

Requirements for Audio (Video) Description – DRAFT FOR REVIEW

BBC, 19th September 2016

Introduction

Audio Description, also known as Video Description, is an audio service to assist viewers who can not fully see a visual presentation to understand the content, usually achieved by mixing an 'audio description' audio track with the main programme audio at moments when this does not clash with dialogue.

This document describes the requirements for documents needed to support audio description exchange throughout the process from production to distribution, and serve as a basis for verifying that any document format(s) intended to support that process is suitable.

The requirements do not specify details of any particular document format however it is anticipated that some simple additions to [TTML 2] would suffice to meet all of the requirements. Furthermore these requirements do not assume or require that a single document format be used for every step of the workflow, however if this is a practical possibility then it appears to be desirable since that reduces conversion step requirements.

More information about what Audio Description is and how it works can be found at [BBC_RD_WHP051].

References

[TTML2] TTML 2 Editor's Draft: <https://rawgit.com/w3c/ttml2/master/spec/ttml2.html>

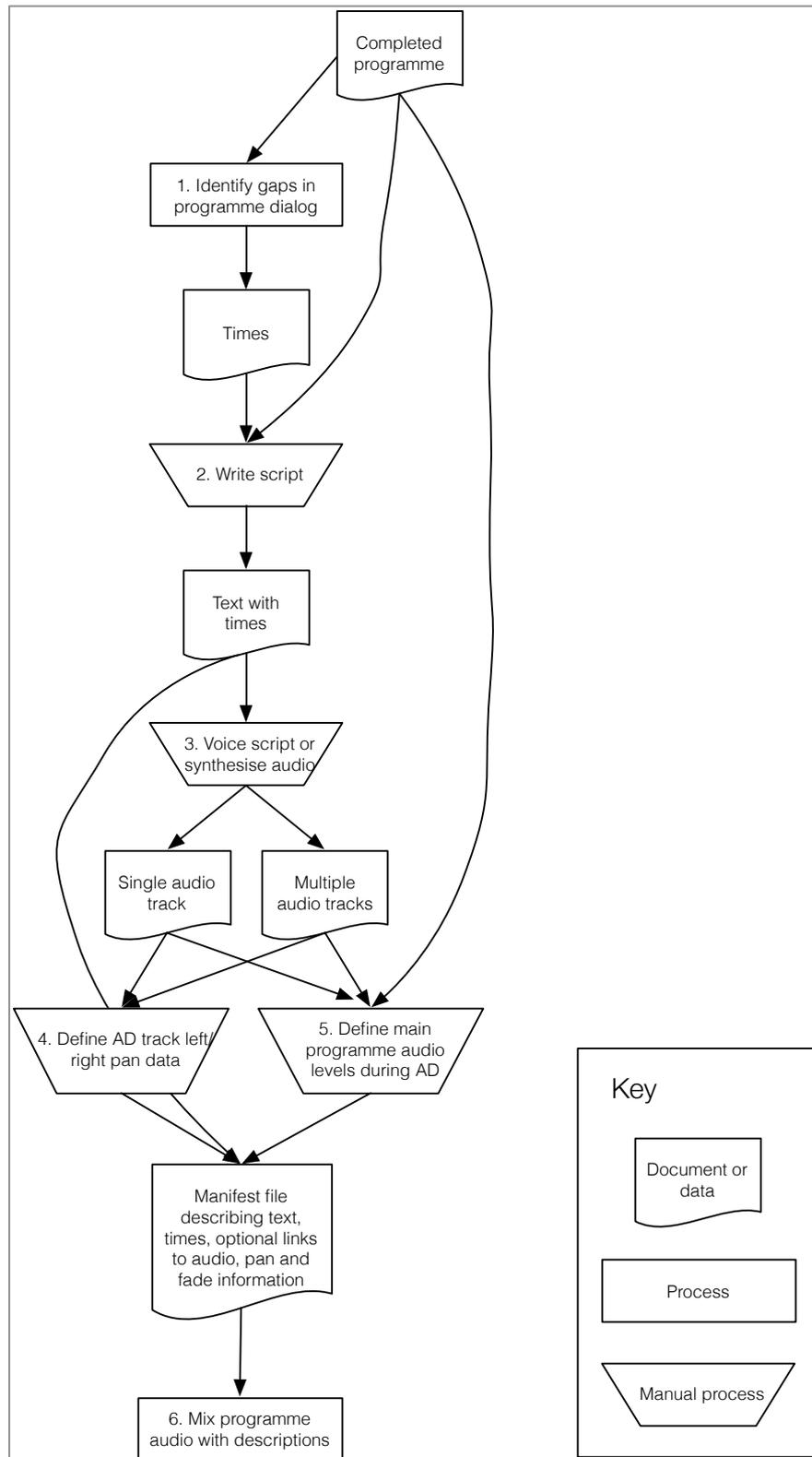
[BBC_RD_WHP051] BBC R&D White Paper WHP 051. Audio Description: what it is and how it works. N.E. Tanton, T. Ware and M. Armstrong. October 2002 (revised July 2004)
<http://www.bbc.co.uk/rd/publications/whitepaper051>

[WEBAUDIO] Web Audio API Editor's Draft: <https://webaudio.github.io/web-audio-api/>

Workflow

The following diagram illustrates the workflow covered by these requirements:

Figure 1 Diagram showing workflow covered by these requirements



At each stage in this workflow the output data may be either inserted into a manifest document such as a TTML document or referenced by it.

Process step	Description
1. Identify gaps in programme dialog	Automatically or manually process the programme audio track to identify intervals within which description audio may be inserted.
2. Write script	Write a set of descriptions to fit within the identified gaps.
3. Voice script or synthesise audio	Generate an audio rendition of the script either by using an actor or voicer and recording the speech or by using a text to speech system. This is typically a mono audio track that may be delivered as a single track that is the same duration as the programme or as a set of short audio tracks each beginning at a defined time.
4. Define AD track left/right pan data	Select a horizontal pan position to apply to the audio rendition of the description when mixing with the main programme audio. This is typically a single value that applies to each description.
5. Define main programme audio levels during AD	Select the amount by which to lower the main programme audio prior to mixing in the description audio. This is typically defined as a curve defined by a set of moments in time and fade levels to apply, with an interpolation algorithm to vary the levels between each moment in time.
6. Mix programme audio with descriptions	Mix the programme audio with the rendered descriptions. This may be pre-mixed (also known as “broadcaster mix”) prior to delivery to the audience, or mixed in real time (also known as “receiver mix”) at playback time; mixing at playback time is a requirement to enable user customisation of the relative levels of main programme audio and descriptions. See [BBC_RD_WHP051] for the reference model for this.

Definitions

The following terms are used in this proposal:

Description	A set of words that describe an aspect of the programme presentation, suitable for rendering into audio by means of vocalisation and recording or speech to text translation.
Main programme audio	The audio associated with the programme prior to any mixing with audio description.
Audio description	An audio rendition of a Description or a set of Descriptions.
Audio description mixed audio track	The output of an audio mixer incorporating the main programme audio and the audio description.

Requirements

The following table lists the requirements at each stage of the workflow:

Requirement number	Process step	Requirement
ADR1	1	The document must be able to define a list of intervals, each defined by a begin time and an end time that are opportunities for adding descriptions.
ADR2	2	The document must be able to incorporate description text to be voiced, each description located within a timed interval defined by a begin time and an end time.
ADR3	2	The document must be able to incorporate additional user defined metadata associated with each description; metadata schemes may be user defined or centrally defined. For example the language of the description may be stored, notes made by the script writer.
ADR4	2	The document must be extensible to allow incorporation of data required to achieve the desired quality of audio presentation, whether manual or automated. For example it is typical to include information about what gender and age voice would be appropriate to voice the descriptions; it is also feasible to include data used to improve the quality of text to speech synthesis, such as phonetic descriptions of the text, intonation and emotion data etc. The format of any extensions for this purpose need not be defined.
ADR5	3	The document must be able to associate a begin time with the beginning of playback of each audio track, for the case that multiple audio tracks are created, one per description.
ADR6	3	The document must be able to associate a begin time with a playback entry time within an audio track, for the case that a single audio track is generated that is the same duration as the main programme audio. The begin time and the playback entry time may be required to be synchronous (coincident values) within the document structure.
ADR7	4	The document must be able to associate a left/right pan value with playback of each or every audio description. This value applies to the audio description prior to mixing with the main programme audio.
ADR8	5	The document must be able to define a fade level curve that applies to the main programme audio prior to mixing with the audio description, where that fade level curve is defined by a set of pairs of level and times and an interpolation algorithm.
ADR9	6	The processor must be able to generate a set of directives to control an audio mixer to generate the desired audio description mixed audio track honouring the pan and fade information within the document. The format of those directives may be implementation dependent.
ADR10	6	The processor may modify the audio mixer control directives under user

		control to customise the relative levels of main programme audio and audio description.
ADR11	6	<p>The audio mixing transitions and semantics must be implementable using [WEBAUDIO], specifically relating to the application of <i>gain</i> and <i>pan</i> as defined therein and the interpolation between values.</p> <p>Note that [WEBAUDIO] specifies three interpolation mechanisms for traversing from one parameter value to another: linear, exponential and linear interpolation between points on a curve, where the default ramp e.g. for <code>setTargetAtTime</code> uses an exponential interpolation.</p>