AUDITORY EXTERNAL REPRESENTATIONS: EXPLORING AND EVALUATING THE DESIGN AND LEARNABILITY OF AN AUDITORY UML DIAGRAM

Oussama Metatla, Nick Bryan-Kinns, Tony Stockman

Interaction Media and Communication research group
Department of Computer Science
Queen Mary, University of London, UK.

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

ABSTRACT

We present an approach that examines the design of auditory displays for accessing graphically represented information in terms of their roles as external representations. This approach describes how a cross-modal translation process should emphasise the semantics of the represented information rather than the structural features of the medium that presents it. We exemplify this by exploring the design of a hierarchical representation to organise relational information encoded in a UML class diagram, and describe two alternative presentation modes to auditorally communicate this structure. We report on an experiment that we conducted to assess the viability of our approach and describe a novel methodological analysis which extends existing evaluation techniques to formally examine how a group of users learn and develop interactive expertise when using this auditory display.

[Keywords: External Representation, Verbal, Nonverbal audio, UML Diagrams, Learning]

1.INTRODUCTION

One comes across a large variation of diagrams in many informal or formal areas of human activity. From the self-sketched sense-making drawings to the more unified notational systems, graphical external representation of information have been proved to form an integral part of the particular cognitive activities they are used for. With the increasing interest in developing auditory tools for accessing visual artefacts, an understanding of the role these tools play, not just as accessibility solutions, but as external representational agents that support and aid cognition becomes crucial.

We propose an approach for designing and evaluating auditory displays that support access to graphical artefacts by assessing the extent to which they succeed to fulfil typical characteristics of external representations. This suggested approach is motivated by two main observations. The first is the dominance of and bias toward visual phenomenon whenever External Representations (ER) are under study. While much has been written about the nature and benefits of interacting with ERs, it is quickly apparent from the literature that the focus is typically on visual and, to a much lesser extent, haptic or tactile artefacts. To our knowledge, the only references to auditory phenomenon when dealing with these sort of issues is limited to either an account for the transformation of speech from the auditory verbal form into a transcribed written representation [10], or when examining parallels between phenomenon which occur in natural language and those occurring in graphical communication [4]. These references hardly reflect the evident and significant potential of using auditory display to communicate and represent rich levels of meaning.

The second observation arises from examining the many reported evaluations of auditory accessibility tools that specifically target graphs and diagrams. Such studies 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 external representations and a powerful means for evaluating the efficiency of such interactive information [8]. However, considerations for learning when evaluating auditory displays for graph-based diagrams seem to be currently neglected and ignored or at best only informally addressed.

This paper describes the design and evaluation of an auditory display that allows a human user inspects and navigates information encoded in diagrams; a common means for external representation of information. We attempt to address issues and aspects of the auditory display that reflect its role as an external representation by (1) considering the relationship between the representation and that which is represented and (2) analysing the development of users expertise through multiple interaction scenarios. We present the later in a form of a qualitative analysis that draws and builds on existing Graphical User Interface evaluation techniques, introduced in [2], to systematically examine the learning aspects and efficiency of an interaction while emphasising a direct relationship between the interactive display and the user.

2.BACKGROUND

2.1.External Representation

External Representations are forms of structured knowledge that are readily available in the environment and which can be directly analysed, processed and manipulated through perceptual sensory motors [10]. Recently, many mainstream cognitive science research has increasingly been directed towards the study of ER and its role in problem-solving activities.

A common trend that emerged in this field specifies and enforces the view that a tight relationship exists between internal and external representations through which complex cognitive tasks are performed. This view emphasises the role of external representations as more than just a form of input to the internal mind or just an external aid to the limits of human memory, but as an intrinsic component in the cognitive process; guiding, constraining, and even determining cognitive behaviour [10]. A number of properties possessed by external representations that give them such an integral role include:

1. *Locational Indexing* which means that effective external representations organise information that tends to be needed for the same inference in adjacent locations, so reducing the amount of search required to find the information [6].

2. *Re-representation* which refers to how different external representation can make the same problem-solving process easier or more difficult even though they represent the same abstract information.

3. *Representational Constraints*, which refers to the imposed level of perceptual constraints, which limit the types of inferences that are permitted about the represented information.

The fact that ERs have now been given much focused attention in the study of human cognition is considered a significant theoretical advancement from traditional cognitive science [8]. Auditory display is being left out of such advancements. Sound is only taken into consideration in the verbal form of spoken language, and even then, it is the transformation from verbal into textual transcriptions that seems to be of interest rather than the auditory display itself. Thus, what we propose is that auditory displays, as a form of knowledge that is processed and analysed directly from the environment, can and should be studied under a representational taxonomy based on the properties of ERs.

Initially, we believe that the above set of features form an accurate and solid framework for evaluating the extent to which an auditory display succeeds or fails in playing the role of an external representation in a given problem-solving context. Whether there are any other features and properties intrinsic to the auditory medium when it comes to external representation is for now an open question.

2.2. Auditory Graphs and Diagrams

There is a growing interest within the Auditory Display community in the non-visual presentation of graph-based diagrams. A number of researchers have suggested accessibility solutions to such graphical information by either combining alternative modalities, typically audio and haptics, or solely relying on the auditory medium. Examples of multimodal displays include the AudioGraf system [5], the TouchMelody system [7] and the TeDUB system [9], all of which provide the user with sufficient interaction to support dynamic access to a variety of graphical representations. Most of these solutions, however, tend to rely on representational models that are directly based on the original graphical artefacts they provide access to. That is, to access the graphically represented information, the user can directly 'feel' the graphs or diagrams through the augmented multimodal displays even though graphs and diagrams are designed to be optimally accessed through vision.

Other suggested accessibility solutions rely solely on audio to display the information encoded in a diagram. More in line with the system we describe in this paper are audio displays that use different representation models to support exploration of such information. Bennett [1] for instance, examined the effect of varying representational models for accessing nodes-and-links diagrams, and showed that different types of tasks are best supported by a matching representation model. This echoes the re-representation property of ERs and can be considered as empirical evidence that such a property can be applied to an underlying data model with its effect observed when such model is presented through sound only.

A hierarchical representation was used by Brown et al in [3] to organise the information encoded in a molecule structure. A certain degree of representational specificity and constraints were apparent in their design in the way hierarchy depth-levels was mapped to user knowledge-levels; only showing high level aspects of a given construct and allowing for a detailed exploration of its constituents if required by the user through a zooming function. This grouping of the displayed components always depend of the molecule structure, which means that every time a new structure is loaded onto the system the user has no prior knowledge of how different parts of such 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 user, as indeed reported by their evaluations.

In the next section we describe our approach for translating an information domain from the graphical to the auditory modality through emphasising the importance of capturing the essence of the represented information, and then using a representational model that implement properties of external representations to organise and structure the encoded information.

3. AN AUDITORY DISPLAY FOR UML CLASS DIAGRAMS

3.1. The Representation

We have focused our investigations on a very common type of diagrammatic representations known as nodes-and-links diagrams. Such diagrams usually describe relational information and are extensively used in the computer science discipline. Examples include Entity Relationship diagrams and the Unified Modelling Language (UML); our current choice of diagrams. Figure 1 depicts a simple class diagram.

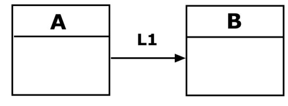

Figure 1: A simple UML class diagram depicting two objects connected by an association relation

A key point in our proposed approach for cross-modal translations is to separate the semantics of the represented information from the structural features of the medium that presents it. Because the diagrams we chose are strictly relational, we discard any visio-spatial information that does not carry explicit semantic value that might effect the encoded relational information. This includes the geometric shapes representing classes and those representing relations. In addition, information about the spatial arrangements of the diagram components is also discarded, all be it that such information influence the ease by which a diagram is visually read.

The components that need to be represented to preserve the relational semantics of a given class diagram are therefore the classes themselves; as existing entities/objects in the information space, and the direction and type of the relations that link these entities to each other. To keep our investigation at a manageable level, we used 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 that represents relational information, 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 in the hierarchy are denoted by the name of the class, and relations by their labels. Figure 2 below shows the higher levels of the hierarchical structure that organise the basic components of the class diagram shown in Figure 1.

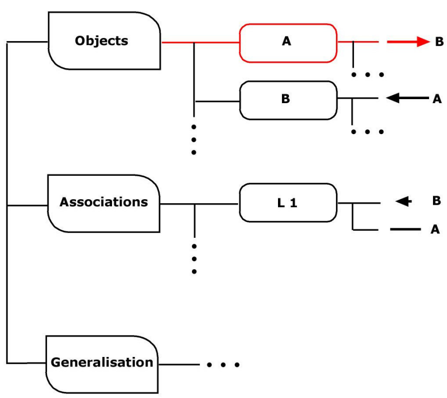

Figure 2: Hierarchical structure used to represent relational information of a UML class diagram

This structure maps relational information to hierarchical depth and provides alternative representations of the same relational information from different perspectives. Take for example the simple class diagram in Figure 1 above. In UML terminology this diagram can be expressed in three different ways; the first and the second emphasise the objects of the diagram, so we say: "Class A has an *association from* Class B"; and say: "Class B has an *association to* Class A", while the third emphasises the relation itself, so we say: "Association L1 is *supplied by* class B and *received by* class A".

Thus the same abstract relational information is expressed in three different levels of complexity, i.e. re-represented to reflect different aspect of the same information in different branches of the hierarchy. Each of these aspect constrains the possible set of inferences that can be made; for instance, only the third expression explicitly states the roles of the objects in a given relation (a supplier and a receiver), which have to be inferred in the first two expressions.

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

To interact with the representation, the keyboard arrow keys can be used in a similar way to that of typical file browsers; where a node in the hierarchy represents a container that can be opened and explored to inspect the list of its children in details, or closed to browse a higher level of the hierarchy, and so on. We also provide the user with a number of shortcut commands to be able to 'jump' around the hierarchy and return to the fixed containers, as well as a *Shift* function to allow quick change of perspective between the different levels of details as exemplified by the three expressions described above.

3.2. The Presentation

The information that needs to be communicated about the representation can be divided into two types; navigational information and content information. Navigational information refers to the system feedback that reflects the 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, etc.

We designed two alternative presentation modes to auditorally display the hierarchy. The two designs differ in the level of verbosity they employ to communicate the different aspects of the representation. We used the same set of abstract sounds to communicate the navigational information in both designs and mapped the depth of the hierarchy to the pitch of the browsing sounds; so the deeper the current list being browsed in the hierarchy, the higher in pitch the browsing sound is. In a *Verbose Mode* of presentation every navigation action was also 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. The example below illustrates an interaction sequence which reflects these differences:

Browsing the hierarchy as highlighted in Figure 2 in the Verbose mode yield 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 + "ASSOCIAITONS FROM OPENED CONTAINS **ONE** ELEMENT"(verbose description of action)*

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)

To highlight the main differences in the auditory design of the two modes, consider the following comparisons. 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 its successful expansion; the ambient sound will continue to be audible until the opened node is collapsed. Three distinct ambient sounds were used for each of the main constructs of the hierarchy (Objects, Associations, and Generalisations).

Secondly, whereas a relation type is verbally described in the Verbose mode (step 6), an abstract sound is used in the Terse mode to communicate 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 to enforce the 'coming into' and 'coming from' 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. We implemented a simple screen to communicate all verbal descriptions in both modes of presentation.

4.EXPERIMENTAL STUDY

4.1.Overview

The properties of external representations described in section 2.1 have been shown to influence the efficiency of interacting with graphical external representations as well as the nature and strategies of problem-solving behaviour when using these representations [10]. The aim of this experiment was thus to assess the extent to which an auditory representation model that implemented these properties influences the interaction with and comprehension of the information it represents. We hypothesised that:

- *H1: "The hierarchical organisation of relational information allows for successful non-visual inspection and navigation of a UML class diagram".*

In addition, we were also interested in examining the effect of varying presentations modes by exploring the extent to which verbal sounds can be extended and replaced by nonverbal sounds to communicate different aspects of the hierarchical representation. Thus, we also hypothesised that:

- *H2: "varying the presentation mode will have an effect on task completion-time and/or diagram comprehension when using a hierarchical representation of a UML class diagram".*

To test the second hypothesis we manipulated verbosity as an independent variable in a between-subjects design factor of presentation mode. In a high-verbosity condition a group of participants used the Verbose mode of presentation to interact with the hierarchical structure, where content information was communicated through verbal descriptions, whereas in a low-verbosity condition, a different group of participants used the Terse mode of presentation to interact with the hierarchical structure, where most content information was communicated through nonverbal descriptions.

4.2.Measurements

We measured task completion time and overall error rates as dependent variables. Errors were divided into three main categories:

1. *Interaction Errors* are errors observed in the participants interaction with the representation including errors that occur when an invalid action is executed.
2. *Comprehension Error* are errors observed in the answers given by the participants to the questions asked in each experimental task.
3. *Efficiency Errors* are errors related to either the choice of strategy for tackling a particular problem, or the efficiency of executing such a strategy to solve the problem.

We relied on a concept known as *Interaction Traps* [2] to identify the third categories of errors. This concept has been used to evaluate Graphical User Interfaces by assessing the interactive relationship between a user and a system in terms of objectives and strategies. It is part of a framework that allow for an analysis of complex interactions where users have multiple objectives, shifting objectives or interleaving tasks [2], which are typical behaviour when interacting with external representations in general.

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 [2], 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 thus extended this concept to cover a wider range of interaction inefficiencies

By extending this concept to fit our assessments, we were able to analyse the development of user expertise as the evaluation scenarios progressed, and assess such development in terms of choice of strategy and the accuracy of executing chosen strategies. This is a new approach that is applied to evaluate the learnability of an auditory display and which, as we shall describe in subsequent sections, allowed us to formally address, analyse and classify the learning behaviour of the participants in our experiment. For this, we classified the execution of an interactive strategy as inefficient, less efficient, or efficient, as follows:

- *Inefficient Strategies* are interaction strategies where one or more interaction traps occurred.

- *Less Efficient Strategies* are all other inefficient interactions not captured by one of the manifestation of interaction traps (as defined in [2]).

- *Efficient Strategies* are instances where the user chooses the optimal strategy and executes it without the occurrence of any of the above.

4.3. Participants and Data Gathering

A total of 20 participants took part in the experiment. All were sighted computer science students and had varying knowledge of UML ranging from low to intermediate expertise. They were briefed that they were taking part in an evaluation study which tested the usability of a non-visual browser of UML class diagrams, and were given a cash incentive for their participation.

We used 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, audio recordings, and questionnaire responses. The participants were also encouraged to use the speak-aloud protocol throughout the experimental sessions.

4.4. Method

Participants were divided into two groups of ten and were randomly assigned to one of the two experimental conditions where they used either a Verbose or Terse mode of presentation to answer a set of diagram-reading tasks.

Experimental sessions were made up of a training phase and a testing phase. In the training phase subject were introduced to the relational concepts of UML class diagrams, and instructed on how to use the hierarchy to inspect and navigate through these concepts. A visual diagram was made available throughout the training phase so participants could refer to it to confirm their findings. This was not allowed in the testing phase, in which participants attempted to solve the experiment tasks (described below) in four scenarios each involving one class diagram. 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 loosing their cash incentive.

The diagrams complexity increased from one scenario to the next, and the order of scenarios was kept constant for all participants. We defined diagram complexity in terms of the number of components that constitute a class diagram as a tuple. The training diagram for instance was of a {5, 3, 2} complexity as it was made up of five objects, three associations, and two generalisations, i.e. of a relatively medium complexity in comparison with the four diagrams used in the four scenarios of the testing phase; these were {3, 1, 2}, {4, 2, 3}, {4, 3, 1}, and {7, 6, 2}.

4.5. Tasks

We assessed the usability of the proposed model in allowing flexible interaction with the diagrams by examining users performance when carrying out four different tasks. The tasks were similar to those described by Bennett in [1] in that they reflect the ability to inspect a diagram from an object perspective (the nodes) or a connection perspective (the links). Gaining an understanding of both these perspectives, we assume, is necessary to achieve full comprehension of any relational diagram.

The participants were asked to retrieve detailed information about a given object in task 1, about a relation in task 2, enumerate the diagrams components in task 3, and graphically reproduce the whole diagram in UML notation in task 4. We observed, captured and analyse their interactions as they completed each task, the results of which are described next.

5. RESULTS AND ANALYSIS

5.1. The Hierarchical Representation

All participants were able to successfully use the auditory display to complete all the tasks presented to them in the evaluation scenarios. 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 6.08). Figure 3 shows the scores for individual participants for all scenarios with scenario one at the bottom and scenario four at the top. There was a maximum score of a 100 for each scenario.

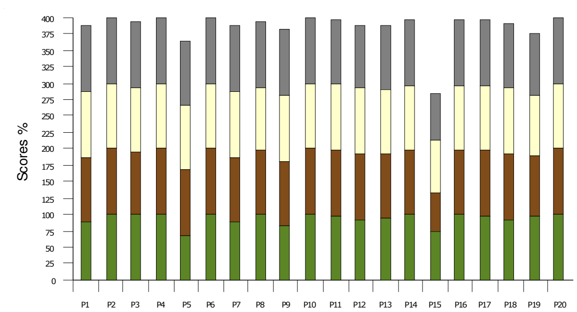

Figure 3: Participant's individual scores across scenarios. P1-10 used the Verbose mode, P11-20 used the Terse mode.

When asked about the mapping, which translates relational information into hierarchical depth, participants found it intuitive and easily accessible; although, as we shall describe in the learning analysis, few of them struggled to grasp the concept at first but performed well as soon as they understood the mapping through practice. There was some individual differences in the amount of practice needed to reach the required level of understanding.

Participants were able to access, navigate and understand the relational information through the hierarchy without the need to visualise the diagram. Some of them reported that they would visualise part of or a whole relation as soon as enough information about it was retrieved, however they would discard the mentally built picture when they move on to other parts in the hierarchy.

Others completely focused on retrieving the necessary information to solve the experimental tasks and did not even attempt or find it necessary to visualise the diagram.

Different branches of the hierarchy were explored to match the current experimental tasks performed. For instance, to solve the first task, which required the retrieval of information about a particular object, participants would explore the *Objects* container, whereas they would explore one of the relations containers to solve the second task. The ability to take shortcuts to different containers from anywhere in the hierarchy, as well as the ability to quickly switch between them were both very well received and used extensively throughout the tasks.

5.2. Presentation Modes

We note here that the amount of time it takes the screen reader to speak 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 (see step 4 and 8 above). Therefore, we have excluded the additional times required by the Verbose mode to describe a given system state that involved these parts and thus calibrate the overall task completion-times for the two conditions.

Figure 4 shows the average task completion-times for each participant across scenarios. A Mann-Whitney test revealed that differences in task completion-times between the two conditions were significant (U=18, z=2.38, p=0.008); participants in the Terse mode spent significantly less time to complete the tasks on each diagram than those who used the Verbose mode. No statistical significance was found when these times were calibrated (U=34, z=1.17, p=0.121).

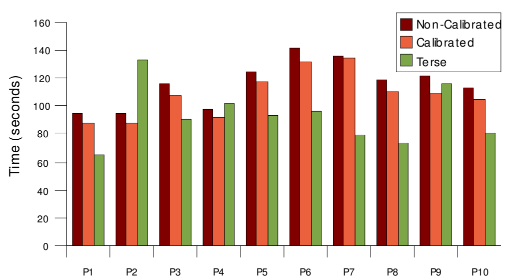

Figure 5: Average task-completion times for each participant across tasks for Verbose (calibrated, non-calibrated) and Terse modes.

Relatively more comprehension errors were made by participants using the Terse mode than those using the Verbose mode of presentation. As Figure 5 shows, these are very low error rates nonetheless, which is in line with the recorded high scores.

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 neither comprehension error rates (U=27.5), interaction error rates (U= 25.5), nor scores (U=41.5). Thus, our second hypothesis (H2) was only partially supported; varying presentation modes did not have an effect on the participants understanding of the relational information encoded in the hierarchy, but interacting with a less verbose display significantly improved their performance in terms of the times it took to complete the diagram-reading tasks.

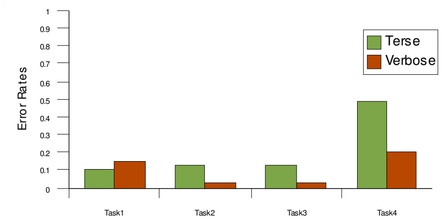

Figure 4: Average comprehension errors on each task for the two conditions

5.3. Sound Design

We could observe three distinct reactions to the sounds used to convey navigational information. Where some navigational sounds were explicitly listened out for, others were only appreciated for their aesthetics, while the rest were completely ignored – although it is worth mentioning that, for some participants, the aesthetically pleasing sounds were later on the ones listened out for to confirm that an action has been completed while their corresponding verbal descriptions ignored. It is also interesting to note that the sounds listened out for were those not at all accompanied by verbal descriptions. The ones ignored however, included the browsing sounds, which meant that mapping their pitch to indicate the depth of the hierarchy was not picked up on by any participant.

This can be attribute to the nature of the information targeted by the depth-to-pitch mapping; while engaged in a problem solving task, participants are directly focused on the content information they are inspecting rather than the browsing information because it is more directly relevant to the ultimate goal of the interaction. It would be interesting to test whether changing the target information of the mapping from the navigational to the content information would yield any different reaction.

The order organisation of the abstract sounds used to communicate the type and direction of a relation in the Terse mode were well received. Participants commented that these "sounded a lot like how the relation would have been drawn". It is also interesting to note here 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 relational 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; they were most aware of them when they changed focus or took a shortcut to a different container. Thus, the ambient sounds seemed to communicate transitional information more appropriately than the positional information they were originally designed for. Participants did not find these annoying or irritating though, which makes their unexpected function worth including in such a design.

5.4.Learning and Expertise Development

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 was increasing. In order to formally examine and assess this obvious expertise development, we classified participants' interaction efficiencies and categorised these using the extended concept of interaction traps.

Similar learning curves were observed in both conditions (see Figure 6 and 7). The percentage of efficient strategies executed across conditions increased from an average 30% in the first scenario to a dominant 75% in the last, while the percentage of inefficient strategies drastically decreased from an average 25% in the first scenario to only 2.5% in the final scenario. The percentage of Less efficient strategies also decreased from a dominant average of 45% in the first scenario to just over 20% in the final scenario.

Overall, 58% of inefficient strategies occurred in the first scenario, where participants had a low level of expertise gained mainly from the instructions given to them in the training phase of the experiment. Most of these inefficiencies were a result of the users misunderstanding some aspects of the representation. For instance, when asked to retrieve information about an object labelled *'sheep'* in task 1 of the first scenario, seven participants out of twenty took a shortcut to the *Objects* container to locate the object of interest, but as soon as they encountered an object labelled *'animal'* they decided to browse this object to locate *'sheep'* within it, under the assumption that all animals would be grouped under the *'animal'* node. This is of course incorrect because both *'animal'* and *'sheep'* are separate entities rather than lists of entities. Here, these participants correctly interpreted their objective, but selected the wrong strategy to achieve it. The extra interaction led to a manifestation of an interaction and hence the categorisation of the interactive strategy as inefficient.

Whereas the above example was a direct result of the modelled domain, a more design related reason which seems to push the user into potential interaction traps was the duplication of node names at different levels of the hierarchy. This was especially the case in the Objects container where objects where contained within each other and 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 are at, which caused momentary confusion in interpreting the relational information. 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 depth-to-pitch mapping to render the content rather than the navigational information.

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 holds and engage in explorative interaction, which we considered a positive contributor to the development of a more thorough understanding of the representation.

Looking at individual distributions of interaction efficiencies for each participant, we could categories three distinct types of learners. Fast learners were participants that quickly picked up on the workings of the representation and manage to execute more efficient strategies in the earlier scenarios. Medium learners took a rather steady pace in developing their expertise and manage to execute relatively more efficient strategies in the final scenarios. Only two participants were classified as slow learners having struggled throughout the four scenarios and managed to execute very few efficient strategies overall.

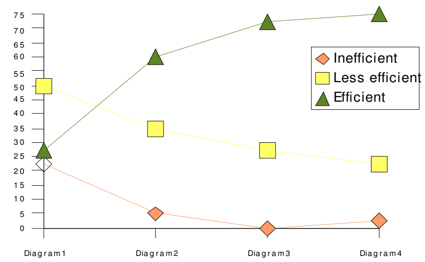

Figure 6: Percentage of efficiency levels of the strategies employed on each diagram in the Verbose mode

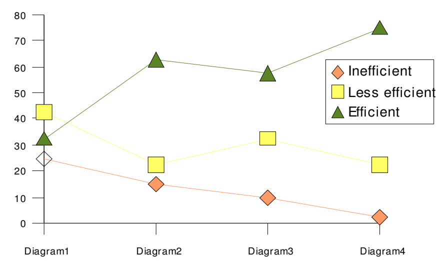

Figure 7: Percentage of efficiency levels of the strategies employed on each diagram in the Terse mode

The reaction to the increase in diagram complexity between the three learner categories also differed. While fast learners were able to efficiently accommodate the challenge by carrying the tasks at relative ease, medium learners' performances were staggered the more complex the diagrams got, and as expected, a slow learner's performance was evidently affected by the increase in complexity. The fact that most participants were categorised as fast learners, however, testifies to the relative ease of usability of the auditory display and the efficiency of the hierarchy in conveying information about relational diagrams, which, coupled with low comprehension error rates and high scores, further confirm and support our first hypothesis (H1).

6.DISCUSSION

The above statistical and qualitative results confirm that the relational concepts encoded in UML class diagrams could be represented hierarchically, and this allowed for such information to be non-visually accessed.

The hierarchical representation was designed to simulate the characteristics of external representations outlined in section 2.1. The positive impact of these properties on the usability of the hierarchy was observed here when sound was used as the main means of communication. Locational indexing was simulated by the organisation of similar components at hierarchically adjacent locations. All objects, for instance, could be found at the same hierarchical level inside the Objects container, and all relations inside the Associations and the Generalisations containers. Similarly, connectivity information specific to a given object or a relation are grouped in the same hierarchical level inside the specific object's or relation's node.

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, got less disoriented and became more efficient at executing different problem-solving strategies.

Furthermore, the fact that each of the three main branches of the hierarchy represent the same abstract information from various perspectives, allowed for the relational information to be re-represented, with each re-representation imposing different levels of constraints on the possible inferences that can be made about it. Participants in our experiment used such re-representations to focus their interaction on different aspect of the information as required by the demands of the particular experimental task, which are similar observations to those reported by Bennett in [1], where different types of tasks were observed to be best supported by a matching representation.

Of course these properties could not have been simulated nor observed if the users could not actively interact with the auditory display. Active interaction allows for this system and other similar accessibility solutions to be potentially studied and evaluated in terms of their role as external representations. It minimise the negative aspects of the temporal and transient nature of the auditory medium of presentation, allowing for the auditorally represented information to be immediately available in the environment. The ability to dynamically navigate and inspect the hierarchy in our system meant that parts of the problem space needed not to be remembered during the problem solving process.

An interesting aspect observed in the participants interactions with the display was the ease by which they interpreted the nonverbal representations of the explicitly relational components of the diagram. That is, relations' types and directions were more efficiently and intuitively accessed when represented non-verbally than through speech. A well designed diagram is said to be one that allows its user to make relatively straightforward mapping between the diagrammatic depiction and the situation it represents [6]. These observations, thus, call for more in depth investigations to identify representational properties that are intrinsic to the auditory medium of presentation. Then, the challenge would be to understand the mechanisms that underlay the question of how such properties are actually realised in this modality. An essential first step towards this direction would be to compare various representational systems in different modalities in order to establish the strengths and weaknesses of each.

7. CONCLUSION

Current advancements in the study of human cognition has put much focus on understanding the role of external representations in supporting problem-solving tasks such as inference, prediction, and problem interpretation. Research into the auditory display of information is providing increasing evidence of the potential of using sound to support such tasks. However, as a modality, the study of the representational properties of sound has been largely neglected. This paper described an approach for translating graphically represented information from the visual to the auditory modality by emphasising the meaning of what is represented rather than the structural tokens of the medium that presents it. We used a hierarchical structure that simulates properties of external representations to organise the relational information encoded in UML class diagrams and reported on an experiment that evaluated its viability. Our results show that replacing verbal descriptions of relational information with nonverbal sounds significantly improved performance without compromising comprehension. These results suggest that properties of external representations can be observed to positively impact interaction with a structuring model that is presented through sound only and highlight the importance of deeper investigations into the intrinsic representational properties of the auditory medium.

8. REFERENCES

[1] D. J. Bennett, “Effects of Navigation and Position on Task When Presenting Diagrams to Blind People Using Sound”. *Lecture Notes in Artificial Intelligence*, 2002, 2317:161

[2] A. Blandford, H. Thimbleby, N. Bryan-Kinns, “Understanding Interaction Traps”. i*n Proc. of HCI 2003: Designing for Society*, Research Press International, UK, 2003, 8-12

[3] A. Brown, S. Pettifer, R. Stevens, “Evaluation of a Non-Visual Molecular Browser”. In *Proc. of the 6th International ACM Conference on Computer and Accessibility*, 2004, 40-47

[4] C. A. Gurr, “Effective Diagrammatic Communication: Syntactic, Semantic and Pragmatic Issues”, *Journal of Visual Languages and Computing*, 1999, 10, 317-342

[5] A. R. Kennel, “Audiograf: A Diagram-Reader For The Blind”. *In Proc. of the 2nd annual ACM conference on Assistive technologies*. ACM Press, New York, NY, 1996, 51-56.

[6] C. Peter, H. Cheng, K. Lowe, M. Scaife, “Cognitive Science Approaches to Understanding Diagrammatic Representations”. *Artificial Intelligence Review* 2001, 15: 79-94

[7] R. Ramloll, S. Brewster, ”*A Generic Approach for Augmenting Tactile Diagrams with Spatial Non- Speech Sounds”*. In Extended Abstracts of ACM CHI2002, ACM Press 2002, 770-771

[8] M. Scaife, Y. Rogers, “External Cognition, Interactivity and Graphical Representations”*, Thinking With Diagrams, IEE Colloquium.* London 1996*,* 1996/010, 8/1-8/6.

[9] The TeDUB-Project (Technical Drawings Understanding for the Blind). Available at http://www.tedub.org

[10] J. Zhang, “The Nature Of External Representations In Problem Solving”. *Cognitive Science,* 1997, *21*(2):179-217