Multilanguage Opera Subtitling Exchange between Production and Broadcaster Companies

Martínez Barbero, Jesús
OEI, E. Universitaria de Informática
Universidad Politécnica
Madrid, Spain
e-mail: jmartinez@eui.upm.es

Bollaín Pérez, Manuel
OEI, E. Universitaria de Informática
Universidad Politécnica
Madrid, Spain
e-mail: mbollain@eui.upm.es

ABSTRACT
Subtitling is used extensively by broadcasters both for foreign-language subtitling and as an access service to help people with a disability to access television programmes. The European Broadcaster Union (EBU) has create a working group in 2009-01-21 for use the standard Distribution Format Exchange Profile [1] (DFXP) for exchange subtitling information in XML with the Material Exchange Format [2] (MXF) file format. At this moment, in order to help the deaf people, broadcaster produces textual data in a file with a time-code referred of the picture, subtitles in opera emissions help to understand the plot. Live subtitle production require speech recognition or special keyboards where the words are presented as a union of several keys pressed. In this paper, we present an initiative based in B2B W3C recommendation for the exchange, production, and broadcast multilingual subtitle for live Opera production.

Keywords
MPEG2, Multimedia, DTV, Subtitling, Broadcasting, XML, DFXP.

1. INTRODUCTION
The organization and structure of TV broadcasting has three major divisions: production, distribution, and exhibition or diffusion. Production is usually performed by mobile units in the same location where the event takes place. The signal is usually uplinked to a satellite so that the downlink can be performed by any of the channels owning the broadcasting rights for each territory. This is a one-to-many transmission. The broadcasters downlink the satellite signal and they package it with the playout system of their channels to broadcast it through the Media owning the rights. Generally this transmission is one-to-many. Typically, the rights of a live TV event are sold to a broadcaster company for each country.
Opera event has inconveniences in the traditional TV broadcasting; most of the people don't understand operas' concerts and difficulty of understanding the words, international transmission adds another inconvenience: language.
With this work, we propose a system for subtitling opera's live events including translating processes without the necessity of live subtitling in each broadcaster headquarters.
This paper is organized as follows: section 2 surveys the different standards and state of art for subtitling and data transmission used in the section 3 where model is presented. Section 4 concludes the paper by describing the benefits.

2. OVERVIEW
We start this section by providing a brief overview of the use of MPEG2 transport stream.

2.1. MPEG 2 Transport Stream
Despite the different encoding formats which have gradually appeared in the market, the most extended video transmission format for the contribution, distribution and broadcasting of professional quality video signals continues to be the MPEG2 standard [1].
In the said standard, two types of formats are specified, the transport stream and the program stream. The first is used for transmission due to its greater robustness concerning noises in the channel and the second is used for production in environments with low error rates.
The various errors which may occur during transmission of the transport stream are corrected at reception, so as to minimize the effect that these may produce in the image. Multiple methods have been developed for this purpose.

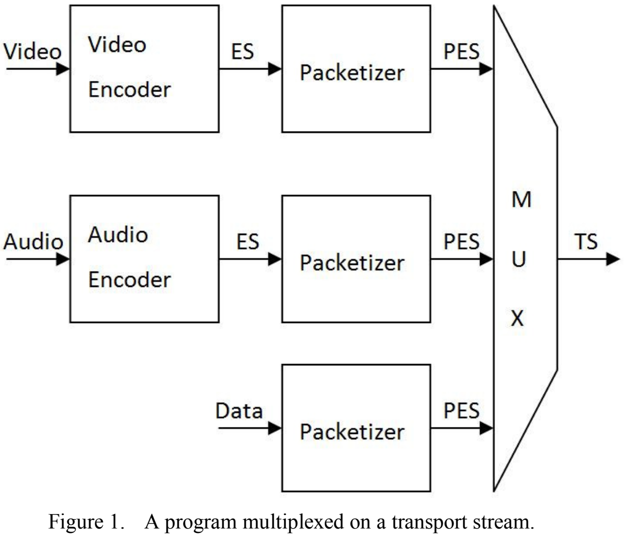

Figure 1. A program multiplexed on a transport stream.

A program comprises several types of data (video, audio, and data) which are encapsulated into elementary streams (ES) and multiplexed in a data stream. Each elementary stream is packaged into Packetized Elementary Stream (PES packet). In order to maintain synchronization between the audio and video data, time stamps are inserted for a correct decoding and displaying of images and sound.
Fig. 1 shows the multiplexing of a video signal, audio signal and other data associated with a program stream as in [4]. The speed of the elementary stream may vary depending on the quality required for the images. In the distribution to broadcaster, the speed may vary from 8 to 50 Mbps. The nature of the images and the transmission purpose will

determine the selected quality. For signal broadcasting, 2.5 to 7 Mbps are generally used.

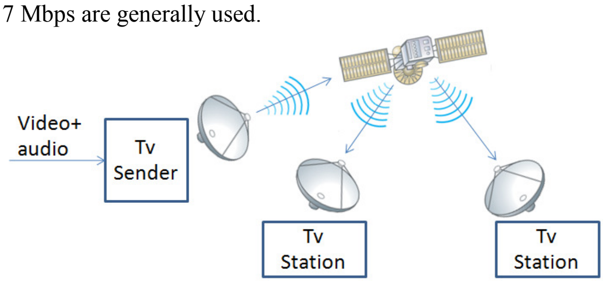

Figure 2. TV distribution.

For the distribution of monitoring channels, either DVB or VoIP, the channels can be compressed to higher rates (i.e. from 4 to 8 Mbps) and several programs can be multiplexed as a single transport stream.
In either case, previous data is accessible from the transport stream which is, thus, generated at the source with the application of inverse operations from the transport stream.
Fig. 2 shows the signal distribution to the different broadcasters.

2.2. File Transfer

Signal distributions are generally performed through dedicated link. This type of video connections is data unidirectional links in charge of transmitting the transport stream. At the reception, it is installed the corresponding demultiplexer.
For file distribution over unidirectional links, there are various file transfer protocols based on retransmission patterns of the same file. The Reliable Multicast Transport (RMT) IETF Working Group deals with the standardization of reliable one-to-many multicast transport protocols.
In [5], a study discusses three types of transfer protocols which can be used in unidirectional networks. The Asynchronous Layered Coding (ALC)[9] does not require any type of feedback from the receivers, and the data are encoded using FEC codes. Repetitions of symbol transfer guarantee the integrity of the file at the expense of diminished effectiveness in the bandwidth.
The Nack Oriented Reliable Multicast (NORM) [10] retransmits only the damaged parts from some of the receptors which send signals of Negative Acknowledgments (NACK) over damaged blocks.
The File Delivery over Unidirectional Transport (FLUTE)[11] [12], based on the ALC protocol, with the extension to be used in any type of transmission channel (unidirectional or not) presents metadata which complete the image signal itself (e.g. File name, codec, etc.). Examples for one-way services can see in [13],[14] y [15].

2.3. Distribution Format Exchange Profile (DFXP)

Currently, EBU has adopted the Timed Text (TT) Authoring Format 1.0 – Distribution Format Exchange Profile (DFXP) and has created a working group for use the standard DFXP for exchange subtitling information in XML.

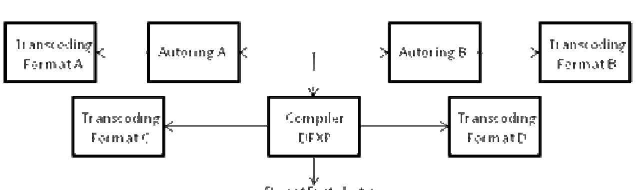

Figure 3. System Model, timed text authoring.

The timed text authoring format is a content type that represents timed text media for the purpose of interchange among authoring systems. Timed text is textual information that is associated with timing information, it serves as a bidirectional interchange format among a heterogeneous collection of authoring systems, and as a unidirectional interchange format to a heterogeneous collection of distribution formats after transcoding or compilation to the target distribution formats as required, and where one particular distribution format is DFXP. Authoring users produces, exchange data, transcode information to different formats and compile to DXFP for distribution to DXFP clients or transcoding to other formats as see in Fig. 2.
A DFXP document contains a header and a body. Header specifies document level metadata, styling definitions and layout definitions; body specifies text content intermixed with references to style and layout information and inline timing information. Below there is an example with DFXP structure, where the head contains the metadata, styling, layout definitions and the body.

```
<tt xml:lang="" xmlns="http://www.w3.org/2006/10/ttaf1">
  <head>
        <metadata/>
        <styling/>
        <layout/>
  </head>
  <body/>
</tt>
```

The body part performs as a container for a sequence of textual content units represented as logical divisions.

```
<body
  begin = <timeExpression>
  dur = <timeExpression>
  end = <timeExpression>
  region = IDREF
  style = IDREFS
  timeContainer = (par|seq)
  xml:id = ID
  xml:lang = string
  xml:space = (default|preserve)
  {any attribute in TT Metadata namespace}
  {any attribute in TT Style namespace}
  {any attribute not in default or any TT namespace}>
  Content: Metadata.class*, Animation.class*, div*
</body>
```

A simple example of content is:

```
<body region="subtitleArea">
        <div>
         <p xml:id="subt1" begin="5s" end="7s">
          How are you?
         </p></div>
</body>
```

Where a subtitle “How are you?” is presented in the image between seconds 5” and 7” when the picture file associated with the subtitle are played. The standard provides more fields in order to indicate other characteristics about the subtitle (position, color, font, etc.).
The text to be displayed is within a “P” element in the DFXP metamodel. A “P” element represents a logical paragraph, serving as a transition between block level and inline level formatting semantics and it has his corresponding identification attribute. This attribute will be used to link original subtitle format and text to the corresponding translated text. There are also two important attributes: begin and duration**.** The XML structure of a “P” element is shown in Fig. 4.

Begin and duration attributes are time expressions that can be a clock time or an offset. The span of time a subtitle is about to be displayed is included in the original file as offset time expressions, setting the begin value to an estimate time and the duration attribute to the corresponding offset (usually, only few seconds more).

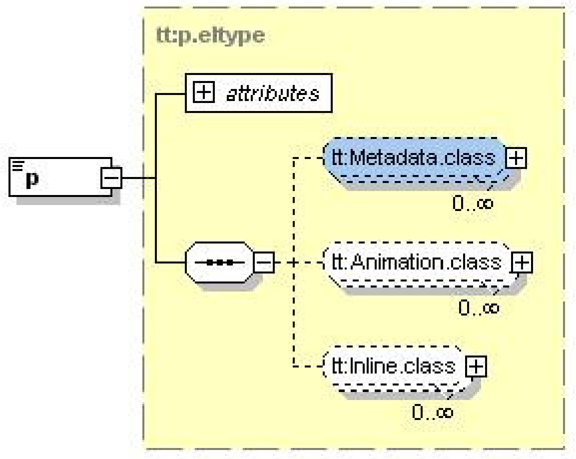

Figure 4. XML structure of a "P" element.

3. PROPOSED SYSTEM

3.1 Workflow

Subtitle live events are usually done both with speech recognition and/or special keyboards, data are sending directly to the playout system in order to add the subtitle data to MPEG2 stream.

New workflow is based in two phases: file and identifier distribution.

1) File distribution. Before the data transmission event producer's send a XML file according with DFXP metamodel to the entire broadcaster involved in the transmission, the file can be transmitted in conventional ways.

Broadcaster translates the information to desired languages; original document can be one of output set. In Fig. 5 we can see the workflow of file from producer to broadcaster in order to translate the text in the desired languages before the transmission date.

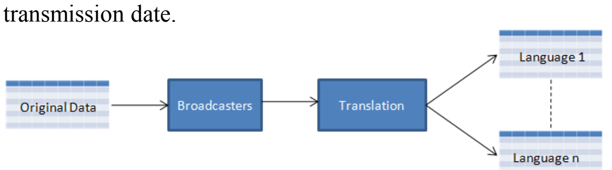

Figure 5. File Distribution workflow.

2) Identification distribution. Identification of each subtitling item is sent by the event producer company as data PES packet inside the MPEG2 transport steam through video link. Each subtitle PES packet can be retransmit several times with FLUTE or NACK protocol in order to avoiding errors in the transmission channel. This protocol has good response with few receptors number [5]. High error rates can cause the impossible to know the identification; the reception can send a NACK to transmitter in order to retransmit the packet.

Live events have to work in real time, if audio are delayed more than 200 milliseconds from video the spectators can notice the delay and the results it's unpleasant but for subtitle case this restrictions is less strong than audio channel, subtitles can be delayed one second maintaining the exposure time.

When the valid identification data is arrived in the TV station, the identification is searching in the language documents for send the appropriated contents depending on the diffusion transmission.

In analogue video transmission, ASCII characters are included in the vertical retrace interval right before the first visible horizontal line. Simple decode circuitry was mandated to be included in all TVs that would provide the extraction and storage of the data and allow the TV user to add the closed caption characters as an overlay to the video in the next fields.

The advent of MPEG compressed video brought with it more possibilities for higher bandwidth closed captioning channel and other supplementary data channels. The European Telecommunications Standards Institute (ETSI) specifies the method by which subtitles, logos and other graphical elements may be coded and carried in DVB bitstreams [16].

Fig.6 shows how Producer Company adds id content into data packets to video stream. Broadcaster extracts the subtitle identifier from transport stream and selects the associated text with its characteristics in order to inserts the subtitle in the image itself and/or send closed captions formats and/or send as data text several language channels for visualize at user demands. Subtitle information can be adapted for each diffusion technologies.

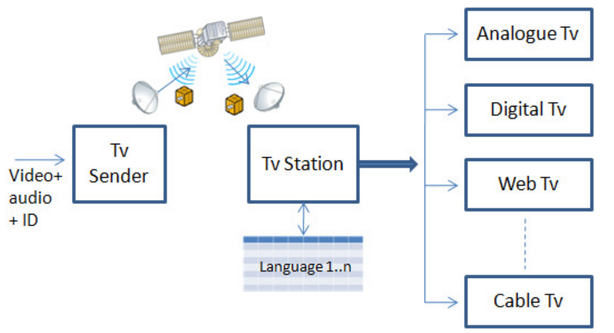

Figure 6. Identification transmission.

3.2 Data Model

For distribution task, Timed Text file include all the necessary data for subtitle but in our case, we don't use *begin* and *end* fields because in live showing the exact moment of subtitle displaying is unknown.

We use the *duration* field in order to indicate the subtitle exposure time. For distribution file, an example of the same subtitle mentioned above is:

```
<body region="subtitleArea">
  <div>
    <p xml:id="subt1" duration="2s"
      Hello Figaro
    </p></div>
</body>
```

The begin field can be added in order to indicate temporal reference, but not for emission task.

```
<p xml:id="subtitle99" begin="30m8s" end="2.0s">
    Are you alone?
  </p>
```

For broadcasting propose in the live transmission, the signal producer sends the *id* field inserted into the video channel as PES packet. Broadcasters receive the identification field and extract the corresponding translated text. Subtitle can be show, with the *dur* exposure time, as a title over the video signal or can be added as closed caption inside the transport stream.

Generally, the shows are stored and we need generate a complete document DFXP into the MXF file. When transmission is in course, begin value will be set in real time according to the show needs, (for example, applause delays or unexpected events). The begin value will be taken from a counter that starts with the show.
The following example shows a processed subtitle to be included, and the next subtitle to be processed:

```
<p xml:id="subtitle98" begin="32m0.76s" duration="1.45s">
    Hello Figaro,
 </p>
```

Note that "subtitle98" is displayed at 32 minutes from the beginning of the show and "subtitle99" is waiting for an appropriate *begin* value. In this example, "subtitle99" *begin* estimate time is "30m8s", that is, the show is delayed two minutes from the estimate timing.

```
 <p xml:id="subtitle99" begin="30m8s" end="2.0s">
   Are you alone?
  </p>
```

The second subtitle will be displayed during two seconds, but the new *begin* value will be set in real time when the subtitle need to be displayed.
The subtitles in different languages are waiting in files, with the same structure than original DXFP document, until their corresponding *id* is received. An example is:
Quien corresponda, busca el id del original (subtitle98) y lo localiza en el fichero de traducción, es decir con:

```
<p xml:id="subtitle98" begin="30m0.76s" duration="1.45s">
    Hola Figaro,
 </p>
```

In this example when the "subtitle98" is received in xml:id
Aquí se nos ha pasado poner el ejemplo para la tradicción. Yo diría que el fichero traducido tiene el mismo contedido que el original, pero con los subtítulos traducios esperando a que se extraigan en su momento con el mismo tratamiento que en ejemplo anterior y pondría el siguiente ejemplo:
Immediately the *begin* value have to be update to the actual time:

```
<p xml:id="subtitle98" begin="32m0.76s" duration="1.45s">
    Hola Figaro,
 </p>
```

Again, "subtitle98" is displayed at 32 minutes from the beginning of the show (two minutes delay) and "subtitle99" is waiting for an appropriate *begin* value. In this translation example, the translated text of "subtitle99" will be extracted from the following translation file item:

```
 <p xml:id="subtitle99" begin="30m8s" end="2.0s">
   ¿Estás solo?
  </p>
```

4. CONCLUSION

TV production and broadcasting model are changing with the introduction of new technologies in distribution processes; the growing of new digital TV channels and the new cheaper communications networks facilitates to share the production between several broadcasters companies.
Event producer companies have the knowledge, broadcaster only retransmitting the video and audio signal with the playout system but usually they don´t adds event information.
Data inclusion inside the video signal help the knowledge transmission between event producers companies to end user while broadcasters only have to adapt video, audio and data content to the different diffusion technologies.
With this work we open an initiative for create a new production services provided from Producers to Broadcaster, this service will help to end user to understand Opera without cost increase.

REFERENCES

[1] W3C Timed Text (TT) Authoring Format 1.0 – Distribution Format Exchange Profile (DFXP). Candidate Recommendation 16 November 2006. Available in: http://www.w3.org/TR/ttaf1-dfxp/
[2] SMPTE 377M-2004. Television — Material Exchange Format (MXF) — File Format Specification
[3] ISO/IEC 13818-1, Information Technology-Generic Coding of Moving Pictures and Associated Audio Information, Part 1: System, April 1996.
[4] Stephan W. Mondwurf, Low Cost MPEG-2 Multiplexing Scheme for Multimedia and Digital TV Distribution Systems, Proceedings of the Fifth IEEE International Caracas Conference on Devices, Circuits and Systems, Dominican Republic, Nov.3-5, 2004.
[5] Neumann, C., Roca, and V. Walsh, R. 2005. Large scale content distribution protocols. SIGCOMM Comput. Commun. Rev. 35, 5 (Oct. 2005).
[6] M. Luby, J. Gemmell, L. Vicisano, L. Rizzo, and J. Crowcroft. Asynchronous Layered Coding (ALC) protocol instantiation, Dec. 2002. Request For Comments 3450.
[7] B. Adamson, C. Bormann, M. Handley, and J. Macker. Negative-acknowledgment (NACK)-Oriented Reliable Multicast (NORM) Protocol, Nov. 2004. Request For Comments 3940.
[8] T. Paila, M. Luby, R. Lehtonen, V. Roca, and R. Walsh. FLUTE - File Delivery over Unidirectional Transport, Oct. 2004. Request For Comments 3926.
[9] U. Reimers, DVB—The Family of International Standards for Digital Video Broadcasting, Proc. IEEE, vol. 94, no. 1, pp. 173–181, Jan. 2006.
[10] U. Reimers, DVB (Digital Video Broadcasting), Springer Verlag Berlin, 2nd Edition 2004.
[11] Bürklin, H., Schäfer, R., and Westerkamp, D. 2007. DVB: from broadcasting to ip delivery. SIGCOMM Comput. Commun. Rev. 37, 1 (Jan. 2007), 65-67.
[12] T. Paila, M. Luby, R. Lehtonen, V. Roca, and R. Walsh. FLUTE - File Delivery over Unidirectional Transport, Oct. 2004. Request For Comments 3926.
[13] E. Pallis, C. Mantakas, G. Mastorakis, A. Kourtis, and V. Zacharopoulos, Digital Switchover in UHF: the ATHENA concept for broadband access, European Transactions on Telecommunications, 17, p.175–182, 2006.
[14] MOBISERVE EU IST project, New mobile services at big events using DVB-H broadcast and wireless networks, http://www.mobiserve.org/.
[15] R. Schatz, S. Wagner, and N. Jordan, Mobile Social TV: Extending DVB-H Services with P2P-Interaction, International conference Digital Telecommunications (ICDT), 2007.
[16] European Telecommunications Standards Institute. 1997 ETS 300 743 Digital Video Broadcasting (DVB); Subtitling systems.