B B C

R&D White Paper

WHP 135

April 2006

Audible ID3v2 metadata for the visually impaired

J.C.W. Newell

Research & Development
BRITISH BROADCASTING CORPORATION

Audible ID3v2 metadata for the visually impaired

Chris Newell

Abstract

Radio programmes are now frequently downloaded by listeners or delivered automatically to media players using a technique called Podcasting. These media players frequently use ID3v2 metadata embedded in the audio files to catalogue and index the programmes, making navigation and selection of the content easier for users.

This white paper proposes a way to embed audible metadata within an audio file that is equivalent to the existing textual metadata. The audible metadata would allow appropriately adapted media players to provide a greatly improved user interface for the visually impaired.

Additional key words: MP3, AAC

Audible ID3v2 metadata for the visually impaired

Chris Newell

1   Introduction

Radio programmes are now frequently downloaded by listeners or delivered automatically to media players using a technique called Podcasting. The audio files usually support a standard called ID3v2 [1] which allows metadata to be embedded within the file. This metadata can include textual information (e.g. title), images, and URLs pointing to related material.

Textual ID3v2 metadata is frequently used by media players to catalogue and index content using parameters such as the show name and broadcaster. The strings carried in these metadata fields are exploited by the user interface to allow the user to navigate easily through the available content. This is important because the high capacity of modern media players would otherwise make it very hard to find specific items.

For the visually impaired this type of user-interface is far from ideal. This white paper proposes a way to provide audible metadata that is equivalent to the textual metadata. The audible metadata would allow appropriately adapted media players to provide a greatly improved user interface for the visually impaired.

2   Approach

One possible solution to this problem would be for the media player to support computer generated speech. However, this technology can be unreliable and would probably require more processing power than is available in a typical media player. A more satisfactory solution would to provide audible metadata in the form of short audio clips that are embedded in the audio file. These clips would each provide an audio description (i.e. a spoken narrative) of an individual item of textual metadata.

Ideally, the audio clips would be recorded and embedded during the production and authoring process. However, where this has not occurred the audio clips could be added at a later stage by the PC used to download and transfer the content to the media player. Two approaches are possible:

1. Audio clips matching the textual metadata could be downloaded from a third-party website (similar to the way that textual metadata used in ID3 tags is frequently obtained from a third-party CD database).

2. A computer generated speech system could be run on the transfer PC, exploiting the higher available processing power compared to the media player itself.

The provision of audio clips corresponding to any fixed text fields used by the user interface would be the responsibility of the media player manufacturer.

Once embedded in the audio file, appropriately adapted media players would be able to play the audible metadata to accompany the presentation or highlighting of the equivalent textual metadata. This would enable a user to navigate through the typical menus of the media player by sound alone.

Although primarily intended for the visually impaired this feature might also be popular with other users particularly if they have media players with limited display capabilities. This possibility may encourage the development of appropriately adapted media players, particularly if content with audible metadata becomes widely available.

3   Proposal

The best way to achieve the proposed approach would be to extend the ID3v2 specifications to provide a standardised way to embed audible metadata within audio files. This section proposes an appropriate extension.

ID3v2 signalling consists of a chunk of data called a "tag" which is placed at the beginning of an audio file. The tag carries a sequence of "frames" each of which carries a particular item of metadata e.g. the "TIT2" frame which carries the title of the content.

A new frame type is proposed called an "Audio-text" frame which would carry the audible metadata. Each instance of this frame would carry a short audio clip that describes the textual information carried by one of the text frames also present in the tag.

The structure of the frame is shown in Figure 1, where "$xx" represents a byte of binary data. The frame is backwards compatible with version 2.3 and version 2.4 of the ID3v2 standards.

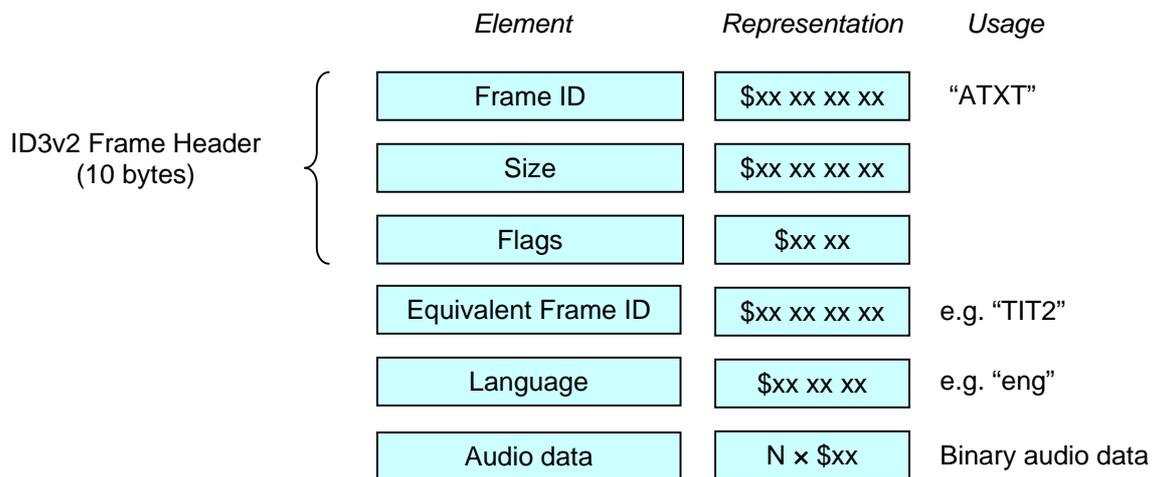|	Element	Representation	Usage
ID3v2 Frame Header (10 bytes)	Frame ID	$xx xx xx xx	"ATXT"
	Size	$xx xx xx xx	
	Flags	$xx xx	
	Equivalent Frame ID	$xx xx xx xx	e.g. "TIT2"
	Language	$xx xx xx	e.g. "eng"
	Audio data	N × $xx	Binary audio data

Figure 1. Structure of the proposed Audio-text frame

- The frame starts with a standard 10-byte ID3v2 frame header with the Frame ID set to "ATXT" using ISO-8859-1 character encoding.

- The Equivalent Frame ID signals the identity of the text frame that the audio-text frame describes (e.g. "TIT2"). This text frame should also be present in the tag.

- The Language field signals the language used in the audio clip. This allows the provision of alternative languages by using more than one ATXT frame to represent a particular text frame.

- The remaining bytes of the frame contain the audio clip.

Each instance of an ATXT frame in an ID3v2 tag would provide an audio clip that is equivalent to a single text frame using one language.

For audible metadata to become usable the proposed ATXT frame would need to be supported by authoring tools, content management tools and media players.

4   Conclusion

An extension to the ID3v2 specification has been proposed which provides support for audible metadata in audio files. This metadata would allow media players to provide improved user interfaces for the visually impaired.

Ideally, the audible metadata would be added during the production process. However, alternative approaches have been suggested that can be used when only textual metadata is available from the original content provider.

Audible metadata is a new concept and may be useful in other situations e.g. in RSS feeds. Further study of potential applications would therefore be worthwhile.

5   References

[1]   http://www.id3.org/develop.html