Research White Paper

WHP 198

August 2011

Audio Description Studio Signal

Trevor Ware

BRITISH BROADCASTING CORPORATION

BBC Research White Paper WHP 198

Audio Description Studio Signal

Trevor Ware

Abstract

This paper describes the audio description studio signal which comprises a mono audio signal and a data stream containing fade and pan information. The paper is the definitive specification of the audio description studio signal and is primarily intended for those implementing systems that use this signal. The paper brings together all the information necessary to deploy the audio description studio signal in the broadcasting environment.

The audio description studio signal was first explained in the IBC 2000 paper entitled “Access Services For Digital Television: Matching The Means To The Requirement For Audio Description And Signing”.

For additional information on audio description in general see BBC Research & Development White Paper WHP051.

Additional key words: AES3, SDI, Manchester code, phase encoding, CRC, Dolby E, AD

White Papers are distributed freely on request.

Authorisation of the Controller of Research & Development is required for publication.

BBC Research White Paper WHP 198

Audio Description Studio Signal

Trevor Ware

1 Introduction

Audio description is an additional audio component in a digital television service that provides a verbal description of the visual scene. It is primarily, although not exclusively, aimed at viewers with visual impairments. It is intended to aid the understanding and enjoyment of the television programme.

The audio description is produced in mono and must be mixed with the main programme audio to produce the final audio described programme. The mixing can take place in the viewers' receiver, in this case the broadcaster transmits the regular programme sound and the mono audio description, and the two components are mixed together in the receiver. This is known as receiver-mix audio description. Or the broadcaster can perform the mixing, and broadcast the programme sound complete with the audio description. This is known as broadcast-mix audio description, in this case the standard programme sound, without the audio description, will also be broadcast for those viewers not requiring the audio described version.

Whichever method of broadcast is employed, there needs to be some data accompanying the mono audio description that will be used by the audio mixing process to attenuate the programme sound during passages of audio description. This attenuation is required because, although there should not be any programme dialogue during the audio description passage, there may be music or sound effects that if not reduced in volume could make the description unintelligible. This fade data is produced when the audio description is being created and from this point in time it must accompany the mono audio signal. There is also pan data that can be used to position the mono audio in a stereo or multi-channel sound stage.

This paper describes the method for carrying the fade and pan data with the audio signal in the broadcaster's domain – the "Audio Description Studio Signal". The audio description studio signal has been designed as a rugged, "audio circuit" compatible signal that will survive high levels of audio processing, including lossy compression systems. The method described in this definitive specification is free to use for those developing, manufacturing and deploying audio description systems.

2 Audio description studio signal overview

The audio description studio signal comprises mono audio and audio description data. By convention the audio is carried in the left-hand channel, and the data carried in the right-hand channel, of an AES3 digital audio signal. Whilst the mono audio samples are synchronous with the AES3 sample clock, the audio description data is sampled asynchronously by the AES3 clock.

The audio description data is carried in the audio space of the AES3 signal as a fixed amplitude "analogue" audio signal with the AES3 Channel Status information indicating that PCM audio is present. The data carried by this audio signal is the AD_descriptor[1] described in appendix A. The AD_descriptor contains the audio description fade and pan data plus some additional signalling used in the receiver-mix broadcast signal. In the audio description studio signal the last 2 reserved bytes of the AD_descriptor are replaced with a 16 bit CRC.

[1] Although called the AD_descriptor, this data is carried in the, fixed length, 16 byte PES_private_data in the PES header of the coded audio of receiver-mix audio description and not for example as a DVB type descriptor.

In the television broadcast environment the AES3 signal would normally be carried as embedded audio in the SDI[iii] and HD-SDI[iv] signal. Audio channel assignments would be at the discretion of individual broadcasters but in the standard definition environment audio channels 3&4 are often used for the audio description studio signal. In the high definition situation audio channels 5&6 are commonly used for the audio description studio signal since audio channels 3&4 often carry multi-channel audio in Dolby E[v] format. The flowchart of figure 1 shows the generation of the audio description studio signal.

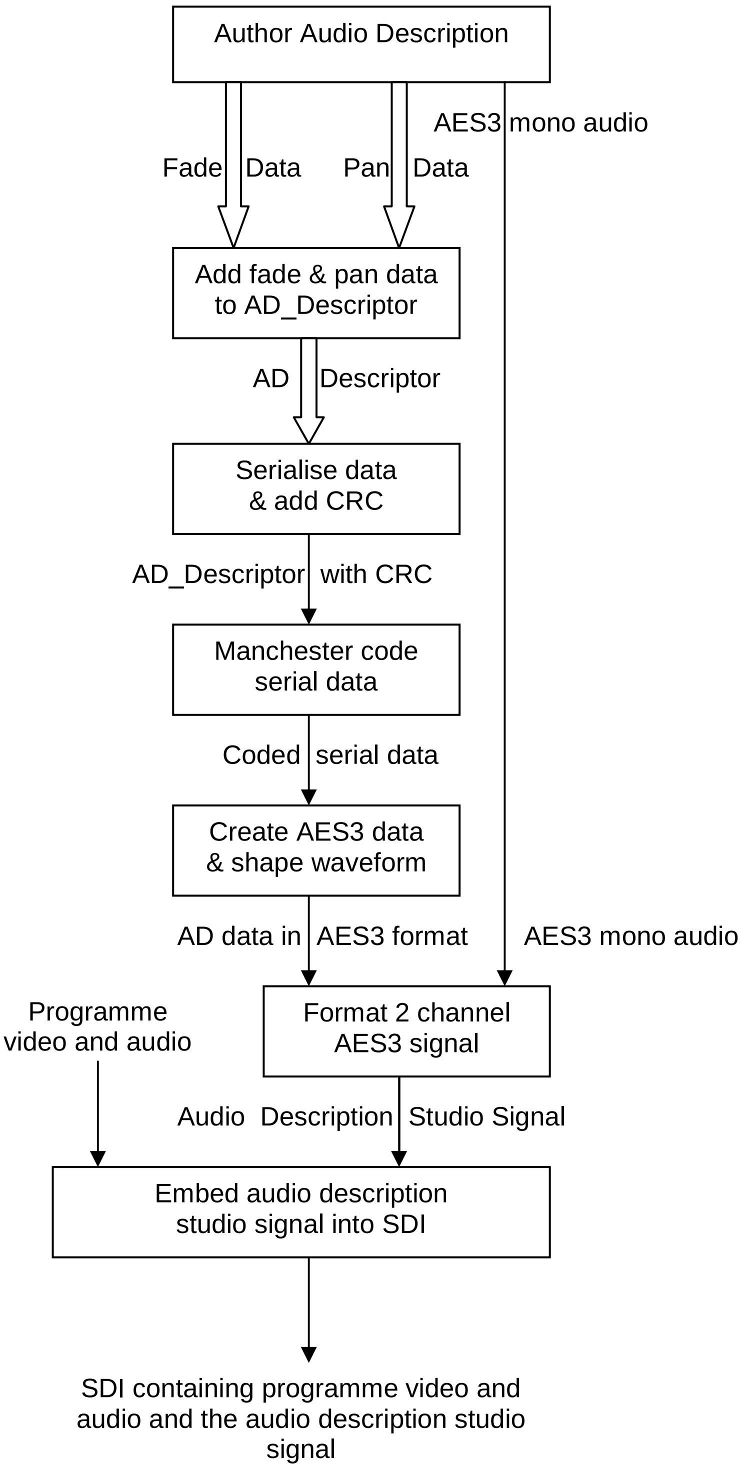


Figure 1. Flowchart showing the generation of the audio description studio signal

3 Serialisation of data and Cyclic Redundancy Check

The data-rate of the serial audio description studio signal is 1.28 kbits/second, this bit-rate is determined by the update rate of the value of the fade byte. To allow the attenuation of the programme sound to be reasonably subtle, with no large jumps in volume during the fade period, the fade value updates at a rate of 10 times per second. As the fade byte, and pan byte, are encapsulated in the 16 bytes of the AD_descriptor (with CRC) this 10 times per second update rate results in a data-rate of 160 bytes/second, thus the total bit-rate of the serialised signal is 1.28 kbits/second.

As stated in section 2, a Cyclic Redundancy Check word replaces the last 2 reserved bytes of the AD_descriptor. The CRC used on the AD_descriptor data is the ITU-T CRC-16[vi] 16 bit generator polynomial represented by $x^{16}+x^{12}+x^5+1$.

To generate the CRC, the last 2 reserved bytes of the AD_descriptor are set to the value 0x0000. The entire 16 bytes are then serialised, with the most significant bit of the first byte first, and passed through the logic shown in Appendix B.

The registers are all set to logic 1 before all 128 bits (16 bytes) of data are clocked into the circuit, with the 128th bit being at position C0. The resulting 2 byte value C15...C0 is then inserted in place of the last 2 reserved bytes of the AD_descriptor data (previously set to 0x0000).

To implement a CRC decoder, the same circuit shown in Appendix B is used with the registers set to logic 1, before all 128 bits of data are again clocked into the circuit. If there are no errors in the data the check value bytes (C15...C0) will be 0x0000.

3.1 Use of the CRC in practical implementations

Equipment processing the audio description studio signal can use the CRC to detect any error in the AD_descriptor data and deal with it appropriately[2]. A receiver-mix emission encoder should check the CRC word and then replace the two CRC bytes with 0xFFFF before inserting the AD_descriptor into the PES header[3].

In a broadcast-mix system the CRC should be checked and the mono audio description mixed with the programme sound using the fade and pan information. Once the programme sound and the audio description have been mixed for the broadcast-mix audio there is no further use for the AD_descriptor information and it may be discarded.

4 Coded fade and pan data

Rather than carrying the serialized AD_descriptor (with CRC) data in the audio domain directly as an NRZ data stream it is first coded. Manchester code (also known as phase encoding) is used, this produces a signal with an average DC level of zero, which can easily be converted to an analogue audio signal if required. This simple coding method also yields a signal that is straightforward to decode - the source clock signal can easily be recovered from the coded data.

---

[2] For example, if a single AD_descriptor was found to have an error the fade and pan values from the previous descriptor could be used.

[3] The AD_descriptor was standardised, without a CRC, before the audio description studio signal was developed and so the CRC used in the audio description studio signal cannot be carried through to the AD_descriptor.

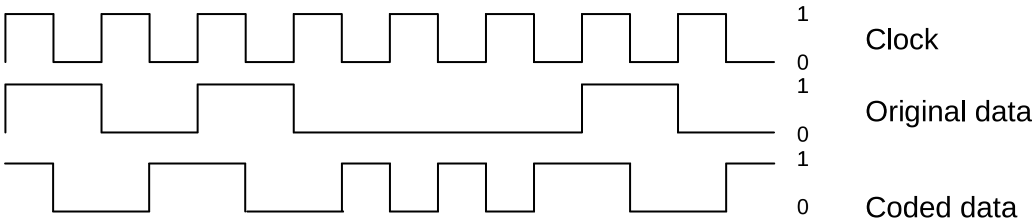


Figure 2. Manchester coding

The waveforms in figure 2 show an example of data[4] that has been Manchester encoded. A “0” in the original data is expressed as a low-to-high transition at the mid point of the data period in the coded data. A “1” in the original data is expressed by a high-to-low transition at the mid point of the data period in the coded data.

Note this is the original convention for Manchester code[vii], another convention[viii] exists where a “0” is represented by a high-to-low transition and a “1” by a low-to-high transition. Practical decoder implementations will check for this potentially inverted signal and decode correctly.

5 Implementation in the AES3 digital audio signal

The audio sample word representing a digital audio sample’s amplitude in the AES3 signal is represented in linear 2’s complement binary form, with the number of bits per sample word specified as 16, 20 or 24 bits.

The coded serial data is represented by 16 bit audio sample words. Logic level 0 is represented by the negative sample value 0xFE00 and logic level 1 is represented by the positive sample value 0x0200. Limiting to these maximum values of audio sample word results in an audio signal with a fixed level of -36 dBFS[5], symmetrical about the 2’s complement zero amplitude level.

The transitions between sample value 0xFE00 and sample value 0x0200 should be shaped by passing through intermediate sample values; this will restrict sideband energy if the data is converted into an analogue form. With a serial data rate of 1.28 kbits/s and assuming an audio sampling frequency of 48kHz, each data bit period of the coded serial data will be represented by 37 or 38, 16 bit audio samples in the AES3 domain. This allows for a number of sample periods around the transitions between 0xFE00 and 0x0200 to be used for intermediate sample values. How the intermediate sample values are generated is not proscribed.

The audio description data will be sampled at the same AES3 digital audio sampling frequency, and will be locked, to that of the mono description audio. This sampling frequency will in turn be locked to the sampling frequency of the programme sound, which is normally locked to the video. In the television broadcast environment the sampling frequency of the AES3 digital audio is typically 48kHz but other rates, such as 32kHz, 44.1kHz and 96kHz, could be used with no detrimental effect to the audio description data.

The oscilloscope waveforms in figure 3 show an actual shaped audio description Manchester coded data stream after digital to analogue conversion.

[4] The original data changes on the rising edge of the clock signal

[5] A signal at -36 dBFS is low enough in level to not be disturbing if reproduced audibly but high enough in level to register on audio level meters.

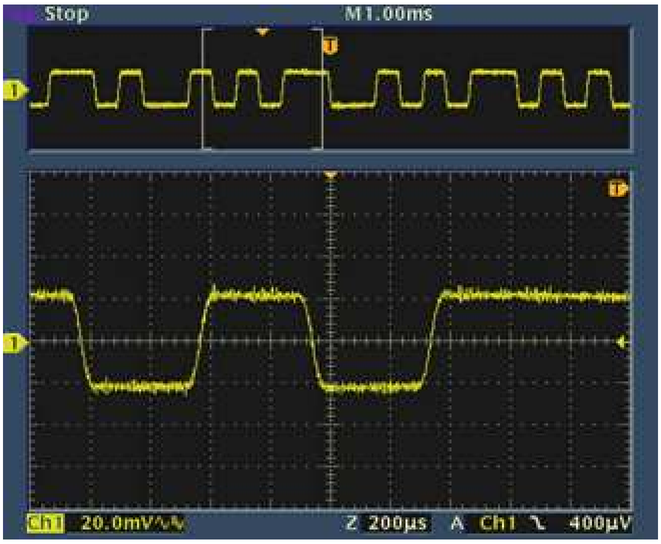


Figure 3. Shaped Manchester coded data

6 Carriage of the audio description studio signal in systems with audio processing

The audio description studio signal is simple and rugged and is not adversely affected by common audio processing that can be found in the broadcast infrastructure, it will survive re-sampling, gain changes, digital to analogue and analogue to digital conversion. The signal is suitable for carriage in lossy bit-rate reduced systems that can be found in distribution and contribution links, including systems that may be used for multi-channel sound such as Dolby E.

7 Summary

This White Paper has described the audio description studio signal with all the technical information necessary for the implementation of the signal in broadcast systems. The audio description studio signal is used by numerous UK broadcasters and equipment that uses the signal is available from many broadcast equipment manufacturers. The specification of the audio description studio signal explained in this paper is free to use.

8 Appendix A

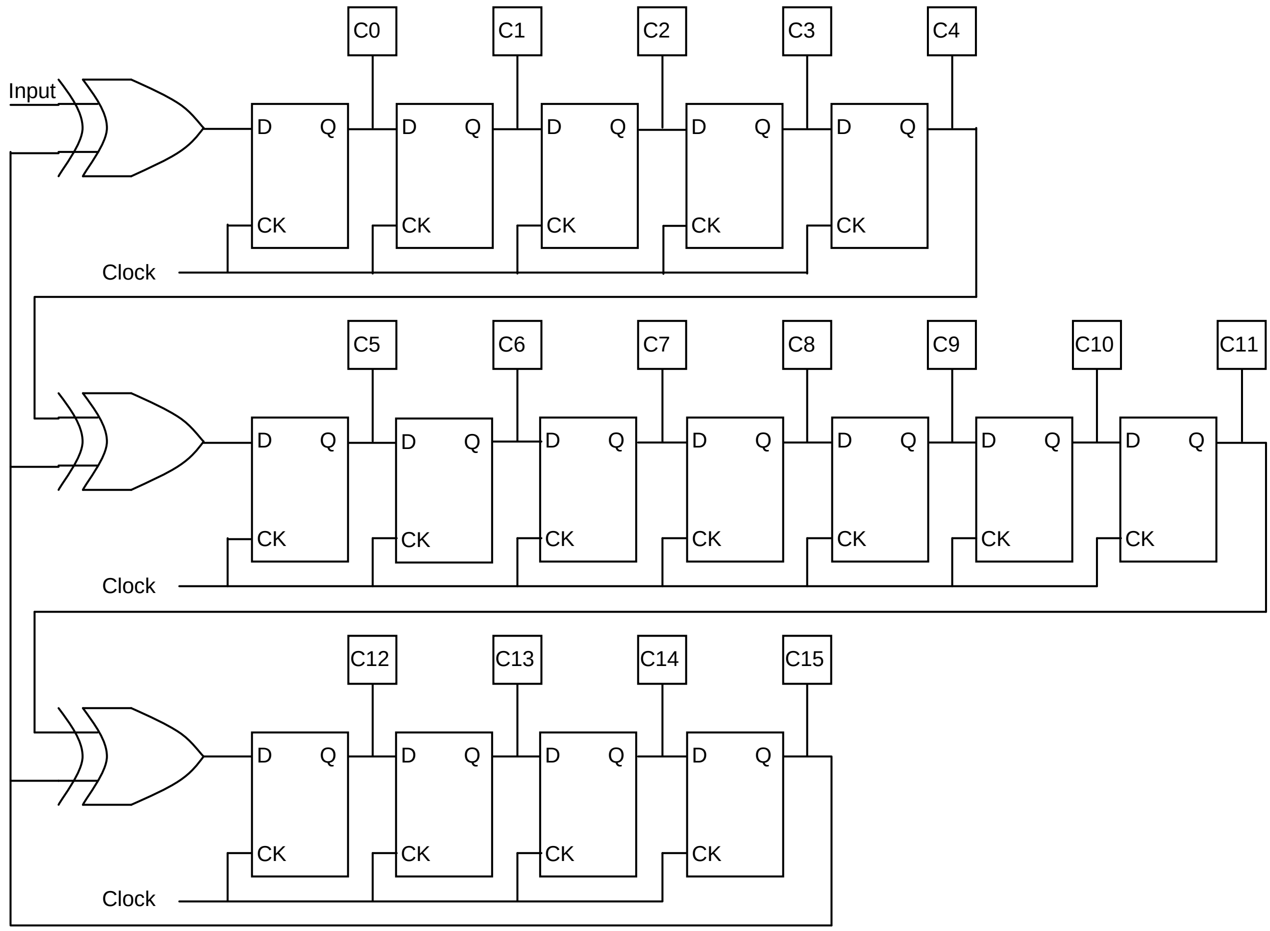
For receiver-mix audio description the fade and pan data is broadcast along with the mono audio description. The data is coded in PES_private_data in the PES (Packetized Elementary Stream) header of the coded audio description component in accordance with ITU-T Recommendation H.222.0 / ISO/IEC 13818-10[ix]. The format of the PES_private_data is specified in ETSI TS 101 154[x] and is as shown here when used to carry just audio description fade and pan data[6].

Syntax	Value	No. of Bits
`AD_descriptor {`		
`Reserved`	`0xF`	4
`AD_descriptor_length`	`0x8`	4
`AD_text_tag`	`0x4454474144`	40
`Version_text_tag`	`0x31`	8
`AD_fade_byte`	`0xXX`	8
`AD_pan_byte`	`0xYY`	8
`Reserved`	`0xFFFFFFFFFFFFFF`	56
`}`		

Note: PES_private_data has a fixed length of 16 bytes per PES packet

[6] The AD_descriptor in ETSI TS 101 154 V1.9.1 has been modified to allow the carriage of data to be used with "clean audio" services. When `Version_text_tag = 0x32` three bytes of reserved data is used for the clean audio information and the `AD_descriptor_length = 0xC`. This is backwards compatible with earlier versions of the specification as the `AD_text_tag` remains unchanged and the syntax and semantics of the fade and pan fields are unaltered.

9 Appendix B



Logic to implement the ITU-T CRC-16 cyclic redundancy check

10 References

[i] Proceedings of the 2000 International Broadcasting Convention. Access Services For Digital Television: Matching The Means To The Requirement For Audio Description and Signing. Nick Tanton, Trevor Ware and Mike Armstrong
http://downloads.bbc.co.uk/rd/pubs/papers/pdffiles/ibc00net.pdf

[ii] BBC R&D White Paper WHP 051. Audio Description: what it is and how it works. N.E. Tanton, T. Ware and M. Armstrong. October 2002 (revised July 2004)
http://www.bbc.co.uk/rd/pubs/whp/whp051.shtml

[iii] SMPTE 272:2004 Television – Formatting AES/EBU Audio and Auxiliary Data into Digital Video Ancillary Data Space

[iv] SMPTE 299-1:2009 24-Bit Digital Audio Format for SMPTE 292 Bit-Serial Interface

[v] Dolby E http://www.dolby.co.uk/professional/technology/broadcast/dolby-e.html

[vi] ITU-T Recommendation V.42: Error-correcting procedures for DCEs using asynchronous-to-synchronous conversion.

[vii] Tanenbaum, Andrew, S. (2002). Computer Networks (4th Edition). Prentice Hall, 274-275. ISBN 0-13-066102-3.

[viii] Stallings, William (2004). Data and Computer Communications (7th ed.). Prentice Hall, 137-138. ISBN 0-13-100681-9.

[ix] ISO/IEC 13818-2:2000 Information technology – Generic coding of moving pictures and associated audio information: Video
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=31539

[x] ETSI TS 101 154 V1.9.1 (2009-09) Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream
http://www.etsi.org/deliver/etsi_ts/101100_101199/101154/01.09.01_60/ts_101154v010901p.pdf