Communicating Graphical Information to Blind Users Using Music: The Role of Context

James L. Alty
LUTCHI Research Centre
Dept. of Computer Science
Loughborough University
Loughborough,
Leicestershire
LE11 3TU, UK
+1509 222648
j.l.alty@lboro.ac.uk

Dimitrios I. Rigas
School. of Computing and Mathematics
University of Huddersfield
Huddersfield,
HD1 3DH, UK
+151 231 2268
d.rigas@hud.ac.uk

ABSTRACT
We describe the design and use of AUDIOGRAPH - a tool for investigating the use of music in the communication of graphical information to blind and partially sighted users. This paper examines the use of the system to communicate complex diagrams and gives some examples of user output. Performance is not as good as expected and it is postulated that context will play an important part in the perception of diagrams communicated using music. A set of experiments are reported which indicate that context does indeed seem to play an important role in assisting meaningful understanding of the diagrams communicated. The implications for using music in auditory interface design are discussed.

Keywords
Blind Users, Music, Graphics, Interface Design, Empirical.

INTRODUCTION
Graphical user interfaces (GUIs) have now become the dominant technique used for interaction between users and computer applications. Such interfaces are said to offer a user-friendly approach with a high utility. One important reason why it is thought that they have been successful is their use of Direct Manipulation techniques, a term first coined by Shneiderman [1] to describe the emphasis on continuous representation of objects, physical actions rather than complex syntax, and rapid incremental reversible actions with instant feedback. There is no doubt that such interfaces have had a beneficial effect to users in general, but they represent a serious step backwards for blind or visually impaired users.

The emphasis on graphical representations has rendered devices such as Screen Readers (on which many blind and visually impaired users have traditionally relied) much less effective. Screen Readers are based on a speech representation of the content of the screen. Because of the intensely serial nature of speech, graphic-intensive screens can be difficult to describe using this medium.

We have been interested in exploring the use of auditory interfaces for some time, both for visually impaired users and for those with normal visual abilities. Indeed, we are particularly interested in creating what we have termed "equal opportunity interfaces" - that is, interfaces which do not make prior judgments about the media capabilities of the user population but offer a variety of communication media, from which the user can select an appropriate mix to match their capabilities and limitations. Such interfaces would not only be capable of adaptation to match users with significant hearing or visual difficulties, they would also allow trade-offs to be made between vision and audio for common applications. The work in this paper explores an extremum of this spectrum (audio only) in contrast to most interface work which is located at the other extremum (visual only).

In comparison with visual media, which have been extensively explored, non-speech use of the auditory medium has been largely ignored. The use of audio in human computer interfaces was first reported by Bly. Gaver has proposed the idea of Auditory Icons [3] and used them in the SonicFinder [4] to assist visually impaired users with editing. Such auditory icons are short bursts of familiar sounds. Edwards has developed the Soundtrack system [5] to assist blind users in editing. Blattner an co-workers has proposed the use of musical phrases called Earcons [6] - short musical melodies which shared a common structure such as rhythm. However, apart from suggesting the idea of sharing a common rhythm, no other musical techniques were suggested. Blattner has also used some properties of musical structures to communicate the flow in a turbulent liquid. During the last few years, some research workers have attempted to use music in interfaces. Brown and Hershberger [7] employed music and sound in supporting the visualisation of sorting algorithms but carried out no experimentation to see how valid their approach was. Alty [8] has developed purely musical mappings for the Bubble Sort and the Minimum Path Algorithms. Recently some progress has been made in assisting program debugging using music, for example, the work of Bock [9] , and Vickers and Alty [10]

There are very few recorded experiments which systematically evaluate the use of music in human computer interaction. Brewster [11] has used music to a limited extent in designing Earcons for Telephone based interfaces and has combined Earcons with graphical output to assist users when making slips in using menus [12]. In both cases some empirical investigations were carried out. Mynatt has investigated the [13] mapping of visual icons into auditory Earcons. She comments on the importance of metaphorical qualities of everyday sounds and how these can be used in designing good Earcons.

This paper reports continuation of our previously published work on exploring the usefulness of auditory media (particularly music) in interface design. Some of our work has concentrated on the audiolisation of algorithms (Alty, 1995) [8] , and program debugging ([10] and [14]). The work reported here is a continuation of our investigation into the use of music to communicate diagrams to blind users (Rigas and Alty, 1997) [15]. This work has already shown that the highly structured nature of music can be used successfully to convey graphical information in diagrams to visually challenged users. The only other related work of which we are aware is the AUDIOGRAF system of Kennel [17]. This system use touch to concentrate upon particular areas of the diagram. Output is by speech or a set of individual notes.

THE RATIONALE BEHIND THE INVESTIGATIONS

The main objective behind the set of experiments reported in this paper is to see if music alone can be used to transfer meaningful information to users on the computer interface. We decided to use an extreme case - to determine if blind users could, by the use of music alone, appreciate the spatial layout of objects in a graphical area using musical feedback alone. Because of this emphasis on music, and the fact that there has been little previous work on the use and evaluation of musical representations, music alone was used both to communicate all information about the interface and for all input control commands. We appreciate that in a real interface, such as a commercial computer application, it would be sensible to use speech in addition to music (particularly for conveying exact information), but we felt that the inclusion of other auditory modes of interaction would risk confusion in the interpretation of the experimental results.

THE AUDIOGRAPH SYSTEM

We therefore constructed the AUDIOGRAPH system, which we employ in our experiments with music for blind users, some of this work has already been described in an earlier paper (Rigas and Alty , 1997) [15] so only a brief

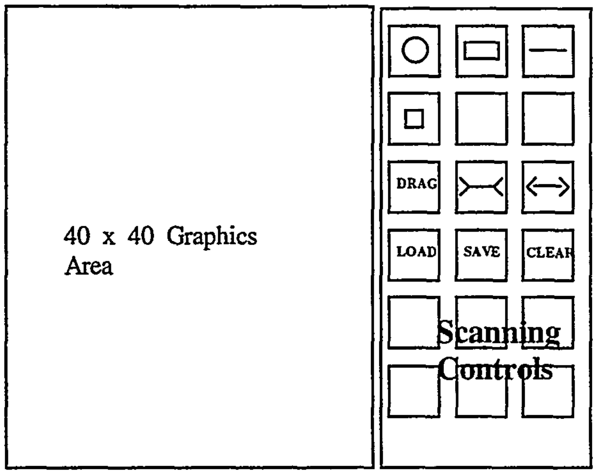


Figure 1 The AUDIOGRAPH System

description of the system will be given here. The AUDIOGRAPH system has both a visual interface (for the experimenter) and a completely auditory interface for blind users. All musical output is produced by MIDI output from Visual Basic, communicated to a stereo sound system using a Creative Labs Soundblaster 16-bit card. The actual sounds are created using a Roland MT32 Multi-timbre output device. All output information and input feedback are communicated using music alone. It is important to note that the AUDIOGRAPH tool is aimed at users with an

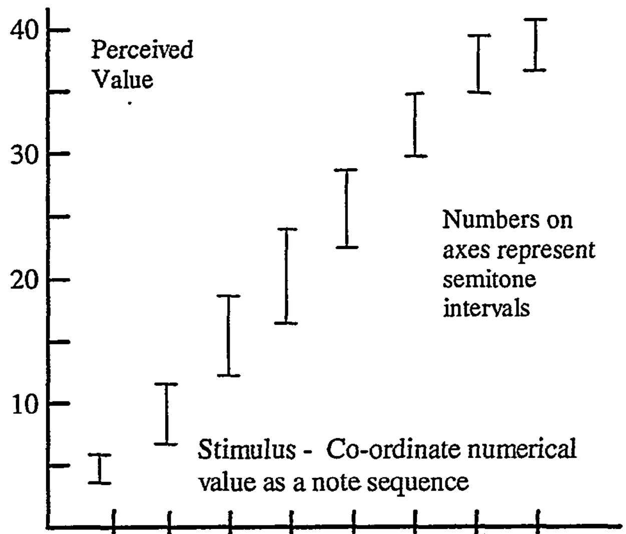


Figure 2 Recognition of Co-ordinate from a Musical Sequence

average musical ability. No special musical ability or training was expected for any of the experiments performed. A musical questionnaire was given to all participants, prior to experimentation, to check their musical experience and knowledge.

The interface (as seen by the experimenter) is shown in figure 1. There are two major areas - the grid on which diagrams are drawn and perceived, and the control area, where buttons are used to select and control actions. Input interaction with the system is exclusively via the arrow keys on the keyboard (our blind users indicated that they did not like using the mouse). These are used for moving the cursor round the graphics area (a 40x40 grid on the left of the Figure) and to select control actions. Other keys (usually any key) are used to confirm actions. The space bar is used to toggle between the graphics area and the control area. The main control actions include shape selection (circle, rectangle, line, square), shape expansion and contraction, dragging, and file loading and saving. There are also a selection of scanning controls which communicate the current contents of the graphical area in different ways.

The following graphical information is communicated using music in the AUDIOGRAPH system (Rigas and Alty, 1997) [15]:

1. the current position of the cursor or the position of a graphical object
2. the nature of a graphical object (e.g. its type, size and shape)
3. The overall position of graphical objects using various scanning techniques

All these used a similar metaphor for implementation - a co-ordinate point is described by using a musical mapping from distance to pitch (a higher note describing a larger co-ordinate value), and X and Y co-ordinates are distinguished by timbre (Organ and Piano). These representations were chosen after extensive experimentation (see [15] for detail). For interest we reproduce the results for pitch interval recognition in figure 2. Subjects were played the sequence of notes from the same lower note (the origin) to the note representing the co-ordinate value, and asked to determine that value.

The different timbres used to represent the X and Y axes (Piano and Organ) were chosen also after extensive experimentation on timbre perception (again see Rigas and Alty). Graphical shapes were described musically by tracing out the shape in sound using this metaphor.

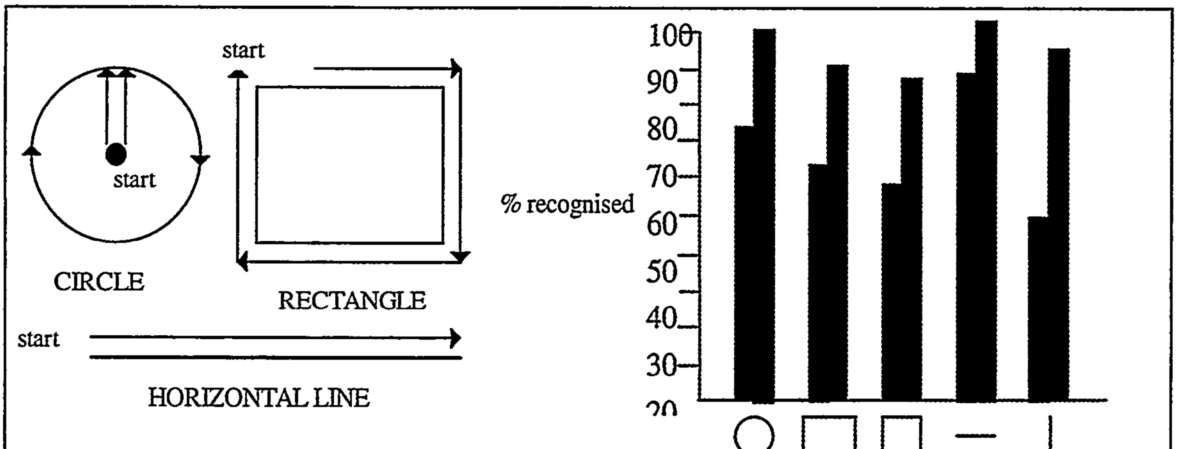


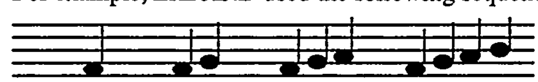
Figure 3 The Musical Description of Graphical Shapes and their Identification by Users before and after Training

Examples are given in Figure 3. Even before training most users recognized the Circle and the two lines. The training session only involved the presentation of 5 examples of each shape. The results from [15] are given in figure 3

Experiments were also carried out to see if users could estimate the size of graphical objects as well as their overall shape. The results indicated that users could estimate size to an accuracy within 10%

Control actions were communicated using meaningful Earcons (many previous Earcons have not had meaningful shapes).

For example, EXPAND used the following sequence:



and CONTRACT was represented by the inverse of this pattern. Users found these completely intuitive and no training was required. The UNDO command was represented by the playing of a tune with a "fault" in it followed by the "correct" tune. At first hearing users were baffled by this, but on hearing the explanation they understood it immediately and had no further trouble recognizing it. This is an example of the importance of context which will be examined shortly. File names were represented as simple tunes.

Three different scanning techniques were provided to enable users to obtain an overall appreciation of the graphical space and to enable them to build up a mental model of the space.

Top-Down Scanning - reads the area musically starting at the top left hand corner and scanning progressively down the area left to right.

Centre Scanning - starts at the centre of the screen and musically scans the area in increasing circles

Ascending Scanning -This scan was provided after user comments. It scanned the objects in the space in ascending order of size.

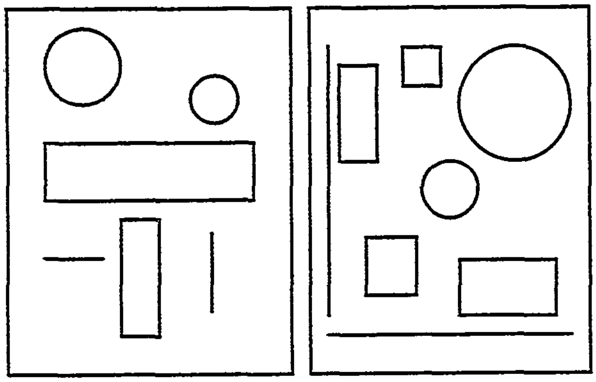


Figure 4. The two stimuli Used in the Arbitrary Graphics Set

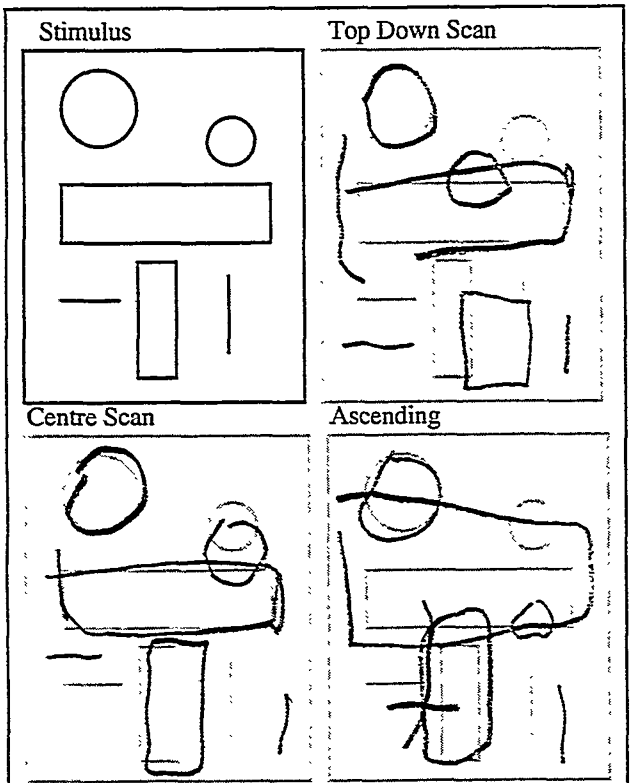


Figure 5. Three Subjects attempts using different scans

RECOGNITION OF AN ARBITRARY SET OF OBJECTS

An experiment was performed to investigate the perception and interpretation of subjects who were presented with a set of arbitrary objects from our display set (lines, circles, squares, rectangles). The two stimuli are shown in Figure 4.

Six subjects took part, three using the first stimulus and three using the second stimulus. Each subject heard the diagram three times and was then asked to draw the diagram. Drawing was carried out using paper with a raised grid on it. This in itself caused drawing problems as when the pencil crossed over a grid, the drawing necessarily became more uneven, and errors in the diagram are a combination of perception errors and drawing errors. It is impossible to devise a scoring mechanism for the output, but the success (or failure) achieved by subjects can be appreciated by examining some actual output. These are shown in Figures 5 and 6 for one subject from each group. It can be seen that subjects have obtained a broad picture of the diagrams, the first subject performing qualitatively rather better than the other.

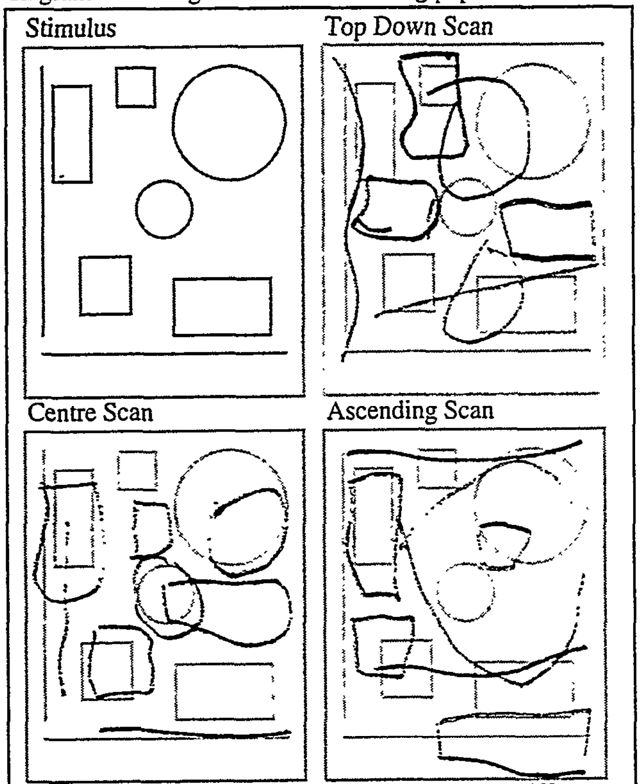


Figure 6. Another three subjects using a different stimulus and the same three scans

Whilst the diagrams generally have captured correct number of objects in the space and the distributions are broadly in agreement, the perception of size is disappointing. For example, in figure 5, Ascending Scan, the subject has realized that one circle is bigger than the other but both have been drawn incorrectly. Our previous experiments in perception and size led us to expect better performance.

The difference between our earlier experiments on individual objects and the experiments on arbitrary sets is the complexity of the latter stimuli. Users need some

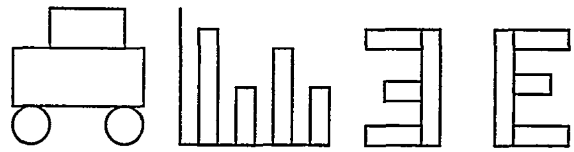


Figure 7 Meaningful Stimuli and Hints

organising principle to assist them in coping with the more complex diagrams. In reality, users rarely have to comprehend meaningless sets of shapes, and their perception is guided by their expectation of what is presented. We therefore decided to carry out a set of experiments with meaningful sets of objects to see if context had a significant effect on comprehension.

RECOGNITION OF A VISUALLY MEANINGFUL SET OF OBJECTS

Twelve blind subjects took part in the experiment. Two different groups of six subjects were constructed from these. The first group (the control group) listened to four different diagrams and were not given any guidance as to the nature of the diagrams. The second group (the experimental group) listened to the same diagrams and were given a hint or semantic guidance as to their nature. Thus, they might be told that the diagram was "a type of vehicle", "a number", "a letter of the alphabet", or "a method of data representation". Both groups knew that the diagram represented some real world concept (i.e. it was not a random collection of shapes). The four diagrams, together with the hints given are shown in figure 7.

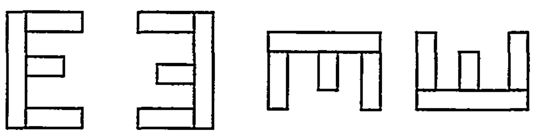


Figure 8 The four diagram variants presented

The diagrams were presented aurally using AUDIOGRAPH in a different order to each subject in each group. All subjects used the Centre-Scanning technique to explore the graphical space.

In those groups given no semantic guidance, no subjects assigned any meaning to the first two diagrams. Two subjects (33%) interpreted the third figure as a "three" and three subjects (50%) interpreted the fourth figure as the letter "E".

In contrast, in the semantic group, all subjects recognized the Car, the Number "3" and the Letter "E", and four subjects (66%) recognized the Data Set. In most cases , for either group, subjects were able to identify the individual components even if they did not get the overall meaning. It is possible that the "Car" was the most representative object in the vehicle set and might well have been guessed, in contrast to "E" and "3" which are not special items of their sets.

The presence of a narrow perceptual context therefore appeared to assist subjects in assigning meaning to the set of graphical objects by using their expectations to bridge perceptual gaps (i.e. distances between objects not properly interpreted). If this is the case, would a shift of perceptual context cause subjects to interpret the same stimulus in a different way ?

Five blind subjects (who were in the control group in the previous experiment) were presented with four diagrams as shown in figure 8 below.

The subjects were presented aurally with these four diagrams in the following order with the following hints (as in Table 1)

Subject	Presentation Order	Semantic Guidance
S1	1,2,3,4	No hint Given
S2	4,3,2,1	Upper Case Letter
S3	2,4,1,3	Upper Case Letter (Backwards or Rotated)
S4	3.1.4.2	Number
S5	4.1.3.2	Number (Backwards or Rotated)

Table 1. Presentation Order and Hints given

The results are presented in Table 2.

Diagram No.	S 1	S 2	S 3	S 4	S 5
1	-	E	E	-	3
2	-	-	E	3	3
3	-	-	E	-	3
4	-	-	E	-	3

Table 2 Performance of Subjects with Semantic Guidance

The results indicate that the communicated graphical information using the musical mapping of the graphical area is interpreted by subjects as a random combination of objects in the absence of a perceptual context. However, in the presence of an expectation, the graphical information communicated is interpreted as a meaningful shape. This implies that the perceptual context has a direct and contributing role in the interpretation of the music used to communicate the graphical objects. Thus although the musical mapping is one variable which contributes to the user's perception and interpretation simultaneously, the creation of an appropriate context in the listener is another important variable. Furthermore, the absence of a meaningful context will often result in a lack of a meaningful interpretation even though the individual elements are perceived. This finding, which has parallels in the use of other perceptual channels (for example the importance of context in visual understanding) has important implications for auditory interface design.

DISCUSSION - IMPLICATIONS FOR THE DESIGN OF AUDITORY INTERFACES

The use of music in an auditory interface to communicate graphical shapes has shown promising results. Users were able to identify shapes and their approximate size, use the tool to move them around the area, and adjust their size. They were able to use the musical controls to expand and contract shapes, file and retrieve them and drag them. Although an actual tool would have used speech output in certain operations in preference to music, the insistence of a full music interface has tested the idea quite thoroughly. User feedback was quite positive (reported in Rigas and Alty [15] but there were some adverse comments, not surprisingly on the length of the musical messages and the effort involved in interpreting them. Such comments would have to be investigated and the implications explored, in the design of any realistic tool. We believe, for example, that much more use of musical abstraction could be used to reduce the length of messages.

However, it is in the experiments on the role of context where we see some interesting lessons being drawn. Firstly, there is much more to auditory interface design than simply producing unique and identifiable mappings. This is a necessary, but not sufficient, condition for a successful design. We call this level the *DETECTABLE MUSICAL MAPPING.* At this level, the designer must allocate the musical structures (e.g. pitch, rhythm, timbre etc..) to domain structures. The target here is to produce a recognizable and distinguishable musical message, to the listener which can be understood in the presence of other musical structures in the application. This level provides a detectable mapping. There are clearly many possible detectable mappings for any problem domain.

The second level we *call CREATING PERCEPTUAL CONTEXT.*. At this level, given a detectable mapping, the designer must create the perceptual context or expectation in the listener because interpretation of the music will depend on the expectation of the listener. At this level, the individual structures are interpreted by the listener in domain terms (music being regarded as a metaphor). Listeners can now assign meaning to individual messages but cannot necessarily reason about the global interaction.

Our results also lead us to suggest the existence of a third level *THE REASONING AND SEMANTIC LEVEL.*. At this level, the listener develops higher level structures to understand the domain from a higher level viewpoint. This level allows users to assign meaning to musical messages without further training or instruction. The very act of mental activity at this level also is likely to increase memorability of the interface.

These ideas can be illustrated with examples taken from the AUDIOGRAPH. The Detectable Musical Mapping is used to identify the important characteristics of the domain which MUST be distinguishable. In the case of the AUDIOGRAPH system, the listener certainly needs to distinguish the following

1. the size of a coordinate
2. an X coordinate from a Y coordinate
3. the different graphical shapes
4. the different control actions

There may be other domain events which need to be communicated depending upon user task, but the above list is certainly the minimum required for basic understanding.

The actual mappings used were:

1. The size of a coordinate is mapped into pitch in the Chromatic Scale (higher numbers having higher note values). Additionally, the notes are grouped into 10s and the longer the sequence, the higher the coordinate value. Thus we provide three basic handles for determining a coordinate value.
2. The X and Y coordinates are distinguished by Timbre (Organ and Piano) and in Time (Y always comes after X). Additionally, we provide a distinctive drum note to prepare the listener for the start of the coordinate sequence.
3. The shapes are derived from geometrical traces of the objects which are shown by experiment to be distinguishable.
4. Control Actions are short distinct Earcons

The required domain differences under these mapping are detectable. We have already shown in [15] that users with average musical ability can distinguish between instruments in the different classes of the orchestra and map pitch (approximately) to numeric difference.

The Perceptual Context is created by use of a common mapping for coordinates, shapes, and cursor movement. The audio versions of the shapes are directly related to their geometric counterpart, in a similar manner to the coordinate description of the space. Thus they are not only detectable, but also can be understood in terms of the metaphor relating pitch to length. The control action Earcons, likewise, are underpinned using a metaphoric interpretation. The EXPAND Earcon is detectable but also does "expand" numerically and is the opposite to CONTRACT. UNDO can be heard to correct a tune and is not just a unique Earcon.

Finally we have observed users employing the third level once they become familiar with the auditory mappings. When one begins to communicate diagrams which have an underlying structure, users are rapidly able to exploit this. For example, if users knows that the graphical scene is a row of houses, and they find windows in the first house, they immediately assume that windows are likely to exist in the others. This aspect is not explicitly presented in the detectable mapping nor in the perceptual support level.

The importance of the perceptual level has been stated in other work. When Mynatt [13] initially designed the auditory icons for Mercator, she had assumed that the use of auditory icons was limited to concrete representations (e.g. sound of a typewriter). Later she was able to represent more abstract actions utilizing the metaphorical qualities of everyday sounds. She then began using sounds for which she did not want the user to identify the original source of the sound. In this respect, this is moving closer to our work in music, where the source of the sound (Clarinet, Trumpet, etc..) is not related to the use of the sound in the representation scheme for the domain. She then critically examines claims about users preferring musical to natural sounds in interface design and points out that this is not because the music somehow sounded "better" but because the musical sounds had been carefully constructed to reflect metaphorically what they were representing (like our EXPAND and CONTRACT Earcons in AUDIOGRAPH or the jagged list in the Bubble Sort). For example, her musical sequence for representing termination of a phone call ended abruptly and sounded as if the phone was being replaced on the hook.

Other musical mappings (particularly Earcons) used in a number of recent studies have relied to some extent on perceptual support. For example Brewster and Crease [12] report the use of auditory Earcons in an experiment investigating user errors when employing menus. Relationships in the menus are reflected in the auditory design (a similar set of timbres are used for items in the same menu). However other mappings were not so perceptually obvious. Moving from one menu item to another was signalled by alternating notes (B2 to E3 in the scale), whereas one might have thought gradually decreasing notes (representing the descent into the hierarchy, might have been closer, perceptually. Also rhythm was used to signal selection and slips. Although such events will be readily detectable, they are not related perceptually to the actions being described. Perhaps a glissando might have been better to describe a slip, for example. We are not suggesting that Brewster and Crease would necessarily have obtained improved results using these mappings, but are simply using the example to illustrate how the perceptual level might be applied.

We have been impressed with the capabilities offered by music in interface design particularly for those who are blind or who are visually impaired. Much more work is required at two of the levels identified in our work - the detectable level and the perceptual level.

At the detection level we need to know more about the capabilities of human beings with an average musical capability. There certainly is a basic musical capability in most people (the existence of the popular music industry is a testament to that). We need to know better how to exploit it, by finding out what people can and cannot perceive musically. Secondly, we need to explore how to support this basic detectability with perceptual support. It is possible that we could learn much here from those who write jingles for commercial television or composers of film music. We also need to know more about possible cultural differences in the ways in which human beings interpret music.

We opened this paper with a plea for the development of "equal opportunity " interfaces, interfaces in which users could decide for themselves, the distribution of information presented between the visual and audio channels and this paper has investigated the audio extremum. A full understanding of this idea will require an appreciation not only of the perceptual support which can be offered in visual and aural interfaces but also how the two interact.

ACKNOWLEDGMENTS

We would like to thank all those subjects from the Royal National Institute for the Blind who spent many hours taking part in the experiments and provided many helpful comments, and to the management of the Institute for their readiness to cooperate.

REFERENCES

[1] Shneiderman, B., "Direct Manipulation: A Step beyond Programming Languages", IEEE Computer, August, 1983, 57 - 62.

[2] Bly, S.A., Communicating with Sound, Proc. of CHI'82, ACM Press, Addison-Wesley, New York, 371-375.

[3] Gaver, W., Using Sound in Computer Interfaces, Human Computer Interaction, Vol. 2, No. 2, (1986), 167-177.

[4] Gaver, W., The SonicFinder: An Interface that Uses Auditory Icons, Human Computer Interaction, Vol. 4, No.1, (1989), 87-94.

[5] Edwards. A.D.N, (1989), Soundtrack: An Auditory Interface for Blind Users, Human Computer Interaction, Vol. 4, No. 1, (1989), 45-66.

[6] Blattner, M.M., Sumikawa, D., and Greenberg, R., Earcons and Icons: Their Structure and Common Design Principles, In Human Computer Interaction, Vol. 4, (1989), 11-44

[7] Brown, M., and Hershberger, J., Colour and Sound in Algorithm Animation, IEEE Computer Vol. 25, No. 12, , (1992), 52-63.

[8] Alty, J.L., Can we Use Music in Human Computer Interaction ?, Proc. HCI'95, People and Computers X,

Kirby, M.A, Dix, A.J., and Finlay, J.E., (eds.), ,Cambridge University Press, 409 - 423.

[9] Bock, D.S., Auditory Software diagnosis Using a Sound Domain Specification Language, Ph.D. Thesis, Syracuse University, Syracuse, (1995).

[10] Vickers, P., and Alty, J.L., CAITLIN: A Musical Program Auralisation Tool to Assist Novice Programmers with Debugging, in Proc. of ICAD'96, (1996), Santa Fe Institute.

[11] Brewster, S., Navigating Telephone-based Interfaces with Earcons, Proc. HCI'87, (Bristol, 1997), People and computers XII, (Thimbleby, H., O'Conaill, B., and Thomas, P., eds.), Springer, Berlin, 39-56.

[12] Brewster, S., and Crease, M., Making Menus Musical, Proc. INTERACT'97, (Howard, S., Hammond, J., and Lingaard, G., eds.), Chapman and Hall, London, 389-396.

[13] Mynatt, E.D., Designing Auditory Cues for Mercator, Proc. INTERACT'97, (Howard, S., Hammond, J., and Lingaard, G., eds.), Chapman and Hall, London, 236-243

[14] Vickers, P., and Alty, J.L., The CAITLIN Auralisation system: Hierarchical Leitmotif Design as a clue to Program Comprehension, in Proc. ICAD'97, (1997), Santa Fe Institute.

[15] Rigas, D., and Alty, J.L., The Use of Music in a Graphical Interface for the Visually Impaired, Proc. of INTERACT'97 (Howard., S., Hammond, J., and Lindegaard, G., (eds.)), Sydney, 1997, 228-235.

[16] Fernstrom, M., and Bannon, L., Explorations in Sonic Browsing, Proc. HCI'87, (Bristol, 1997), People and computers XII, (Thimbleby, H., O'Conaill, B., and Thomas, P., eds.), Springer, Berlin, 117-132.

[17] Kennel, A.R., AUDIOGRAF: A Diagram-Reader for the Blind. In Proc. ASSETS'96, Vancouver, ACM Press, (1996), 51 - 56.