10 dBfs in both channels, but if the voice is panned to one channel it can still only peak to -10 dBfs in that channel. Thus to retain compatibility, the pan law for the mono AD signal needs to operate by attenuating one or other channel whilst keeping the gain of the other channel constant.

The line up of an AD decoder can be summarised in the following requirements.

- Reference level on the main stereo channels and reference level on the AD channel should both appear at the same level on the output when the AD is panned centrally and the user mix adjustment is at the default (0 dB gain) setting.
- Reference level on the AD channel should appear at the same level as a centrally panned signal in one channel and be attenuated by at least 20 dB in the other channel when the description is fully panned to one side and the user mix adjustment is at the default setting.
- Throughout the pan from left to right channel, reference level in the AD channel should appear at the same level in the louder channel and at a lower value in the other.

AD pan law [7]

To implement the AD pan in a way which tracks the 1.4° steps across the stereo image, the required relative levels for this movement can be calculated from the "stereophonic law of sines[8]".

For loudspeakers placed at ±30° this gives the formula

$\sin \alpha = (A - B) / 2.(A + B)$ where α is the angle from centre.

Therefore for the right half of the stereo image (where $0° \leq \alpha \leq 30°$) the required levels of the two channels become

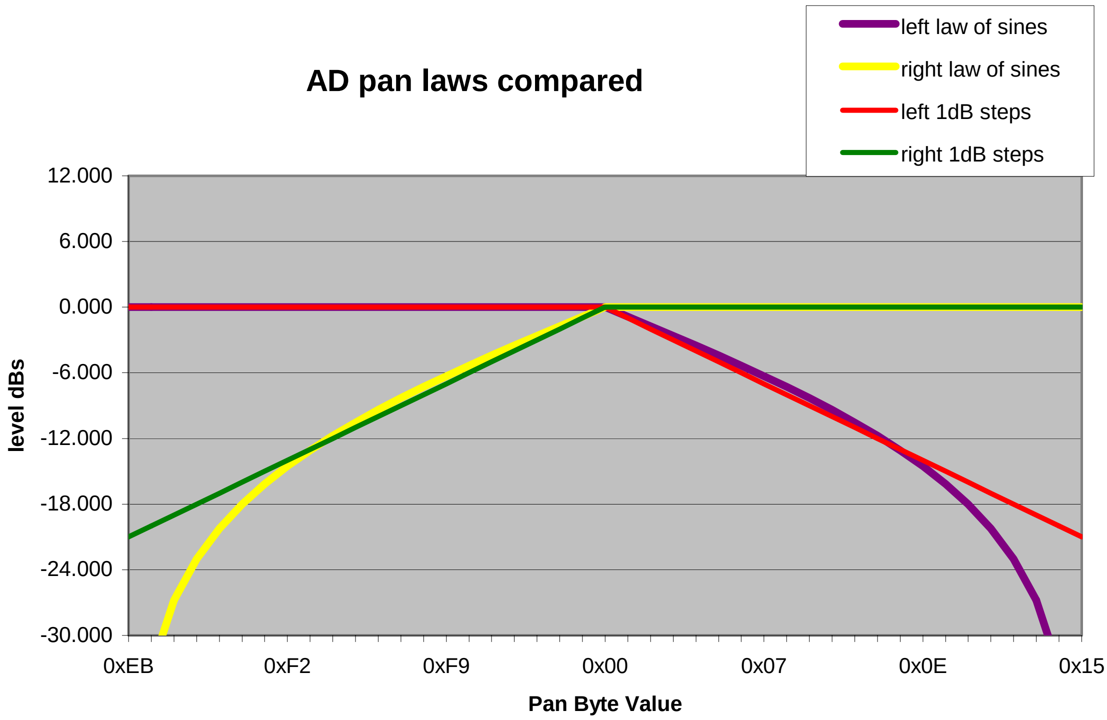
$A = M.(1 + 2.\sin \alpha) / (1 - 2.\sin \alpha)$ and $B = M$

and for the left half of the image

$L = M$ and $R = M.(1 - 2.\sin \alpha) / (1 + 2.\sin \alpha)$.

Annex 1 shows the appropriate levels for each value of pan byte from 0xEB - 0xFF & 0x00 - 0x15.

For a practicable implementation of AD pan, a reasonable match to these values can be obtained by attenuating the signal in one channel at a rate of 1dB per pan value. This gives an accuracy of better than ±1dB over the range of ±20° and a maximum attenuation of 21 dB which is sufficient to move the position of the sound into the louder of the two speakers [1].



---

7 The pan law described here supercedes other strategies previously suggested (such as constant power).

8 Quoted from "Phasor Analysis of Some Stereophonic Phenomena" B.B. Bauer, JASA, vol 33, no. 11, 1961.

Signalling the presence of audio description

The MPEG-2 System syntax defined in ISO/IEC 13818-1 [2] provides an optional simple fixed length field for **PES_private_data** with a total capacity of 16 bytes (128 bits) per packetised elementary stream (PES) packet.

Fade and pan control information is therefore coded in PES_private_data within the PES encapsulation of the coded AD component.

The structure and syntax of this field are as follows.

```
AD_descriptor(){
    reserved                           1111     4 bslbf
    AD_descriptor_length               1000     4 bslbf
    AD_text_tag                0x4454474144    40 bslbf   (5 bytes)
    revision_text_tag                  0x31     8 bslbf
    AD_fade_byte                       0xYY     8 bslbf   (FADE byte)
    AD_pan_byte                        0xYY     8 bslbf   (PAN byte)
    reserved               0xFFFFFFFFFFFFFF    56 bslbf   (7 bytes)
}
```

The semantics are as follow.

AD_descriptor_length : the number of significant bytes following the length field (i.e. 8).

AD_text_tag : a string of 5 ASCII characters forming a simple and unambiguous means of distinguishing this from any other PES_private_data.

A receiver which fails to recognise this tag should not interpret this audio stream as audio description.

revision_text_tag : the **AD_text_tag** is extended by a single ASCII character version designator (here “1” indicates revision 1).

Descriptors with the same **AD_text_tag** but a higher revision number shall be backwards compatible with this specification – the syntax and semantics of the fade and pan fields will be identical but some of the reserved bytes may be used for additional signalling.

AD_fade_byte : takes values between 0x00 (representing no fade of the main programme sound) and 0xFF (representing a full fade)

Over the range 0x00 to 0xFE one lsb represents a step in attentuation of the programme sound of approximately 0.3dB giving a range of about 77 dB. The fade value of 0xFF represents no programme sound at all.

The rate of signalling and the expected behaviour of a decoder to changes in fade byte are described below.

AD_pan_byte : takes values between 0x00 representing a central forward presentation of the audio descriptor and 0xFF, each increment representing a $^{360}/_{256}$ degree step clockwise looking down on the listener (ie. just over 1.4 degrees, see figure 3).

The rate of signalling and the expected behaviour of a decoder are described below.

reserved : the remaining 7 bytes are set to 0xFF and reserved for future developments if and when required.

A PES-packet from an audio stream [3] carrying audio description will therefore typically commence thus :

```
packet_start_code_prefix        0x000001           24 bslbf

stream_id                       0xC0               (0xC0-0xDF ≡ audio streams)

PES_packet_length               0xYYYY             (as appropriate)

'10'                            10                 2  bslbf
PES_scrambling_control          YY                 2  bslbf   (as appropriate)

PES_priority                    Y                  1  bslbf   (as appropriate)
data_alignment_indicator        Y                  1  bslbf   (as appropriate)
copyright                       Y                  1  bslbf   (as appropriate)
original_or_copy                Y                  1  bslbf   (as appropriate)

PTS_DTS_flags                   10                 2  bslbf   (PTS present)
ESCR_flag                       0                  1  bslbf
ES_rate_flag                    Y                  1  bslbf   (as appropriate)

DSM_trick_mode_flag             0                  1  bslbf
additional_copy_info_flag       0                  1  bslbf
PES_CRC_flag                    Y                  1  bslbf   (as appropriate)
PES_extension_flag              1                  1  bslbf

PES_header_data_length          0x10               8  uimsbf  (16d)

'0010'                          0010               4  bslbf
PTS[32..30]                     YYY                3  bslbf   (as appropriate)
'1'                             1                  1  bslbf
PTS[29..15]                     YYYYYYYYYYYYYYY    15 bslbf   (as appropriate)
'1'                             1                  1  bslbf
PTS[14..0]                      YYYYYYYYYYYYYYY    15 bslbf   (as appropriate)
'1'                             1                  1  bslbf

if (ES_rate_flag ==1'1') {etc.}
if (PES_CRC_flag ==1'1') {etc.}

PES_private_data_flag           1                  1  bslbf
pack_header_field_flag          0                  1  bslbf
program_packet_sequence_counter_flag  0            1  bslbf
P-STD_buffer_flag               0                  1  bslbf
reserved                        111                3  bslbf
PES_extension_flag_2            0                  1  bslbf

AD_descriptor(){

  reserved                      1111               4  bslbf
  AD_descriptor_length          1000               4  bslbf   (16d)
  AD_text_tag                   0x4454474144       40 bslbf
  revision_text_tag             0x31               8  bslbf
  AD_fade_byte                  0xYY               8  bslbf   (FADE byte)
  AD_pan_byte                   0xYY               8  bslbf   (PAN byte)
  reserved                      0xFFFFFFFFFFFFFF   56 bslbf
}

for (i=0; i<N1; i++) {stuffing_byte}                          (if required)
```

Thereafter commences the coded audio elementary stream data.

The presentation time-stamp (PTS) is the ISO/IEC 13818-1 mechanism for synchronising the presentation of a decoded stream by the decoder. The PTS for each time-critical component of a service is referenced to the service programme clock reference (PCR). One result of this temporal binding is that good audio-vision synchronisation can be maintained. The PTS field encapsulated in a PES packet refers to the first "access unit" [9] (AU) which commences in that PES packet.

The maximum rate of signalling of fade and pan values is determined by the number of audio PES packets per second for that AD stream. For bit-rate efficiency it is usual to encapsulate several access units of audio within one PES packet. Note that the fade and pan values in each AD_descriptor are deemed to apply to each of the AUs encapsulated within, and which commence in, that PES packet.

It is usual for audio AUs to be aligned with the PES packet and for the PES packet to encapsulate an integer number of AUs. In this case the PTS refers to the first AU within that PES packet and the fade and pan value are deemed to apply to all of the AUs therein. In principle this integer number could also vary from one PES packet to the next .

It is also possible (but relatively unusual) for the PES packetisation to be asynchronous with the AUs and for the PES packet to commence with and/or to end with an incomplete AU. In this case the PTS refers to the first AU which commences within that PES packet and the fade and pan value are deemed to apply to all the AUs which commence therein. Decoders should be capable of accommodating each of these forms of PES packetisation.

In practice the encapsulation of several AUs within one PES packet means that fade and pan values are transmitted typically every 120ms to 200ms. This allows the programme provider to have some control over the attack and decay of a fade and for fades to be reasonably gentle (i.e. taking several intermediate values between no-fade and the final target) where the gap in narrative permits.

In the UK the rate at which fade and pan values are transmitted shall never exceed 10 per second but successive PES packets may well convey different fade and pan values (e.g. during a fade) which must be accurately represented in the AD decoder output.

An AD decoder must maintain the relative timing between the decoded description signal and the decoded programme sound signal ***and*** between the appropriate fade and pan values and the decoded description signal.

As noted above, the description signal is mono speech and will typically be coded at a relatively low bit-rate (e.g. 64 kbits/s). Whilst broadcasters tend to use 48 kHz sampling for digital audio in their studio infrastructures, other sample rates are possible for digital television services [4]. A simple practicable constraint is that ***the sampling rate of the AD audio shall be identical to that of the programme sound for that TV service*** **[10]**. AD decoders should therefore be capable of decoding MPEG1 layer II or MPEG 2 mono signals at bit-rates between 64 kbits/s and 256 kbits/s and of supporting the audio sampling rates relevant to applicable digital television services.

AD stream components with DTT services in the UK are distributed unscrambled regardless of any scrambling which might be applied to other service components.

For the duration of a described programme the AD for DTT is transmitted as described above. During programmes for which there is no description however there is little point in transmitting an AD stream of continual silence; during these periods the bit-rate accorded to AD could be reassigned for other purposes [11]. In an 18 Mbit/s DTT multiplex with 4 linear TV services this could yield 300 kbits/s or more of reusable bit-rate during periods when none of these services is being described. Decoders should therefore be able to respond promptly to the addition of an AD component at the start of a described programme.

---

[9] An "access unit" is a unit of coded data – e.g. a frame of video or a 24 ms "frame" of MPEG1 layer II audio.

[10] This is a new but entirely practicable constraint designed to simplify receiver implementations.

[11] e.g. by using null-packet harvesting and using opportunistic data insertion for time non-critical application data such over-the-air receiver software upgrades etc..

Signalling AD in the PSI & SI

ISO/IEC 13818-1 (MPEG-2) and DVB-SI rules provide straightforward methods of referencing and labelling the individual stream components of a digital television service. For such a service on DTT the audio description component is signalled in the Program Map Table (PMT) of the Programme Specific Information (PSI) in a similar manner to the signalling for the programme sound component.

The streams for programme sound and for audio description are distinguished in the PSI by the use of the **ISO_639_language descriptor.** The **audio_type** field within the descriptor associated with programme sound is assigned the value 0x00 ("undefined") whilst the equivalent descriptor associated with audio description has its audio_type field assigned the value 0x03 ("visual impaired commentary").

This is illustrated below from a typical DTT PMT

```
…..
//  main programme audio details
{
    stream_type             0x03      ; Audio MPEG1
    reserved                111b
    elementary_PID          0x0259    ; PID for programme sound for this service
    reserved                1111b
    ES_info_length          0x009
  {
    descriptor_tag          0x52      ; stream identifier descriptor
    descriptor_length       0x01
    component_tag           0x02
  }
  {
    descriptor_tag          0x0A      ; ISO 639 language descriptor
    descriptor_length       0x04
    ISO_639_language_code   "eng"     ; English
    audio_type              0x00      ; undefined
  }
}
…..
// audio description details
{
    stream_type             0x03      ; Audio MPEG1
    reserved                111b
    elementary_PID          0x025A    ; PID for AD component for this service
    reserved                1111b
    ES_info_length          0x009
  {
    descriptor_tag          0x52      ; stream identifier descriptor
    descriptor_length       0x01
    component_tag           0x03
  }
  {
    descriptor_tag          0x0A      ; ISO 639 language descriptor
    descriptor_length       0x04
    ISO_639_language_code   "eng"     ; English
    audio_type              0x03      ; visual impaired commentary
  }
}
```

In the event that a service has AD in several languages (e.g. English and Welsh) the PMT reference to each stream would have the appropriate ISO_639_language_code and the AD decoder would discriminate between them on the basis of the preferred language chosen in the user settings.

If a service includes several AD streams (e.g. for different languages) ***and*** if there is no mechanism for the user to select between these streams then the AD decoder should decode the first AD stream referenced in the PMT for that service.

Decoder behaviour in the presence of AD

In the presence of a valid AD descriptor in the encoded description signal for the selected service the AD decoder should present the appropriate mix of programme sound and description signal to the user.

AD decoders should be capable of decoding MPEG1 layer II or MPEG 2 mono signals at bit-rates between 64 kbits/s and 256 kbits/s and of supporting the audio sampling rates relevant to applicable digital television services.

The relative timing of programme audio, description audio ***and*** the signalled fade/pan values should be maintained correctly at all times.

When the fade value is 0x00 or in the absence of an AD stream the programme sound level should be unattenuated.

In the presence of a valid AD descriptor, the AD decoder should attenuate the programme sound by 0.3 dB per fade value increment. If the AD decoder cannot support such small steps then the implemented attenuation should match the intended attenuation as closely as possible. For example if only 1 dB steps are possible then fade values of 0x00 and 0x01 should map to 0db, 0x02, 0x03 and 0x04 should map to1dB, 0x05, 0x06, 0x07 & 0x08 to –2db etc..

In a stereo environment the AD decoder should interpret any pan values outside the ranges 0xEB..0xFF and 0x00..0x15 in the following manner. Pan values from 0x16 to 0x7F inclusive should be mapped to the value 0x15 (i.e. stereo hard right). Pan values from 0x80 to 0xEA should be mapped to the value 0xEB (i.e. stereo hard left).

If, whilst listening to a described programme, the user selects a new programme, the AD decoder should mute the decoded description signal and restore the programme sound to the unfaded state. This restoration should not be abrupt - it is recommended that under such conditions the value of fade and of pan are ramped to the default values (0x00) over a period of at least 1 second.

AD decoders should present to their VCR output a mix of programme sound and description modulated as appropriate by fade and pan but before any attenuation applied by the user control of overall volume control shown diagrammatically in figure 1.

Decoder behaviour in the presence of errors

If the AD decoder detects an error in, or absence of, the AD descriptor in the encoded AD signal, it should have a strategy which leads to muting the decoded description signal, restoring the programme sound to its default unfaded amplitude and setting the effective fade and pan values to 0x00[13].

When the AD signal is suddenly lost or regained the AD decoder behaviour as experienced by the user should never be abrupt. It is recommended that in the event of an error or the absence of AD signal the value of fade and of pan implemented by the AD decoder be ramped from the signalled values to the default values (0x00) over a period of at least 1 second [14]. Equally, on recovery from an error or on the reappearance of the AD signal the value of fade and of pan should be ramped to the signalled values from the default values (0x00) over a similar period.

[13] e.g. the AD decoder might flywheel through isolated errors caused by uncorrected transmission errors but respond appropriately to successive instances of loss caused by problems at the “headend”.

[14] Under normal operation the AD decoder should follow at signalled speed any changes in the transmitted fade and pan values

Decoder user indications

As many potential users of AD will be visually impaired, the user interface should not rely solely on visual clues (lights or on-screen display logos) to indicate status information such as the presence or absence of description. Audible indications are desirable and designers should consider how to distinguish different states using, for example, contrasting tones.

The user should be able to interrogate the decoder (e.g. using a remote control) and be given an indication that all is well (e.g. a recognisable “beep” and flashing LED). This will indicate presence in the stream of decodable AD, even when the description may at that particular moment be silent and the fade value 0x00. Equally a distinguishable indication of the detected absence of the AD signal is highly desirable (e.g. a using a second style of “beep”).

Other desirable controls with distinctive audible indications include the ability to mute the combined sound and to adjust the description and overall volumes.

Any user tones applied to the headphone or hi-fi outputs of the decoder should not be added to the decoder VCR output.

Test streams

The UK DTG Test Centre maintains a suite of test streams to help developers ensure correct behaviour of DTT AD decoders.

These streams provide tests for include service selection, audio level, relative timing, fade and pan response and behaviour in the presence of AD signalling errors.

Standardisation

The syntax and signalling for receiver-mix AD described in this white paper is as defined in annex G of the ETSI DVB guidelines document TR 101 154 v1.5.0 (2004-01).

Currently however TR 101 154 does not explicitly define a preferred pan characteristic – indeed it suggests (as a footnote example) a constant power model.

The preferred pan law for AD described above has been developed since the text for annex G of TR 101 154 was written. It is straightforward to implement, produces manageable peak levels in the receiver and can also be used to produce a fully broadcast-compatible pre-mix for broadcaster-mix AD under M6 line-up.

This pan law has also been validated in a commercial product (see below) and should be the pan characteristic used for future decoder implementations of receiver-mix AD [15].

---

[15] The pan characteristic described in this document is also the appropriate characteristic for equipment used to generate “broadcaster-mix” AD on DTV platforms where that particular signal-form is used.

Implementations

PCMCIA card

The UK DTT multiplex operators represented by TDN commissioned the development of a PCMCIA card module which fulfils the requirements of an audio description decoder and plugs into any DTT receiver which has a working “Common Interface” (CI) socket [5]. It was designed and built by SCM under contract to TDN with technical oversight and functional testing provided by BBC R&D. The ADM has been successfully demonstrated with an integrated digital TV receiver and with a suitably equipped DTT set-top-box.

The design was initiated at a time when UK DTT first carried a number of encrypted services. Contemporary commercial sensitivity about including decryption in the AD module (ADM) functionality resulted in a design in which the programme sound input to the module is analogue and from the receiver phono sound output whilst the audio description signal (broadcast by agreement “in the clear”) is demultiplexed, decoded and processed within the ADM itself.

The ADM also manages the fade and pan processing together with the user interface. This interface is via a separate and simple infra-red remote controller and external IR receiver pod which affixes to the top of the receiver.

Separate outputs are available for headphones, hi-fi and VCR. The remote control provides means of adjusting the description level and of the level of the overall programme-sound/AD mix, of querying the status of AD on the selected programme and of muting the headphone and hi-fi outputs. Distinctive tones are also added to these outputs to provide audible confirmation of the remote control keystrokes and of the AD status.

figure 4: The TDN Audio Description Module

Netgem i-Player

Shortly before Christmas 2003, Netgem launched a version of their **i-Player** DTT set-to-box product in which the AD decoding and mixing is performed in software using existing resources within the original product.

Netgem have clearly considered the usability of this product. Control of the description volume is integrated on the existing remote control and audio tones are used to distinguish between presence or absence of AD as the user selects a new service. The decoder also enunciates the name of the service by playing and decoding short MP3 speech files stored in the product.

Nebula

Nebula Electronics have implemented AD decoding and processing in software on both the PCI card and USB versions of their DigiTV DTT decoder product.

Philips

Philips Semiconductors have demonstrated AD decoding and processing in software on their existing PNX1300 ("TriMedia") chip could implement receiver-mix AD in future DTV decoding products.

References

[1]	BBC RD Report 1979/7	
[2]	ISO/IEC 13818-1	MPEG 2 system syntax
[3]	ISO/IEC 11172-3	MPEG 1 audio syntax
[4]	TR 101 154 v1.5.0 (2004-01)	Digital Video Broadcasting (DVB); Implementations guidelines for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream.
[5]	EN 50221	Common Interface Specification for Conditional Access and other Digital Video Broadcasting Decoder Applications

Glossary of acronyms

Like many contemporary technologies, digital television generates acronyms at an alarming rate. This glossary aims to unzip just the acronyms used in this white paper.

AD	**A**udio **D**escription	
DTT	**D**igital **T**errestrial **T**elevision	
MPEG	**M**oving **P**ictures **E**xperts **G**roup	*(ISO common interest group)*
DVB	**D**igital **V**ideo **B**roadcasting	*(EU industry body)*
PSI	**P**rogramme **S**pecific **I**nformation	
SI	**S**ervice **I**nformation	
DTG	**D**igital **T**elevision **G**roup	*(UK industry body)*
DSat	**D**igital **S**atellite	
DCable	**D**igital **C**able	
PES	**P**acketised **E**lementary **S**tream	
VCR	**V**ideo **C**assette **R**ecorder	
PCMCIA	**P**ersonal **C**omputer **M**emory **C**ard **I**nternational **A**ssociation	*(industry body)*
CI	**C**ommon **I**nterface	*(DVB standardised interface)*
TDN	**T**he **D**igital **N**etwork	*(UK industry body)*

Annex 1 – Pan values which follow the law of sines

pan byte	pan value	pan angle degrees	left sine law	right sine law	left level dB	right level dB
0xEB	-21	-30.0	1.000	0.000	0.000	-∞
0xEC	-20	-28.5	1.000	0.022	0.000	-33.061
0xED	-19	-27.1	1.000	0.046	0.000	-26.784
0xEE	-18	-25.7	1.000	0.071	0.000	-23.000
0XEF	-17	-24.2	1.000	0.097	0.000	-20.233
0xF0	-16	-22.8	1.000	0.126	0.000	-18.022
0xF1	-15	-21.4	1.000	0.156	0.000	-16.159
0xF2	-14	-20.0	1.000	0.188	0.000	-14.534
0xF3	-13	-18.5	1.000	0.222	0.000	-13.082
0xF4	-12	-17.1	1.000	0.258	0.000	-11.759
0xF5	-11	-15.7	1.000	0.297	0.000	-10.537
0xF6	-10	-14.2	1.000	0.339	0.000	-9.393
0xF7	-9	-12.8	1.000	0.384	0.000	-8.312
0xF8	-8	-11.4	1.000	0.432	0.000	-7.283
0xF9	-7	-10.0	1.000	0.484	0.000	-6.295
0xFA	-6	-8.5	1.000	0.541	0.000	-5.340
0xFB	-5	-7.1	1.000	0.602	0.000	-4.413
0xFC	-4	-5.7	1.000	0.668	0.000	-3.506
0xFD	-3	-4.2	1.000	0.740	0.000	-2.616
0xFE	-2	-2.8	1.000	0.819	0.000	-1.738
0xFF	-1	-1.4	1.000	0.905	0.000	-0.867
0x00	**0**	**0.0**	**1.000**	**1.000**	**0.000**	**0.000**
0X01	1	1.4	0.905	1.000	-0.867	0.000
0x02	2	2.8	0.819	1.000	-1.738	0.000
0x03	3	4.2	0.740	1.000	-2.616	0.000
0x04	4	5.7	0.668	1.000	-3.506	0.000
0x05	5	7.1	0.602	1.000	-4.413	0.000
0x06	6	8.5	0.541	1.000	-5.340	0.000
0x07	7	10.0	0.484	1.000	-6.295	0.000
0x08	8	11.4	0.432	1.000	-7.283	0.000
0x09	9	12.8	0.384	1.000	-8.312	0.000
0x0A	10	14.2	0.339	1.000	-9.393	0.000
0x0B	11	15.7	0.297	1.000	-10.537	0.000
0x0C	12	17.1	0.258	1.000	-11.759	0.000
0x0D	13	18.5	0.222	1.000	-13.082	0.000
0x0E	14	20.0	0.188	1.000	-14.534	0.000
0x0F	15	21.4	0.156	1.000	-16.159	0.000
0x10	16	22.8	0.126	1.000	-18.022	0.000
0x11	17	24.2	0.097	1.000	-20.233	0.000
0x12	18	25.7	0.071	1.000	-23.000	0.000
0x13	19	27.1	0.046	1.000	-26.784	0.000
0x14	20	28.5	0.022	1.000	-33.061	0.000
0x15	21	30.0	0.000	1.000	-∞	0.000