ACCESS SERVICES FOR DIGITAL TELEVISION : MATCHING THE MEANS TO THE REQUIREMENT FOR AUDIO DESCRIPTION AND SIGNING

Nick Tanton, Trevor Ware and Mike Armstrong

BBC R&D UK

ABSTRACT

Digital television offers excellent opportunities for the provision of access services such as audio description for the visually impaired and signing for the profoundly deaf.

The goal of "closed" access services (where the user can select whether to benefit from the service or not) is commercially attractive but the mechanism chosen must be practicable both for the service provider and for the user and, *most importantly*, must deliver a quality of service which is acceptable to the user. This paper describes the technical development of practicable cost-effective techniques for authoring, programme distribution and delivery of audio-description services for UK Digital Terrestrial Television. It also addresses the complex technical and operational issues which attend the provision of a signing signal component to be associated with existing digital television programmes.

INTRODUCTION

Digital television offers excellent opportunities for the provision of access services such as subtitling, audio description for the visually impaired and signing for the profoundly deaf [1].

To this end the UK Broadcasting Act of 1996 mandated that subtitling, audio description and signing be made available with UK Digital Terrestrial Television (DTT) services.

The development of DVB subtitling in the UK was described at IBC '99 [2]. This paper describes how audio description has been implemented in the UK and outlines some of the challenges which must be overcome in the future provision of closed signing services.

Whilst DTV allows closed access services to be delivered as additional service components, the overall bitrate available on each multiplex places constraints on the practical bitrate available for these ancillary services, particularly on DTT. Furthermore the handling of ancillary signals within the studio infrastructure always needs to be simple, straightforward and cost-effective. Some ingenuity has therefore been necessary to provide an end-to-end solution which supports the user requirements and delivers an appropriate service in a simple and bit-rate frugal manner.

AUDIO DESCRIPTION

The User Requirements

Audio description (AD) conveys salient contextual information about the scene or the action. As such AD is intended to add to the programme narrative such detail as is only conveyed by picture detail and visual clues. Description passages are therefore best confined to gaps in the speech. One type of programme (e.g. a "soap") may leave little opportunity (and arguably less need) for description than another (e.g. a documentary with strong uncommented visual content). As the programme sound level varies throughout a programme it is also desirable to be able to change the relative level of programme sound during description passages as heard by the AD user, reducing the programme sound more during loud passages (e.g. during an action sequence such as a car chase).

A brief UK experiment during the early 1990s transmitted low bit-rate coded mono audio description conveyed as teletext data in the analogue PAL signal to a small number of special decoders. The experiment demonstrated the benefit of a closed audio description service for the visually impaired but also showed that the service provider should have dynamic control of the main programme sound during description passages whilst the user should control the volume of the description and of the overall mix (see figure 1).

The Transmitted Signal

To satisfy these requirements for UK DTT, the transmitted AD signal conveys fade and pan information embedded in **PES_private_data** within the coded AD stream [3]. The mono AD is coded as MPEG-1 layer II; in practice 64 kbits/s produces a suitable match between the voice description and the stereo programme sound which has usually been coded at 256 kbits/s.

The 1-byte fade value tells the AD decoder how much to fade the programme sound before adding the AD. Between description passages the fade value is 0, short descriptions over quiet programme sound may require no fade whilst for longer passages and/or over loud programme sound the fade may be increased in a controlled manner, returning to zero at the passage end. The fade value can change from PES packet to PES packet to allow fades to be shaped.

To allow the programme provider to position the description within the resulting stereo image a pan value is also transmitted. For programmes with stereo sound the position of the describer will be restricted to $\pm$ 30° but the mechanism does allow for positioning anywhere within a multi-channel sound presentation.

The Studio Signal

The basic requirements for the studio signal are :

- to convey mono description signal plus 2 bytes (fade and pan) ten times a second
- to survive existing infrastructure
  - recording (DAT, Mini-Disc, digiBeta etc.)
  - audio processing (sync., rate-change etc.)
  - MPEG-2 distribution network [4]
  - Gain change and inversion
  - A-D and D-A conversion
- to be simple to code and decode
- to be simple and cost-effective to monitor

The mono audio description is conveyed directly as the left-hand channel of a standard AES-3 signal. Fade and pan information are conveyed in the right-hand channel as multi-level samples of a shaped bi-phase encoded signal. Simple bi-phase (Manchester) coding is used (data "1" represented by a bit-cell with a "1" > "0" transition in the centre of the bit-cell and a data "0" by a "0" > "1" transition). The signals transitions are suitably shaped to restrict sideband energy.

The multi-level samples are 16-bit values, symmetrical about the 2's complement zero of the AES-3 signal. The amplitude used is chosen pragmatically. Very small amplitudes (1 lsb) can be used if the signal remains in the digital (AES-3) domain and if it experiences no non-linear processing such as severe MPEG coding-loss. On the other hand a somewhat higher level is desirable if the signal is to survive analogue processing. In practice we have used a signal amplitude of –36 dBfs (`0x0200` to `0xFE00`).

The data encoded comprises the first 14 bytes of the AD Descriptor as described by the D-Book [3] terminated by a 16-bit cyclic redundancy check (crc.) performed over the first 14 bytes.

```
0xF8          reserved & descriptor length
0x4454474144  Text tag
0x31          version number "1"
0xYY          AD_fade_byte
0xYY          AD_pan_byte
0xFFFFFFFFFF  reserved
0xXXXX        crc. value
```

This data structure is serialised most-significant-bit first before bi-phase coding. The bit-rate for the data signalling is 1.28 kbits/s; thus there are 37 or 38 samples for each bit-cell for 48 kHz sampling, 34 or 35 for 44.1 kHz sampling and 25 for 32 KHz sampling.

The signalling is asynchronous – this is of little consequence in this application because the whole data structure is conveyed 10 times a second. In practice fades should commence before the description starts and return to full programme sound after it finishes; closer synchronisation to the modulated mono channel is thus unnecessary.

Insert edits will corrupt a single signalling structure but the effect can be entirely mitigated by placing a requirement on any downstream AD equipment to maintain the fade and pan information for the previous (crc correct) data structure if it detects a single crc error. Multiple errors are usually handled by muting the description and ramping the fade and pan values to zero over several signalling periods.

Whilst this is not an efficient use of AES-3 bits experiments have shown the studio transport mechanism to be entirely resilient when cascading many of the systems noted in the requirements above.

Practical Implementation

In the initial absence of off-the-shelf equipment BBC R&D designed prototype equipment to author, monitor and handle the AD signal in the studio. This has allowed various issues to be resolved and signal ruggedness to be confirmed prior to the BBC and other UK DTT providers specifying commercial equipment including encoders. Having already specified the transmitted signal and generic decoder behaviour [3] a small group from The Digital Network (TDN) representing UK DTT operators have specified and commissioned the commercial development of AD decoder modules which will plug into the common-interface socket of existing DTT set-top boxes and integrated receivers. When these modules become available the AD user will be able to listen to programme sound plus description whilst the rest of the family listen to programme sound. Meanwhile BBC R&D have developed an emulator which in conjunction with a professional ird allows monitoring of transmitted AD as it will be heard by the user. The BBC commenced experimental transmission of AD in this form in late Autumn 1999; given the complexity of the BBC distribution network [4] and the concurrent addition of significant regional programming to BBC services on DTT earlier this year this has allowed essential system integration work to be performed with low risk. AD services have been more generally available on DTT since mid-May 2000.

SIGNING

Sign Language in the UK

British Sign Language (BSL) is the first or preferred language of around 55,000 people in the UK many of whom have never had any useful hearing. BSL differs from English in grammar syntax and vocabulary, has many local variations and no written form. It can import words from English using a system of finger spelling, but finger spelling is more difficult to follow.

BSL is a language of hands, face & body; it incorporates a number of “articulators”, including both hands, the eyes, the lips, the head and the body posture. Image detail and motion rendition are therefore very important.

Sign language has developed almost completely as a live and interactive form of communication where the viewer has to watch the face of the signer [5]. The deaf audience has no choice but to continuously watch the image of the signer and cannot look away without missing part of the programme. Hearing viewers, on the other hand, are free to look away and follow the programme via the soundtrack.

These points have significant implications for the successful video coding of signing.

Open signing vs closed signing

Open signing, unlike open subtitling or even open audio description, is not comprehensible to most viewers. Increasing the proportion of programme material which are signed and providing increased sign-language access to peak-time viewing thus makes closed signing increasingly attractive. Until a practicable and acceptable method of closed signing is available however the BBC will continue to provide open signing with a proportion of its programmes on all its channels.

The requirements of a closed signing system

Any candidate closed signing method must deliver a credible service to the community it is to serve. It must be

- **practicable** - delivering a sensible consumer proposition, a user-acceptable quality of service and must be simple to originate, distribute, receive and use;
- **sustainable** – coping with an increasing proportion of signed programmes, with linguistic variation and changes in BSL and avoiding legacy problems for broadcaster and user and
- **affordable** – for users (cost of receiving equipment), broadcasters (cost of production & delivery) and multiplex operators (bit-rate frugal).

Desirable features include platform independence (or ease of interworking) if signing is made mandatory for all platforms and the ability to make use of technology with other possible entertainment applications .

The system should deliver legible signing images with minimal viewer-fatigue. The broadcaster needs a system with low capital and revenue costs, compatible with existing procedures and infrastructure and supporting live and pre-recorded operation. The system must also be independent of interpreter, linguistic variation and signing style and be reliable and straightforward to monitor. It is also desirable to have interoperability between open and closed signing.

The multiplex operator needs a system with bit-rate frugal coding which is able to support regionality and uses open standard signalling mechanisms.

The process model

Any closed signing system must encompass the entire broadcast chain from commissioning through playout to the viewer and be practicable throughout the signal path.

- **Scheduling** will be constrained by the suitability of a programme for signing and the availability of interpreters & technical facilities. It is also a function of the bit-rate required since the less required the more easy it will be to schedule signing for programmes popular with the deaf community
- **Capture** will be greatly simplified by the use of conventional television production processes. However, there is a need for additional information (positional etc.) to be stored and conveyed with the signing image if this is to be combined with programme video in the receiver.
- **Coding** requires maximum transparency of the signing image for minimum bitrate. Storage and signal distribution around the studio may also impact on legibility.
- **Storage & replay** media for closed signing should be compatible with existing production and post-production techniques.
- **Distribution** of a signing signal to and through regional sites will need to make use of pre-existing infrastructure and retain its logical and temporal relationship to the main programme components. For a broadcaster such as the BBC with regional and national programming the signing signal must be compatible with standard regional activities including time-shifting and opting.
- **Decoding & display** of the signing image at the home receiver must resolve issues of synchronisation, image precedence and

positioning if overlaying the signing image on video and possible subtitling.

- **Monitoring** provides a particular challenge with such an ancillary service, since at any stage the service component may be lost without most viewers noticing or for a signing stream the signing content may be incorrect and yet go unnoticed by operational staff.

The technical and operational challenges

Many of the main technical and operational challenges involved in providing a signing service are illustrated in figure 4. The issues shown on the left hand side apply to both open and closed signing services, whilst those on the right are largely issues associated with closed signing alone. All of these, and more, will have to be addressed in designing a system for a closed signing service.

The interconnecting lines indicate issue-interdependence. For example, any change to the coding system would require a change in the decoder design, whilst changes to the recording format will have an impact on the methods of editing signing and issues like time shifting at regional centres. No one issue can be tackled in isolation and the coding method is only a small part of the overall system design.

Methods of Coding Signing

To retain compatibility with standard broadcast practice and to sustain open signing where necessary it is now safe to assume that a closed signing system will be originated as video.

MPEG-2 is the most obvious choice of video coding as it is readily available in hardware form. However, tests at BBC R&D with a commercial MPEG-2 coder optimised for low bit rates have shown that even with carefully shot video of a signer with black, detail-free clothing against a completely clean background, coding a full height signing image may require at least 750kbits/s for legible signing.

MPEG-2 requires that the whole active picture be coded – much of this would be discarded in the final display and uses valuable bitrate resource. It is also difficult to recover a clean keying signal from the decoded video to enable it to be overlaid on the original video without keying noise.

MPEG-4 has been designed to allow parts of a scene to be transmitted separately allowing the viewer to chose between different display options. It therefore provides a more sophisticated tool set for coding certain types of image, of which a signer is one appropriate example.

Using MPEG-4 we can code non-rectangular images as with the image of the signer without the background At the same time we can define and transmit the shape of the signer for use in composing the combined image in the display. The use of a binary outline, rather than a graduated outline, save bit rate and simple edge-shaping in the receiver is shown to generate a clean key (see figure 5).

Using a software MPEG-4 coder and decoder, initial results have given excellent full-height full-resolution pictures at about 350kb/s and clear and legible pictures at around 280kb/s.

MPEG-4 is one of a number of possible coding mechanism being investigated. It remains to be seen what functionality (e.g. software decoding) will be available in future DTT decoders to facilitate providing a bit-rate frugal closed signing service with a quality of service acceptable to the user.

Ongoing work

Work continues to develop a closed-signing system for UK DTT. Having determined the requirements and identified many if not all of the challenges it remains to demonstrate a full end-to-end system which matches them.

REFERENCES

[1] Mornington-West, A. and Silver, J. Audio description, signing and subtitle services with a new font. Proceedings of the 1998 International Broadcasting Convention pp. 197-205.

[2] Tanton, N.E.& Weitzel, P. DVB Subtitling in an open environment. Proceedings of the 1999 International Broadcasting Convention pp. 320..326.

[3] UK DTG 1998. Requirements for Interoperability ("the D-Book").

[4] Tanton, N.E., Whiting, G.P., Marsden, R.P. & Jackson, J.D. The design and implementation of a system for the UK-wide distribution of BBC digital television service. Proceedings of the 1998 International Broadcasting Convention pp. 498..504.

[5] Collins J., Alker D. & Kyle J.. Content of the Sign Language Server. Centre for Deaf Studies, June 98. (http://www.bris.ac.uk/Depts/DeafStudies/SignWorks/d361.htm)

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues in various departments within the BBC for their help and encouragement in this ongoing work They would also like to thank the BBC for permission to publish this paper.

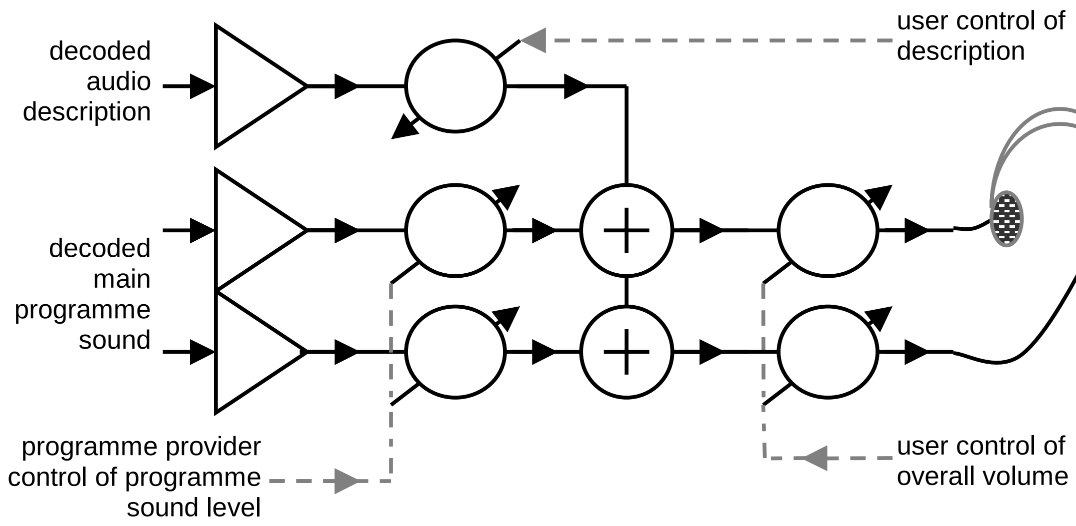


Figure 1 : Reference Model for Audio Description Control

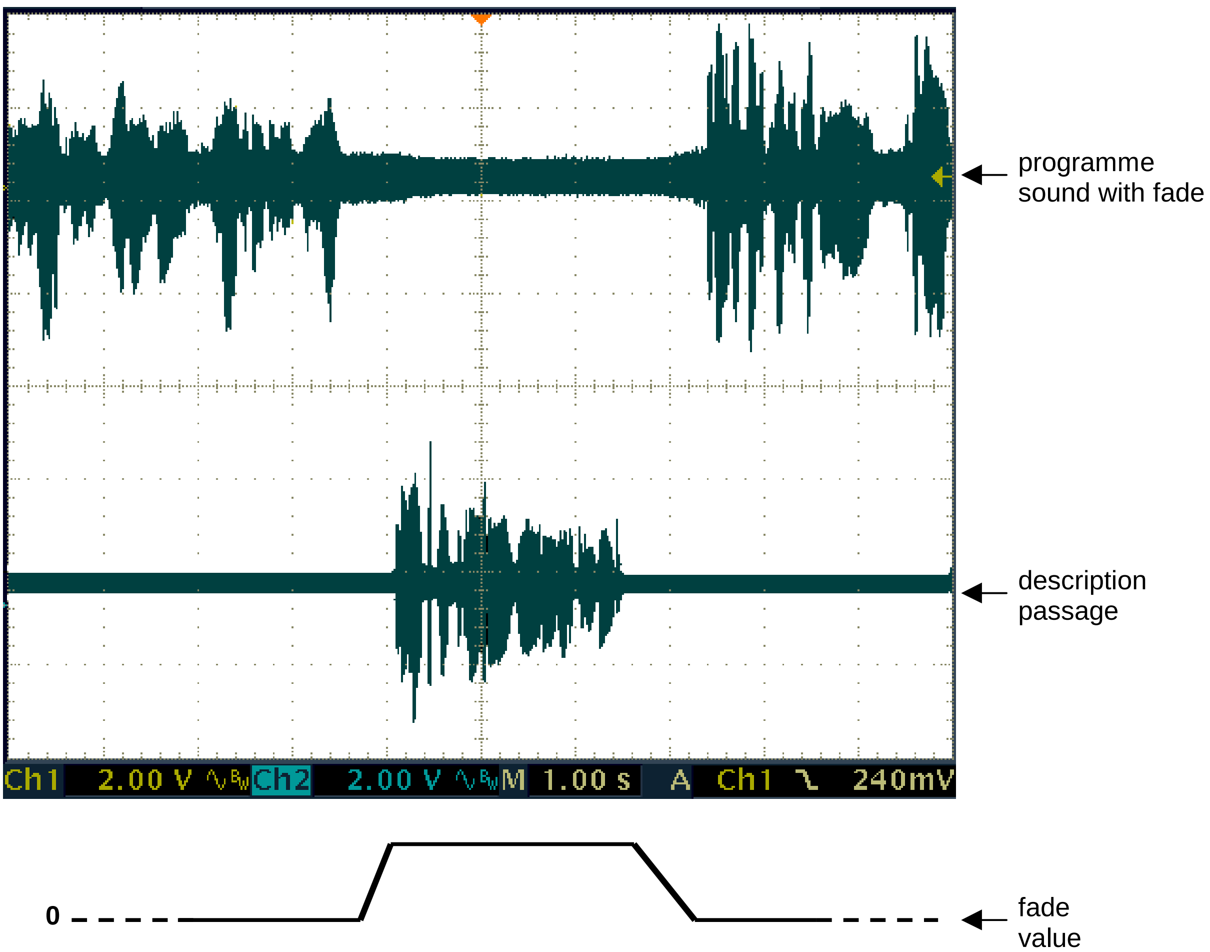


Figure 2 : Oscillograph of description passage showing decoded programme sound (with fade) and description sound prior to mixing

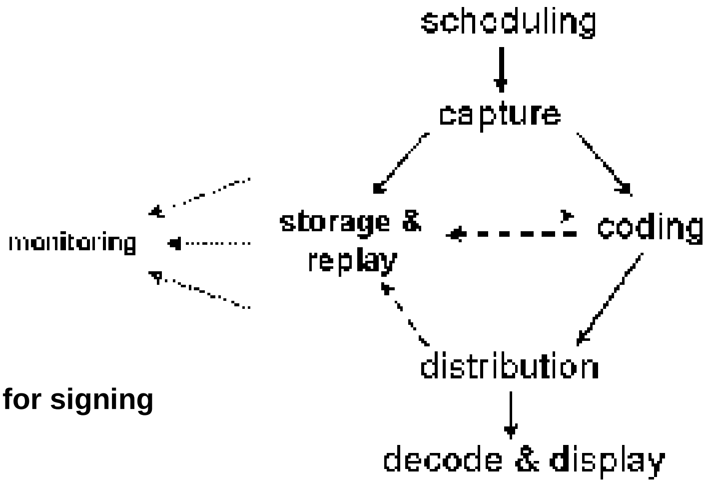


Figure 3 : Process model for signing

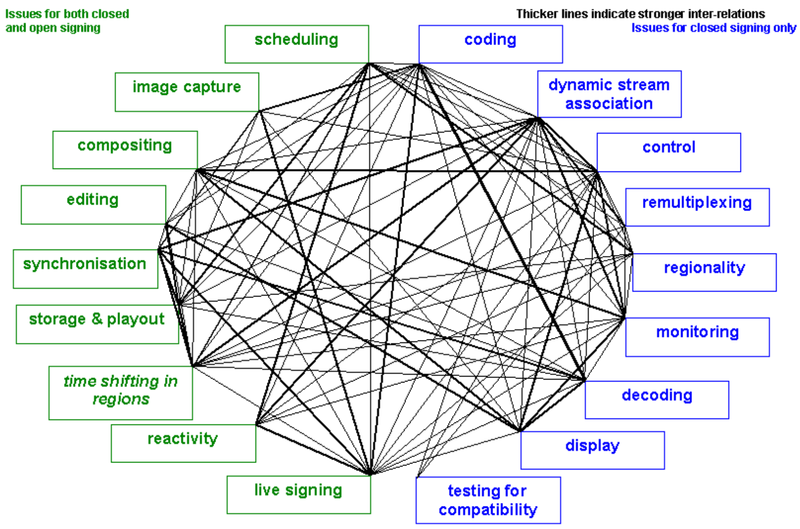
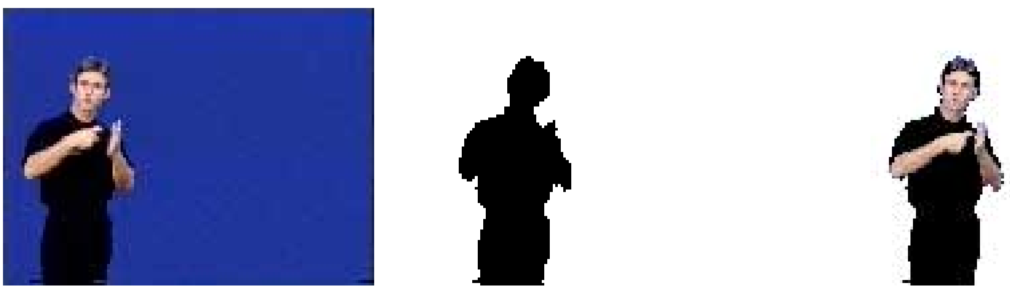


Figure 4: Signing Issues and their interelationships



Source image **Shape signal** **Coded object**

Figure 5 : Coding a signer as a visual object