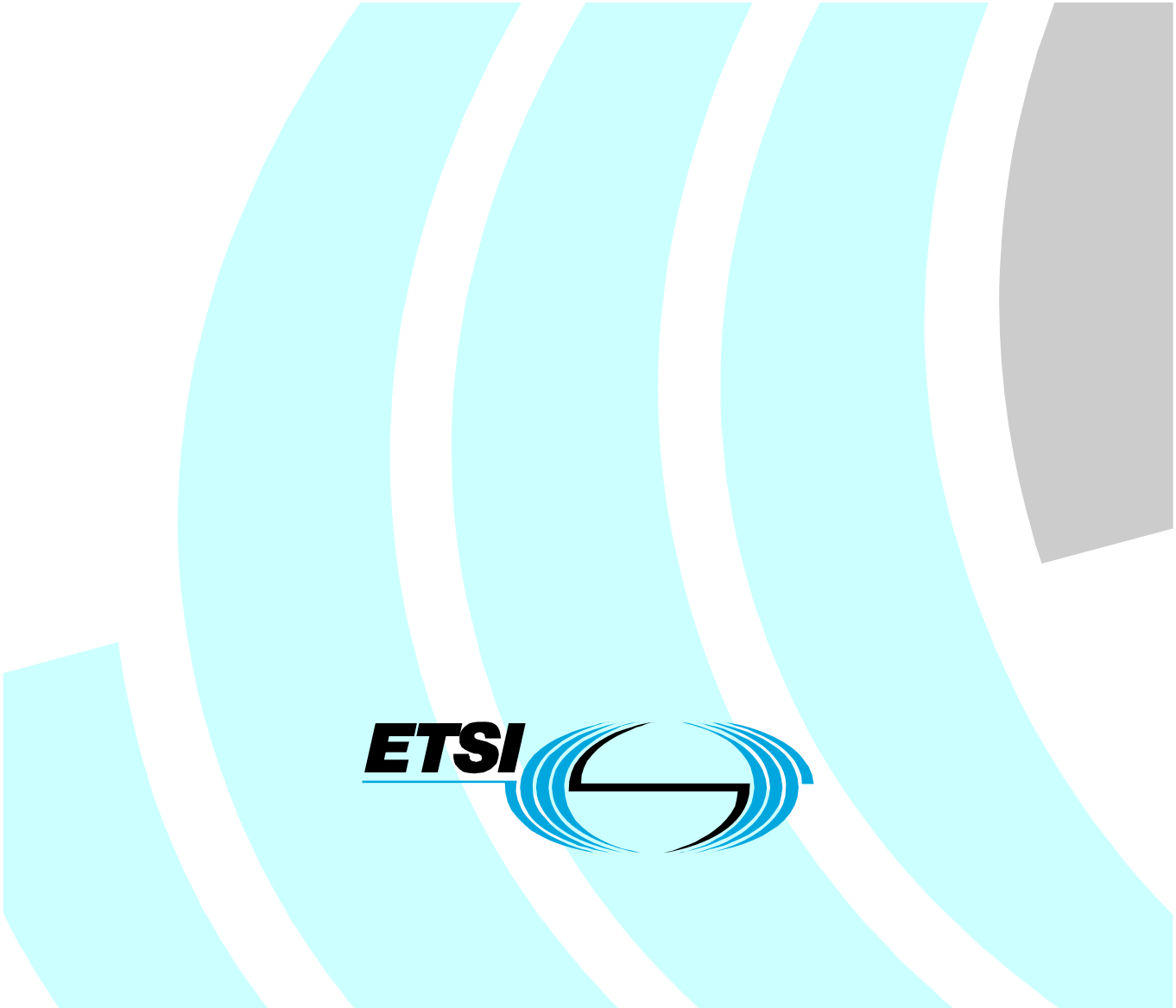
ETSI TR 102 989 V1.1.1 (2011-05)

Technical Report

Media Content Distribution (MCD); Subtitles distribution, situation and perspectives



Reference
DTR/MCD-00012

Keywords
access, distribution, teletext, transmission

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Media Content Distribution (MCD).

1 Scope

The present document is an analysis of the situation in the distribution of subtitling information of associated television services.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the reference document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are necessary for the application of the present document.

Not applicable.

2.2 Informative references

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] SMPTE 0259M: "Television - SDTI Digital Signal/Data - Serial Digital Interface".

[i.2] SMPTE 292: "1.5 Gb/s Signal/Data Serial Interface".

[i.3] ETSI EN 300 743: "Digital Video Broadcasting (DVB); Subtitling systems".

[i.4] ETSI EN 300 706: "Enhanced Teletext specification".

[i.5] CEA-608: "Line 21 Data Services".

[i.6] EBU Tech. 3264-E, Specification of the EBU Subtitling Data Exchange Format, European Broadcasting Union, February 1991.

NOTE: Available at: http://tech.ebu.ch/docs/tech/tech3264.pdf.

[i.7] Association media for all.

NOTE: Available at: http://www.mediaforall.eu/.

[i.8] ETSI TS 102 796: "Hybrid Broadcast Broadband TV".

[i.9] Directive 2007/65/EC of the European Parliament and of the Council of 11 December 2007 amending Council Directive 89/552/EEC on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the pursuit of television broadcasting activities.

[i.10] Directive 2002/19/EC of the European Parliament and of the Council of 7 March 2002 on access to, and interconnection of, electronic communications networks and associated facilities (Access Directive).

[i.11] Directive 2002/21/EC of the European Parliament and of the Council of 7 March 2002 on a common regulatory framework for electronic communications networks and services (Framework Directive, (FwD)).

[i.12] Directive 2002/22/EC of the European Parliament and of the Council of 7 march 2002 on universal service and users' rights relating to electronic communications networks and services (Universal Service Directive, (USD)).

[i.13] Directive 1999/5/EC of the European Parliament and of the Council of 9 March 1999 on radio equipment and telecommunications terminal equipment and the mutual recognition of their conformity (R&TTE Directive).

[i.14] ETSI TR 102 688-3: "Media Content Distribution (MCD); MCD framework; Part 3: Regulatory issues, social needs and policy matters".

[i.15] WHATWG WebVTT, extracted from the WHATWG HTML specification.

NOTE: Available at http://www.whatwg.org/specs/web-apps/current-work/webvtt.html.

[i.16] W3C Recommendation: "Timed Text Markup Language (TTML)".

NOTE: Available at: http://www.w3.org/TR/ttaf1-dfxp/.

[i.17] Free TV Australia Operational Practice OP- 47: " Storage and distribution of teletext subtitles and VBI data for high definition television".

[i.18] SMPTE 2031: "Carriage of DVB/SCTE VBI Data in VANC".

[i.19] IETF RFC 3629: "UTF-8, a transformation format of ISO 10646".

[i.20] EBU R 110: "Subtitling on digital TV services".

[i.21] ETSI TS 101 547: "Digital Video Broadcasting (DVB); Frame Compatible Plano-Stereoscopic 3DTV".

[i.22] ETSI ES 202 432: "Human Factors (HF); Access symbols for use with video content and ICT devices".

[i.23] ITU-T Recommendation Y.1901: "Requirements for the support of IPTV services".

[i.24] Directive 2002/21/EC of the European Parliament and of the Council of 7 March 2002 on a common regulatory framework for electronic communications networks and services (Framework Directive, (FwD)).

[i.25] Directive 98/34/EC: "Procedure for the provision of information in the field of technical regulations and rules on information society services".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

anti-aliasing: technique of minimizing the distortion artefacts appearing when representing high-resolution objects such as fonts to a low-resolution display such as a TV screen

captions: real-time on-screen transcript of the dialogue as well as any sound effects

NOTE 1: This service can be provided by means of either textual or graphical supplementary content. The captions and the dialogue are usually in the same language. The service is primarily to assist users having difficulty hearing the sound. Ideally, users may have some control over the position and size of the presentation. Different speakers are distinguished, usually by different colours.

NOTE 2: This is based on ITU-T Recommendation Y.1901 [i.23], clause 3.2.4.

content provider (CP): actor making available any kind of content where it has editorial responsibility or represent those with editorial responsibility

electronic communications service (ECS): service normally provided for remuneration which consists wholly or mainly in the conveyance of signals on electronic communications networks, **including telecommunications services and transmission services in networks used for broadcasting**, but excluding services providing, or exercising editorial control over, content transmitted using electronic communications networks and services; it does not include information society services, as defined in Article 1 of Directive 98/34/EC [i.25], which do not consist wholly or mainly in the conveyance of signals on electronic communications networks

NOTE: This is based on the FwD [i.24], Article 2, Definitions, (c).

electronic communications service provider (ECSP): actor offering ECS; in the context of Media Content Distribution (MCD) means those that offer the contents available from Content Providers but do not have editorial interference on it

encode: convert content from one format (typically raw) to another (typically lower bit rate)

sign language: language that uses a system of manual, facial, and other body movements as the means of communication

subtitler: person in charge of creating subtitles for a content

teletext: data delivery system within television transmission

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

3D	Three Dimensional
ANSI	American National Standards Institute
ASCII	American Standard Code for Information Interchange
ATSC	Advanced Television Systems Committee
CBR	Constant Bit Rate
CC	Closed Caption
CP	Content Provider
DFXP	Distribution Format Exchange Profile
DVB	Digital Video Broadcasting
DVD™	Digital Versatile Disc
ECI	Experts Community Integrated production
ECSP	Electronic Communications Service Provider
HD	High Definition
HDTV	High Definition Television
HTML	Hyper Text Markup Language
IP	Internet Protocol
IPTV	Internet Protocol TeleVision
MXF	Metadata eXchange Format
OCR	Optical Character Recognition
PID	Packet Identifier
SD	Standard Definition
SDI	Serial Digital Interface
SDTV	Standard Definition Television
SMIL	Synchronised Multimedia Integration Language
VANC	Vertical ANCillary data space

VBI	Vertical Blanking Interval
VBR	Variable Bit Rate
VOD	Video On Demand
W3C	World Wide Web Consortium
XML	Extensible Markup Language

4 Subtitling in the media content distribution context

4.1 Subtitling vs. captioning

In general the term caption is more related to original capture of all information (noise, sound, surrounding conditions) associated to an event and the term subtitling rather used in relation to the textual (in a picture or as a sequence of words and letters) expression describing or reproducing the same event. Using this textual expression, different types of presentation techniques (text overlapped on the image in one or several languages, the language of signs or even audio description) may be used.

However "captioning" is more common in the U.S.A. and is typically used for same language transcription, for instance for the hearing impaired; "subtitling" often refers to language translation of foreign programmes.

The terms "open captions" and "closed captions" are also very common:

- Open captions (also called "burnt-in" "in-vision", or "hardsubs") are encoded as a permanent part of the video image, irreversibly merged into the frames. They cannot be disabled, and no specific "track" is needed to convey them. Also, no special equipment or software is required for playback, and any character set (including pure graphics) can be used. Consequently this technique allows for very complex transition effects and animation (for instance karaoke song lyrics). However it does not allow for multiple user-selectable variants of subtitling (such as English, English Hard of Hearing, and French).

- Closed captions are viewer-selectable and carried over a separate "track" or "component" (which may be collocated inside the video track, but not in the video data itself). Examples of closed captions include the Line 21 Closed Captioning system in the USA (carried as text), Teletext subtitles in Europe (text as well) or DVB subtitles (graphics). The carriage of this separate, discrete data stream may be problematic in some contexts.

Nowadays both the subtitling and the captioning terms are disappearing in favour of the "timed text" expression. In the other clauses of the present document, "subtitling" refers to both "subtitles" and "closed captions"; open captions are carried inside the video data and do not require a specific handling.

In other categorization, digital video subtitles are sometimes called internal, if they're embedded in a single video file container along with video and audio streams, and external if they are distributed as separate file (that is less convenient, but it is easier to edit/change such file).

4.2 Bitmap and textual subtitles

Clause 4.1 shows that subtitling data can be carried on a different track than video data, which requires special player support, but allows the user to disable the subtitles, or choose a specific variant. Two types of presentation formats can then be used:

- A bitmap, pre-rendered format, such as DVD Sub-Picture Units or DVB subtitles: the subtitles are stored as images (generally not the same codec as the video, but a simpler algorithm, with minimal bit rate and colour depth), and the player overlays these images over the video frames.

- A textual format, such as Teletext (for distribution) or an authoring format: the subtitles are stored textually with instructions, usually a specially marked up text with time stamps and stylistic information (position, colour, weight, etc.). They are rendered by the player (or converted to another format) and displayed over the video at the specified time stamps.

The choice of bitmap vs. textual formats depends on a lot of factors, which are examined in more details in clause 9.1. Bitmap subtitles are currently more popular for the distribution step because they historically allow for a better-looking, consistent between receivers, user experience.

There are many flavours of textual formats, especially for the production domain, but they are in general reciprocally convertible. Typically, textual formats are easier to create, change, convert and re-use in other applications; they are thus frequently used for fansubs (subtitles fans created by viewers). Bitmap subtitles are difficult to convert or re-use, though special OCR software exist to convert bitmap subtitles to textual format.

4.3 Presentation techniques

4.3.1 On-screen techniques

When producing subtitling content, the first choice to make is the display mode:

- Pop-up mode: traditional display, scheduled, as in movie theatres.

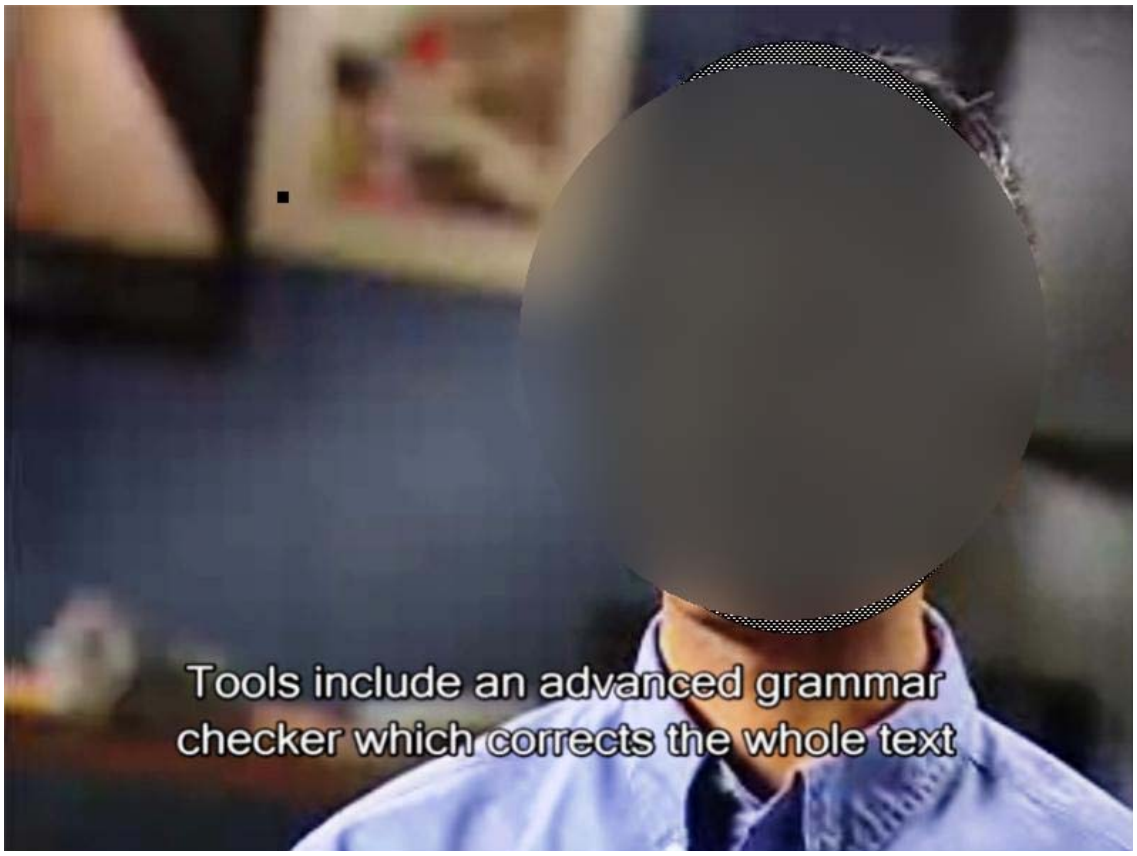


Figure 1

- Cumulative mode: word by word and sentence by sentence, also called Roll, Push or Snake.

Figure 2

The cumulative mode typically takes a higher bandwidth than the pop-up mode. The presentation technique also greatly depends on the nature of the programme, and the amount of time allocated to the task of subtitling.

NOTE: Regardless of the mode of display of the subtitles themselves, ETSI has already standardized the way for an end-user to enable subtitling, via standardized icons ([i.22], clause 5.1).

4.3.2 Techniques complementary to subtitling facilitating other functionalities

While not technically "subtitling", subtitling data can also be used as a source for other presentation techniques.

Among the European deaf population, approximately 12 % of the people are illiterate and cannot read subtitles. These persons often need the sign language to understand the transmitted message. It normally requires a human translation made by someone correctly perceiving the original message and reproducing it in the sign language. However, there are new techniques to synthesize so-called "Avatars", "Talking heads" or "3D puppets", which may be three-dimensional images of a face that reproduces lip movements. The goal is to allow hearing-impaired people to follow a speaker, even when the lips are hidden, or the speaker is off-screen. In order to facilitate lip reading, a hand is also represented to add information on the consonants and vowels, which makes it possible to eliminate ambiguities between sounds corresponding to the same shapes of the lips, such as "p" and "b". Also, some avatars actually reproduce the simplified spoken language, the most common sign language, which conveys meaning by simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expression.

Those systems can take advantage of subtitling data; indeed they combine a phoneme recognition tool and a complex driving system based on a model that makes it possible to associate a shape of the lips and face, and its computer graphics translation, with each sound. Examples of these research projects are the ISA and PAROLE project teams of INRIA Lorraine/LORIA, in partnership with the DATHA association (development of technological aids for hearing-impaired persons), and Artus/Arte (shown below).

Figure 3: 3D puppet reproducing lips movement (Artus/Arte project)

Another presentation technique, which is not strictly-speaking subtitling, is Braille, available on special keyboards capable of activating a matrix of dots reproducing the message. It is especially useful for original version subtitling for vision-impaired people. When growing old, a person may have difficulties reading the original version subtitles on screen, and may use a Braille device for that purpose.

With the progress of automatic translation machines and voice synthesizers, it is easy to imagine that the text associated to subtitling might in a near future be used as the base for:

- the presentation of the subtitles in several languages, possibly selectable by the user; or
- audio description facilities used for blind or elderly people with visual impairments.

Textual formats facilitate the use of subtitling data by these presentation techniques; however the current trend is to encourage bitmap subtitling.

4.4 3D-specific challenges

With 3D, the subtitling process becomes more complex because of the depth of the picture: subtitles have to appear in front of the objects behind them. If the subtitles are flat, it is impossible to read them because they would appear behind the action. Also, some people prefer the subtitles to be always 1 pixel in front of the most forward object/the object of interest, and that means that the subtitles depth can be dynamic.

The subtitle operator has to be in charge of this. If the depth is not respected, the picture or the subtitle will be visually damaged, or uncomfortable for the viewer.

A new DVB specification [i.21] summarizes the normative requirements by DVB to carry and render 3D plano-stereoscopic content. DVB subtitles [i.3] have been extended to define subregions within regions, and assign them a depth (disparity) indication. The disparity can also be temporally adjusted.

4.5 Necessity of subtitles distribution

Subtitling is a very important feature for users with hearing difficulties. These difficulties may have genetic origin, result from an accident (e.g. acoustic shock) or from the common biological human ageing process or be associated to the fact that the user is listening to a language or a dialect that is not his own. The overall percentage of the population suffering from this impairment is increasing since the life expectance is increasing, the number of migrants also and the circulation of audiovisual contents as well.

Governments are getting more and more concerned with the fate of impaired users, and a lot of regulation authorities have taken measures to enhance the integration of the affected groups of the society. Some authorities force the bigger content providers to implement subtitling features on a significant part of their programs and communications service providers and network operators to appropriately support these services. These actions are normally in the context of or resulting from studies in the area of e.inclusion and more specifically e.accessibility. TR 102 688-3 [i.14] explains the European regulatory and policy environment and refers to this area of issues in clauses 9.6.2, 9.6.4 and 9.6.7.

The legislated requirements of digital broadcasting in Europe (European directive 2007/65/CE [i.9]) have prompted an increase in the number of captioned programs of up to 100 % at the end of 2015 for the latest European country. In France the law 2005-102 requires broadcasters whose mean audience is over 2.5 % of the population to caption 100 % of their programmes by 2010. The bases of these measures are cited in TR 102 688-3 [i.14], clause 6.2.2.

At a level of communications services, the e-communications directives and particularly the framework [i.11], the universal service [i.12] and the access (and interconnection) [i.10] directives were revised to underline the importance of the support to services for people with the disabilities. This is explained in TR 102 688-3 [i.14], clause 7.1 and further clauses of clause 7.

The EU directive covering telecommunications terminal equipment (the R&TTE directive [i.13]) also considers in article 3.3.f the need to support "*certain features in order to facilitate its use by users with a disability*". This is referred to in TR 102 688-3 [i.14], clause 8.2.1 but has not been yet translated into specific measures in the European regulatory environment at the present.

Regulatory measures are more often addressing real time broadcasting, but the general recognition of the importance of subtitling services is very likely to impose it as a general market request for all types of distribution of media contents, including on-demand services, over managed and non-managed networks, using broadcast, unicast and multicast techniques.

Subtitles are also used in a lot of countries to display movie contents with original audio; original audio is a plus for premium channels that is more and more often required.

Therefore subtitling is nowadays considered an essential component, along with the video and audio. Different technologies coexist at several levels to convey the subtitles; however manipulating them is not as simple as manipulating audio tracks, in a lot of cases.

5 Subtitles flow

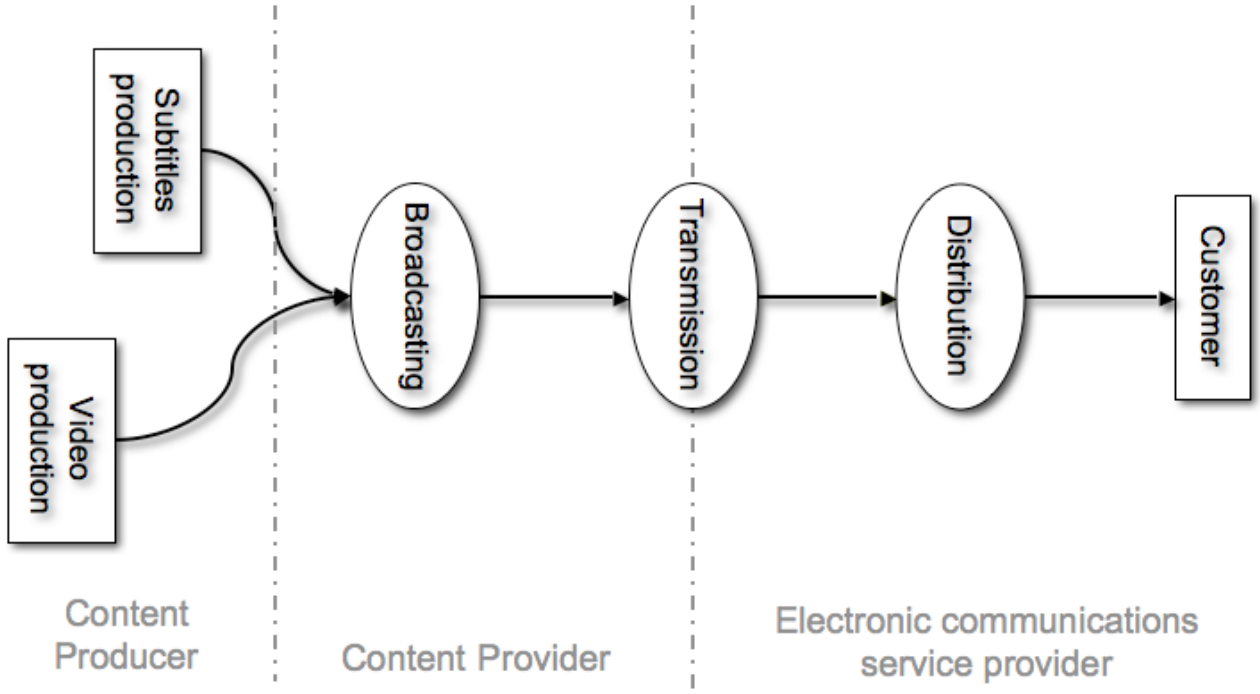


Figure 4: Steps in signal distribution

It is very difficult to find unambiguous terms to describe steps in this area. For the purpose of this document, the following steps are defined as follows:

- **broadcasting:** Literally, to broadcast is to cast or throw forth something in all directions at the same time; a radio or television "broadcast" is a scheduled content that is transmitted over airwaves or other means for public reception by anyone with a receiver configured accordingly. However for a few years, the "broadcaster" has become the company which manages the contents and "plays" them with automation, video servers and so on, rather than the operator which manages the transmission and broadcasts effectively on the air or through the networks.

- **transmission:** The "transmission" is the common area of competence between the content provider and the electronic communications service provider. It consists in putting the content in a format suitable for use by the ECSP (generally the competence of CP), and transporting it to the ECSP's place of distribution (generally, the competence of the ECSP). In some cases one or both of the tasks are subcontracted to a third company.

- **distribution:** The term "distribution" is in fact the old name for "broadcast". Nowadays, the distribution term is preferably used to refer to the last network before the end user. Depending on the countries and the needs, the frontier between transmission and distribution can be thin, but more and more, every section from the TV or radio station to the end-user becomes specific, and new needs are re-defining existing terms.

Those steps are more or less reflected in the different entities or "actors" managing them:

- **content producer:** The content producer creates the content and normally its affiliated data (including subtitles or the sign language), either live or off-line.

- **content provider:** This actor determines and manages the programmes of linear channels, or packages non-linear contents.

NOTE 1: Either directly the content producer, or the content provider may aggregate the original content to its associated data.

- **electronic communications service provider:** This actor is in charge of the carriage of the content to the appropriate users.

NOTE 2: Earlier "broadcasters" assumed all the above mentioned functions, but at present there is an increasing specialization determined by the technology evolution. However, the limits between actors and steps are still very thin and imprecise.

Different formats can be used for the same purpose. For instance, a raw picture format can be carried over SDI [i.1] or HD-SDI [i.2]; for the transmission, either the raw signal is used, or an MPEG2- or MPEG4- encoded signal over IP. And of course at the distribution level, a variety of codecs is available. However, the following trends emerge:

- SD signals are being replaced by HD signals; thus, HD-SDI supersedes SDI.
- Production facilities are becoming file-based (e.g. replacing HD-SDI connections with MXF files).
- The transmission tends to be in encoded format for practical reasons.

For the moment, we have examined the case of a linear television channel. The same scheme can be applied to non-linear contents as well: the content is produced from tape or film and encoded to a mezzanine file, which serves the same purpose as the "transmission" step above. The mezzanine file is then transcoded to the target distribution format (VOD file, DVD™, Blu-Ray™, etc.). The same issues will be encountered, with subtle differences which will be noted when present.

5.1 In an SD-only content distribution

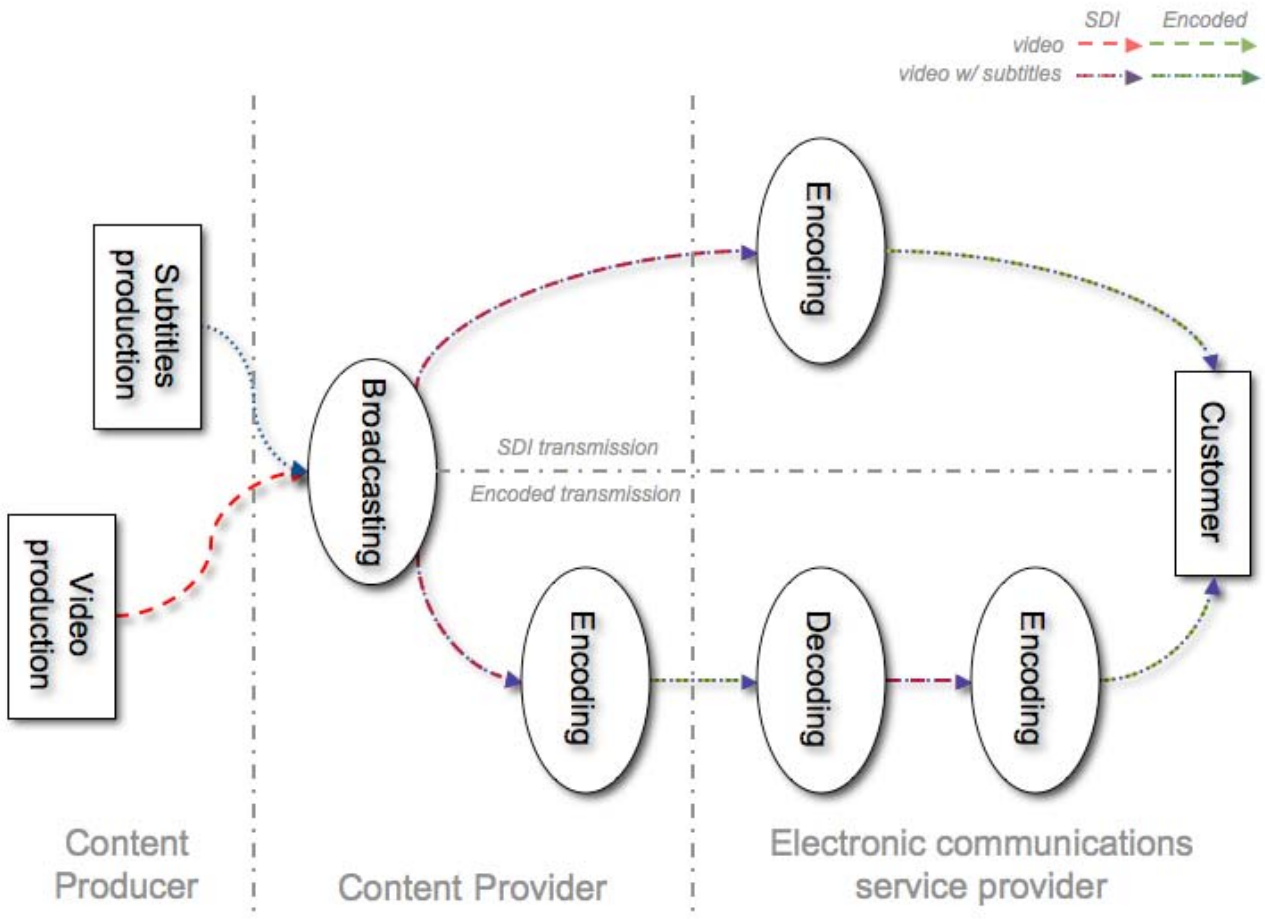


Figure 5: Subtitles flow throughout the distribution of an SD TV channel

SD formats have the particularity that the subtitling information can be embedded into the video signal, either in the VBI of raw signals, or in a specific PID of an encoded signal. The flow is therefore quite simple and straightforward, even with an encoded transmission.

5.2 In an HD content distribution with HD-SDI transmission

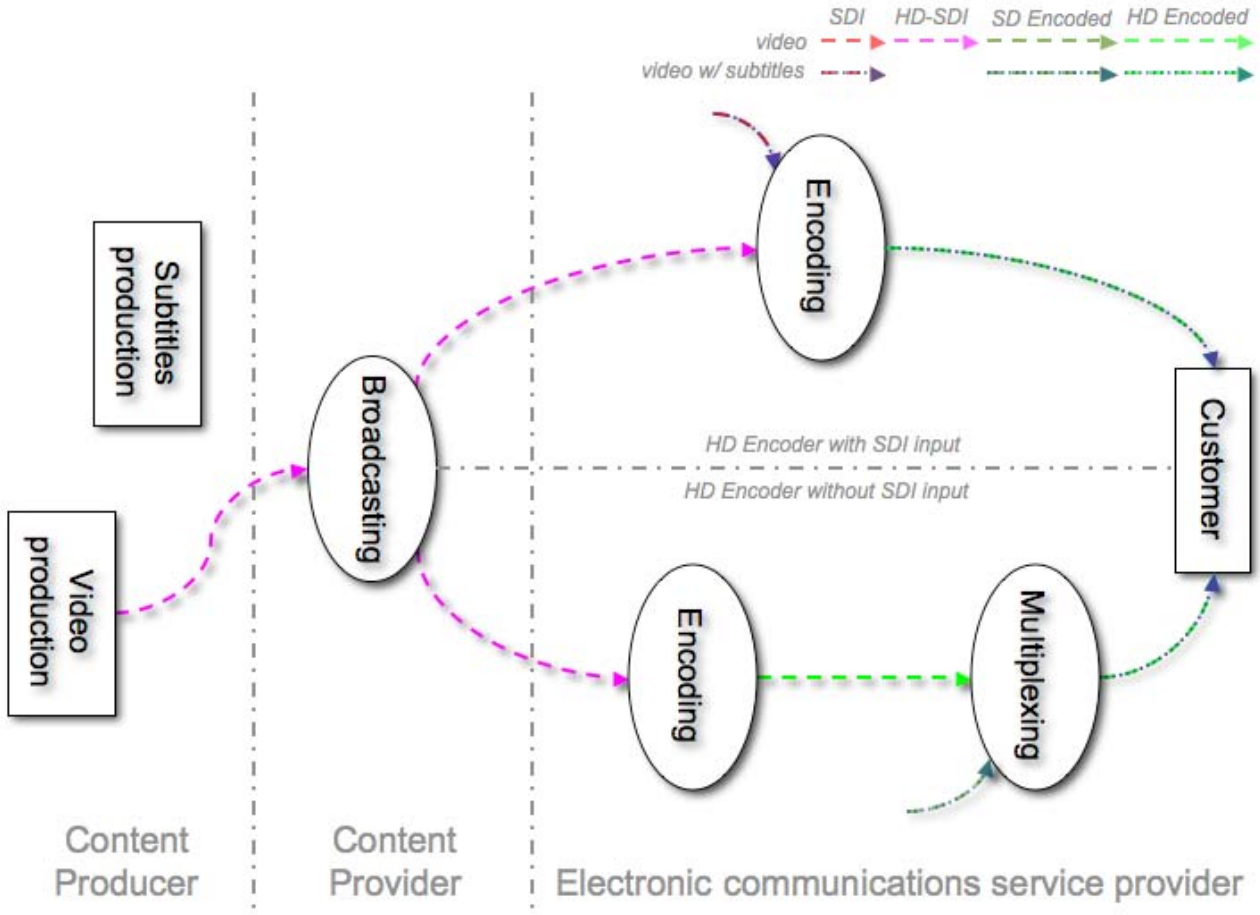


Figure 6: Subtitles flow throughout the distribution of an HD TV channel (HD-SDI transmission)

In the HD-SDI raw format, subtitles cannot be embedded. Therefore a large part of the flow shows the video without subtitling information. Subtitles can be embedded in encoded formats, but they have to be inserted at some time. Operators have two options:

- Use an HD encoder which features an extra SDI input, and connect it to the SDI-delivered signal of the channel; the video and audio are encoded from the HD-SDI input, but the subtitling data are extracted from the VBI of the SDI input. Since the SDI and HD-SDI signals of a given channel are synchronized, video and subtitles are in-sync.
- Use a multiplexer which extracts the subtitling packets from an SD encoded version of the channel, and embed them into the HD encoded signal. However, SD and HD signals do not have the same latency, so a synchronization delay is applied, sometimes manually with a limited precision.

5.3 In an HD content distribution with encoded transmission

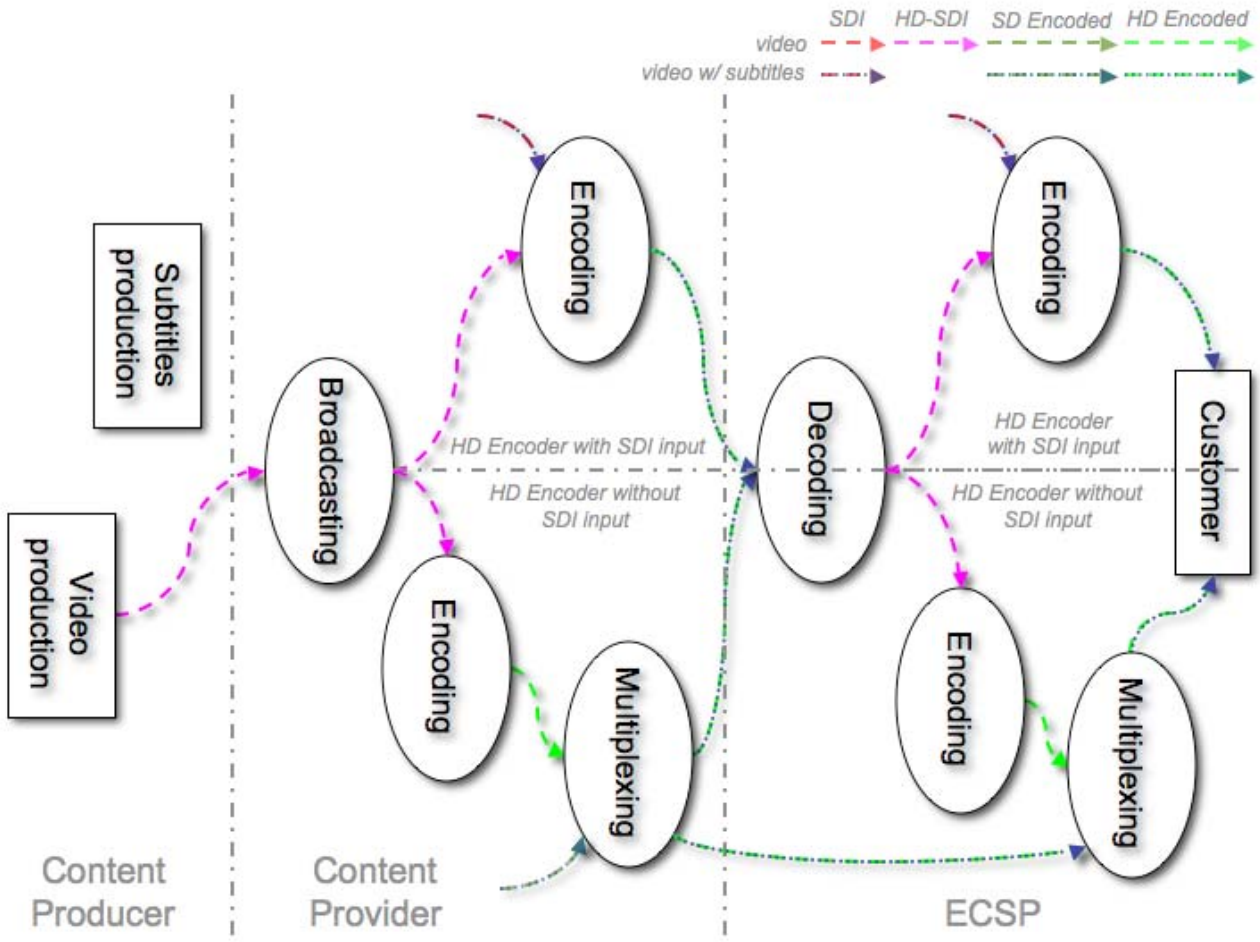


Figure 7: Subtitles flow throughout the distribution of an HD TV channel (encoded transmission)

In this case the content provider delivers an encoded HD signal with subtitling information embedded in a dedicated PID. The electronic communications service provider uses a decoder in order to re-encode the signal to the target distribution format. However, the decoder outputs HD-SDI, and HD-SDI cannot embed the subtitles data. So they are inserted afterwards, either from an SDI signal of the same channel, or with a multiplexer and a delay.

5.4 In a content distribution with late component binding

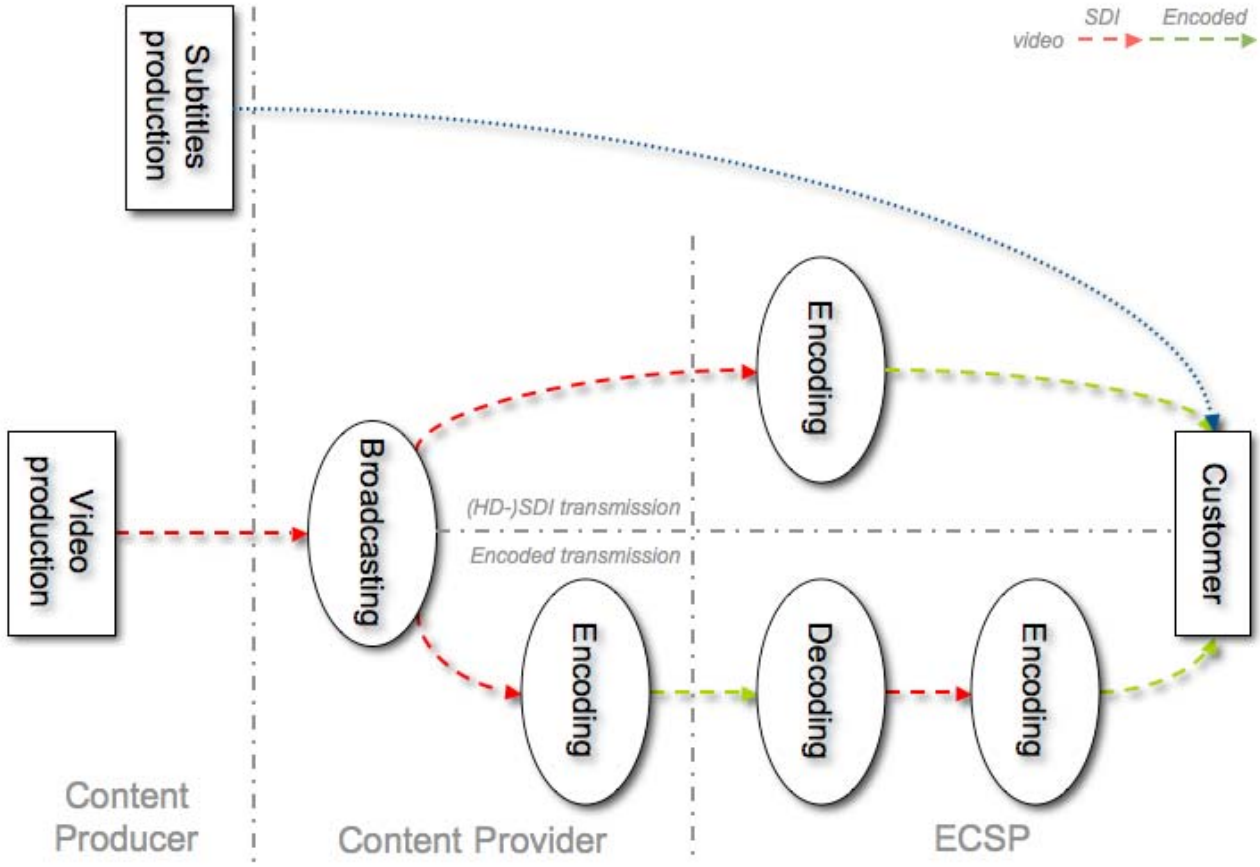


Figure 8: Subtitles flow throughout the distribution of TV channel, with late binding of the subtitles

In this case the subtitling information is not embedded into the video stream, neither in the (HD-)SDI signal, nor in the encoded signal. The subtitles are transmitted to the end-user equipment via an independent means, and only on-demand. The protocol and format of distribution are specifically adapted to allow the late-binding of components, with features like common presentation dates. This enables the operator to modulate the bandwidth required for a particular programme: you can have multiple audio and subtitles tracks without wasting bandwidth for the tracks which are not asked by the customer, unlike to what happens with traditional MPEG-2 multiplexes where all components must be in the same stream. Also the additional tracks can be monetized differently.

This scheme is not currently in use in the traditional broadcast world. Some compatible protocols are now available, but are mainly used for Internet TV.

6 Production

6.1 Off-line vs. live subtitling

Off-line subtitling is the traditional way to proceed: a subtitler creates the subtitles by viewing previously recorded material, and building a file with text (the current dialogue, its translation in a foreign language, or noise information for the hearing impaired) and associated time codes. As this work cannot be done before the end of the editing, making and adding subtitling is always the last thing to do before delivering a video file. Off-line subtitling usually produces superior results because the subtitler can use as much time as needed to make sure that the translation/transcription is accurate with no spelling mistakes, and that the subtitles are timed to coincide precisely with the dialogue. Also, special care can be taken to position subtitles to avoid obscuring other important onscreen features.

However it is a very slow and expensive work: about 20 times the real-time. Also it is not suited to live or near live events such as news, sport and special events, for which accessibility requirements more and more apply. Different technologies of live subtitling can be used depending on the nature of the programme, and the availability of text inputs, for instance: script, teleprompter, stenographic typists, etc. In that case, the subtitles can be built by repurposing text sources and cueing them at the appropriate time.

Another technology which is soaring is the voice recognition, available for a lot of languages; however there are some languages for which no voice recognition software is available (for instance at the time of writing, Polish). The best results are obtained with a process using a re-speaker, that is a person who has been trained to use the voice recognition system and who speaks again what is said in the content, with an appropriate pace or tone. In general a custom dictionary is prepared in advance to cover specialist words, last names or place names that are expected to be used in the content. This technology allows a 90 % accuracy rate [i.7], but still requires a verification stage.

In the case of a re-broadcast of an recorded live programme, either the live subtitles initially produced are re-used as is, or new, off-line subtitles are created using the live subtitles as input source.

Systems using this mode of production are able to insert subtitles in real time in the signal. Live subtitling often uses the cumulative mode of display.

6.2 Production formats for off-line contents

6.2.1 Historical formats

Historically, the EBU defined in Technical Reference 3264-E [i.6] an open format intended for subtitle exchange between broadcasters (.STL file format). It is however an old specification lacking basic features such as Unicode support.

Consequently, vendors of subtitle insertion systems and software promote their own file formats, such as: .PAC, .RAC, .CHK, .AYA, .890, .CIP, .CAP, .ULT, .USF, .CIN, .L32, .ST4, .ST7 and .TIT.

In addition, the production of subtitles now takes into account subtitles created for specific media players or by specific PC software; it is becoming more and more common to deliver such files. Again, lots of proprietary formats are available, in three categories:

- Text-based: .AQT, .ASS, .JSS, .PJG, .PSB, .SSA and .SUB.
- HTML or XML-based: .GSUB, .RT, .SMI, .SSF, .TTXT (MPEG-4 Timed Text) and .USF.
- Image-based: VobSub (.SUB + .IDX) and XSUB (embedded into .DIVX).

6.2.2 WebVTT

The WebVTT format (Web Video Text Tracks) [i.15], formerly known as WebSRT (Web Subtitle Resource Tracks) is a WHATWG specification. It is designed for marking up external timed track resources.

WebVTT is based on the historical SRT format, which was a very simple format (ASCII text). Not only is this a great way to foster tool development and support in non-browser clients, it is also good for authors: timed tracks are often tweaked in a simple text editor.

Implementers of Mozilla™, Apple™, and Opera™ have all committed to support the SRT format at a minimum; none of them have an interest in TTML. SRT content is quite widespread already and is implemented by many players.

WebVTT extends SRT to support an HTML syntax, while retaining its simplicity. It also takes advantage of Unicode UTF-8 [i.19], but like the historical format, it does not embed any information on text positioning. The lack of this very important feature makes it unsuitable for broadcasters.

6.2.3 TTML

The TTML (Timed Text Markup Language) [i.16] is a W3C recommendation. It is designed to address two purposes:

- interchange among authoring systems;
- direct distribution of subtitles.

The commonly used DFXP (Distribution Format Exchange Profile) term refers to a profile of TTML intended at interchange among distribution systems.

6.2.4 European Broadcasting Union

The EBU has a Group working on subtitling formats. The Group (ECI-DFXP) is verifying if a profiled version of TTML (DFXP) can be used as a follow up of EBU STL.

Currently (Dec 2010) a draft list of requirements and a draft structure for a new EBU Recommendation are available in the EBU Group.

The work is driven by two main developments:

- Broadcasters switching to file-based production.
- Broadcasters creating more diverse 'web-oriented' content in various formats, including HbbTV.

6.3 Production formats for live subtitling

There are two elements to consider: the live transmission, and the result of the production, which could be reused in a more traditional work-flow.

The task of subtitling live programmes is often subcontracted to an external company, which can be in a different location. The link between the producer of live subtitles and the broadcaster is usually IP over network: the producer directly controls the equipments of the broadcaster through an IP or serial protocol; the protocol itself is proprietary, vendor-dependant. The equipments can embed Teletext into a live (HD-)SDI feed, or multiplex DVB subtitles [i.3] into an encoded signal, depending on the content provider's choice of transmission or distribution format.

When producing live subtitles, every word typed has its own time code. At the end of the process, the producer can export the result of its work to various file formats, which can be reused by the broadcaster for new, off-line, transmissions.

7 Broadcasting

During the broadcasting operation, the signal is usually output over an SDI or HD-SDI link. Regarding subtitles, the equipment ensures that the subtitles are kept in sync with the video and audio.

7.1 SD signal

In Standard Definition, the subtitles were traditionally embedded in the VBI data of the video, according either to the Teletext standard (European countries) [i.4] or Closed Caption line 21 (ATSC countries) [i.5]. This VBI data was found in the SDI signal of live transmissions, but also on professional videotape media (such as Digital Betacam). It was therefore very convenient to manipulate and was transmitted and copied alongside the video and audio component; it is a straightforward way to make sure that the subtitles are correctly in sync.

7.2 HD signal

High definition video is in that respect a regression compared to simple Standard Definition; the VANC of HD-SDI signals can no longer be used to carry such data. There is currently no standardized solution to embed subtitling information into an HD-SDI signal, and HD media such as HDCAM cassettes. There have been several attempts, for instance the Australian OP47 initiative [i.17] and SMPTE 2031 [i.18], but they are all based on embedding Teletext, CC line 21 or even older formats like Newfor (which is not an industry standard). Such old formats exhibit a number of flaws:

- Limited number of languages (no more than 1 or 2).
- Very dependent on the character set (ASCII, ANSI); no support for Unicode.
- Limited number of fonts (1 or 2); non-proportional (unlike Times New Roman , an I takes up the same width as a W).
- Limited support for styles (no Bold, no Underline).
- Very few colours (maximum 2 to 8).

7.3 SMPTE-TT

This new format is, at the time of writing, still in preparation.

The SMPTE draft ST 2052-1 "SMPTE Timed Text":

- creates an SMPTE profile of W3C TTML;
- defines a TTML-compatible "tunnel" for transport of raw data streams (e.g. CEA 608), for possible waveform reconstruction to support CE devices that need legacy composite signals;
- provides a mechanism for graphical timed text (e.g. Asian languages).

Another SMPTE draft, RP 2052-10, translates legacy CEA-608 essence into SMPTE-TT:

- for processing in contemporary broadband devices and extensibility;
- proponents are welcome for other format RPs.

8 Transmission

8.1 (HD-)SDI transmission

This is the traditional method for delivering a television signal: raw picture format. The electronic communications service provider either places its final encoders at the television channel's premises, or uses devices to carry (HD-)SDI over fibre optics or IP networks.

Again, HD-SDI does not carry subtitling data, so in practice an addition SD-SDI signal is needed to carry the whole of the television signal.

This method tends to be superseded by the encoded transmission, because the number of electronic communication service providers rose dramatically over the past years: from the historical analogue broadcast operator to satellite, cable, IPTV and mobile, a typical television channel may deliver its signal to a dozen operators. The master control room generally does not have enough space to accommodate for the equipments of the operators; an encoded IP delivery is therefore simpler.

8.2 Encoded transmission

In this method the content provider chooses to deliver a transport stream with encoded video and audio over multicast IP. Electronic communication service providers connect to this closed, private network in order to receive the channel(s).

The exact specifications of the coded signal vary. In general SD video is encoded using MPEG-2 video coding at a bitrate of 8 or 15 Mbit/s. HD video uses either MPEG-2 at 60 to 80 Mbit/s, or H.264 around 30 Mbit/s; some channels are now using 4:2:2 chrominance format. Audio uses the simple MPEG-1 layer II format at 256 kbit/s or 384 kbit/s.

An encoder is used to create the delivery signal. SD encoders have the ability to capture the subtitling information from the VBI of the video, and insert them as teletext into the output signal (see next clause). However this is not possible with HD encoders as HD-SDI does not carry subtitling data (and if it were, current encoders would not take advantage of it).

It is nowadays common to find high definition encoders featuring also an SD-SDI input; the SD-SDI input is used only to extract subtitling data from the SD VBI, in order to embed it into the HD output signal. Another (very common) method is, at the very end of the flow, to use a multiplexer to copy, with a certain delay (SD and HD encoders do not introduce the same latency), the subtitles PID from an SD stream to the HD stream. Both solutions are 'ad-hoc' solutions which work in the field, but show that the situation is considerably worse than with SD television.

9 Distribution

The distribution is the final operation in the delivery of media content. The signal is encoded to a specific format, adequate to the delivery network:

- DVB-style network (satellite, cable, terrestrial): DVB Teletext and DVB Subtitling are both used; DVB Teletext is more common but considered as "legacy"; DVB subtitling is now the recommended solution [i.20].
- IPTV: It uses transport stream at its core, sometimes with DVB extensions; Teletext and DVB subtitling are commonly used.
- Internet TV: Subtitling is not yet commonly available, but provided more and more. In the Netherlands, the original Teletext data for example is reused for the (Silverlight™-based) VoD service "Uitzending Gemist". This provides an effective closed subtitling service for Deaf & Hard of Hearing.
- Mobile: Subtitling is very rare considering the size of the screen, and when present burnt into the video.
- Physical medium (DVD™, Blu-Ray™): They use a bitmap format.

9.1 Bitmap and textual distribution formats

During the production, broadcasting and transmission phases, the subtitles format is textual: they are carried as text with stylistic information, position, synchronized with the video. In the distribution phase they are often converted to bitmap formats for the following reasons:

- Textual subtitles are rendered by the TV screen or the set-top-box; therefore, their look depends on the implementation of the end device and can sometimes be a/ deceptive or b/ divergent (different sizes or positions depending on devices).
- Bitmap formats are rendered exactly the same on all screens; they can embed different fonts, non-European character sets, stylistic information, colour or pictograms more easily.
- Bitmaps have a better look than most fonts embedded into TV sets.

It is a long trend in the industry to replace textual subtitling with bitmaps. However, this approach has several drawbacks, including:

- High peak bandwidth use.
  - At first glance, the average bandwidth use is comparable between DVB subtitles and textual subtitles like Teletext. It may sound striking, but DVB subtitles have the advantage to be only transmitted once, when there is a need to display a new subtitles. On the contrary, Teletext is constantly transmitted (it usually requires to transmit two lines of the VBI data, amounting to 37 600 bits/s), whether there is subtitling data or not; also, each frame carries its subtitling data, so a textual information is typically transmitted 100 times over the air. Consequently, the average bitrates for both technologies are equivalent.
  - However, the advantage of only transmitting the subtitle once when needed soon disappears with the live subtitling techniques, where words appear one by one; and live subtitling techniques are getting more and more popular, because they are much cheaper to produce. In such circumstances, DVB subtitles often take up as much as 150 kbits/s.
  - Also, the DVB subtitles bit stream features peaks of bandwidth with each new subtitling data; the DVB subtitling system uses a very variable bitrate, on the contrary to Teletext, which has a strictly constant bitrate. Though this is not a problem with large statistically-multiplexed transport streams, where the DVB subtitles bandwidth can be shaped and re-used by opportunistic data when it is unnecessary, it is not the case with low multiplex rate and single programme transport stream, such as xDSL TV. There, the bandwidth for the worst case has to be provisioned, and therefore a lot of space is wasted.
- Bitmaps are not reusable, e.g. it is difficult to extract the subtitles to derive another target format (see clause 4.3.2 for other presentation techniques).
- Bitmaps are not scalable by the end user, which is problematic for the visual impaired people.
- Quite often bitmap formats are limited to a 4 bit colour space, which does not allow anti-aliasing techniques; the embedded text renderer of nowadays TV sets and set-top-box feature a much higher quality.
- One bitmap format is rendered by delivery target: The size of the bitmap will not be the same for SDTV, HDTV or mobile.

9.2 Available formats

There are several solutions to distribute subtitling information to the end-user equipment:

- Teletext, a remnant of the analogue world, is a text format; as such, it can be displayed over different picture size with an end-user-equipment-dependant quality (usually very poor) and uses limited bandwidth (37,5 kbits/s are generally enough); it lacks support for a unified character set such as Unicode, and flexibility in the presentation of data.
- Closed Caption line 21 is the ATSC equivalent for Teletext for subtitling; it is simpler but suffers from the same flaws.

- DVB subtitles have gained some support from the manufacturers and regulation authorities; DVB sub is a bitmap format, and as such is displayed identically, with a generally excellent quality, on all equipments; however it is VBR and its peak bandwidth requirement can be high, especially in live mode (80 to 150 kbits/s per language); also it is not scalable.

NOTE: The DVB subtitles specification also defines objects containing strings of character codes; however it says that the specification alone is not enough to be able to render such glyphs, and that private agreements are needed to enable this feature. In practice this feature is not implemented in end-user equipments. However, it has already been used by the BBC™ to transmit subtitling information on a contribution link, with dedicated professional equipments.

- DVD™ and Blu-Ray™ use a bitmap format; on DVD™ the colour space is limited to 4 bits, which makes anti-aliasing difficult.

Table 1: Summary of available subtitling formats for distribution

	Format	Rendering	Bandwidth
Teletext	Text	Equipment-dependant	Low (37,5 kbits/s), CBR
CC line 21	Text	Equipment-dependant	Very low, CBR
DVB subtitles	Bitmap	Good, limited anti-aliasing	VBR, up to 150 kbits/s
DVD™ SPU	Bitmap	Good, limited anti-aliasing	
Blu-Ray™ PGS	Bitmap	Good	

10 Synthesis and conclusions

10.1 For a better balance between text and bitmap formats

Many people have praised DVB subtitles as the next-generation subtitling technology. However, the diversity of distribution means make it difficult to use a single format of subtitling, since DVB subtitles are authored for a specific definition: they will not scale. Different tracks are authored for SD, HD, and mobile phone broadcasts. It is a waste of bandwidth and an operational concern to create those tracks at the same time.

Also, clause 4.3.2 shows that bitmap subtitles can hardly be used for other accessibility purposes, such as avatars and Braille. The user also cannot scale the subtitles.

Text-based solutions are more suited to today's world. The processing power of current TV sets and set-top-boxes now allows to display text with a good quality (many platforms now embed an HTML-derived browser). The advent of HTML-based interactivity standards like HbbTV [i.8] could also open new ways for the subtitling distribution.

Standard organizations should invest time into making good text-based standards, which could be used in every step of the subtitles flow.

10.2 For a modern timed text format

In whichever environment it is used, a modern subtitling standard should have the following characteristics:

- Support for Unicode character sets.
- Extended stylistic information (fonts, size, style, colours).
- Precise positioning.

- Dynamic depth information.
- Scalability for various screen sizes.
- Re-usability for other accessibility purposes.

10.3 Lack of a production standard

The content producers suffer from the heterogeneity of the proprietary file formats that are trying to surpass the limitations of the old EBU subtitles standard. The creation of an open standard should be supported.

The EBU seems to be already tackling this task of creating a new standard, based on TTML.

10.4 Lack of a broadcasting standard

Starting with the broadcasting step, it is necessary to convey video, audio and subtitling together in order to ensure proper synchronization. The SMPTE standard for raw HD video, HD-SDI, currently has no support for subtitles embedding. Initiatives to extend it are based on the venerable Teletext, which has a lot of limitations.

It is technically possible to carry any kind of digital data in the vertical and horizontal ancillary spaces. A standard allowing to embed subtitling data in an HD-SDI signal should be encouraged.

The SMPTE appears to be working on new standards in this area.

10.5 Lack of a distribution standard

The trend is to distribute subtitles in a bitmap format. Clause 9.1 details the drawback of these technologies. However, the only available distribution format of textual subtitling data is, again, a packetized Teletext.

An initiative should therefore be started to assess the possibility of defining a modern textual subtitling standard that would be suitable for embedding in an ISO MPEG or DVB stream.

An option could be to participate in defining character objects of DVB subtitles, so that they could be implemented in an interoperable way in end-user equipments. TTML could then be a candidate for the styled text format embedded into DVB subtitles character objects.

10.6 Wide application for a standardized timed text

As explained in clauses 4.5 and 4.3.2, a standardized time text solution is not only already very much needed today but opens the possibilities for many more functionalities with significant value particularly in the context of 'total communication for all'.

In fact, while a picture or a bit map integrated or transmitted separately from the main picture and sound tracks is difficult to be transposed to several presentation techniques, a text message may be the source for:

- Subtitles.
    - in different languages;
    - using different types of letters;
    - with different backgrounds;
    - as far as the terminal device supports such transpositions (believed to be increasingly easy with the convergence between PC and TV displays and video processing techniques); this may in the future dispense the 'bit map' solution when the level of graphic presentation possibilities in the terminal overcomes the advantage of the 'bit map' presentation.

- Sign language (in 2D or 3D, SD or HD techniques) generated by the user terminal or by a server included in the distribution chain.
- Information to be received over Braille keyboards.
- Audio description for visually impaired people, as far as terminals provide audio synthesizer functions.
- Other facilities.

It is interesting to note that all the above discussion was much centred on the rich heritage of broadcasting associated services, it is having an increasing impact in media distribution offers, like multimedia news on demand offered by the 'printed press' (and, in general, on demand offers) and is very likely to evolve in the future to facilitate, expand and reduce costs of the relay telephony services, since the presently widely human operated relay telephony centers may gradually be completed by well performing machines allowing higher traffic with lower social costs. These relay machines are expected to recognize in a first step the received speech audio signal, transform it to a text and to deliver the text signal or transform it into a language (sign or another language) or message type to enhance the quantity and quality of information addressed to the user.

Annex A: Bibliography

- Editions Rodopi: "Media for all : subtitling for the deaf, audio description, and sign language" ISBN: 978-90-420-2304-8.
- CEA-708: "Digital Television (DTV) Closed Captioning".

History

Document history		
V1.1.1	May 2011	Publication