

The production and delivery of DTV

Access services

Peter Olaf Looms

Danish Broadcasting Corporation (DR)

This is the second of three articles on television “access services” to be published in EBU Technical Review.

This installment explores some of the *whats* and *hows* – what do prepared subtitling, live subtitling, audio description, spoken subtitling and visual signing entail ... what kinds of issues emerge when offering access services ... and how can these production and delivery issues be addressed by broadcasters.

The previous article on e-Inclusion discussed the *whys* – the rationale of offering access services for television. It also provided an overview of the mature and emerging access services that should be considered at regional or national level to improve the e-Inclusion of digital television.

Some improvement to e-Inclusion can be achieved *without* access services, simply by being careful with the design of the video, graphic overlays and audio. This point was brought home forcibly during the 2010 Football World Championship in South Africa when a solution had to be found to cope with Vuvuzela trumpets to make the commentary intelligible.

In most cases, however, the goal of e-Inclusion will also require access services. This article explores some of the *whats* and *hows* – what prepared subtitling, live subtitling, audio description, spoken subtitling and visual signing entail, what kinds of issues emerge when offering access services, and how these production and delivery issues can be addressed by broadcasters.

The term “broadcasters” is used here quite deliberately, as broadcasters are at various stages in the development of their work with access services. The experience, means and regulatory circumstances of a major public-service broadcaster like the BBC are very different from, say, RAI in Italy and a small public service broadcaster in Estonia. Rather than suggest quick fixes, the article builds on experience across Europe and identifies critical issues that will need to be addressed when working on the production and delivery of digital television access services.

Frans de Jong and EBU Project Group P/AS provided an excellent review of the access services situation for television in 2004 [1][2]. Their work covered both analogue and digital access services. As many countries in Europe will complete the transition to digital delivery by 2012, this article focuses on *digital* television only, the aim being to highlight challenges that still remain to be addressed, and those that have emerged since the 2004 report.

Subtitles

The impact of viewer expectations

Broadcasters who are about to introduce subtitling for the deaf and hard of hearing need to be aware of the prior knowledge and expectations that viewers have about subtitles. Both of these can have a big impact on the success of the proposed subtitling service.

Table 1
Overview of the production and delivery of digital television access services

Authoring	Exchange/Contribution	Delivery	Presentation on device
Subtitling			
Prepared subtitling Live subtitling	Tape VBI, DVD File-based	DVB subtitling (EN 300 743) DVB Teletext (EN 300 472)	DVB subtitling receiver DVB Teletext receiver
Audio/spoken subtitling			
Subtitles turned into synthesized speech	Synthesized speech with fade information	Pre-mixed DVB audio (broadcaster mix) DVB audio with fade info (receiver mix)	Any DVB receiver DVB receiver capable of mixing
Audio Description			
Script and fade information turned into additional audio	Description signal & control track synchronized with video (tape, file-based)	Pre-mixed DVB audio (broadcaster mix) DVB audio with fade info (receiver mix)	Any DVB receiver DVB receiver capable of mixing
Visual Signing			
Signer interprets speech and action in vision	Signer on tape, file-based or live	Signer superimposed Additional vision component Simulcast IP delivery of signer superimposed on signal	Digital receiver Digital receiver with mixing capability IP-capable receiver (HbbTV or IP device - broadband)

Source: Adaptation of summary in EBU Project Group P/AS report 144-2004

Firstly, in countries where viewers are already familiar with subtitles to translate foreign language programming, there is little social stigma associated with same-language subtitles. In dubbing countries, however, where subtitles are uncommon, subtitles are associated with hearing impairments. Until attitudes towards subtitling become more relaxed, potential users and their families may have to overcome prejudice and resistance in order to benefit from same-language subtitles.

Secondly, there is the challenge of creative subtitling that can help as many people as possible. Being as inclusive as possible depends on a number of factors and trade-offs, including the fidelity of the subtitles and viewer's reading speed.

Viewers cannot be expected to know how subtitles are produced. For this reason, many assume that subtitles are verbatim transcriptions of the soundtrack. They may be suspicious of linguistic condensation that "changes" the intended meaning, rather than accepting that editing the text can help bring the required reading speed down to an acceptable level. In user studies from Belgium, Denmark and the UK, viewers have asked for such verbatim transcriptions even though very few of the viewers in question would be able to read fast enough to benefit from them.

When introducing a subtitling service, it will be necessary to:

- listen carefully to the audience;
- set up some kind of consultation mechanism with the organizations that represent the deaf and hard of hearing;
- assess the impact of a given solution on the various groups of users concerned, and;
- be as transparent as possible about the production trade-offs made.

A useful introduction to viewer expectations of subtitles can be found in a study by OFCOM in the UK [3].

Scaling up – from pre-produced subtitles to subtitles for all TV programmes

Subtitles have been with use since 1929, two years after the emergence of films with sound. On digital television, we have two delivery mechanisms that have been widely implemented [4]:

- **DVB Teletext** (ETSI standard EN 300 472) which delivers subtitles via DVB bitstreams in a form that can be easily transcoded at the receiver and inserted into the Vertical Blanking Interval for display on TV sets capable of displaying Teletext;
- **DVB Subtitling** (EN 300 743) which offers a more sophisticated and flexible set of features than Teletext subtitling, whilst also being efficient in terms of required bitrate.

As tools and workflows for authoring inter-lingual subtitles have been with us for decades, one would think that same-language subtitles should not constitute a major challenge. Consequently, when it comes to prepared or pre-produced subtitles, there **are** no major technological problems. However, for those embarking on a subtitling service, thought needs to be given to production costs.

The cost of producing and “spotting” subtitles for pre-recorded television programmes remains largely unchanged since the 2004 report (usually in the range of 4 - 15 Euros per programme minute, depending on the language, the programme genre and country in question). It is also relatively easy to calculate the annual production costs for this kind of television programming using the template provided. For general TV channels of major broadcasters, the additional production costs of subtitles are less than 1% of the production budget for the programme itself.

But for niche channels using archive and third-party programming, or for broadcasters in small countries, the figure may be as much as 25-30% of the channel’s production budget. In such cases, the cost of offering subtitles for the deaf and hard of hearing cannot be readily absorbed in the production budget, and the business model for subtitling needs to be discussed in depth.

The challenge does not stop there. For many television broadcasters, live programming accounts for an increasing proportion of overall output.

When the production of subtitling goes beyond pre-recorded television programmes into the realm of live broadcasts, additional challenges emerge. Three of these are:

- Finding viable solutions for the authoring of subtitles for live television programmes;
- Mitigating viewer problems caused by the delay between the dialogue in a programme and the subtitles appearing on the screen.
- The higher costs of live subtitling compared with prepared subtitling (live subtitling is reported to cost 15% more per programme minute than prepared subtitles).

Viable production solutions for subtitling live programmes

Stenography or keyboard-based approaches to live subtitling are costly in terms of training. Since the 2004 report, there has been a marked swing away from these solutions towards the use of re-speaking systems for producing live subtitles. Re-speaking involves a human subtitler who watches and listens to the programme and then dictates the subtitles with the necessary linguistic compression of the dialogue. Voice recognition is used to turn the re-spoken input into subtitles that can be displayed in the usual way.

Vendor concentration in the field of re-speaking systems has become noticeable as IBM and Philips sold their solutions to the dominant player, Nuance. While there are open-source re-speaking solutions in use, there is little or no choice when it comes to the less-widely spoken European languages. Whereas there is a healthy market for re-speaking systems for professional use (e.g. to reduce administrative staffing levels in hospitals by getting doctors to dictate their findings into patient journals), general-purpose re-speaking systems are both more difficult to develop and have fewer clients who need to pay for the initial capital costs of setting up the language database and the dictionary of items used for speech-to-text generation.

It is unrealistic to expect a vendor to develop a suite of re-speaking systems for European languages in the absence of a reasonable return on investment. Yet broadcasters cannot be held to account in terms of production targets moving towards 100% subtitling without re-speaking tools that allow them to do the job at a reasonable price. Concerted political effort both nationally and at European level will be required to break this logjam.

The delay between the dialogue in a programme and the subtitles appearing on the screen

Until 2008, the debate on re-speaking focused on the errors that were created by speech-recognition and the extent to which this influences the viewer's appreciation of television. In 2008, anecdotal evidence emerged indicating that the delay between the dialogue in a programme and the subtitles appearing on the screen also caused viewers problems.

The production of live subtitles has a *human* component and a *technological* one. The subtitler handling live subtitling has to do some or all of the following tasks:

- Listening to what is said;
- Deciding whether to repeat verbatim (state funeral – slow delivery rate – infrequent comments) or edit what is heard (live political debate – rapid speech);
- Editing the text where necessary (condensing information to keep up with the speakers – on the one hand, faithfully reflecting the content but, on the other, not making excessive demands of the viewers' reading abilities);
- Dictating the resulting text into a microphone linked to a speech recognition system);
- Time and system permitting, correcting any errors;
- Inserting the correct punctuation;
- Adding colour codes to clarify who is speaking;
- Positioning the subtitles so as not to obscure interesting on-screen information;
- Presenting subtitles in a way that the viewer always has time to read them.

These processes will inevitably create a delay. Depending on the programme genre, the production method and the number of subtitlers, there will be a delay of at least 3 seconds.

The production and transmission infrastructure also has latencies that are caused by:

- The nature of the signal feed that the operator is using (is he/she watching an internal feed, a feed delivered by some kind of contribution system, or actually watching the transmitted signal?);
- The encoding of the voice input;
- The language model used to generate text;
- The presentation module to format subtitles (either scrolling or block);
- The presentation – editing and positioning module (where to position, use of colours, adding punctuation);

Abbreviations

AD	Audio Description	HbbTV	Hybrid Broadband/Broadcast Television
DVB	Digital Video Broadcasting http://www.dvb.org/	PVR	Personal Video Recorder
ETSI	European Telecommunication Standards Institute http://pda.etsi.org/pda/queryform.asp	RNIB	Royal National Institute for the Blind (UK)
		RNID	Royal National Institute for the Deaf (UK)
		VBI	Vertical Blanking Interval

- The feed from the presentation module to the encoder for transmission.

In the best of cases, the production infrastructure up to the DVB encoder will introduce latencies of about 1 second, so overall the total is usually at least 4 seconds and in some cases is nearer 7 seconds (Denmark) or 14 seconds (Spain).

The issue came up in a study by the UK's Royal National Institute for the Deaf (RNID) that was conducted in 2009 and published in 2010 [5]:

“Although [the study] was focused generally on TV access, participants identified subtitling as the main issue they wanted the RNID to campaign on. Almost 80% of the participants had had problems with subtitles and more than half had to stop watching a programme as a result. The two main problems were the delay of subtitles with regard to the audio (25%) and their inaccuracy (17%), identified as more important factors than having no subtitles available (7%). In other words, viewers seem to prioritise now quality over quantity and, judging by their main concerns (delay and accuracy), it is the quality of live subtitling they are most worried about.”

Given that the RNID survey had a predominance of participants with a university education (and thus above-average reading skills) one can safely say that the true level of dissatisfaction among those with hearing impairments may well be higher than the figures quoted.

Additional light was shed on the impact of the delay in a study conducted in 2009 by DR [6]. Although the Danish study was exploratory only, the number of people who reported having problems with live subtitles was about two-thirds of the sample. Many subjects had to resort to either reading only, or (in the case of those with moderate hearing impairments) to adjusting their hearing aids or wireless TV audio connection – they chose to listen without recourse to the subtitles.

Nearly all those who found it difficult or impossible to follow live subtitles did not comment on their inaccuracy as they simply did not read them.

The delay in the presentation of live subtitles is a universal problem. There are two main options for doing something about it:

- Centrally introducing a delay to the video and audio – but not the subtitles – when playing out a programme with live subtitles in order to “re-synchronize” them.
- Decentrally introducing a delay in the receiver and/or personal video recorder (PVR) to re-synchronize the live subtitles.

There are examples in the Netherlands where the first option is used, i.e. introducing delays of 10 seconds (news), or of more than 1 minute (a live debate programme), to give subtitlers time to improve the quality of subtitles. Generally speaking, however, there may be logistic or even legal reasons why introducing a central delay of the video and audio is impracticable.

The second option – of introducing a delay decentrally in the digital receiver or personal video recorder – can be implemented in several ways. All of the options require some kind of solid-state buffer so that the various television assets (video, audio, live subtitles, prepared subtitles) can be stored for up to 15 seconds before being presented in a re-synchronized form on a television screen.

The DVB standard makes use of presentation time stamps. Each asset has a time stamp introduced in the encoder that allows the receiver to assure that the assets are synchronous when the programme is viewed. Changing the presentation time stamp for live subtitles is not feasible. An alternative involves the flagging of a programme expected to contain live subtitles, flagging of the live subtitles themselves and an indication of the required offset.

In a receiver or PVR that is capable of this kind of decentral buffering, an application would respond to the “programme with live subtitling” flag. There would be an initial freezing of the programme for the desired offset after which the programme would be shown with the live and pre-prepared subtitles displayed synchronously. Freezing the visual component of a programme is a considerable design challenge. Any delay of more than 2 seconds is known to create viewer dissatisfaction. On the other hand, the only viewers using this feature would be individuals for whom watching a pro-

gramme with live subtitles is critical to their being able to follow the programme. Those without this need would not be aware of the problem.

During sequences with conventional, prepared subtitles, the presentation time stamps would not be affected, and these sequences would be shown normally.

During live sequences with live subtitles, however, the application would subtract the offset from the time-stamped subtitles. (In the case of DR, it would show them 7 seconds earlier).

At the end of the programme, the buffer would have to mask the fact that 7 seconds would be lost as the receiver switched from displaying via the buffer to normal reception. The feasibility of this approach has not been assessed.

The need for a buffer in television receivers may become more widespread with the emergence of hybrid broadcast/broadband displays using the BBC's *YouView* (formerly "CANVAS") or the ETSI HbbTV standard. Where assets are delivered to the same display device by two different distribution mechanisms (broadcast and broadband), it is likely that there will be disparities in the availability of assets that need to be shown together. At some point, having a buffer would allow for the resynchronization of assets not only for live subtitles but also for programmes delivered using such hybrid infrastructures.

Higher costs of live subtitling

Live subtitling requires that the necessary software for speech recognition is available for the language in question. There are a number of European languages where off-the-shelf solutions do not currently exist. In such cases, a significant investment in producing the language model and speech recognition modules will be required.

For languages where re-speaking can be outsourced, live subtitling is reported to cost 15% more per programme minute than prepared subtitles. Where production volumes are low, this may reflect the additional costs of maintaining the re-speaking solution itself. An additional source of costs is "Doing things twice" – the economic implications of having to re-do live subtitles to accompany repeats. For some television genres such as debates, the re-speaking subtitles can be regarded as a "best effort" solution, given the time constraints of an unscripted live broadcast. Such subtitles can be improved and inevitable semantic errors corrected prior to their re-use when the programme is broadcast again.

What this section on subtitles shows is that live subtitling of programmes for the deaf and hard of hearing needs to be regarded as "work in progress". Although pre-produced subtitles can be a cost-effective way of addressing viewers' hearing impairments, much remains to be done in order to meet the target of 100% subtitling. As we move from live broadcast to multiple distribution modes, service definition for subtitles requires us to find production solutions that are independent of distribution infrastructures.

Audio Description (AD)

In the EBU report from 2004 mentioned earlier, most of the AD services operating in Europe at that time used the *broadcast-mix* to deliver an additional stereo channel containing the audio description.

Since then,

- Audio description has been introduced in additional European territories;
- Target service levels in countries already offering AD have increased;
- The move from stereo to multichannel audio, as part of High Definition (HD) broadcasts, has begun;

- There is less bandwidth available on broadcast networks for access services since the introduction of HD;
- Digital receivers that are capable of mixing a mono AD channel with the existing audio have become quite widespread in many European countries as a consequence of the move to **receiver-mix** in the UK;
- Broadcasters need to think about AD services in connection with Personal Video Recorders, “catch-up” TV and other on-demand services.

These changes, when viewed together, constitute incentives for the broadcaster to move from broadcast- to receiver-mix on broadcast television.

What we are seeing at broadcasters who are offering AD are workflows that increasingly lead to mono file(s) with the necessary mixing metadata for fades. Depending on the delivery network, the service can be formatted to meet the requirements of that specific network.

The move to producing mono for receiver-mix rather than broadcast-mix saves money, if care is taken with production metadata that handles fades throughout the production and delivery systems to the point where mixing takes place. It also offers improved flexibility as we move from stereo to multichannel audio to accompany HD broadcasts.

The major challenges have less to do with the production of AD than conformance issues at the receiver end, especially in free-to-air digital terrestrial broadcasting. The experience of introducing AD using alternative audio or defining AD as a virtual channel around Europe highlights the need for testing the chosen solution in collaboration with the platform operator, the consumer electronics manufacturers and the bodies responsible for informing the general public about which receivers are known to work and which may not be suitable.

As with subtitling, metadata about the presence of AD will be critical for users dependent on access services. If the availability of AD is not mentioned in printed programme guides and EPGs, to all intents and purposes the service does not exist.

Audio (spoken) subtitling

In 2004, spoken subtitling was offered only in the Netherlands and Sweden. Since then, solutions for digital television have emerged that build upon receiver mix capabilities in digital receivers.

The main need addressed by spoken subtitling is foreign language programming normally offered with inter-lingual subtitles. Having subtitles read aloud using speech synthesis either centrally at the broadcaster or decentrally in the television receiver allows for two very different user propositions. Central solutions confined themselves to television programmes and the service is produced using a speech synthesizer per channel. Having speech synthesis on the receiver holds the promise of offering aural interfaces for the blind and those with serious visual impairments and the capability of “transcoding” DVB subtitles on any channel that can be received.

Figures from YLE in Finland suggest that the capital costs per channel and the operating costs of spoken subtitling are well within the means of most broadcasters. To get started, the broadcaster needs a server and software that can generate a mono audio track for selected programmes using the pre-produced subtitles and has the necessary bandwidth (64 kbit/s) in the multiplex. The initial start-up costs are approximately EUR 100,000 per channel and annual operating costs including depreciation are less than EUR 20,000.

The decentral solution for spoken television interfaces has been championed by the Royal Institute for the Blind (RNIB) in the UK who have had a working prototype operational since 2008. Ideally, if a fully DVB and D-Book version could be launched, this would be a useful mechanism for helping both the blind to navigate TV and for helping those with visual and reading impairments to understand foreign programming with subtitling through speech synthesis in the national language.

Visual signing

At the time of the 2004 report, visual signing for the deaf invariably took the form of a window containing the signer overlaid on the conventional picture with or without a border, as can be seen in *Fig. 1*. Here, the signer occupies more than half of the screen height and her movements are clearly visible.



Figure 1
Visual signing of a children's programme

Source: Red Bee Media



Figure 2
Visual signing on DR2 (Denmark)

Source: DR

Unfortunately, the use of *open* visual signing is subject to complaints from viewers with normal hearing (with the possible exception of RTP's use of visual signing of early evening programming in Portugal). While there is a clear will to offer visual signing, neither advertising-funded nor licence-funded broadcasters want to alienate parts of their audience by forcing them to watch visual signing if this has direct, quantifiable implications for their market share.

To this end, various approaches have been tried to make visual signing an opt-in service. One of these is visual signing as a temporary channel with a signer – full-screen – standing to one side of a flat panel display showing the channel as shown in *Fig. 2*.

In Denmark, there is a virtual channel shared by three channels from two broadcasters who coordinate their transmission schedules from late afternoon to early evening so that there is never more than one programme with visual signing being offered at a time. It is also bandwidth-intensive, and requires statistical multiplexing to “shoe-horn” the additional 3.5 - 4 Mbit/s into the multiplex from 16:00 to 19:30 hours, when programming with visual signing is offered.

The challenge of this solution is that it only makes sense in a Standard Definition universe where competing channels can agree to share a virtual channel that results in a drop in picture quality for the other channels in the multiplex while the visual signing channel is operating. When the Danish multiplex in question changes to MPEG-4 encoding in 2012, with the general trend towards HD, a new solution will need to be found.

Alternative, opt-in solutions that are being validated include:

- A quarter screen or widget stream with the signer delivered via the Internet (or in the same multiplex) which is overlaid on the programme that requires visual signing.
- An Internet simulcast of the programme with the visual signing added that can be viewed by viewers with a good broadband connection on a flat-panel TV with an Internet input (which could also include receivers complying with the ETSI HbbTV standard).

One of the challenges facing visual signing in particular is the propensity of regulators to determine not only the nature and level of the service, but also the delivery mechanism to be used. While this has the advantage of assuring that the solutions agreed upon are available to nearly all their potential users, it has the inadvertent drawback of being excessively conservative.



Peter Olaf Looms is a senior consultant in the commissioning body of DR in Denmark. For the last few years he has been working on access services for digital television so that as many people as possible can get the full benefit from watching television. Much of this work is related to the EU project called DTV4ALL.

In addition to his work at DR, Mr Looms teaches digital media on postgraduate programmes at several universities.

As a result of service definition that also prescribes the delivery network, broadcasting continues to be used as narrowcasting to deliver access services to thousands or tens of thousands of users. This prevents the use of broadband and hybrid broadcast/broadband devices that are more cost-effective. It would be in the public interest to be more flexible in terms of access services and explore the potential for using savings in distribution costs to increase service levels.

In the third and final article, a number of these over-arching themes to do with access services will be re-examined and suggestions will be made to improve efficiency and effectiveness without having to make dramatic increases in production budgets.

References

- [1] Frans de Jong (EBU Project Group P/AS): **Access Services for Digital Television**
EBU Technical Review, October 2004.
- [2] EBU Technical - Information I44-2004: **EBU report on Access Services** [includes recommendations] 2004.
- [3] OFCOM: **Subtitling – An Issue of Speed?**
6 January, 2005
<http://stakeholders.ofcom.org.uk/binaries/research/tv-research/subt.pdf>
- [4] EBU Recommendation R-110-2004: **Subtitling for Digital Television Broadcasting**
- [5] RNID (2010): **A summary of the TV access survey results**
http://www.rnid.org.uk/howyoucanhelp/join_rnid/_member_community/volunteering_campaigning/volunteering_campaigning_news/tvaccessresults.htm#results
- [6] Rander, Anni and Peter Olaf Looms: **The accessibility of television news with live subtitling on digital television**
Pages 155-160. Proceedings of the 8th international interactive conference on Interactive TV & Video, Tampere, Finland June 09 - 11, 2010. Association for Computing Machinery, Inc. NY, USA. ISBN 978-1-60558-831-5.

This version: 27 September 2010

Published by the European Broadcasting Union, Geneva, Switzerland

ISSN: 1609-1469

Editeur Responsable: Lieven Vermaele

Editor: Mike Meyer

E-mail: tech@ebu.ch



**The responsibility for views expressed in this article
rests solely with the author**