Television For All

BBC Research & Development

EBU•UER

INTRODUCTION

2003 has been designated the European Year of Disabled People by the European Commission to raise awareness of the rights of disabled people to equality and full participation in everyday life.

Broadcasters have their part to play in providing enhancements that enable people with visual and hearing impairments to access existing broadcast services.

BBC R&D has taken a major rôle in developing technologies to make this happen and this suite of demonstrations illustrates three distinct ‘access services’.

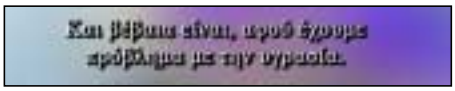
The first is subtitling, where digital broadcasting opens new opportunities for broadcasters to optimise the clarity, flexibility and usefulness of subtitles and where new tools are now available to simplify and reduce the cost of subtitle authoring.

The second is audio description, for which digital broadcasting allows additional audio tracks to be added to a programme to enable the visually impaired better to understand the action in a scene.

The third is the use of ‘signing’ to allow deaf sign-language users to understand the programme content. In a digital environment a signing component might in future be made a selectable addition to some programmes.


DVB SUBTITLES

1



2

3

1 An example of English subtitles.

2 Greek and other scripted languages can be subtitled more easily.

3 Using bit-mapped graphics means that special characters, unusual fonts, and new symbols can be made into subtitles.

Subtitles provide simple and straightforward access to television for the hard-of-hearing and for deaf people who can read. They also offer means of clarifying indistinct dialogue and providing multilingual explanation.

DVB subtitling, as defined in the ETSI standard EN 300 743 v1.2.1, allows television subtitles to be delivered more effectively than alternative methods.

Content of DVB subtitles

DVB subtitles support bit-mapped graphics. The most significant benefit of this is that the broadcaster can guarantee a consistent *look-and-feel* on all compliant receivers in terms of font, of letter-spacing, of position on the screen and of timing. Any font can be chosen; in the UK we use 'Tiresias' which has been designed by the Royal National Institute for the Blind specifically for its legibility on electronic display screens. The use of anti-aliassed characters and of proportional spacing further increases legibility.

By using bit-mapped graphics, more complex images can easily and effectively be created. Thus logos, script languages such as Hindi, Arabic and Hebrew or special characters such as mathematical symbols are all supported.

Other benefits of bit-mapped graphics include a reduction in receiver complexity and in receiver legacy issues because the graphics are created by the broadcaster, not in the receiver.

DVB subtitling supports live and prepared subtitles and provides a simple mechanism for maintaining accurate timing.

DVB subtitle implementation

Existing infrastructure can be used to deliver DVB subtitles. In this case the DVB subtitle encoder receives Teletext subtitles in the vertical blanking interval (vbi) and transcodes them to DVB subtitles. This has a number of benefits including:

- common authoring and playout when simulcasting on analogue and digital channels;
- straightforward use of archive material;
- use of existing subtitle authoring systems;
- ease of monitoring before coding.

Where transcoding from vbi teletext is used no further signal delay is involved; the relative timing of subtitles with vision/audio is managed by the MPEG-2 timestamp process.

Alternatively, subtitles can be delivered direct as data to the DVB subtitle encoder. This approach offers possible benefits in scalability for multiple channels and multiple languages and can allow simpler management of equipment redundancy.

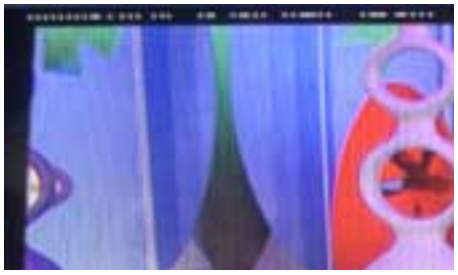
Bit-rate requirements for DVB subtitles

It is a common misconception that DVB subtitles must consume a lot of bit-rate. Whilst DVB subtitles typically require about 100 kb/s peak bit-rate per service, even when using anti-aliassed fonts and multiple colours the *average* bit-rate required will be much less (typically less than 10 or 15 kb/s for prepared subtitles). This compares very favourably with the 38 kb/s or so which in practice is needed to support delivery of subtitles via DVB teletext.

As DVB subtitles are delivered on a once and just-in-time basis unused bit-rate allocation can be harvested for use by opportunistic data – across a multiplex this may yield considerable benefit.

Decoding DVB subtitles

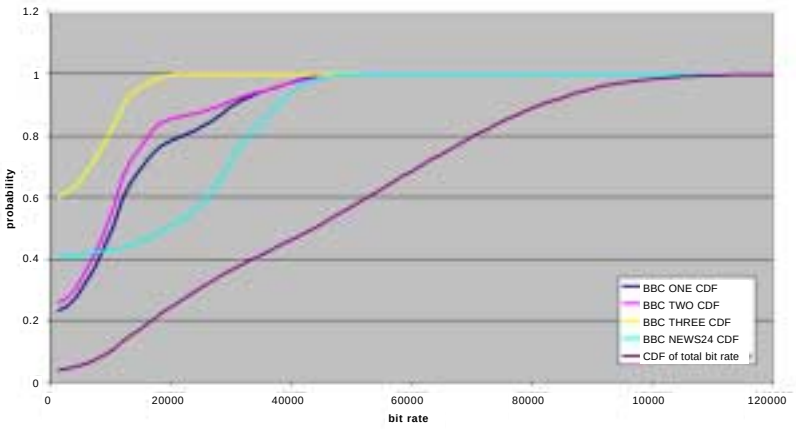
There is nothing particularly hard about decoding DVB subtitles. Most manufacturers of iDTVs and of set-top boxes are delivering products capable of decoding DVB subtitles: these include Hitachi, Nokia, Pace, Panasonic, Philips, Pioneer, Sony, Thomson, Toshiba, Daewoo, LG, Matsui, Hauppage, Netgem, Labgear, Lidcom, Nebula.



4

4 Teletext subtitles can be stripped from the appropriate line in the video blanking interval ready for transcoding into DVB subtitles.

Bit rate CDF, five seconds moving average. Wednesday 4 June, 24 hours.



5

5 Each of these services has been allocated 100 kb/s. They all use less than one third of their allocation averaged over any five seconds.

(CDF: cumulative probability).

NEW APPROACHES TO AUTHORING SUBTITLES

1

1 Creating live subtitles using speech recognition.

In the UK, as in many other countries, the proportion of television programmes that will be subtitled is set to rise significantly over the next five years. This trend can already be seen; for example, five years ago the BBC subtitled about 50% of its programmes but this has now reached 80%. Furthermore, this growth is set to continue: the BBC's commitment is to provide subtitles for all its programmes by 2008. Similar growth can be expected in the USA, where legislation requires that, by 2006, all new productions must be captioned.

This increasing demand for subtitling, exacerbated by the rise in the number of television channels, raises the question of how this can be achieved cost effectively. We clearly need to see how advances in technology can be exploited to give improved efficiencies so that what would otherwise be escalating costs, are kept under control.

What does subtitling entail?

There are two categories of subtitling: live and pre-recorded. For live subtitling, the subtitle text is created as the programme is broadcast, leaving little time, if any, for prior preparations. The majority of this type of subtitling is performed by skilled stenographers who use special keyboards, and, typing 'phonetically', can subtitle in real-time at up to 250 words per minute. An alternative to this is fast typing on a standard QWERTY keyboard, with two people working together to maintain the subtitling speed necessary.

In contrast, the preparation of pre-recorded subtitles is carried out in advance of the broadcast allowing the subtitler to pay careful attention to the layout of the subtitle text and how it matches the programme content. This is a labour intensive task requiring the subtitler to work through the programme, phrase by phrase, to create each subtitle in turn. This can take between 12 and 16 hours for each hour of programme. Furthermore, the different release formats, e.g. teletext, line-21, DVD, or as web-pages, have differing capabilities and this means that subtitles created for one frequently need to be re-authored to meet the requirements of the others.

How can new technology help?

Speech recognition is the obvious technology to turn to for help with this task but what can it offer? Unfortunately, the goal of using speech recognition to create subtitles directly from the broadcast soundtrack still remains some way off. Even with a speech recogniser trained for the task, the variety of speakers, acoustic conditions, music and sound effects that might be encountered gives highly variable results.

Despite these limitations, if applied appropriately, speech recognition technology can still offer extremely worthwhile advantages for both live and pre-recorded subtitling.

Live subtitling

Over the last year or two, speech recognition technology has offered an increasingly attractive alternative to the traditional approaches for live subtitling. For 'speech subtitling', the subtitler first carefully trains a speech recogniser with their voice; then, during the broadcast, listens to the programme on headphones and re-speaks the words into the recogniser. Because the recogniser only hears the voice to which it has been trained and the subtitler speaks clearly and in a quiet environment, results can be obtained which are comparable in accuracy to stenography.

At BBC R&D we have developed a live subtitling system that allows subtitles to be created either by using speech recognition or stenography. The design offers the particular benefit of allowing subtitlers to work on any broadcast service from anywhere in the BBC, or even from home. Any number of broadcast services can be catered for (the BBC has nearly 30, including its regions), and the routing is done automatically with its complexity hidden from the subtitler, who just chooses a service from a pop-up list of names.

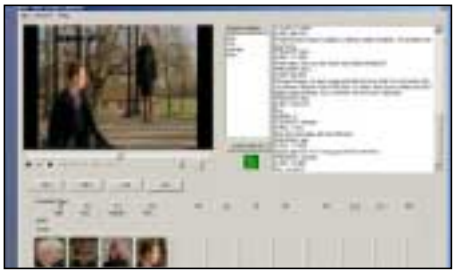
The deployment of this new live subtitling system throughout the BBC is now well advanced and it has already been used on-air for several months. With the versatility that the design offers, we hope to achieve valuable improvements in the efficiency of our live subtitling; these savings will be crucial in helping to achieve the BBC's aim of subtitling all its output by 2008.

Pre-recorded subtitling

For pre-recorded programmes, speech recognition can also be used to good effect albeit in an unusual way. We have developed a technique called Assisted Subtitling in which a speech recogniser is used to track through the script of a programme matching the written dialogue to the programme soundtrack.

Having the script available beforehand means that the speech recogniser no longer needs to work out what is spoken but instead automatically assigns a timing marker to each word in the script. Once this timing stage has been completed, the script is, in effect, synchronised to the programme soundtrack and from this, fully synchronised subtitles can be produced automatically.

2



3

2 Pre-recorded subtitles produced automatically using a speech recogniser to match the programme script to the dialogue.

3 Creating a programme script by re-speaking the content.

Further analysis of the script allows text colours, which are used to distinguish one speaker from another, to be assigned automatically whilst shot change detection ensures that the appearance of subtitles near abrupt picture changes can be avoided.

This system has been in use in the BBC for over three years and has proved to be very effective, saving 30 to 50% of the time taken to prepare subtitles for scripted programmes.

To extend the range of programmes that can benefit from this approach, we are developing methods of producing scripts for programmes, again using speech recognition. ScriptCapture uses an approach which is similar to that for live subtitling, but with the user re-speaking the content of the programme as they replay it from a video server.

The text from their speech recogniser becomes the script and this can be processed subsequently, using the assisted subtitling approach, to create the subtitles. Being able to produce a script reasonably quickly in this way significantly increases the scope for efficiency saving in the pre-recorded subtitling process.

AUDIO DESCRIPTION

What is audio description ?

For understandable reasons most television programmes rely on visual content to tell their story. 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 with visual impairments.

Description is typically confined to gaps in the normal programme narrative. Opportunities to describe a scene are dependent on the programme genre and on the editing of the main programme sound. Some programmes are naturally more suited to description than others. News for example provides little opportunity for description and is generally self-documenting. Science and informative programmes, on the other hand, often have relatively long narrative gaps associated with purely visual content; these can provide ample opportunity to fully describe the concurrent visual images. Action drama and 'soaps' are edited more tightly and typically provide only brief windows in the dialogue allowing only concise descriptions; nevertheless the dramatic purpose is greatly enhanced by brief descriptions (e.g. 'she glances at him').

Digital television offers ample flexibility for the delivery of new services or service components and so provides an excellent opportunity for adding AD to appropriate television services.

BBC Research & Development have pioneered the technology for delivering audio description on UK Digital Terrestrial Television.

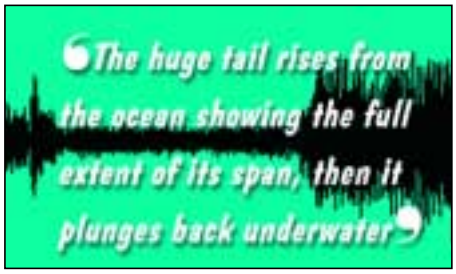
Audio description – what are the requirements ?

Since AD won't suit everyone, it is provided as a 'closed' system in which the user chooses to hear the additional description.

Gaps in narrative often include loud sound effects or music, which may make a description hard to discern. There is thus a need, on a description-by-description basis, to adjust the relative level of programme sound and description in the mix, which the AD user hears. The depth and timing of this fade is best determined by the describer under controlled conditions.

An experiment of audio description on analogue television found that the users wanted to be able to control the volume of the AD.

1



2

1 A scene from *Blue Planet* at which there is a gap in the programme narrative.

2 Audio description is used to describe the visual images.

3

4

5

3&4 The audio description module provides a method of receiving audio description while using any current model of DVB digital receiver, both terrestrial and satellite, that has a suitably implemented CI slot.

5 Separate outputs are available for headphones, hi-fi and VCR.

This is especially useful for the visually impaired user who also happens to have a hearing impairment.

Successive description passages may in practice be minutes apart. The AD user who has selected a described programme will therefore need to determine whether a temporary absence of description is intentional or is the result of the programme not being described or of a fault in scheduling, transmission or decoding.

The user interface needs to be suitable for users that have visual impairments.

For the broadcaster bit-rate is always at a premium. It thus makes sense to broadcast the voice-only description component in the most economical or bit-rate frugal fashion possible. The combined programme sound and description signal as heard by the AD user will however need the normal audio bandwidth appropriate accurately to convey the programme sound.

In summary the user requirements for AD are:

- a closed system;
- ability to adjust relative volume of description;
- ability to promptly determine that a programme is currently being described;
- bit-rate frugal delivery of the service;
- a delivery mechanism that uses existing and open standards (e.g. ISO/IEC 13818-x, DVB etc.)

Audio description for digital terrestrial television

The UK DTT multiplex operators represented by 'The Digital Network' (TDN) have commissioned the development of a PCMCIA style module which fulfils the requirements of an AD decoder and plugs into any DTT receiver which has a working 'Common Interface' (CI) socket. The Audio Description Module (ADM) has been designed and built by SCM Microsystems under contract to TDN with technical oversight and functional testing provided by BBC R&D.

The ADM demultiplexes and decodes the AD component of the selected programme and mixes it with the main audio from the receiver. Fade data that the broadcaster transmits with the AD component is used by the ADM to reduce the level of the programme sound during descriptions. Pan data is also transmitted, this can be used to position the AD in the stereo sound stage of the resulting mix. The ADM also manages the user interface. This interface uses a separate and simple infra-red remote controller and external IR receiver pod which fixes to the top of the receiver.

The design was initiated at a time when UK DTT carried some encrypted services and commercial sensitivity precluded any decryption in the ADM. This resulted in a design in which the programme sound input to the module is the analogue audio from the receiver's phono sound output. The audio description sound has always been broadcast un-encrypted.

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

Audio Description – the future

The audio description module provides a method of receiving AD while using any current model of DVB digital receiver, both terrestrial and satellite, that has a suitably implemented common interface slot. Upcoming digital receivers will have the functionality of the ADM embedded within them. This should reduce both the cost and the complexity for the consumer.

Audio Description Demonstration

The demonstration shows an Audio Description Module working with a Nokia 221T receiver with a signal from a local DTT multiplex. Four genres of programmes are shown each with audio description.

SIGN LANGUAGE AND DIGITAL TELEVISION

1

2

1&2 Examples of programmes with signing.

What is sign language ?

Sign languages convey content and meaning through movement of the signer's face, hands and body. They are the natural form of communication between people who cannot hear and often represent the primary language of the deaf community.

Sign languages are characteristically very different from the local spoken languages and do not have established systems of notation.

Why use sign language on TV ?

As it is difficult to learn a written language that you cannot hear, those born deaf are often less fluent in reading their local written language. This means that subtitles cannot sufficiently convey the content and meaning of a TV programme for this section of the viewing audience. Sign language thus becomes of vital importance to the deaf viewer.

One approach is to make TV programmes in sign language specifically targeted at the deaf community; most broadcasters will however only make few such programmes, typically weekly and in a magazine format.

The more inclusive approach is to add sign language interpretation to programmes made for a general audience. The image of a sign language interpreter is added to the original programme, usually with the programme video reduced in size so that the interpreter obscures less of the main picture. The interpreter will also convey subtleties such as tone-of-voice and provide cultural translation so as to present the full message targetted at the deaf audience.

Why closed signing ?

Because sign language is not comprehended by the majority of TV viewers and because it occludes part of the video image it is not liked by general viewers. So programmes re-versioned with sign language are often broadcast at unsocial times.

Thus a 'closed' or selectable system is needed where the image of the sign language interpreter may be added to the main programme in the viewer's receiver. This would enable broadcasters to provide

sign language interpretation even during peak viewing times with no impact on general viewers.

Digital television offers the flexibility to deliver new services and service components and so again provides a suitable opportunity to add sign language to television services.

BBC Research & Development have been investigating technologies to create an end-to-end solution for the provision of a closed sign language service on UK Digital Television.

What are the requirements for a closed system ?

Any credible candidate method of closed signing must be:

- practicable
  - deliver a sensible consumer proposition;
  - provide a user-acceptable quality of service;
  - be simple to originate, distribute, receive;
  - work for both live and pre-recorded programmes;
- sustainable
  - cope with an increasing proportion of signed programmes;
  - handle linguistic changes in the target sign language;
  - avoid legacy problems (for broadcaster and user);
- affordable
  - for users (cost of receiver);
  - for broadcasters (cost of production and delivery);
  - for multiplex operators (bit-rate frugal).

Another desirable feature would be platform-independence.

Possible delivery mechanisms for closed signing

The practicable options for delivering sign language to a digital receiver fall into two categories, (i) as a coded video image of the sign language interpreter, or (ii) as motion information to drive an avatar which is then rendered locally to a video image. In both cases it has to be remembered that the viewer who is following the sign language interpretation has to watch the signer continually, whereas the hearing viewer can sometimes look away and continue to follow the soundtrack. This means that the perceived quality of the signing image presented to the viewer needs to be good enough for sustained critical viewing.

In both cases the broadcaster must generate, store and broadcast the signing signal alongside the other programme components using, as far as possible, existing infrastructure.

Video coded delivery for closed signing

Video coding is directly compatible with the production of open signing where, typically, the interpreter stands in front of a chromakey background. Instead of adding the signer to the programme in the studio, the image of the signer would be sent as a separate closed component to be added to the programme in the receiver.

3

4

3&4 Using MPEG-4 video coding with shape signal, a signer can be coded separately from the main video image. The shape signal enables the receiver to key the signer over the main video.

The drawback is the bit-rates required to produce a reasonable quality image at TV resolution. These are typically:-

Open signing/signing channel	~ 4,500 kb/s
MPEG-2 signer only	~ 600 kb/s
MPEG-4 signer only with shape coding	~ 300 kb/s
H264 signer only	~ 180 kb/s

Note, these low bit rates are only achieved if the signer wears plain clothing against a clean digitally-generated background or separated from the background.

Motion capture for avatar delivery of closed signing

The development of a suitable motion capture studio suitable for the capture of sign language is a considerable technical challenge. The motion data must be captured so as to encapsulate the meaning of the language in a transparent manner. This includes the difficult problem of capturing and conveying subtleties of facial expression and eye gaze. The process of capture must be sustainable for perhaps up to an hour of continuous signing without discomfort to the signer. Low 'shooting ratios' are also important. Furthermore, the avatar in the receiver must be rendered in a manner which conveys the sign language to the viewer in an accurate and acceptable manner.

BBC R&D are following up the work on avatars and motion capture for signing by the Eu ViSiCAST project, to bring together practicable marker-based techniques for sustained live motion capture of signing and to develop techniques for image based motion capture of the face.

The transmission of signing by motion capture offers the prospect of significantly lower bit rates, e.g. 50 to 100 kb/s, which would enable broadcasters to add sign language components to all their services.

5

5 An example of signing using an avatar courtesy of the recent Eu ViSiCAST project.

Signing Demonstration

The signing demonstration shows examples of programmes made in sign language, and with open signing, followed by demonstrations of closed signing using MPEG-4 vision coding and of the Eu ViSiCAST avatar.

CONTACT DETAILS

BBC Research and Development
Kingswood Warren
Tadworth
Surrey
KT20 6NP

http://www.rd.bbc.co.uk/

email: info@rd.bbc.co.uk

ACKNOWLEDGEMENTS

The assistance of BBC Broadcast, David Allen, Screen Subtitling Systems Ltd., ITFC, 20th Century Fox, DICAS, Televirtual and other partners in the recent Eu ViSiCAST Project and of TDN in the preparation of these demonstrations is gratefully acknowledged.

BBC Research & Development
Kingswood Warren
Tadworth
Surrey
KT20 6NP

http://www.rd.bbc.co.uk/

email: info@rd.bbc.co.uk