

Title: Universal Interfaces to Multimedia Documents

Authors:

Helen Petrie¹, Ine Langer², Gerhard Weber³, Wendy Fisher¹, Keith Gladstone⁴, Cathy Rundle⁴, Liesbeth Pyfers⁵

¹Centre for Human-Computer Interaction Design, City University, London, U.K.

{h.l.petrie|w.fisher}@city.ac.uk

²Harz University of Applied Studies, Wernigerode, Germany

ilanger@hs-harz.de

³University of Kiel, Multimedia Campus, Kiel, Germany

g.weber@mmc-kiel.com

⁴Royal National Institute for the Blind, Peterborough, United Kingdom

{KGladstone|CRundle}@rnib.org.uk

⁵Pragma, Hoensbroek, The Netherlands

pragma@home.nl

Contact author:

Helen Petrie

Professor of HCI

Centre for Human Computer Interaction Design

City University, Northampton Square

London EC1V 0HB

email: h.l.petrie@city.ac.uk

phone: +44 (0)20 70408481

fax: +44 (0)20 7477 8859

Universal Interfaces to Multimedia Documents

Helen Petrie, Wendy Fisher
Centre for HCI Design,
City University, London, U.K.
{h.l.petrie|w.fisher}@city.ac.uk

Gerhard Weber
University of Kiel, Multimedia Campus,
Kiel, Germany
g.weber@mmc-kiel.com

Ine Langer
Harz University of Applied Studies,
Wernigerode, Germany
ilanger@hs-harz.de

Keith Gladstone, Cathy Rundle
Royal National Institute for the Blind
Peterborough, U.K.
{KGladstone|CRundle}@rnib.org.uk

Liesbeth Pyfers
Pragma, Hoensbroek, The Netherlands
pragma@home.nl

Abstract

Electronic documents theoretically have great advantages for people with print disabilities, although currently this potential is not being realized. This paper reports research to develop multimedia documents with universal interfaces which can be configured to the needs of people with a variety of print disabilities. The implications of enriching multimedia documents with additional and alternative single media objects is discussed and an implementation using HTML + TIME has been undertaken.

1. Introduction

Vast amounts of information, from novels to recipes, from encyclopaedia to gardening manuals, are now available in electronic formats. This trend to provide information in an electronic format is set to continue and increase, with many new forms for electronic provision appearing. Theoretically, the increased provision of information in electronic formats has great advantages for everyone, particularly people with a variety of print disabilities. The presentation of the information and modes of interacting with it can be adapted to provide universal access for people with differing needs. In practice, a growing number of people have serious problems reading electronic information. These problems involve issues such as orientation in the information space, navigation through the information and usability of the information.

The MultiReader Project is investigating the problems

of orientation, navigation and usability of multimedia documents for both mainstream and print disabled readers. On the basis of a detailed understanding of readers' needs, the project is developing innovative multimodal interfaces incorporating navigational support mechanisms to meet the practical needs of all readers whilst maintaining the pleasure of reading. The MultiReader Project will demonstrate these ideas by developing and evaluating a series of multimedia documents with an adaptable multimodal interface system.

2. User requirements

To understand the reading and navigation needs of both adult mainstream and print disabled readers, a series of focus groups and interviews were conducted. Three focus groups were conducted with a total of 18 participants: one with mainstream readers (comprising 3 women and 4 men, all undergraduate or graduate students); one with totally blind readers (comprising 3 women and 4 men, all used both Braille and audiotape for accessing reading materials); and one with partially sighted readers (comprising 2 women and 2 men, who used a variety of text enlargement technologies for reading). In addition, interviews and questionnaires were conducted with 75 mainstream, blind, partially sighted, deaf¹ and dyslexic readers and with 31 experts in visual impairment, deafness

¹ In the MultiReader Project, the deaf readers were all individuals who were born totally deaf or who became deaf very early in life, and who use a sign language as their first or preferred language.

and dyslexia (the experts included teachers of pupils with special educational needs, rehabilitation officers and producers of alternative media materials). The interviews and questionnaires covered the strategies used and problems experienced by print disabled readers, and the state of the art of technological assistance.

From these data sources, user requirements for the different target reader groups were established:

2.1 General requirements for all readers

- support different reading strategies for different activities
- provide a bookmarking system
- provide easy navigation between table of contents/index and content

2.2 Blind readers

- speech output for all text
- control of speech output: e.g. rewind, skip forward, pause
- volume and rate control
- good descriptions of images and graphics (complete verbal descriptions of these objects)
- audio descriptions in videos (additional commentary to explain visual-only components in the video, to be embedded in the audio stream of the video)

2.3 Partially sighted readers

- vary font style and size
- vary text and background color to increase contrast
- enlargement of images, graphics and video
- speech output to supplement visual output (see also control requirements for blind readers)
- good navigational aids to facility movement around enlarged screens

2.4 Deaf readers

- text or graphic output for all speech, sound effects, music and auditory signals
- extensive use of pictorial, graphic and video material
- online dictionary – in text and Sign Language
- translate text into Sign Language
- vary font style and size
- vary text and background color to increase contrast
- increase line spacing, line length
- word-by-word or sentence-by-sentence highlighting of text
- presentation of information in short and simple “bite sized” chunks for ease of reading and comprehension

2.5 Dyslexic readers

- vary font style and size

- vary text and background color to increase contrast
- increase line spacing
- moving word-by-word or sentence-by-sentence highlighting of text
- speech output to supplement visual output (see also control requirements for blind readers)
- navigation by images/graphics rather than by words
- presentation of information in short “bite sized” chunks for ease of reading and comprehension

Although there is considerable overlap in the reader requirements for the different groups, each group has a particular set of requirements, and within each group there are considerable individual differences. For example, one partially sighted reader may have particular problems in dealing with color contrasts due to color blindness, another partially sighted reader may have no difficulties with color but requires enlargement of the text and graphics. This diversity of reader requirements lends itself to the construction of different *reader profiles* that can be used to configure the multimodal interaction and multimedia presentation for particular readers. In addition, reader profiles can also be used to enhance navigational support through electronic information spaces for the different reader groups and individuals.

3. Multimedia documents

As suggested by ISO 14915 [10], a multimedia document should be designed by structuring it into smaller units and indicating parallel and sequential constraints of each single media object (SMO) (see Figure 1). Thus time dependent (e.g. audio, video) and time independent (e.g. text, graphics) media may be combined.

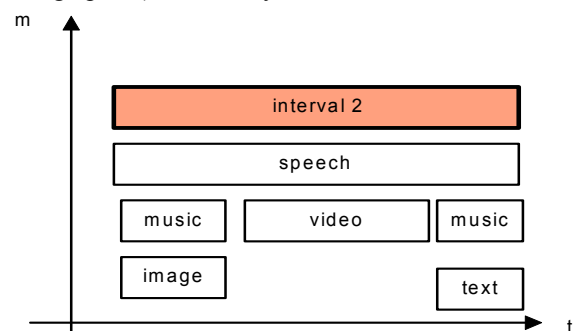


Figure 1: Temporal layout of multimedia titles

The approach proposed by ISO 14915 is well suited to enriching multimedia documents with additional and alternative SMOs to address the needs of readers with a variety of print disabilities. In order to meet the user requirements set out above, the enrichments provided will include:

- descriptions of images and graphics for blind and partially sighted readers
- audio descriptions in video for blind and partially sighted readers
- Sign language translation/interpretation for deaf readers
- text description of speech and sound objects for deaf readers
- images replacing text for dyslexic readers

Note that some of the SMOs which require enrichment are time independent (e.g. images and graphics) and others are time dependent (e.g. video) and that the enrichments potentially belong to the opposite category. Thus, a video (time dependent) may be enriched with a text description (time independent) or by an audio description (time dependent); similarly, a graphic (time independent) may be enriched by a text description (time independent) or by an audio description (time dependent). In addition to these forms of content enrichment, when cross-category enrichment is provided, we will show below that it is also necessary to provide further forms of temporal enrichment in order to support effective navigation and form-based interaction. Temporal enrichment refers to the identification of the context for interaction. Whereas activation of hypertext links is only driven by the timing provided by the reader, time dependent media have their own timelines, as illustrated in Figure 1. Enriching SMOs by other SMOs may require the interpretation of user input (key strokes, mouse clicks or speech commands) in a context-dependent manner based on temporal enrichments. This is necessary to identify the appropriate timeline and therefore effect its ongoing presentation.

Even in within-category enrichment, temporal dependencies of the SMOs may be effected. For example, if a Sign Language explanation is provided for deaf readers of a particular sound SMO, the duration of the signing may exceed that of the original sound object. Thus, the sound SMO will leave a pause which will require filling or it can be delayed in order to ensure parallelism between the two different media streams (see Figure 2). Other SMOs not shown in Figure 2 will in turn be effected by such changes in the timeline.

The enriched document is a truly multimedia document, but presenting all possible SMOs to all readers would render it unusable and counteract the intention of universal usability. For example, presenting dyslexic readers with videos showing Sign Language may be distracting and actually confusing.

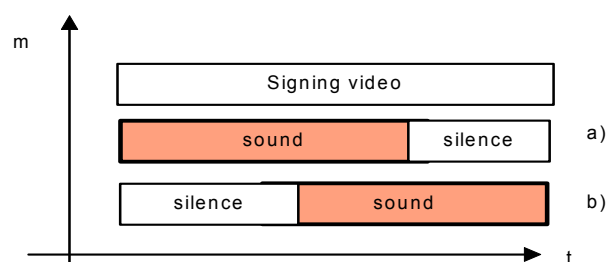


Figure 2. Effects on SMOs: a) pausing b) delaying

Interactive control of additional SMOs such as font variation or color settings as identified in section 2 above adds further to the complexity as the user interface needs to become more universal. Screen readers for blind people are a good example of the enhancement of user interfaces. By providing additional interaction techniques they allow the user to read and navigate through text using speech synthesis, (a time dependent SMO) whereas sighted users read and navigate through text and graphics (time independent SMOs). Screen readers are built for a multimodal user interface in order to allow separate control of the extra interaction objects with speech output and those of applications, for example a browser.

Hence a more flexible approach to multimedia document development is needed, allowing generation of different presentations, depending on identification of the reader profile. This requirement disqualifies some popular technical formats such as TV broadcast, Quicktime, MPEG, Flash, Shockwave, as these do not allow for generation of a document for a particular reader profile. Standard TV broadcast format requires degrading the quality of visual and auditory presentation (see [6] for a discussion of the use of Sign language on television) and hence is undoubtedly not acceptable to mainstream viewers. The only exception to this lack of flexibility is the MPEG approach, which uses extra information channels to provide subtitles for deaf movie viewers and to add audio description to movies for blind people. But MPEG is unable to support other enrichments such as the highlighting of text for dyslexic readers and the enlargement of images for partially sighted readers.

4. Navigating multimedia information spaces

Navigation through electronic information spaces is difficult, whether one has a print disability or not. Over a decade after the phrase was coined [1], people are still getting "lost in hyperspace". Information spaces (physical or electronic) can be complex in structure, regardless of the complexity of the content. For example, everyday materials such as cookery books, tourist guides and gardening manuals can have complex structures that require sophisticated navigational techniques (e.g. going

from the listing of ingredients for a recipe that includes eggs, to the method for using the eggs in that particular recipe, to a text about storage and freshness of eggs in another part of the recipe book, and back to the method for the recipe – this is a trivial task to an experienced recipe book user, but a complex navigation through numerous components of an information space, and probably only a text space at that).

In physical books we have many landmarks and cues to assist us in navigation – intentional landmarks placed by that author such as tables of contents, indices, and headings, or placed by the reader, such as margin notes, and unintentional ones, such as remembering that a particular piece of information is on the top left of a page about one third of the way into the second chapter of the book. One of the problems of electronic information spaces is that we have not yet recreated interfaces with the richness of the landmarks and cues that are available in physical books. A multimodal interface may well address this problem. Multimodal interfaces may well address some of these problems.

Numerous navigational support tools for electronic information spaces have been developed, particularly on the WWW, from simple text-based features such as breadcrumb trails [12] to complex graphic features such as fish-eye views [2]. Although there seems to be a clear need to such support, the effectiveness and usability of such aids is not well established [4]. A further problem for people with print disabilities, particularly those with visual disabilities, is the fact that navigational support tools are almost always highly visual. Even a simple text tool such as a breadcrumb trail, actually works, if it works effectively, because the user can glance at the trail and immediately click on the place to which they wish to return. If the temporal mark-up of multimedia documents is modified by a universal reading interface, navigation is affected as well and hence a document's usability. Transforming landmarks, cues, breadcrumb trails, etc. for implementation within multimedia documents are expected to increase their efficiency and allow for more universal use.

5. Interaction in multimedia documents

Navigation in multimedia hypertext documents is an important interaction technique and has become possible using the Synchronised Multimedia Integration Language (SMIL) [3, 13]. SMIL is an XML application which has been available for some time, the latest version is 2.0 and implemented by Real and Microsoft (see below), amongst other commercial developers, in their latest browsers. Moreover, SMIL 1.0 has been reviewed by the Web Accessibility Group of the World Wide Web Consortium. As a result it allows enrichment of multimedia material

and the selection of SMOs to configure a particular multimedia presentation on the basis of the particular reader's profile. Navigation in SMIL is based on mark-up of links, which may occur in time-independent and time-dependent SMOs [3, 9]. Links to targets within the same document are still possible as well.

5.1 Enrichment of multimedia using SMIL

SMIL is a mark-up language specifying sequential and parallel presentation of SMOs. As is common for most XML applications, mark-up can be nested. A unique feature of SMIL is its ability to specify the timing of SMOs, both in a relative and an absolute manner. Relative timing is based on XML's ID and IDREF attribute values. Figure 3 indicates that a video is delayed by 1.4 seconds in an absolute manner, while the audio starts relative to the video after 0.5 seconds. We refer to this as temporal mark-up.

```
<par>
  <text src="leader.html" region="r1" dur="5s" />
  <video src="v.mpg" region="r2" begin="1,4s" />
  <audio src="c.aiff" region="de" begin="id(r2) (0,5s)"/>
</par>
```

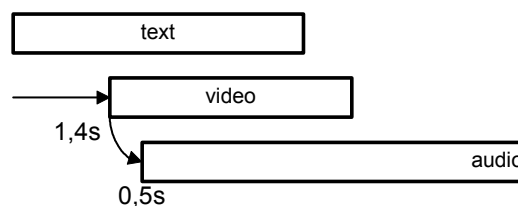


Figure 3: Temporal mark-up in SMIL

If, for example, sign language is inserted at a later point during document development, relative temporal mark-up can be modified locally and more easily.

In order to configure the multimedia document for a particular reader, it is necessary to identify the reader's profile through a *switch* tag and select the appropriate branching of presentation possibilities through attributes such as `systemCaptions` or `systemAudioDesc`. Using these attributes, temporal mark-up and composition of SMOs can be different for different readers, for example deaf readers will have captions and hearing readers will not.

5.2 Mixing Mark-up

Navigation through multimedia documents constructed with SMIL is possible by the use of hyperlinks. Links occur in SMIL multimedia documents as a mark-up element (a-tag) similar to those used in HTML. SMIL links can be valid over a particular time period and act as hotspots in video or audio media streams.

An advanced feature of XML is the use of namespaces

in order to mix document type definitions (DTDs). Internet Explorer is a browser mixing HTML and SMIL through such declaration of namespaces. Starting with IE 5.5, TIME has been included as a partial implementation of SMIL 2.0. As HTML is widely accepted for mark-up of text and images in multimedia documents, the combination of the two DTDs will also allow handling of video and audio in such documents.

In order to introduce advanced interaction techniques, navigation in HTML +TIME can be based on HTML forms. Figure 4 a) shows a sample page with one link labelled "Market Place" and enabled. Reading about sight seeing places through a sequential presentation of link labels is useful for people interested in specific places. Every 4s the link is replaced by another link (see Fig. 4 b)). But readers unfamiliar with the town may start just a tour of their choice by selecting the button. However, even without any input the browser changes the default between Fig. 4a) and b). If the return key is pressed when looking at Fig. 4a) the link "Market Place" is selected. If the return key is pressed while Fig 4b) is shown, then a tour is started erroneously. Figure 5 shows the source code partially.

Figure 4. A dynamically enhanced form

```
<form action="test.php" method="post">
<t:par>
<t:seq repeatCount="indefinite">
<p id="ip0" class="time" dur="4" fill="remove" >
<a href="test.php?place=market">Market Place</a>
</p>
<p id="ip1" class="time" dur="4" fill="remove" >
<a href="test.php?place=castle">Castle of
Wernigerode</a></p>
</t:seq>
<t:seq>
<p> <input type="radio" name="place"
value="1">historic tour</input>
<input type="radio" name="place" value="2">quick
tour</input> </p>
</t:seq>
</t:par>
<p> <input type="submit" name="submit" /> </p>
</form>
```

Figure 5. HTML+TIME code for the form in Figure 4

An analysis of temporal mark-up in Figure 5 identifies the parallel presentation of two different input sensitive time lines. One is alternating the labelled links; the other deals with form input. Both time lines each make up a separate context for input events. Link selection is no longer deterministic and Event Interval Response Systems should be used instead [11].

As a consequence of the limitations identified above, an appropriate implementation of multimedia documents has to ensure suitable interaction techniques according to the reader's profile in addition to provision of redundant SMOs even if standardized DTDs are used.

6. A multimedia tourist guide

To demonstrate the potential of enrichment of multimedia documents, we have developed a tourist guide for the German town of Wernigerode using HTML + TIME suitable for all the reader groups discussed above. In the following section we describe some of the design issues encountered during development and functional testing of this document.

6.1 Tourist guide for blind readers

HTML 4.0 has many mark-up elements in order to provide additional formatted textual contents for blind readers through a screen reader. For example, a table listing fares for various train destinations is read cell by cell enriched with row and column headings. For enrichment of images, we have inserted text through the alt-text attribute. A screen reader using synthetic speech renders this information. Sound SMOs needed no enrichment.

6.2 Tourist guide for deaf readers

Enrichments for deaf people are based on TIME using MS Windows' accessibility options to identify the deaf reader's profile. Figure 6 shows a snapshot of a video with subtitles enabled.

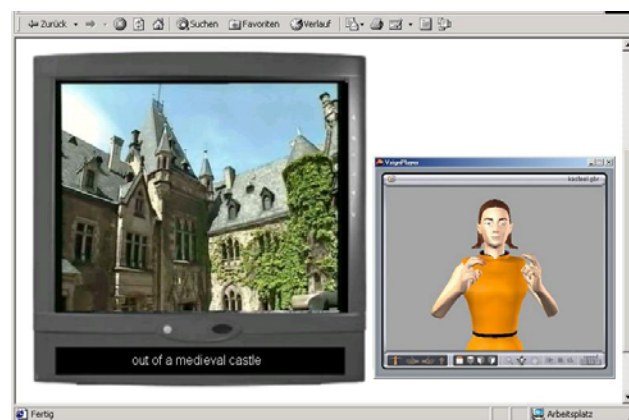


Figure 6. Subtitles and Sign language created in TIME

Subtitles are an enrichment based on HTML and CSS text formatting. Hence single-line and multi-line subtitles are possible using different font styles. The SMIL-based tool Magpie was used to determine timing for subtitles.

6.3 Tourist guide for partially sighted readers

Partially sighted readers may work without an assistive device for screen magnification as these zoom only text appropriately and produce blurred pixels for images and do not enlarge videos. We have enriched HTML through CSS style sheet. Thereby background colour and foreground colours can easily be adjusted for colorblind readers. We developed Scalable Vector Graphics (SVG) versions of images (scanned at 300 dpi) and especially maps. Zooming by variable factors between 1 and 4 hence does nearly not degrade the quality while restricting file size. Videos can be enlarged by factor 2 through features of the playback codec used.

6.4 Tourist guide for dyslexic readers

Font style modifications and line distance variations are possible through CSS. However, a screen reader with text highlighting is still necessary for dyslexic readers of the tourist guide. We have added no text to pages containing videos to make these most attractive to this user group. Videos partially summarize long text pages. Pages describing sightseeing stops have at least one large meaningful photo. Background information on history is considerable and would need further enrichment by videos if readers require it.

A full user-based evaluation will be conducted in the near future of the usability and accessibility of the tourist guide by all the target reader groups.

7. Conclusion and outlook

Multimedia documents with multimodal interfaces have the potential to provide people with print disabilities access to electronic information sources. Readers with different print disabilities have their own profiles of requirements which are needed for successful interaction with and navigation through multimedia documents. SMIL and HTML provide suitable formats for the development of suitable adaptive interfaces and contents which can be configured via reader profiles to meet these requirements. In particular, multimedia documents can be enriched with additional and alternative SMOs to provide the content which readers with print disabilities will require. We have discussed some of the complexities and implications of enriching both time independent and time dependent SMOs.

Navigation in multimedia documents is complex and may lead to non-deterministic behaviour of links if standard mark-up languages are mixed. Taking this into

account for a range of reader profiles requires dynamic generation of interaction techniques. However, we suggest that additional mark-up can provide multimedia documents which are accessible and usable to a range of readers with print disabilities as well as mainstream readers.

Acknowledgements

The MultiReader Project is supported under the IST Programme by the Commission of the European Union (Project IST-2000-27513). The Consortium consists of City University London (United Kingdom), Katholieke Universiteit Leuven (Belgium), Royal National Institute for the Blind (United Kingdom), Federation of Dutch Libraries for the Blind (the Netherlands), Harz University of Applied Studies (Germany), University Kiel – Multimedia Campus (Germany), and Pragma (the Netherlands).

References

- [1] Dillon, A., McKnight, C. and Richardson, J., "Navigation in hypertext: a critical review of the concept", in D. Diaper et al (Eds.), *Human-Computer Interaction: INTERACT '90*, Elsevier, Amsterdam, 1990, pp. 587-592.
- [2] Furnas, G.W. (1986). Generalized fisheye views. *Proceedings of CHI '86*, SigCHI Bulletin, 1(4), ACM Press, New York, 1986, pp. 16-23.
- [3] L. Hardman, D.C.A. Bulterman, and G. van Rossum, "Adding time and context to the Dexter model", *Communications of the ACM*, 31, 1994, pp. 514-531.
- [4] L. Nielsen, *Is navigation useful?*, 2000, Available at: <http://www.useit.com/alertbox/20000109.html>.
- [5] K.R. Page, D. Cruickshank, and D. DeRoure, "It's about time: link streams as continuous metadata", *Proceedings of Hypertext '01*, ACM Press, New York, 2001, pp. .
- [6] Prillwitz, S., Services for deaf people in TV and their reception (*Angebot für Gehörlose im Fernsehen und ihre Rezeption*, in German). Unabhängige Landesanstalt für das Rundfunkwesen, Kiel, 2001.
- [7] Pyfers, L., *Guidelines for the production, publication and distribution of Signing Books for the Deaf in Europe*, www.signingbooks.org, 2000.
- [8] G.G. Robertson and J.D. Mackinlay, "The document lens", *Proceedings of UIST '93*, New York: ACM Press, 1993, pp. 101-108.
- [9] N. Sawhney, D. Balcom, and I. Smith, "Hypercafe: narrative and aesthetic properties of hypervideo", *Proceedings of Hypertext '96*, New York: ACM Press, 1996, pp. 1-10.
- [10] Sutcliffe, A. *Designing multimedia presentations*, Tutorial 7, HCI International '99, Munich, Germany, 1999.
- [11] Weber, G., *Temporal modelling of multimedia interactive systems* (in German), Shaker, Aachen, 2000.
- [12] C.D. Wickens, C. D. "Frames of reference for navigation", in D. Gopher & A. Koriats (Eds.), *Attention and performance XVII: Cognitive regulation of performance*, 1999, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 113-144.
- [13] World Wide Web Consortium (2001). <http://www.w3c.org>