

## DEVELOPMENT AND EVALUATION OF A CROSS-MODAL XML SCHEMA BROWSER

*Dena Al-Thani*

Department of Computer Science,  
Queen Mary, University of London  
Mile End Road, London E1 4NS, UK  
dea1@dcs.qmul.ac.uk

*Dr Tony Stockman*

Department of Computer Science,  
Queen Mary, University of London  
Mile End Road, London E1 4NS, UK  
tonys@dcs.qmul.ac.uk

### ABSTRACT

We describe the development and evaluation of across-modal XML (Extensible Mark-up Language) schema browser. The aim of developing the system is to investigate cross-modal collaboration between users. The browser provides an audio representation of XML schema documents in a way that preserves the structure of documents and supports multi-level navigation. The project has two principle objectives: 1) to overcome the difficulties faced by visually impaired users and sighted people using small screen devices when browsing XML schema files, 2) To explore usability issues when users collaborate using the auditory and visual interfaces of the system. The paper also examines differences between sighted and visually impaired users of the developed auditory interface.

The overall results of the usability evaluations demonstrate that both sighted and visually impaired users were able to perform tasks using the audio modality efficiently and accurately, and the same was true of sighted users interactions with the GUI.

The use of the system to support collaboration where each user employs a different mode (audio or visual) of the system clearly demonstrated that cross-modal collaboration is effectively supported, enabling users to collaborate and successfully complete a complex shared task.

### 1. INTRODUCTION

It has been more than a decade since XML was first introduced as a standard by the W3C (World Wide Web Consortium). XML is used in many applications, the primary one being data exchange between computer applications over the web. It is also used to store data in semi-structured databases. It plays an important role in modern web searching and in the transferr of data to portable devices such as mobile phones and PDAs (Personal Digital Assistants).

As XML has grown in popularity, it has become necessary to provide a schema language that ensures that XML documents satisfy a pre-specified structure. One of the most popular solutions to this problem has been the development of XML editors that provide a graphical tree representation of XML schema documents [1].

Using tree representations, XML editors provide the user with an overview of the XML document based on the schema that is to simplify the process of creating XML documents.

Screen readers present a linear representation of information, and have few mechanisms to provide overviews of information or to facilitate the exploration of data at different levels of detail [2] [3]. This project seeks to address these shortcomings by providing an auditory representation of XML schema information, and enable its efficient exploration at different levels.

### 2. RELATED RESEARCH

The way most visually impaired users browse web pages falls into one of two categories: either by using a screen reader such as Jaws or Window-eyes to render the output from a mainstream browser such as Internet Explorer or Firefox, or using a specially developed audio browser. Most audio browsers such as IBM Home Page Reader, Lynx, pwWebSpeak, etc are almost entirely speech based, and very largely loose information about the spatial layout of page elements. In general screen readers render the information on web pages in a very linear fashion [4].

A number of research efforts have examined in detail the role that non-speech sound can play in preserving the spatial information of web pages and improving the bandwidth of computer-human communication. A study by James [5] which examined the presentation of HTML in audio showed that when presenting hierarchical structures, such as heading levels within an HTML document, earcons proved to be more effective than a simple change in sound level. Petrucci Et Al. [6] developed WebSound, an auditory Web browser for blind and visually impaired users. They demonstrated that the use of non-speech sound in graphical interfaces can increase the bandwidth of computer output. They further demonstrated that a 3D immersive virtual sound environment, combined with haptic manipulation of the audio environment, can enable blind users to construct a mental representation of the spatial layout of HTML documents. James [7] developed the AHA Browser, in which auditory icons are combined with musical cues and speech processing to render web pages in which visual formatting is preserved. Murphy et al. [8] developed a multimodal browser plug-in with audio and haptic feedback, to explore how basic concepts in spatial navigation can be conveyed to web users with visual impairments. Using multimodal cues, users were able to successfully navigate a sequence of screens with directions from a sighted user.

The above research projects focused on how non-speech sound, in some cases in combination with haptics, could

be used to preserve spatial layout and improve the audio presentation of html based web pages. In this project, we examine how non-speech audio can improve the accessibility of the audio presentation of XML schema documents, in comparison with the speech-based approach of current screen readers. We are not aware of any other studies that have applied non-speech sound to XML document presentation. However, the Auditory Display literature provides some guidance on how the different elements in such an auditory display might be chosen. The work by Brewster [9], on the use of hierarchical earcons, suggests that earcons may provide a good candidate for the representation of the tree structured XML schema documents we wish to represent. In this work, Brewster examined the use of earcons in “communicating hierarchical information”. They also investigated how much a user can recall sound representations of hierarchical structures. The results of these experiments indicated that users can recall earcons with a high degree of precision, and so were able to know their position within a hierarchy after only a short amount of training.

Comparative studies of auditory icons and earcons have shown that users can react more quickly to auditory icons than earcons, but the structured nature of earcons enables them to represent more complex information [10]. Clearly speech will continue to play an important role in the display of the XML schema documents in our system, where the specific names or values of XML schema elements must be rendered.

Mynatt [11] synthesized a set of principles and guidelines regarding a non-visual representation of a GUI interface. These can be summarized as follows:

1. Mynatt 1: All the functionality accessible to sighted people using GUI interfaces must be accessible by visually impaired people. That includes icons, images, buttons and spatial location of GUI objects.
2. Mynatt 2: apply good GUI design principles wherever possible such as direct manipulation when implementing a non-visual representation of a GUI.
3. Mynatt 3: Change any interaction device used in a GUI which is not appropriate for use in an auditory interface.
4. Mynatt 4: Mechanisms should be provided to support mutual awareness, i.e. users should be aware of the focus of attention and actions of co-users
5. Mynatt 5: Both non-visual and visual interfaces must support the same mental model. However, Winberg & Bowers [12] argued that having non-visual and visual interfaces which are coherent and have a similar mental model does not guarantee success.

In the design section we will examine the way in which we used the above guidelines to determine the rationale for the design of the XML schema browser.

### 3. SYSTEM DEVELOPMENT

#### 3.1 Choice of schema language

The XML schema language produced by the W3C was chosen because it is widely used and supports the definition of complex document structures.

#### 3.2 Choice of schema style

XML schema documents are written in numerous different styles. To keep the development work within manageable bounds, while providing an adequate test of the approach, it was decided to develop the cross-modal browser for one specific style of XML schema representation. A number of recent studies aimed to classify these styles and identify their strengths and weaknesses [13]. Among the most popular XML schema organization styles are Russian Doll, Salami Slice, Venetian Blind, and Garden of Eden. These styles differ mainly in the way in which they define complex elements. [13] The chosen style is the Salami Slice. In the salami slice style, the complex elements can only contain references to simple and complex elements which are defined in the first level of the XML schema document. The reason for choosing the Salami Slice style is that according to the W3C schools website, it is the most widely used [14]. Additionally, studies have shown that it is conceptually simpler than the other organizational styles of XML schema. Further, it supports the reuse of elements within the document [13] [15].

## 4. DESIGN

We first describe the visual interface design. We then detail the design of the auditory interface, before examining the role that the guidelines developed by Mynatt [11] played in assisting the process of mapping from the visual to the auditory design.

#### 4.1 Visual Interface Design

In the graphical interface, two main components need to be displayed. The first is the graphical representation of the XML schema. The second is the control panel which contains the available functionality to support browsing. The screen is divided in to three parts. The top part of the screen contains the buttons for loading an XML file, getting help and listing to sound samples. The lower right part of the screen contains the graphical representation of the XML schema document, and the left part of the screen contains the buttons that are used to get node details and to move from one node to another. Figure 1 shows the initial GUI design.

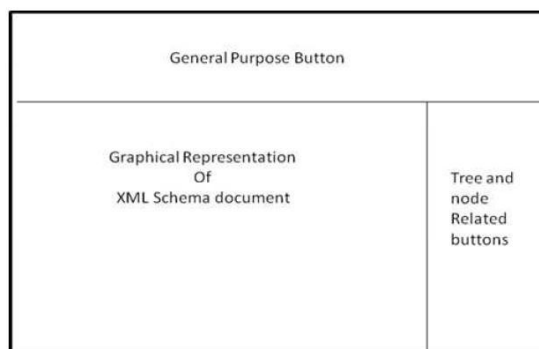


Figure 1: The Visual Interface Design

#### 4.2 Auditory Interface design

Because of the complexity of the documents to be represented, we determined to support two views of each schema document:

1. An overview of the XML schema document, to give the reader a sense of the document's size and complexity. This is particularly important given the serial nature of audio. Providing an overview of complex data has been the topic of several research papers [16] [17] in which it was emphasized that it should not be necessary for users to listen to long streams of audio to gather an overall view of the major features of the represented data.

2. A detailed view of the XML schema document in order to present its structure and detailed contents. Audio samples and video demo of the XML schema browser can be downloaded from (<http://www.megaupload.com/?d=8W9QV5DU>)

#### 4.2.1 Use of Speech

Speech is an essential component in this application, as this is the only way to represent certain types of essential information using sound such as the names of nodes in XML schema documents. Speech was used substantially in both the overview and detailed display components of the auditory interface of the browser. In order to help the user keep track of their current depth in the tree, the pitch of the speaker's voice is modified to represent the level of the element within the XML schema tree. As the user navigates to lower levels of the tree structure, the pitch of the spoken voice is decreased, and visa versa.

#### 4.2.2 Non-speech sound

For Non-speech sound, two representation techniques were used auditory icons and earcons. Samples of the elements used in the auditory display are included with the paper and will be presented at the conference.

*Auditory Icons:* Auditory icons have been used in many parts of the application to represent a number of components of an XML schema document. In particular these were used because their sounds are very distinct and life-like, and where for example there was a direct word association with the schema element being represented. Examples of the use of auditory icons are as follows:

- a) The sound of a car braking is used to represent restrictions and limitations associated with simple elements and attributes.
- b) The sound of keys clinking is used to represent ID, IDREF, and IDREFS types in an XML schema. ID usually represents a primary key while IDREF or IDREFS represent foreign keys in a database.
- c) The sound of a water bubble is used to represent an attribute; Attributes are always linked with complex elements. In order to differentiate between the child elements of a complex element and its attributes the water bubble sound is used through its association with the bubble symbol often used in visual diagrams. It can be argued that this is a less direct association for a visually impaired user, who may or may not know of the use of the bubble symbol, but it is an attractive and memorable sound which should be relatively intuitive to sighted users of the system using PDAs.

*Earcons:* Auditory icons are easier to recall, but studies [9] have shown that in some cases, performance becomes more efficient when using earcons. Earcons are used to inform the user about the number of child elements of a complex element. They are played in prior to the complex element name. They are produced at runtime. The numbers of

musical notes in an earcon represent the number of child elements of the complex elements. In this project Earcons are also used to notify the user that the end of a tree branch has been reached. Earcons were produced using a combination of Csound and Audacity.

Concurrent presentation of auditory information is used to reduce the pace differential between browsing in audio and visual modes. The application provides concurrent audio and visual feedback when a button is clicked. In addition, while traversing within the XML schema tree, concurrent feedback of speech and non-speech sound is heard by the user in order to overcome the serial nature of audio information. For instance, when users navigate to the next element in the tree, non-speech sound heard before the element name indicates that it is a complex element. This non-speech sound also indicates the approximate number of child elements that exist.

#### 4.3 Applying Design Guidelines

As mentioned in the Related Research section, Mynatt 1997 represents one of the most detailed attempts to provide general guidelines about designing auditory interfaces that deliver equivalent functionality and usability as their GUI counterparts. We examine below how we applied Mynatt's [11] guidelines in the development of our system:

- Mynatt 1: All the functionality accessible to sighted people using GUI interfaces must be accessible by visually impaired people. That includes icons, images, buttons and spatial location of GUI objects. This was achieved by mapping visual objects to appropriate auditory objects, principally auditory icons, earcons and static and dynamic speech elements. Through the overview mechanism we tried as far as possible to provide a summary of schema documents which would give some idea of their size and complexity, providing some of the characteristics available to a sighted user when over-viewing the document on screen.
- Mynatt 2: apply good GUI design principles wherever possible. The position we took on this guideline was that the strengths and weaknesses of audio and graphical representations are very different, and that what works well in a GUI will not necessarily translate intuitively to an auditory interface. For example, brackets are widely used in schema specifications to indicate nesting, but these were not reproduced directly in the auditory display, but audio users are provided the equivalent information through the audio context described in terms of speech, auditory icons and earcons.
- Mynatt 3: Change any interaction device used in a GUI which is not appropriate for use in an auditory interface. We adhered to this principle by substituting the keyboard for the mouse when navigating schema documents and ensuring common navigation options are supported by hot key combinations.
- Mynatt 4: Mechanisms should be provided to support mutual awareness, i.e. users should be aware of the focus of attention and actions of co-users. We adhered

to this principle by ensuring that the presentation of schema information is always synchronized between the visual and audio interfaces.

- Mynatt 5: Both non-visual and visual interfaces must support the same mental model. We adhered to this principle by using the tree structure as the basis of schema representation in both the audio and visual interfaces. In the audio interface, users start from the top level, and have a choice either to continue to navigate the schema at the level of complex elements, or whether to open up successive amounts of detail on demand. Navigation of the tree structure and synchronous cross modal presentation of information is supported by presenting the element name and highlighting it when the position has changed. The user is also able to change the position at anytime using buttons or equivalent keyboard shortcuts.

Concurrent presentation of auditory information is used to reduce the pace differential between browsing in audio and visual modes. For example, while traversing within the XML schema tree, concurrent feedback of speech and non-speech sound is heard by the user in order to overcome the serial nature of audio information. For instance, when the users navigate to the next element in the tree, non-speech sound heard before the element name indicates that it is a complex element. This non-speech sound also indicates the approximate number of child elements that exist.

## 5. IMPLEMENTATION

The system was implemented using Java on a PC platform, using the Java Speech API (FreeTTS) and the DOM (document object model) API for representing XML schema documents as tree structures. The earcons were created by calls to the Java Sound API (MIDI), while auditory icons were presented by playing pre-recorded sounds using the Java Sound API. In a number of situations requiring only static speech, we pre-recorded the speech and made use of better quality TTS engines such as Verbose by NCH Swift Software and VoiceMax by Tanseon systems.

In the case of speech sound, Free TTS is used to represent the runtime data which is the XML schema tree. It is used to give information about the current element. When a child element of a complex element is represented, the pitch of the Free TTS voice is slightly lowered to differentiate between a child element and its parent element. From pre-recorded sound software, two voices were chosen. The voice of a female was used to represent indicators in XML schema documents and the voice of a male was used represent the buttons. Echo was added to the male voice in order to distinguish between the male voice that represents the buttons and the male voice of Free TTS.

For non-speech sound, MIDI was used to represent complex elements. Two audio representations of complex elements are designed for this purpose. For both representations, the MIDI sound is played prior to the name of the complex element.

The first represents complex elements with child elements that are less than or equal to three. In this

representation, the number of the repetitions of the MIDI notes indicates the number of child elements, thus allowing the users to know the number of child elements without needing to go to each child element.

The second represents complex elements with more than three children. In this representation, a major chord of four notes is played using two instruments. The reason for coming up with an alternative solution to represent complex elements with more than three children, is that while prototyping the first representation with participants, it was noticed that repeating the MIDI notes helped the participants to know the number of child elements. However, when the number of child elements are larger than three it started affecting the user's performance time. As the number of notes increased, the time of playing the MIDI sound increased. Therefore, a better alternative was needed.

## 6. EVALUATION

We were fortunate in having ready access to users and Formative evaluations of early prototypes guided the design of the system, but the results described here come from a more detailed, summative evaluation.

The main goals of the usability experiments are to find out whether the audio representation of the XML schema documents is able to provide a way for visually impaired users and sighted users who used small devices such as mobile phones and PDAs to work with XML efficiently, and whether the audio and visual interfaces together can support cross modal collaboration. In addition, while conducting the usability experiment we also aimed to compare the auditory XML browser interface with a screen reader that visually impaired users use to read XML schema documents.

### 6.1. Auditory Interface Usability Evaluation

#### 6.1.1. Hypotheses

Since the goal of using this approach is to determine the usability of the system, two hypotheses were defined:

*Hypothesis 1:* Using this interface, users are able to obtain a useful understanding of the nature, application area and major components of a schema document.

*Hypothesis 2:* Using this interface, users are able to navigate efficiently to appropriate parts of the schema document in order to perform tasks such as information seeking and compare schema elements.

To test the first hypothesis, we needed to examine whether the auditory interface allows the user to have an effective overview of the information presented in the XML schema document. This is tested by asking participants to listen to two audio presentations of the schema by the system, and asking them a set of general questions about the schema. These questions ask about the size and application area of the given schema as well as the numbers of complex and simple elements. The second hypothesis is tested by investigating the efficiency of the navigational features of the auditory interface. This is achieved by allowing users to navigate around the schema as much as they wish, while asking them a set of questions focusing on low level details of the schema, such as finding the

details of specific elements, determining the number of IDREFs, and navigating to child elements.

### 6.1.2 Procedure

Nine sighted and four visually impaired participants were recruited. They were all given a sufficient amount of training prior to conducting the evaluation which took from 15 to 40 minutes. The primary factor behind the variability in training time was the user's previous knowledge of XML.

The participants were asked to use the interface to answer two sets of questions. The first set of questions was used to determine the participant's ability to get an overview of the schema, therefore the questions were quite general such as the number of simple elements, the number of complex elements and the domain the XML document is related to. Whereas the second set of questions determines was used to determine their ability to understand the schema in details. It contains questions that ask the participants to navigate to a certain node and write down the names of its child elements, attributes or primary key. The schema given to the participants in training and evaluation ranged from medium to large, where we defined a medium schema to have from 5 to 10 complex elements and from 5 to 10 simple elements, and a large schema we took as having more than 10 complex and more than 10 simple elements. The maximum schema that was given has 17 complex elements and 25 simple elements.

### 6.1.3 Analysis of Results and Discussion

#### General Observation

The participants' overall performance was fairly good with a very low error rate in both set of given questions. For the first set of questions the visually impaired participant error rate was 0% and the sighted participants' error rate was also extremely low, 3.8%. For the second set of questions which examine the participants' ability to understand the schema details the error rates were also low for both groups of participants. The average error rate for visually impaired participants was 20% and the average error rate for sighted participants was 13.8%.

Training times ranged from 15 minutes to 40 minutes while the overall task performance time ranged from 20 to 50 minutes. From direct observation of the interactions and discussion with users, there were a number of external factors that affected the individual training time and performance. These were as follows:

1. Computer literacy: Computer literacy played an important role in the overall user performance time. Two visually impaired and one sighted participant, who were less familiar with HTML and XML than the other participants, took longer to perform the tasks.
2. Experience with auditory display: Sighted Participants who had not previously used an auditory interface showed some hesitation and confusion at the start of the training session. This was expected as studies have shown that representing complex data needs in-depth training (Brewster, 1994) (Vickers and Alty, 1996). In addition, some sighted participants had problems remembering some of the non-speech sounds, in particular the attribute sound, whereas visually impaired participants had no difficulties remembering them.
3. Familiarity with XML documents: Participants who were less familiar with XML documents had some difficulties in differentiating between simple elements and attributes.

However, they developed a better understanding as they worked through the tasks. Additionally, it was noticed that participant performance in a given task can sometimes be affected by not having prior knowledge of XML. For instance, in some cases the participants forgot that a complex element has child elements.

Apart from the above factors affecting performance times, the participants performed as expected. All participants were able to relate auditory icons to the XML schema components which they were intended to represent. By listening to the earcon which sounded before the complex element name, they were able to identify the number of child elements belonging to the complex element. Both visually impaired and sighted participants made use of the relationship between the pitch of the voice and their current level in the tree.

#### Summary of the results of Experiment 1

The main findings revealed that both visually impaired and sighted participants performed well on both the overview and detailed navigation tasks. Both were able to develop a good overall understanding of the XML document with low error rates. This supports the first hypothesis, as all the participants developed an adequate understanding about the schema size and nature of the schema. It was clear that the differences between participant performance times were due to the external factors described in the general observations section.

When examining the participant's ability to get a sufficient understanding of the XML tree details, there are a number of variables that affected the performance of the users and therefore had a bearing on the second hypothesis. Firstly, the scores were affected by the external factors explained previously. Secondly, training played an important role in this part of the experiment. Thirdly, the learning curve had an influence on the subjects' performance. Figure 2 shows the results of three participants in three trials. In each trial the participant was given a different schema, but they were of the same level of complexity. Complexity is defined here as a composite measure incorporating the number of complex elements, simple elements, attributes, and restrictions within an XML schema document. For the three participants the scores in the second trial were higher.

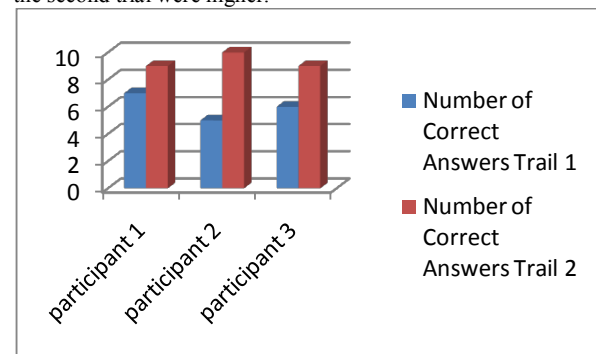


Figure 2: Three Participants Results in Two Trials

## 6.2 Cross-modal Interface Usability Evaluation

### 6.2.1 Hypothesis

Since it is a cross-modal system the aim was to enable users using the audio interface only and users using the visual interface only to be able to work together and have a similar

mental model of the system. In other words, their understanding about the main components of an XML schema document should be similar and so allow them to work together coherently. The hypothesis to be tested was as follows:

*Hypothesis 3:* That the two users are able to collaborate and to develop an accurate representation of the XML schema they are browsing.

#### 6.2.2 Procedure

For this experiment we recruited participants in pairs. Three pairs of sighted participants were recruited. Participants were trained on either the visual or audio interface.

The participants were then asked to use the interface on which they were trained until they felt comfortable using it.

Following this, Participants worked in pairs. For each pair, one participant only used the auditory interface and the other participant only used the visual interface. In the experiment, both participants using the different modalities were given the same schema. Their task was to work together to create an XML document that satisfied the structure defined by the schema.

They used different computers and were seated with their backs to each other as they were also informed that they cannot view the other participant's task sheet at anytime and that the only way to communicate with the other participant is via direct conversation. Their conversations were then recorded and analyzed. It was made clear to the participants that they can plan their work collaboratively in the way that suits them, as the experiment is mainly focused on the result of the collaborative work rather than the process of their collaboration.

#### 6.2.3 Analysis of Results and Discussion

Generally, the collaborative task was performed well, with an average time to complete the task of 10.5 minutes. It was observed that the participants in the collaborative work did not face any difficulties while trying to explain information related to a specific element in the schema tree.

An interesting observation was that in all three experiments, the participants using the visual interface started the conversation first and tried to lead the collaborative work.

However, around the middle of the process the participant using the visual interface stopped leading the work and both participants started working together more evenly. The reason might be that both participants were not familiar with auditory interfaces. Therefore, the participant using the audio interface was more hesitant at the start than the participant using the visual interface, allowing the participant using the visual interface to lead the work. Once the audio interface participant became familiar with the interface and had gained more confidence, then both participants took part in the work more evenly.

It was clear that individual and collaborative performances improved with time. The participants' collaboration work became better towards the end of the process, as they became more familiar with the system.

The most important observation relates to the result of the overall task, as both participants were able to create XML documents correctly. Even though participants using the audio interface had not seen the XML schema structure, they were able to create an XML document that satisfied the given schema.

However, the XML documents created by non-XML users were, not surprisingly, less efficient, but nonetheless they did

demonstrate a good understanding of the structure presented to them through the interface of the schema browser. An example of the kind of error these users made was in differentiating between attributes and simple elements.

### 6.3 Usability Experiment comparing schema reading using the XML browser with reading schemas using a Screen Reader.

#### 6.3.1 Experiment Design Research Hypothesis: (Hypothesis 4)

The use of speech and non-speech sound in the schema browser to represent XML schema documents is more efficient compared to reading schemas with a screen reader.

#### 6.3.2 Procedure

Four visually impaired participants were recruited. All participants are experienced JAWS screen reader users and all are computer literate. However, they have little knowledge of XML. They were all given 20 to 40 minutes training on the schema browser's auditory interface.

The participants were asked to review three XML schemas using the JAWS screen reader, and three other XML schemas using the audio interface. The XML schemas reviewed using the screen reader are different than the ones reviewed using the XML browser. That is to avoid any bias results, as reviewing the XML schema using one tool, will effect the participants performance when reviewing the same XML schema using the other tool. After reviewing each XML schema, participants were asked to describe the XML schema. They were also asked to give their comments regarding both tools.

#### 6.3.3 Analysis of Results and Discussion

Due to the fact that all the participants are experienced screen reader users and have a modest knowledge in XML, external factors such as computer literacy, and knowledge in XML were fairly constant across the 4 users. From the timings collected in the experiment, it was clear that the time that it takes a participant to review and fully understand a schema using a screen reader was longer than the time it takes another participant to review the same schema using the audio interface. The figure below (figure 3) shows the time it took the participant to review a schema using both tools. Schemas were numbered according to their complexity, with one being the most complex with larger numbers of complex and simple elements and six being the smallest. From the figure it is evident that with larger schemas the average time taken to review using a screen reader was almost double the time it took the participants to review the same schemas using the XML schema browser.

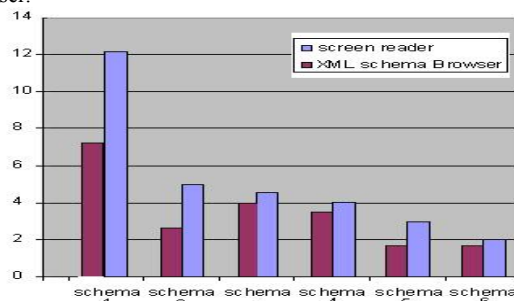


Figure 3: The average time taken using the XML schema Browser and the Screen Reader.

All participants were able to describe the content of schemas with differences depending on the tool they used. It was noticed that with complex schemas participants when using screen readers seemed to be less clear on the overall structure of the schema, whereas participants who were using the XML browser on the same schema demonstrated a good understanding about the structure of the schema as well as its details.

The above outcomes support the hypothesis that within the small number of participants involved, the combination of speech and non-speech sound provided in the schema browser is able to enhance the users' performance in comparison with screen readers.

## 7. DISCUSSION AND CONCLUSIONS

This paper introduced a novel approach to represent XML schema documents in audio. It also examined a number of ways of representing complex data in audio. Evaluations have examined the use of the audio interface of the browser used alone, and the use of the visual and audio interfaces by 3 pairs of users performing a collaborative, cross modal task.

The results of the evaluations demonstrate that the audio interface was successful in supporting audio browsing and cross modal collaboration for the relatively small numbers of users involved in the trials.

The benefit of the approach taken can be summarized as follows:

- 1) The audio XML browser helps to overcome the problems visually impaired XML users face when using screen readers. Given the serial nature of sound, screen reader rendering of XML schema contain a number of repetitive and unnecessary symbols that can overload the user's short term memory, which may affect the user's understanding of the structure of the data.
- 2) The browser enables rapid identification of the XML schema structure, and gives the user the option to get more details on demand.
- 3) The use of auditory icons and earcons provide a concurrent presentation of the properties of the elements which help to improve use of the communication bandwidth between the computer and human, rather than presenting these elements serially as they are when read with a screen reader.
- 4) Rather than representing the data serially, the data is represented in three different levels, leaving the user to match the level of detail to the task.

Additionally, Evaluation of the system for cross modal collaboration suggests that once users of the auditory interface have become comfortable with its use, they are able to take a full part in the collaborative task and that both users are able to form a sufficient mental model of the structure of the xml document described by the schema to be able to synthesize a document that accords with the underlying schema.

As well as performing the experiments described here with more users, an important remaining experiment is to test the collaboration between visually impaired and sighted users. However, given the results so far, we anticipate that the results of this experiment will be at least as good as the collaborative experiment described here, as visually impaired users in general

will not have the difficulties of lack of familiarity with the auditory interface experienced by sighted users in the early part of the collaborative experiment described above.

## 8. ACKNOWLEDGMENT

Many Thanks goes to all who helped us from our colleagues.

## 9. REFERENCES

- [1] D. Mertz, *XML Matters: A roundup of editors. IBM's Developer work*.2001. Website: <http://www.ibm.com/developerworks/library/x-matters6.html>.
- [2] L.Brown, S. Brewster, R. Ramlohl, M. Burton, M. and B. Riedel. "Design Guidelines For Audio Presentation Of Graphs And Tables". *Proc. of Conf. on Auditory Display (ICAD 2003)*, Boston,USA, July2003.
- [3] T. Stockman, G. Hind,C. Frauenberger. "Interactive Sonification of Spreadsheets". in *Proc. of the 11<sup>th</sup> Int. Conf(ICAD)*. Limerick, Ireland. 2005. pp 134-139.
- [4] T. Stockman and O. Metatla. "The Influence of Screen Readers on Web Cognition", *Proc. of Accessible Design in the Digital World*, 2008.
- [5] F. James, "Presenting HTML structure in audio: User satisfaction with audio hypertext," *In Proc. of Conf. on Auditory Display( ICAD) ,Xerox Parc,1996*.
- [6] L. Perrucci, E. Harth , P. Roth , A. Assimacopoulos and T. Pun, "WebSound: ageneric Web sonification tool, and its application to an auditory Web browser for blind and visually impaired users", In Proc of the 6th Int. Conf. on Auditory Display (ICAD 2000), 2-5 April 2000.
- [7] F. James, 1998 "Lessons from developing audio HTML interface", *Proc of the 3<sup>rd</sup> Int. ACM Conf. on Assistive technologies*, Marina del Rey, CA, United States, pp.27-34. April 15-17, 1998
- [8] E. Murphy, R. Kuber, P. Strain, G. McAllister and W. Yu. "Developing sounds for a multimodal interface: conveying spatial information to visually impaired webusers", *Proc. of Int. Conf. (ICAD)*, p356-363. 2007.
- [9] S. A. Brewster, "Using Non-Speech Sounds to Provide Navigation Cues". *In Proc. Of ACM Transactions on Computer-Human Interaction* ; 5, 3: 224-259. 1998.

- [10] M.P. Bussemakers and A. de Haan, "When it Sounds like a Duck and it Looks like a Dog... Auditory icons vs. Earcons in Multimedia Environments", in *Proc of the 6th Int. Conf. on Auditory Display(ICAD)*, Atlanta, US, 2000.
- [11] E. Mynatt, "Transforming graphical interfaces into auditory interfaces for blind users", *Human-Computer Interaction*, 1997, vol. 12, pp7-45.
- [12] Winberg, F & Bowers, J. (2004) Assembling the Senses: Towards the Design of Cooperative Interfaces for Visually Impaired Users. In *Proc. of CSCW'04*, Chicago, Illinois, USA, November, 2004.
- [13] R. Lämmel, "Style normalization for canonical X-to-O mappings". *Proc. of the 2007 ACM SIGPLAN symposium on Partial evaluation and semantics-based program manipulation*.
- [14] <http://www.w3schools.com/>
- [15] Khan, A., and Sum, M., Introducing Design Pattern in XML schemas.2006. Retrieved on: 25th July, 2009. From sun developer network. Website : [http://developers.sun.com/jserver/archives/nb\\_enterprise\\_pack/reference/techart/design\\_patterns.html](http://developers.sun.com/jserver/archives/nb_enterprise_pack/reference/techart/design_patterns.html)
- [16] Zhao, H., Plaisant, C., Shneiderman, B. and Duraiswami, R. "Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study". In *Proc of the Int. Conf. on Auditory Display (ICAD)*, (2004).
- [17] Finlayson J, and Mellish, C., (2005) "The 'Audioview' – Providing a Glance at Java Source Code," in *Proc of Int. Conf. on Auditory Display (ICAD)*, Limerick, Ireland, July 2005, pp. 127-133.