

# Using Hierarchies to Support Non-Visual Access to Relational Diagrams

Oussama Metatla, Nick Bryan-Kinns, Tony Stockman

Interaction Media and Communication Research Group

Department of Computer Science

Queen Mary, University of London

Mile End Road, London E1 4NS, UK

{oussama, nickbk, tonys}@dcs.qmul.ac.uk

## ABSTRACT

This paper describes an approach to support non-visual exploration of graphically represented information. We used a hierarchical structure to organize the information encoded in a relational diagram and designed two alternative audio-only interfaces for presenting the hierarchy, each employing different levels of verbosity. We report on an experimental study that assessed the viability of our proposed approach as well as the efficiency and learnability of each interface. Our results show that the relational information encoded in a diagram could be non-visually navigated and explored through a hierarchy, and that substituting verbal descriptions of parts of such information with nonverbal sounds significantly improve performance without compromising comprehension.

## Categories and Subject Descriptors

H.1.2 [Models and Principles]: User/Machine Systems - *human information processing*; H.5.1 [Information Interfaces and Presentation (e.g., HCI)]: Multimedia Information Systems - *audio input/output*; H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces - *auditory (non-speech) feedback, evaluation/methodology, interaction styles, voice I/O*.

## General Terms

Design, Experimentation, Human Factors.

## Keywords

Diagrams, auditory display, accessibility, hierarchical navigation, verbosity, learning.

## 1. INTRODUCTION

Graphically represented information is extensively used in both formal and informal areas of human activity. Diagrams in particular are prevalent in intellectual endeavour. In many cases diagrammatic representations have become common standards for expressing specialised aspects of the particular discipline they are employed in; for example, meteorologists use weather maps, architects use floor plans, and computer scientists use nodes-and-links diagrams almost indispensably.

As external representations, diagrams have been and still are the subject of numerous empirical investigations, providing

accurate specifications of many properties that give this form of graphical representation such an integral role in human cognition [17]. These include graphical constraining, where elements of the graphical representation influence the kinds of possible inferences about the represented information [14], information indexing, which eases search and recognition within the information space [10], and transitivity [8], which allows for certain conclusions to be directly read off a diagram.

The very graphical nature of these representations, however, makes them partially or totally unusable by visually impaired and blind individuals. Such populations of users rely primarily on screen-reader technologies to access and interact with computers. Screen-readers allow for the contents of a given file or document to be verbally communicated as well as linearly navigated using the computer keyboard. But while they are somewhat reliable for accessing sequentially presented information, they fall short of providing flexible access to graphical information. When graphical components are encountered within a document, expensive computational efforts need to be invested in order to process their verbalised descriptions. This hinders the users comprehension of the depicted concepts since it is mostly dependent on the accuracy and thoroughness of the available description, if one is actually available.

There exist several technologies for converting graphics into other accessible representations. Tactile displays are particularly common and used in the form of raised drawings produced through some embossing kits or stereocopiers. The main problem with these solutions is that their tactile product can quickly become cluttered with information, especially if the graphics are too complex or accompanied with textual annotations, which is typical of most diagrammatic representations.

Research on computer-based non-visual access to visual information attempts to overcome the many shortcomings of the less flexible, manually produced solutions. Many such efforts have been directed towards promoting accessibility to graph-based diagrams, producing promising solutions for supporting access to mathematical diagrams such as line graphs [11], pie charts [7] and other numerical applications such as spreadsheets [15]. The audio-based solutions achieve this by defining appropriate mappings of data parameters into one or more acoustic parameters such that, for example, changes in data values alter the pitch of a sound to represent variances in the data.

While graphs typically convey numerical data, a diagram can more generally depict concepts that are based on other types of data. As such, no direct rendering of data parameters into acoustic parameters can be used to represent such information auditorially. The challenge in these cases is to produce intuitive

mappings to deliver perceptually accessible means of representation to convey the concepts described in the given diagram.

This paper describes an approach for translating relational diagrams into an audio accessible form of representation. We discuss the design of two audio-only interfaces that rely on hierarchical structures to organise relational information, and report on an experimental evaluation that assessed the efficiency of the proposed approach. In the following sections we describe some relevant background, the design of the hierarchy and its audio display, and conclude with a discussion of the findings from the experimental evaluation.

## 2. BACKGROUND

There is a growing interest in the non-visual presentation of visual information. A number of researchers have suggested accessibility solutions to graph-based diagrams by exploiting other sensory channels. Touch and hearing are most popular and are either used in synthesised multimodal interfaces or in isolation as haptic-only or audio-only displays. Evaluations of such technologies are providing increasing evidence of their benefits and potential in making up for the shortfalls of other manual accessibility techniques both in terms of cost and efficiency.

### 2.1 Multimodal Displays

Examples of multimodal displays include the AudioGraf system [9], in which an audio-tactile mode of diagram exploration combines a touch sensitive panel with synthetic speech and non-speech sounds. The TouchMelody system [12] overlays tactile-based diagrams with non-speech 3D spatial audio to augment and enhance accessibility to line-graphs. Recently, Cohen et al. [1] described a non-visual presentation technique to communicate relational graphs. This system projects a graph-based diagram on a tablet PC interface and allows direct interaction with the graph through the use of a stylus which would generate an audio cue when an element of the graph is encountered.

A more specialised multimodal interface was used in the TeDUB system [16] to allow blind users access semantic information of technical drawings, such as digital circuitry and some UML diagrams. Evaluations of the system showed that users could successfully complete diagram-reading tasks, which included diagram exploration and information searching.

All of these techniques provide the user with sufficient interaction to support dynamic access to various forms of graphical information. They do, however, rely on representational models that are directly based on these graphical artefacts. That is, the use of the tactile medium of presentation allows a human reader to directly ‘feel’ the shapes of the graphics and their spatial arrangements, thus employing sound as a complementary source of information to augment the haptic displays. This direct mapping can be inadequate if no consideration is taken to account for the difference in perceptual bandwidth between different human sensory capabilities. The system we describe here explicitly avoids direct translation of the graphical information, and instead defines a structuring model that captures the essence of the information being translated.

### 2.2 Audio Only Displays

More in line with our approach are displays that rely on audio as the main means for structuring and communicating graphical information. The work by Bennett [2], for instance, investigated

navigational strategies specific to node-and-link diagrams and explored the effect of varying navigational models of the same information on problem-solving behaviour. His findings show that different types of tasks are best supported by a matching representation model.

Brown et al. [5] described an audio-only interface to support non-visual browsing of molecular diagrams. A hierarchical structure was used to organise the relational information encoded in a molecule, thus allowing for an ordered grouping of the components of the structure based on the concept of visual chunking and hierarchical analysis.

The organisation of this structure is, however, dynamic enough to cancel any form of user expectation and anticipation of possible information indexing. That is, the grouping of the displayed components always depends on the molecule structure, which means that every time a new diagram is loaded onto the system the user has no prior knowledge of how different parts of such a structure will be grouped. Only when they actually browse to each construct will such distribution be discovered, which can cause orientation problem especially for the novice users, as indeed reported by their evaluations. Perhaps further higher-level groupings could have been beneficial to enforce the navigation model of the system. Such a concept is employed in our design for supporting audio access to relational diagrams; as we shall describe in the next section, we imposed a partially fixed hierarchical structure to organise the encoded relational information to support orientation and structured navigation.

Examining the many reported evaluations of these accessibility tools, they frequently describe a certain degree of user learning and improvements taking place as a result of the practice users gain through the course of the evaluations. Incidentally, such development of expertise forms an important aspect of interacting with diagrammatic representations and a powerful means for evaluating the efficiency of such interactive information [14]. However, considerations for learning when evaluating non-visual solutions for graphs and diagrams seem to be currently neglected and ignored, or at best only informally addressed. Part of the evaluation of our non-visual application was to use a more methodological approach to examine users’ expertise development and learning.

## 3. A NON-VISUAL INTERFACE FOR A RELATIONAL DIAGRAM

A key point in our proposed approach for designing alternative displays to a visual representation is to focus a given cross-modal translation process on carefully selecting the set of information that needs to be communicated in order to deliver the appropriate meanings carried in the representation being translated. Thus, a translation process should emphasise and target the essence of the represented information independently from the structural tokens of the medium that presents it. This, coupled with considerations of the properties of the new medium of presentation, should produce an equally efficient representation.

Our solution is thus based on a hierarchical structure that organises information encoded in relational diagrams. The aim is to support flexible navigation and exploration of such information by providing intuitive means for information indexing and grouping.

We focused our investigation on UML class diagrams as a case study of the general nodes-and-arrows type of diagrams. We explicitly note that what we are interested in is the

communication of the concepts encoded in these diagrams rather than the particularities of UML as a notation.

### 3.1 The Hierarchical Representation

#### 3.1.1 Model Design

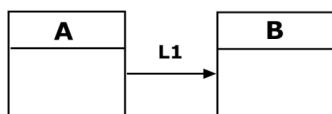
The first step in a translation process guided by this approach is to discard any information not carrying explicit meaning that might affect comprehension of the essential information encoded in a given diagram.

Because the diagrams under investigation are strictly relational, we discard all visuo-spatial information that is strict structural features of the visual medium of presentation. This includes the geometric shapes representing classes – square and rectangular boxes are typically used to represent classes or objects – and those representing relations. Figure 1 is an example class diagram.

While these geometric shapes (squares, lines, circles etc.) may carry a certain cultural context in the visual mode of presentation, they lose any significant meaning when they are described in the auditory mode. For example, if a class diagram is to be described verbally, unless a non-expert is describing it, one would not refer to it as square boxes with a label inside each. Rather, the represented semantic of the square box, or the arrow, would be used for description (i.e., a class, an association relation etc.) It is this information that we try to capture through our proposed hierarchical structure.

These geometric shapes provide a visual convenience by exploiting visual conventions and standards. An example of a visual convention typically used in relational diagrams is the mapping of direction to the geometric shapes that make up an arrow. To follow our approach we discard the line and the triangular shapes of an arrow and retain the information about the direction. This information is then mapped to hierarchical depth (Figure 2) because such information affects the relational information encoded in a class diagram.

In addition, information about the spatial arrangements of the graphical components of the diagram is also discarded, all be it that that spatial information influence the ease by which a diagram is visually read, it can be considered as a mere function of the graphical presentation mode rather than part of the encoded relational information.

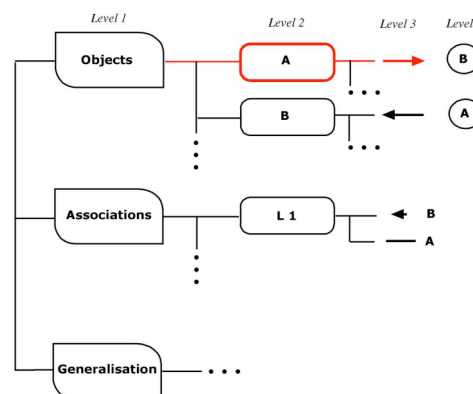


**Figure 1. A simple class diagram depicting two classes, A and B, connected by an association relation labelled L1.**

The second step in the translation process would be to define a way to organise the preserved information to allow for appropriate possibilities for accessing and navigating such information. In the case of UML class diagrams, the preserved components that need to be represented are therefore the classes, or the *Objects*, as entities which exist in the information space, and the types and direction of the *Relations* that link them to each other.

To keep our investigation at a manageable level, we decided to use a reduced version of class diagrams in which we only model two types of relations: *Associations* and *Generalisations*. We thus used these three main components as bases of a hierarchical structure, where an *Objects* container would hold information about all the objects of the diagram, an *Associations* container would hold information about all associations of the diagram, and so on. Individual objects are denoted by the name of the class, and relations by their labels. Figure 2 below shows a hierarchical representation of the relational information encoded in the diagram in Figure 1.

Notice how this hierarchy provides alternative representations of the same relational information from different perspectives. Take, for example, the simple class diagram in Figure 1. In UML terminology this diagram can be expressed in three different views. The first and the second emphasise the objects of the diagram and the type of connection relating each object to the other; so we say: “Class A has an **association from** Class B”; and say: “Class B has an **association to** Class A”. The third view emphasises the relation itself; so we say: “Association L1 is **supplied by** class B and **received by** class A”.



**Figure 2. The hierarchy used to represent relational information of a class diagram.**

The design of our model essentially maps these three different perspectives to different branches of the hierarchy. Thus, the same abstract relational information is expressed in two different levels of complexity; the node-based (the objects) and the connection-based (the relation, associations or generalisations). Each level constrains the possible set of inferences that can be made. For instance, only the third expression in the above example explicitly states the roles of the objects in a given relation (supplier and receiver), which have to be inferred in the first two expressions.

This model of information organisation is similar to that employed by Brown et al. in their non-visual molecule browser [5]. The difference in our representation is that we employ a fixed higher-level construct representing the main three components comprising the diagram (*Objects*, *Associations*, and *Generalisations*) to enforce anticipation of how the components of a given diagram will be organised. Thus, we push down the dynamic components, which are specific to a particular diagram, to a deeper level in the hierarchy that can be progressively accessed as required by the particular demands of a given task.

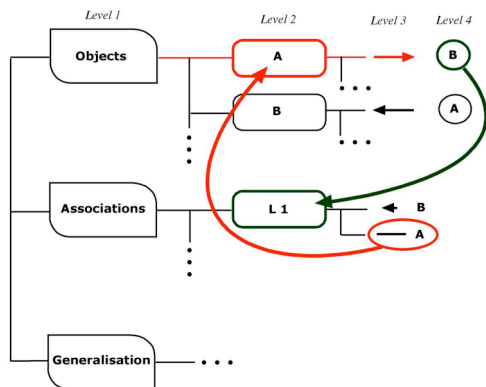
#### 3.1.2 Interaction Design

Overlaying this structure with an appropriate navigation model allows for a given diagram’s components to be hierarchically

inspected and explored. In our system, browsing is supported in a similar way to that of typical file browsers using the keyboard arrow keys; where a node in the hierarchy represents a container that can be expanded and explored to inspect the list of its children in detail, or collapsed to browse a higher level of the hierarchy, and so on.

A number of shortcut commands are provided for quick browsing and transportation around the three main components containers of the diagram; users can be transported to either the Objects, Associations, or Generalisations containers from anywhere in the structure by pressing the 'O', 'A', and 'G' keys respectively. This was included to allow for easy reorientation and navigation of the hierarchy; in an optimum performance, users can relocate themselves through a minimum number of interaction steps if they get disoriented, and quickly change their focus by transporting to a different branch in the hierarchy if their information needs change half way through an interaction sequence.

A further and more specialised shortcut key is provided, which we refer to as a *SHIFT* function. This function allows users to 'shift' between the three different views, described in the above example. Figure 3 depicts this concept. Users can shift their perspective by a single keystroke, back and forth between these branches, to access and emphasise different aspects of the same relational information. For instance, having browsed along the highlighted path in Figure 3 and reached the node B in level 4 of the hierarchy, the user has now established that an object A has an arrow pointing outwards at object B. Activating the SHIFT function will allow the user to retrieve the label of this relation in the Association container. There the user will be able to access detailed information about this particular relation from a connection-based perspective (i.e., L1 is supplied by B and received by A).



**Figure 3. The SHIFT function allows the user to emphasise different aspects of the same relational information.**

Finally, a feedback command is provided to allow the user to request detailed positional information that verbally described the current node, parent node, and next item available for inspection.

### 3.2 The Audio Presentation

We designed two alternative presentation modes to auditorally display the hierarchy. The two designs differ in the level of verbosity they employ to communicate different aspects of the encoded information.

The main information that needs to be communicated about this representation can be divided into two types; navigational and

content information. *Navigational* information refers to the system feedback that reflects the interactive actions performed by the user, such as browsing between nodes, expanding or collapsing a node, taking a shortcut etc., whereas *Content* information refers to the actual information contained in each node, such as the names of the classes, the types of the relations, the role of an object in a relation, and so on.

We used a set of abstract sounds to communicate the navigational information in both presentation modes, where a unique sound indicates the different actions performed by the user:

- an *Expand Sound* is played when a node in the hierarchy is opened; we used a mixture of frequency and amplitude modulation on a basic pulse oscillator to produce a sweep that ends with a bell like sound as an analogy for a successful expansion.
- a *Collapse Sound* is played when a node in the hierarchy is closed; we applied the opposite modulations to the reverse sequence of the Expand Sound as an analogy for a successful closure.
- a *Browse Sound* is played to communicate the movement between nodes within a list; we used a single beep sound that got higher in pitch the deeper in the hierarchy the current list being browsed is.
- an *End Of List Sound* is played when the user attempts to browse further down a list where no more children exist. We used a double beep with similar pitch mapping to that of the Browse Sound.
- a *Shortcut Sound* is played when the user executes one of the provided shortcut commands. Similar pitch mapping was used on this sound as well.
- an *Error Sound* is triggered whenever a user attempts to perform an action that has already been executed or when an action is not possible; we used a short and distinct sound burst that was louder than any other sound in the display.
- a *Request Information* is played when the user requests more detailed navigation information. A short burst of white noise was used as an analogy to a walkie-talkie verbal exchange.

We implemented a basic screen-reader to communicate content information; its level of verbosity varied between presentation modes. In a *Verbose* mode of presentation every navigation action was accompanied by a verbal description of that action as well as the content of the current node under focus. In a *Terse* mode of presentation, these verbal descriptions were replaced by nonverbal sounds. We also introduced further nonverbal sounds to completely replace some verbal descriptions. The example below illustrates an interaction sequence that reflects these differences; text within quotation marks is a verbal output, bold text is nonverbal.

Browsing the hierarchy highlighted in Figure 2 in the Verbose mode yields the following interaction sequence:

- 1 User: <press browse> (*object A selected*)
- 2 System: **Browse Sound** + "A" (*verbal description of content*)
- 3 User: <press open>
- 4 System: **Expand Sound** + "A OPENED" (*verbal description of action*)
- 5 User: <press browse>
- 6 System: **Browse Sound** + "ASSOCIATIONS FROM" (*verbal description of content*)

7 User: <press open>  
 8 System: **Expand Sound** + "ASSOCIATIONS FROM OPENED CONTAINS ONE ELEMENT" (verbose description of action)

Browsing the hierarchy highlighted in Figure 2 in the Terse mode yields the following interaction sequence:

1 User: <press browse> (object A selected)  
 2 System: **Browse Sound** + "A" (verbal description of content)  
 3 User: <press open>  
 4 System: **Expand Sound** + container sound (nonverbal description of action: continuous ambient sound)  
 5 User: <press browse>  
 6 System: **Browse Sound** + association from sound (nonverbal description of content)  
 7 User: <press open>  
 8 System: **Expand Sound** + "ONE" (less verbose description of action)

The main perceptual differences between the two presentation modes in the above interaction sequence are underlined. Firstly, whereas a node is verbally described as having been opened in the Verbose mode (step 4), a continuous ambient sound is used in the Terse mode to reflect the successful expansion and current location in the hierarchy; the ambient sound will continue to be audible until the opened node is collapsed. Secondly, a relation type (*Associations From* in the above example) is verbally described in the Verbose mode (step 6), whereas in the Terse mode it is presented by an abstract sound that communicates its type and direction.

Different timbres are used to communicate different types of relations; direction, on the other hand, is communicated by combining a short and a long sound together to form one abstract sound, where the short sound represents the arrowhead, and the long sound represents the line part of the arrow. Thus the order in which these two sounds are arranged reflects the direction of the arrow as pointing inwards (short first then long) or outward (long first then short) from an object. We also used amplitude modulation on the line part of the arrow, making the sound increase and decrease in loudness to enforce the effect of direction.

Finally, when a list node is expanded (step 8) the Verbose mode provide a full verbal description of the action whereas the Terse mode only communicates an enumeration of the list.

## 4. EXPERIMENTAL STUDY

### 4.1 Overview

The aims of this experiment were twofold: to evaluate the viability of representing relational information encoded in a relational diagram using the proposed hierarchy, and to assess the impact of varying presentation modes on users' interaction with the hierarchy. We therefore aimed to test two hypotheses:

- H1: The hierarchical organisation of the relational information encoded in a relational diagram allows for non-visual navigation and exploration of such information, and
- H2: Varying presentation modes will have an effect on users' interaction with, and comprehension of, the relational information when such information is hierarchically organised.

To test the second hypothesis we manipulated verbosity as an independent variable in a between-subject design factor of

presentation mode. In a high-verbosity presentation condition participants used the *Verbose* interface to interact with the hierarchy, where content information was communicated through verbal descriptions, whereas in a low-verbosity presentation condition participants used the *Terse* interface to interact with the hierarchy, where most content information was communicated through nonverbal descriptions.

## 4.2 Experiment Design

### 4.2.1 Participants and Setup

We initially intended to recruit 20 subjects, with 5 visually impaired and 5 sighted to be assigned to each experimental condition. Unfortunately, we only managed to get responses from two visually impaired individuals and thus we decided to use the help of 20 sighted users instead and exclude this subgroup of users from the statistical analysis of the results. All participants were undergraduate and master computer science students and had varying knowledge of UML ranging from low to intermediate expertise. They were randomly assigned to one of the experimental conditions and briefed that they were taking part in an evaluation study to test the usability of a non-visual browser of UML class diagrams. They were given a cash incentive for their participation.

Experimental sessions were made up of a training phase and a testing phase, each lasting 45 minutes to an hour. The participants sat in front of two computer speakers positioned to the left and right of a computer keyboard that they used to interact with the audio display. They faced the experimenter and were not blindfolded (Figure 4). In the training phase, the participants were introduced to the relational concepts of UML class diagrams and the particular interface used for accessing information depending on which condition they were assigned to. Once familiar with the system, they were presented with an example (audio) diagram and a set of training tasks similar to those used in the testing phase. With assistance from the experimenter, they were shown how to solve such tasks to ensure that they had a good understanding of the hierarchical structure and the sounds used to convey information. Participants could refer to the actual (visual) diagram as they were attempting to solve the tasks to confirm their findings, but this was not allowed in the testing phase.

In this phase participants attempted to solve a set of tasks, without any assistance, in a total of four scenarios each involving a different class diagram. Because most subjects had virtually no previous experience with screen-readers, they were allowed to ask clarification questions, which the experimenter answered by repeating what was spoken by the system more articulately. No time limit was given for answering the questions, although participants were made aware that they could give up a task or a scenario and move on to the next, or withdraw from the whole experiment at any point without losing their cash incentive.

The complexity of the diagrams increased from one scenario to the next, and the order of scenarios was kept constant for all participants. We defined complexity in terms of the number of components that constitute the diagrams as a tuple. The training diagram for instance was of a [5, 3, 2] complexity because it was made up of five classes, three associations, and two generalisations. This was of a relatively medium complexity in comparison with the four diagrams used in the testing phase; these were [3, 1, 2], [4, 2, 3], [4, 3, 1] and [7, 6, 2]. The first three diagrams varied mostly in terms of relational complexity rather than the number of components, as complex relational concepts, such as the involuted relation, were progressively

introduced (see Figure 5). The diagram used in the fourth scenario was more complex in terms of both relational complexity and the overall number of components.

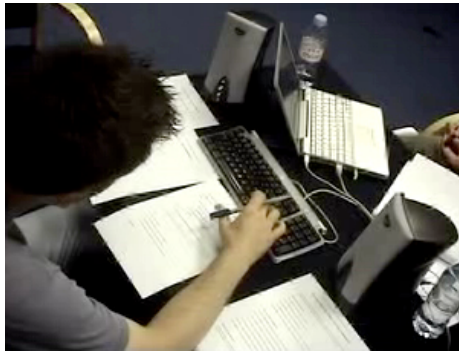


Figure 4. Experiment setup.

#### 4.2.2 Tasks

To assess the efficiency of the proposed model in allowing flexible navigation and exploration of a relational diagram we observed and measured users' performances when carrying out three different tasks. The tasks were similar to those described by Bennett [2] in that they reflect the ability to inspect a relational diagram from an object perspective (the nodes) or a connection perspective (the arrows). We assumed that gaining an understanding of both these perspectives is necessary to achieve full comprehension of any relational diagram. Participants answered all three tasks on a pre-formatted answer sheet:

1. Task 1 required the users to locate an object within a diagram, explore its connections and find out how it links to other objects.
2. Task 2 required them to locate a particular connection, determine which objects are linked through it, and its direction.
3. Task 3 required them to enumerate the diagrams components; that is how many objects and relations.

In addition to these three tasks, participants were asked to reproduce graphically the diagrams in a fourth task using the UML notation. They were not allowed to review their previously answered questions and were given a separate sheet to draw on. No explicit instruction was given in the training phase on how to go about achieving this final task.

#### 4.2.3 Data Gathering

We decided to use a number of data gathering techniques to collect the maximum amount of data for a thorough analysis. Participants were asked to sign consent forms for anonymous subsequent use of interaction logs, video and audio recordings, and questionnaire responses. They were also encouraged to use the speak-aloud protocol; together with the video logs this provided good insights into the intricacies of their interactions, which we later compared against the interviews and questionnaires conducted at the end of each experimental session.

#### 4.2.4 Measurements

We measured task completion-times and overall error rates as dependent variables. We defined three categories of errors; the first and second would inform analysis of the efficiency of the hierarchy in organising and communicating relational information. The third category would inform analysis of the

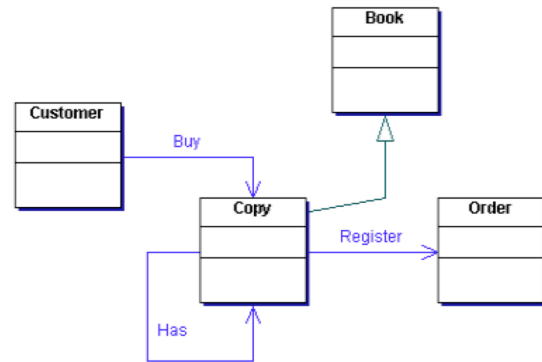


Figure 5. A UML class diagram of a [4, 3, 1] complexity used in the third scenario of the experiment containing an involuted relation (i.e. An object pointing to itself).

efficiency of the participant in using the hierarchy to complete the diagram-reading tasks. The categories are:

1. Interaction errors, which occur when users attempt to carry out an action that is not allowed in the current system state; this would trigger the system to produce the error sound.
2. Comprehension errors, which are observed in the answers given by the participants to the questions asked in the tasks; this would reflect their understanding of the represented information.
3. Efficiency errors are related to either the choice of strategy for answering a task or the efficiency of executing such a strategy.

Whereas the first and second categories are quantitatively measured, we relied on a concept known as *Interaction Traps* [4] to identify and measure the third category of errors. This concept has been used to evaluate graphical user interfaces by assessing the interactive relationship between a human user and a system in terms of strategies and objectives. It is part of a framework that allow for an analysis of complex interaction where users have multiple objectives, shifting objectives, or interleaving tasks [4], all of which are typical behaviours when interacting with external representations such as diagrams [13]. A manifestation of an interaction trap results in an inefficient interaction, and is considered a direct indication of a mismatch between the participant's understanding of the representation model and the system's representation of its state.

In order to fit our assessments, however, we extended this concept to cover a wider range of interaction inefficiencies. While the concept is mainly based on analysing the efficiency of users' understandings of the achievability of an objective and/or how to go about achieving it [4], we observed other instances of inefficient interactions that occurred even when an objective seemed to be well understood and the system's states well interpreted. We therefore introduced a new category of interaction efficiency, which we refer to here as *less efficient* strategies, to capture such instances.

We thus classified the execution of an interaction strategy as inefficient, less efficient or efficient where:

1. Inefficient strategies are instances of an interaction where a trap or more occur.
2. Less efficient strategies are all other inefficient interaction not captured by one of the manifestations of interaction traps [4].

- Efficient strategies are instances where none of the above occurs.

Extending the concept of interaction traps in this manner allowed us to formally address and analyse the learning behaviour of the participants. By observing and classifying their interactions as they progressed from one experimental scenario to the next, we were able to plot their learning curves and analyse these in relation to the rest of the data. We discuss our findings in details in the following sections.

## 5. RESULTS

### 5.1 The Hierarchy

Overall, all participants responded positively to the system, and the concept of hierarchically organising relational information was well received. All without exception were able successfully to use the audio interface and complete all the tasks presented to them in the testing phase of the experiment. High scores were recorded with a 96% mean of correct answers across diagrams for the average participant. Individual scores varied between 70% and 100% with a normal distribution (standard deviation of 6.5). Figure 6 shows the scores for individual participants for all scenarios with diagram 1 at the bottom and diagram 4 at the top. There was a maximum score of a 100 for each scenario.

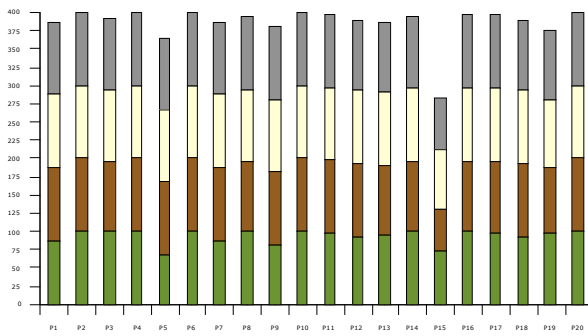


Figure 6. Participant's individual scores across diagrams.

The ability to take shortcuts to different containers from anywhere in the hierarchy was particularly well received and used extensively throughout the tasks. Participants explicitly commented that they felt they could focus on different aspects of the diagram and ignore others. The fixed depth of the hierarchy proved useful too, in that it usually took the participants no more than two to three browsing steps to reorient themselves within the structure in both presentation modes, even without including a shortcut step.

### 5.2 The Presentation Modes

A Mann-Whitney test revealed that participants in the Terse mode spent significantly less time to complete tasks on each diagram than those who use the Verbose mode ( $U=18$ ,  $z=2.38$ ,  $p1=0.009$ ). We note here that the amount of time it takes the screen-reader to verbalise some parts of the information in the Verbose mode and the equivalent nonverbal description of the same parts in the Terse mode are equal in duration.

However, as the example in Section 3.2 highlights, there are other verbal descriptions in the Verbose mode which were completely discarded in the Terse mode. Therefore, we have accounted for these differences and calibrated the total task-completion times. This was done by excluding the additional time taken by the screen-reader in the Verbose mode to describe

a given system state that involved these parts (Figure 7). The differences between the two conditions for the calibrated times were not statistically significant ( $U=34$ ,  $z=1.17$ ,  $p=0.121$ ).

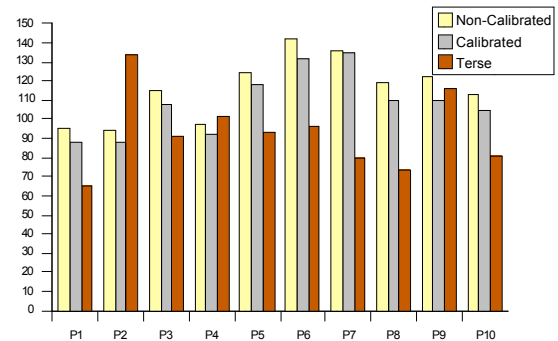


Figure 7. Average task-completion times for each participant across tasks for Verbose (calibrated and non-calibrated) and Terse modes of presentation.

Median comprehension and interaction error rates as well as scores from the four scenarios were also averaged for each participant. Mann-Whitney U values were again computed to test whether the differences between these variables in the two conditions were significant. Contrary to what was expected, there was no significance for comprehension error rates ( $U=27.5$ ), interaction error rates ( $U=25.5$ ), and scores ( $U=41.5$ ).

Thus, our second hypothesis (H2) was only partially supported; varying verbosity did not have an effect on the participants' understanding of the relational information encoded in the hierarchy, but interacting with a less verbose interface significantly improved their performance in terms of the time it took to complete the given diagram-reading tasks.

### 5.3 Learning Rates

As typically reported by similar evaluations studies, a noticeable improvement in performance was observed with all participants. They commented on feeling more at comfort and ease with using the display as the evaluation tests progressed. This was also confirmed by the fact that performance levels and scores were kept relatively constant across scenarios even though the complexity of the diagrams used in each scenario was increasing. Thus, in order to formally assess and analyse this obvious expertise development, we classified participants' interaction strategies for each task and categorised these based on the extended concept of interaction traps described earlier.

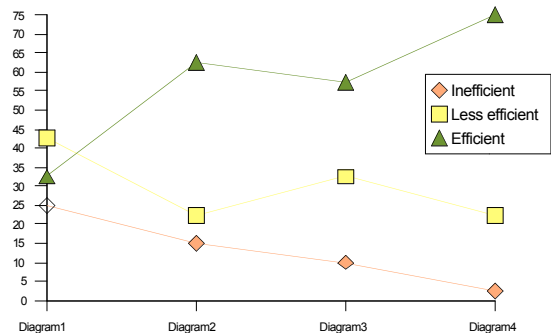
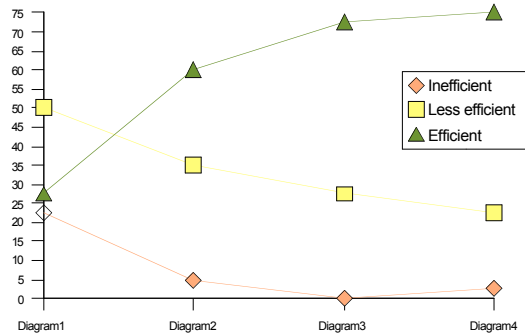


Figure 8. Interaction efficiency levels of the strategies employed on each diagram in the Terse mode.



**Figure 9. Interaction efficiency levels of the strategies employed on each diagram in the Verbose mode.**

Overall, similar learning curves were observed in both conditions. There was a slight difference in the third scenario where, as shown in Figure 8 and 9, participants in the Terse mode employed more inefficient strategies than participants in the Verbose mode. By the end of the final scenario, the percentage of efficient strategies employed by participants in both conditions reached an equal 75%, while the percentage of inefficient strategies reached an equal 2.5%.

Averaging across the two conditions, the percentage of efficient strategies was observed to increase from 30% in the first scenario to a dominant 75% in the last. Inefficient strategies, on the other hand, drastically decreased from 24% in the first scenario to just 2.5% in the last, while less efficient strategies decreased from a dominant 46% to just over 22% in the last scenario. This confirms the participants' comments that they felt they were improving as the study progressed.

## 6. ANALYSIS

### 6.1 Sound Design

Observing video logs and analysing data gathered through interviews and questionnaires, we classified three distinct reactions to the sounds used to convey navigational information. Where some navigational sounds were explicitly listened out for, such as the End of List and Error sounds, others were only appreciated for their aesthetics, such as the Expand and Collapse sounds, while the rest were completely ignored. As the study progressed, however, the aesthetically pleasing sounds were later on the ones that were listened out for by some participants to confirm that their action had been completed, while the verbal descriptions were ignored.

It is also interesting to note that the sounds explicitly listened out for were those not accompanied by verbal descriptions. The ones ignored, however, included the Browse sound, which meant that mapping their pitch to indicate the depth of the hierarchy was not picked up by any participant. Instead, participants relied on both content information and the Error sound to determine their position. This can be attributed to the nature of the information targeted by the mapping; while engaged in a problem-solving task, participants are directly focused on the content information they are inspecting rather than the browsing information, especially when they get more familiar with the system. It would be interesting to test whether changing the target information for the mapping from the navigational to the content information would yield any different reaction.

Together with the amplitude mapping, the order organisation of the abstract sound used to present the type and direction of a

relation in the Terse design was well received. Participants commented that the relations “sounded a lot like how they would have been drawn”. An interesting observation, which is in line with these comments, is the fact that participants using this presentation mode ‘drew’ their answers on the answer sheet, unlike participants using the Verbose mode, who wrote down their answers in English as spoken by the screen reader. The former seemed to struggle less with determining direction, which had to be inferred from the spoken output in the Verbose mode; for example, that an ‘Association From’ is an arrow pointing outwards from a class. In this instance, at least, nonverbal sounds were superior in representing direction over verbal sounds.

Finally, not all participants listened out for the continuous ambient sounds which were intended to reflect the current state of the container being browsed. Participants noted that they were most aware of them when they changed focus or took a shortcut to a different container. Thus, the ambient sounds seem to communicate transitional information more appropriately than positional information, which they were originally designed to communicate. Participants did not find these annoying or irritating though, which makes the unexpected function of the ambient sounds worth including in such a design.

Perhaps more important is the fact that while the differences in task-completion times were only significant for non-calibrated times, and comprehension errors were not affected by varying verbosity, non-verbal sound could effectively replace verbal descriptions without compromising - and even enhancing - the problem-solving behaviour. As noted by our observations and qualitative results gathered from questionnaires and interviews with the participants, it was much easier to interpret the relational information encoded in the representation when nonverbal sounds were used to communicate them.

The UML terminology used to describe relational information is a little counterintuitive, which required the participants to invest more effort to infer the appropriate meaning and describe it in their answers when it was spoken in plain English. For instance, it was much easier to describe – verbally, graphically, and in this case through a non-verbal audio cue – an association that “it points to object A”, than it was to describe it as “supplied by object A”. The nonverbal mapping used to communicate direction facilitated the formulation of the first expression. Of course, such differences would be affected by the levels of participants' expertise, both with the UML domain knowledge and in the use of the provided auditory representation.

### 6.2 Interaction Efficiencies

58% of inefficient strategies occurred in the first scenario where participants had low levels of expertise in using the system - gained mainly from the instructions given to them in training phase. Most of these inefficiencies were a result of the users misunderstanding some aspects of the representation. For instance, when asked to locate an object labelled *Sheep* and retrieve its connections in Task 1 of the first scenario (i.e., what type of arrows point inward and outward of it), seven out of twenty participants took a shortcut to the *Objects* container, browsed through it with the intention of locating the *Sheep* class, but as soon as they encountered an object labelled *Animal* decided to explore this object to locate *Sheep* within it. This incorrect move was based on the assumption that all animals would be grouped under the *Animal* node. This is incorrect because both nodes represent separate entities rather than lists of entities, and in order to explore the connections of a specific

entity in the hierarchy it has to be located at level 2 of the hierarchy and not level 4 (see Figure 2). Attempting to expand a node at the fourth level would produce an error sound. Here, these participants correctly interpreted their objective, but selected the wrong strategy to achieve it.

Where the above example was a direct result of the modelled domain, a more general design related reason which seems to push the users into such potential interaction traps is the duplication of node names at different levels of the hierarchy. This was especially the case in the Objects container where objects were contained within each other. This is a direct result of mapping relational concepts to hierarchical depth. Some participants would, as a result of this duplication, misinterpret the level of the hierarchy they were at, which caused some momentary confusion in interpreting the relational information. They usually recovered quickly by browsing a step upwards or hearing the Error sound and correctly realising that they had reached the deepest level of the hierarchy. Nonetheless, more audio design consideration can make up for this shortcoming, for instance, by using different voices to verbalise the content information at different levels in the hierarchy, or changing the pitch mapping to render the content rather than the navigational information as discussed earlier.

Instances of less efficient strategies occurred when relatively longer interaction paths were followed to achieve a certain objective. This, however, does not include interactions where the user puts the objective on hold and engages in explorative interaction, which we considered a positive contributor to the development of a more thorough understanding of the representation. Other occurrences of less efficient strategies were observed when a user either misunderstood the objective itself, and as a result only partially completed a task, or did not realise that an objective had been achieved, and as result carried on with unnecessary interactions. In both cases, lack of domain knowledge seems to have been a direct cause of the misunderstandings.

### 6.3 Learning

Overall, classifications of interaction efficiencies allowed us to categorise three types of learners: fast, medium and slow. Fast learners are participants that quickly pick up on the workings of the representation and manage to execute more efficient strategies earlier in the scenarios. Medium learners take a rather steady pace in developing their expertise and manage to execute relatively more efficient strategies in the final scenarios. Slow learners, on the other hand, struggle throughout the four scenarios and manage to execute very few efficient strategies overall.

Twelve out of the 20 participants were classified as fast learners, having employed more efficient strategies in the earlier scenarios and throughout. Six out of the remaining 8 were classified as medium learners, and the last two as slow learners. Table 1 below shows an example efficiency distribution from each category of learners. We note here that both visually impaired participants were also classified as fast learners, which is encouraging given that they represent the population of user who would directly benefit from such a system.

As diagram complexity increased from one scenario to the next, the observed reaction of the three learner categories to these changes was also clearly different. While fast learners were able to overcome the increasing complexity and efficiently accommodate the challenge by carrying out the tasks with relative ease, medium learners' performance was slightly

staggered the more complex the diagrams get. And, as expected, the slow learners' performance level was evidently affected by the increase in complexity.

	Fast	Medium	Slow
<b>Diagram1</b>			
TASK1	(i)	(L)	(i)
TASK2	(E)	(E)	(L)
TASK3	(E)	(L)	(L)
DRAWING	(L)	(i)	(i)
<b>Diagram2</b>			
TASK1	(E)	(L)	(i)
TASK2	(E)	(E)	(E)
TASK3	(E)	(E)	(E)
DRAWING	(E)	(E)	(L)
<b>Diagram3</b>			
TASK1	(E)	(i)	(E)
TASK2	(E)	(L)	(L)
TASK3	(E)	(E)	(L)
DRAWING	(E)	(E)	(i)
<b>Diagram4</b>			
TASK1	(L)	(L)	(E)
TASK2	(E)	(E)	(L)
TASK3	(E)	(E)	(E)
DRAWING	(E)	(E)	(L)

**Table 1. Example of interaction efficiency distributions across the four scenarios from each category of learners, (E): Efficient, (L): Less Efficient, and (i): Inefficient strategies.**

The fact that most participants were classified as fast learners, however, testifies to the ease and usability of the interface, and, more generally, to the efficiency of the hierarchy in communicating the relational information encoded in a UML class diagram. We believe that this, coupled with the low comprehension error rates and high recorded scores, supports our first hypothesis (H1) that hierarchically organising the relational information encoded in relational diagrams supports efficient non-visual access to such information.

## 7. DISCUSSION

One important and crucial aspect of this system and many of the alternative tools being researched and developed for supporting non-visual access to graphical information is interactivity. In the case of audio-only interfaces, interactivity seems to minimise the negative aspects of the temporal nature of the auditory medium, allowing for the auditorally represented information to be immediately available in the environment on request. With relation to the approach proposed in this paper, the ability to dynamically navigate and inspect the hierarchy meant that parts of the problem space need not to be remembered during the problem-solving process. Furthermore, when encountered, some parts of the representation seem to have acted as memory triggers for retrieving relevant prior knowledge when this was required by the interactive strategy under execution. Moreover, the use of a partially fixed hierarchical structure helped minimise the computational efforts required for keeping track of where a given diagram's components could be. In the offset, a user could anticipate where different parts of information will be available in the representation by relying on the fixed decomposition of the diagram constructs, thus easing the search process for a particular component.

An important implication of this structured and fixed organisation, we observed, was that users could anticipate where items of interest would be located, as well as what would be heard after executing an interactive action. If the verbal or nonverbal sounds heard did not match their expectations, they

could easily interpret the unexpected feedback to reason about location and how to go about repairing erroneous interactions. This meant that it usually took them no more than two to three browsing steps to reorient themselves within the structure, and as they gained more expertise, they got less disoriented and became more efficient at executing different problem-solving strategies.

The participants in our study could explicitly focus their attention on a single aspect of the diagram separately from the others by browsing one of the main containers. This behaviour was also observed to match the particular demands of a given task, for example, when the participants browsed to either the objects or one of the relations' containers to solve the object-oriented or the connection-oriented tasks appropriately. These are similar findings to those reported by Bennett [2], where different types of tasks were observed to be best supported by a matching representation model. Our results extend these findings by suggesting that matching interactive possibilities to the different views of the represented concepts allow for more flexible and efficient access to such information. That is, the shortcut commands including the SHIFT function allowed the participants to interact more intuitively with the represented information. The visually impaired participants used such a function extensively and both commented that adding a feature to customise this function would be even more beneficial, for example by allowing a bookmarking feature that could be applied to any part of the hierarchy, which they could then return to as required by a particular navigation strategy.

All participants were able to function under the hierarchical model to reason about the modelled domain without feeling the need to internally visualise the diagram, quickly and efficiently navigating and retrieving the encoded information driven by their information needs. Many of them reported that they would visualise a relation as soon as enough information about it was retrieved, however they would discard the built picture when moving on to other parts in the hierarchy.

While this testifies to the strength and flexibility of the proposed hierarchy, it compromised their comprehension of the overall structure of the diagram. As diagram complexity increased, participants found it increasingly difficult to build a 'bigger picture' of the diagram, until they eventually completely gave up attempting to build such global understanding in the final diagram – of a [7, 6, 2] complexity where they focused explicitly on answering the questions asked in the experimental tasks. When reaching such complexity, the hierarchical representation seemed to have hit the limit of how much overall impression of the diagram it could convey. A possible solution to this shortcoming is similar approaches to the algorithm developed by Brown et al. [6] to convey an impression of the topology of node-and-link diagrams, although higher complexity graphs may still cause problems. This generally forms a fruitful avenue for future research for any researcher interested in audio interface design, particularly those dealing with graphs and diagrams.

Another interesting aspect observed in the participants' interactions with the non-visual interface was the ease by which they interpreted the nonverbal representations of the explicitly relational components of the diagram. Relations' types and directions were more efficiently and intuitively accessed when represented non-verbally than through speech. Participants responded very positively to the mapping used to convey such information. In fact, this became so intuitive that they drew their answers almost synchronously with the displayed sounds. This is an encouraging finding that calls for more in-depth investigations of strengths and weaknesses of the

representational properties of audio, which would certainly inform the design of more intuitive audio-only and multimodal interfaces. Although relevant, these findings are unfortunately still limited to the fact that all participants in our study were sighted individuals; including partially or visually impaired individuals in future work along this avenue is necessary.

Finally, by using and extending on existing graphical user interface evaluation techniques (mainly [4]) we were able to address formally the learning behaviour of our participants and their expertise development. Systematic classifications of learners' behaviour allowed us to categorise different types of learners and observe their varied reactions to different experimental settings while relying on an audio-only interface to access a diagrammatic representation. Such techniques can be informative for ways to better the design of data presentation and provide insights into accommodating adaptivity in auditory display design.

## 8. CONCLUSION

We have presented a novel approach for translating graphical information from the visual to the auditory medium of presentation. This approach emphasises the meaning of the representation being translated, rather than the structural tokens of the medium that presents it. We exemplified this by using a hierarchical model to organise the information encoded in a relational diagram, and assessed the impact of varying audio presentation modes on the comprehensibility and efficiency of interaction with this model. Our results show that the proposed hierarchical structuring model allows for the relational information encoded in a diagram to be non-visually accessed and navigated. Furthermore, the substitution of verbal descriptions of parts of the diagram with nonverbal descriptions was found to improve performance times significantly without compromising users' comprehension of the represented concepts. These findings highlight the importance of deeper investigations into the representational properties of the auditory medium. Such investigations will certainly improve our understanding of the role played by such media in human cognition and hence inform the design of more intuitive tools to support it.

## 9. REFERENCES

- [1] Cohen, R. F., Yu, R., Meacham, A., and Skaff, J. PLUMB: Displaying graphs to the blind using an active auditory interface. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility (ASSETS 2005)*. ACM Press, New York, NY, 2005, 182-183.
- [2] Bennett, J. D. Effects of navigation and position on task when presenting diagrams to blind people using sound. In *Proceedings of the second international conference on Diagrammatic representation and inference (Diagrams 2002) - Lecture Notes in Computer Science, 2317*, (2002), 307-319.
- [3] Bertel, S. Show me how you act on a diagram and I'll tell you what you think. In *Proceedings of the AAAI spring symposium on Reasoning with mental and external diagrams: Computational modelling and spatial assistance* (Menlo Park, CA, 2005). 2005.
- [4] Blandford, A., Thimbleby, H., and Bryan-Kinns, N. Understanding interaction traps. In *Proceedings of HCI 2003: Designing for Society – Volume 2*. Research Press International, Bristol, UK, 2003, 57-60.

- [5] Brown A., Pettifer S., and Stevens R. Evaluation of a non-visual molecular browser. In *Proceedings of the 6<sup>th</sup> international ACM SIGACCESS conference on Computer and accessibility (ASSETS 2004)*. ACM Press, New York, NY, 2004, 40-47.
- [6] Brown A., Stevens R., and Pettifer S. Audio representation of graphs: A quick look. In *Proceedings of the 12<sup>th</sup> international conference on Auditory display (ICAD 2006)*. (London, June 21-23, 2006). 2006.
- [7] Franklin. M. K., and Roberts. C. J. Pie chart sonification. In *Proceedings of the 7<sup>th</sup> international conference on information visualisation (IV 2003)*. IEEE Computer Society, Washington, DC, 2003, 4-9.
- [8] Gurr, C. A. Effective diagrammatic communication: Syntactic, semantic and pragmatic issues. *Journal of Visual Languages and Computing*, 10 (1999), 317-342.
- [9] Kennel, A. R. Audiograf: A diagram-reader for the blind. In *Proceedings of the second annual ACM SIGACCESS conference on Assistive technologies*. ACM Press, New York, NY, 1996, 51-56.
- [10] Larkin, J. H., and Simon. H. A. Why a diagram is (sometimes) worth ten thousands words. *Cognitive Science*, 11, 1 (1987), 65-100.
- [11] Mansur, D. L., Blattner, M., and Joy, K. Sound-graphs: A numerical data analysis method for the blind. *Journal of Medical Systems*, 9 (1985), 163-174.
- [12] Ramloll, R., and Brewster, S. *A generic approach for augmenting tactile diagrams with spatial non-speech sounds*. In *CHI 2002: Extended abstracts of the SIGCHI conference on Human factors in computing systems*. ACM Press, New York, NY, 2002, 770-771.
- [13] Scaife, M., and Rogers, Y. External cognition: How do graphical representations work. *International Journal of Human-Computer Studies*, 45, 2 (1996), 185-213.
- [14] Scaife, M., and Rogers, Y. External cognition, interactivity and graphical representations. In *Proceedings of the IEE Colloquium: Thinking with diagrams*. (London, UK, 1996). Digest No: 1996/010, 8/1-8/6.
- [15] Stockman, T. The design and evaluation of auditory access to spreadsheets. In *Proceedings of the tenth meeting of the international conference on Auditory display (ICAD 2004)*. (Sydney, Australia, July 6-9). 2004.
- [16] *The TeDUB-Project (Technical Drawings Understanding for the Blind)*. Available at <http://www.tedub.org>, accessed February 2006.
- [17] Zhang, J. The nature of external representations in problem solving. *Cognitive Science*, 21, 2 (1997), 179-217.