Using Non-speech Sounds to Improve Access to 2D Tabular Numerical Information for Visually Impaired Users

Rameshsharma Ramloll, Stephen Brewster, Wai Yu & Beate Riedel

Glasgow Interactive Systems Group, Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK

Tel: *+44 141 330 4256*

Fax: *+44 141 330 4913*

Email: *{ramesh,stephen,rayu}@dcs.glasgow.ac.uk, beate@psy.glasgow.ac.uk*

URL: *http://www.dcs.glasgow.ac.uk/research/gist/*

We investigated two solutions for numerical (2D) tabular data discovery and overview for visually impaired and blind users. One involved accessing information in tables (26 rows x 10 columns containing integers between and including 0 and 100) by this target user group using both speech and non-speech sounds. The other involved accessing similar information in tables of the same size through speech only by the same user group. We found that opportunities to access data through non-speech sounds result in a highly significant decrease in the overall subjective workload, more specifically in the mental, temporal, performance and frustration workload categories. This subjective workload assessment was supported by our quantitative results which showed a highly significant decrease in the average time taken to complete a given data comprehension task and a significant increase in the number of successfully completed tasks.

Keywords: data visualisation, sound graphs, subjective workload assessment, non-speech sounds, 2D tables, speech output.

1   Introduction

This paper describes the design and evaluation of a tool for improving access to numerical information in 2D tables by visually impaired and blind people. Our early experimentation with homegrown prototypes for accessing data in tables suggests that tabular data discovery and overview rapidly becomes overwhelming especially for large tables when access occurs solely through the speech medium. We propose that that this is because it takes an impractically long time to listen to the whole data set in speech and secondly, it is hard to construct meaningful meta-information, e.g. trend information, about the data presented in this manner because of the limitations of human working memory (Gomez et al., 1994). In this paper, we report the results of our evaluation carried out to verify formally these insights.

1.1   Issues Relevant to Cell-by-cell Access in Speech

The present de-facto standard for accessing numbers in 2D tables is through synthetic speech. Many blind and partially sighted people use screen reader applications to listen to the contents of a table on a cell-by-cell basis. The speech synthesiser reads out the cell, selected through cursor key presses. Discussions with visually impaired users accessing tables using this approach reveal three central issues:

Awareness of their current position within the table: Users get quickly discouraged when they are lost. The 'Where am I?' information has to be made readily available to avoid any unsettling uncertainties about their current location in the table.

Overloaded speech feedback resulting from their navigational moves: The common approach of associating every cursor key press with speech feedback describing the current row index, column index of a cell followed by its contents, may be inappropriate. All this information is not needed all the time and loads unnecessarily the short-term memory of users.

Lack of overall picture of the data structure containing information of interest: Users need to be aware of the overall size of the table and need to have some idea of peripheral cell contents in addition to that of the currently selected one. This concern about the importance of the context of a local piece of information on its role in the overall data representation is in general relevant to most representation media (Kurze, 1998; Alty & Rigas, 1998).

The three issues just mentioned reflect clearly the users' need to remember more than what is immediately accessible at any point in time. According to Ben Shneiderman's (1994) mantra of "overview first and filter to reveal detail", data browsing scenarios typically involve shifts between data overviews and data inspections. Unfortunately, producing data overviews in the auditory medium is currently not as clear as it is in the visual medium. A number of designers have tackled this problem through the use of non-speech sounds but not specifically for 2D tables containing numbers (Gardner et al., 1998). While the core of our approach is not radically different, right from the beginning, we aim at evaluating our application at the level of the real world tasks they are designed for rather than at the level of narrowly defined ones more suitable for psycho-acoustic experiment designs.

1.2 Motivation for Using Non-speech Sounds

There have been a number of earlier attempts at providing users with access to trend information about a sequence of numbers (Mansur, 1975). A common approach is to represent a given number using the pitch of non-speech sounds. The use of pitch to represent data was shown by Pollack & Ficks (1954) to have a lower error rate in comparison to other auditory dimensions such as sound duration, repetition rate, or loudness. A central aspect of our approach is the assumption that if overview information becomes important at some point during a data comprehension task, then opportunities should be provided to allow numerical values to be accessed in pitch. When a numerical sequence is accessed as a series of pitches, the resulting 'signature tune' can presumably inform the listener about its contents quicker than if its contents were to be read out one at a time in speech. An early investigation showed that there is a natural tendency to perceive a pitch that is higher in frequency to be coming from a source that is vertically higher in space when compared to some lower tone (Roffler & Butler, 1968). This result encourages us to map low numerical values to low pitches and to increase the latter as the numerical values increase perhaps to maintain some congruity with the visual graphical representations we are accustomed to.

This approach is not without problems. While the human ear is very good at detecting small differences of pitch at a high temporal resolution, the ability to make absolute judgment is a very rare skill (1% of the population have perfect pitch) (Moore, 1982). Fortunately, we are not interested in replacing the representation of numerical values with pitch. We do not expect users to make correct numerical value estimations based on pitch perceptions. In addition, experiments have shown that there is no linear relationship between pitch perception and the frequency of the auditory signal (Stevens et al., 1937). Indeed, any relationship that exists is user specific. This non-linear relationship between the number and mapped pitch makes accurate meaningful comparisons between pitches hard. To clarify this issue further with an example, it is quite difficult to listen to two pitches, perceive the first to be twice higher than the second and deduce that the value associated with the first one is 2 times higher than that associated with the second one.

In our design we do not make use of pure sine waves to generate the tones. Instead we use sounds produced by a MIDI synthesiser. The main advantage of this approach is that the tones produced have intrinsically a more complex spectrum and therefore are more easily distinguishable. In addition, Brewster et al. (1993) found they are more pleasant and this is known to have important implications for long-term use. It is interesting to note that we are using the familiar technique used by Mansur to map line graphs to sound graphs to represent numerical values in tables. Thus, traditional boundaries between data representations in the visual medium (e.g. graphs and tables are commonly accepted to be in different data visualisation categories) are not maintained in the auditory medium.

1.3 Our Evaluation Strategy

We are interested in finding out whether there is any value in exploiting the fairly widespread ability to perceive pitch differences for improving the accessibility of tables containing numerical information to the visually impaired people. So far, there

is little evidence showing convincingly whether this strategy can work for realistic data comprehension tasks tackled by visually impaired or blind users on their own. In order for our evaluation to produce useful results we adopted the following policies:

- The participants of our experiments have to be representative of our target users. In that respect, we contacted institutions (Royal National Institute for the Blind in Peterborough and Royal Blind School in Hereford) engaged in the education of the visually impaired students to draw our pool of participants.

- Our design has to be tested under realistic data comprehension task conditions. In principle, the investigation cannot focus on only one intermediate step in the data comprehension process that we assume to be most relevant.

Our main reason for adopting the second policy is that narrowly focused evaluations often examine isolated aspects of an application. For example, one could devise experiments where participants are made to listen to a sequence of tones and asked to draw a sound graph. From a human computer interface evaluation perspective, the success of such an exercise does not necessarily extrapolate to a successful technique for tackling data comprehension in the auditory medium. For example, while providing users with the ability to browse data mapped to non-speech sounds, the interface of the application has to be made more complex in order to accommodate the facilities that enables the user to choose the browsing mode, e.g. speech or non-speech, she wants to be in. A more complex interface inevitably makes its learning curve steeper. In addition, the user will require more careful planning before accessing information because she now has to decide the relevant mode she wants to be in prior to the data access. Before the experiments, we had no idea how these various factors would play out and we did not have enough evidence to make assumptions about the impact of these unavoidable modal shifts on the usability of our application.

2 Apparatus and Stimuli

The Microsoft SAPI4.0 (Speech Application Program Interface) was used to implement the speech functionalities of the application. The application sends MIDI messages to a synthesiser to generate non-speech sounds. Participants have access to the auditory messages through headphones and navigate the table through key presses of a numeric keypad. Table 1 illustrates the typical table that is accessible in the auditory medium using our application. In this paper, we use the following conventions to simplify our descriptive and explanatory stories. $X_1 \ldots X_C \ldots X_M$ are the labels of the columns, X_1 is the first column, X_C refers to the current column and X_M refers to the last. $Y_1 \ldots Y_C \ldots Y_N$ are the labels of the rows, Y_1 is the first row, Y_C refers to the current row and Y_N refers to the last. The pair (X_C, Y_C) refers to the current cell of interest i.e. the cell in focus.

2.1 Sound Mapping Strategy

The mapping of numbers to musical note in the version of the application used in our experiments is straightforward. Every number (integer between and including 0 and 100) in a cell of a table is used as a MIDI pitch parameter. A more general

	Column referents					
	X_1	X_2	...	X_C	...	X_M
Y_1						
Y_2						
...						
Y_C				X_C, Y_C		
...						
Y_N						

(Row referents label on the left side)

Table 1: A Typical 2-dimensional table that can be accessed using our application.

Note	A	A#	B	C	C#	...	F	F#	G
MIDI pitch	0	1	2	3	4	...	125	126	127

Table 2: MIDI pitch parameter to musical note mapping.

mapping from integers to musical notes implemented in the latest version of our application is given by the following function:

$$y = \left(\frac{x - S_{min}}{S_{max} - S_{min}} \right) \times 127$$

where y is rounded to the nearest integer and is the value of the MIDI pitch parameter representing an integer x in an arbitrary set of integers ranging from S_{min} to S_{max}. We select the MIDI piano instrument for rendering the notes mainly because of its wide range of audible notes in our MIDI synthesizer.

The mapping also involves panning of the representative non-speech sound sources. The latter are localised along a line joining the left and right ears and positioned according to the position of their relevant cell in its row or column. When moving down a column, the first value is associated with a pitch heard in the left ear, the subsequent values are associated with pitches localised on a line between the left and right ear, and finally, the last value of the column is perceived in the right ear. The same effect is obtained while navigating across a row. The first value is heard to the left, the last value is heard to the right and the intermediate values are heard in between the left and right ears.

We use the coarse panning instructions to place the sound sources in the appropriate locations. Coarse panning in MIDI ranges from 0 to 127, with 0 representing left and 127 right. The following simple function is used to set the location of the sound sources; $P = \left(\frac{R}{N} \right) \times 127$ rounded to the nearest integer, where P is the panning number, R the rank (starting with 0) of the cell in the current column or row and N is the number of cells in the current column or row. Thus every cell is associated with two panning numbers, one based on the row and the other on the column on which the cell is found. During navigation along a row, row panning numbers are triggered and during navigation along a column, column panning numbers are triggered.

2.2 Interface (Re)Design Based on Output from Initial Pilot Studies

Before carrying out our experiment, 16 sighted pilot participants tried accessing the tabular information through the auditory medium alone in order to identify early on any obvious problems with the application and with the data comprehension tasks. Our choice of a 2×2 factorial experiment design meant that 4 treatment groups were required. Since we planned to allocate 4 subjects to each group, we therefore needed 16 participants in all. The pilot study was the dress rehearsal for the real experiment to be carried out with the visually impaired and blind users.

2.2.1 To Cross or Not To Cross Table Boundaries

Participants were initially allowed to cross table boundaries and move out of them. Once this happened the application triggered speech messages to navigate them back into the table. This strategy was intended to inform users of the limits of every row and column. Once participants were out of a table, pressing a key to reveal their current location caused the 'not in table' message to be read out. When users wandered away from the contents of the table, clicking on an arrow keys produced messages such as 'move left', 'move right', 'move up' and 'move down' which was intended to guide them back into the table. Trials showed that this strategy was confusing to the users. Thus, instead of allowing the user to venture out of a table, we restricted any navigation outside the table boundaries. In the version used for our experiments with the visually impaired, whenever users attempted to cross a table boundary an appropriate non-speech message was generated to highlight the fact that a boundary has been reached. Any further key presses for navigation in the inappropriate direction repeated the boundary sound without any change in their current position.

2.2.2 Optimising Speech Feedback

Most users found speech information that they really can do without annoying. Our initial design was judged to be over verbose because the application read out the name of the current row and column before presenting its contents in speech whenever a cell was brought into focus. Unnecessary information was distracting and overloaded unnecessarily the short-term memory of users. Our application therefore makes use of two different speech modes to optimise the speech output. In the 'labels' mode, whenever a cell of the table is brought into focus, its column name and row name (the cell referents) are read out only when the key 5 is pressed, the column name is only read out with vertical (Up and Down) cursor key presses and row name with horizontal (Left and Right) cursor key presses. In the 'value' mode, only the content of the cell in focus is read out. In the current design, the user does not need to wait for an auditory speech message to complete before initiating a key press. Ongoing auditory messages are simply flushed once a key press is detected. These simple strategies prove to be effective in providing the user with better control on the amount of speech feedback.

We also provide users with redundant information about the mode they are currently in by making voice characteristics (e.g. pitch and speed) mode specific. For

Key	Function	Output	Cell information revealed by key press
/	Enter labels mode	Speech	Either row, column or both read out
*	Enter value mode	Speech	Cell content read out in speech
-	Enter pitch mode	Pitch	Cell content mapped to pitch

Table 3: Modal keys to select 'value', 'pitch' or 'labels' mode.

example, in a given mode, the speech feedback sounds feminine; while in another, the feedback sounds masculine.

During our pilot studies, we found that the spatial cues we provided using the MIDI coarse panning parameter helped users identify approximately their current position in the current row or column. Pilot users also did not report any confusion with the mapping and thus cleared some concerns that the spatial encoding might not be readily understood during a task.

2.2.3 Key Descriptors in Speech to Facilitate the Learning of the Interface

In order to facilitate the learning of the interface, and the recall of the functionality of input keys in various modes, we develop a key descriptor facility. Our application includes a help mode triggered by holding down the Insert key. This mode allows users to press any key of interest and to be informed of the key functionality in speech. These messages also change to match the relevant mode in which the user is currently in as the functionality of the keys is mode dependent. This facility has proved to be very useful to participants during the training phase of the experiment.

3 Functionality of Numeric Pad Keys

Pressing the Delete key produces a high level description of the data, e.g. "table contains information about number of reported cases in London, columns represent type of crime and rows represent years". The keypad keys have been grouped into groups: modal keys (Table 3), navigation keys (Table 4), shortcut keys (Table 5) and overview keys (Table 6). Note that (X, Y) indicates that the contents of the cell in column X and row Y read out in speech, $[X, Y]$ indicates that the contents of the same cell is represented as a non-speech sound. The Key-functionality map we present here is not the first one we have come up with. The inclusion of three navigation modes in our design namely, labels, value and pitch is the result of a number of refinements to provide users with better control on the auditory output.

It is worth mentioning here that in any of the modes keeping the keys pressed (with the exception of key 5) will cause the cell in focus to change repeatedly. Thus in the pitch mode, keeping the cursor keys pressed will generate an overview similar to that obtained when the overview keys are pressed. The main difference is that when cursor keys are used to generate overviews, the user can control the direction and span of the overview. When the overview key is used the current position of the user is not changed in table so that the user can fall back to where he was before the onset of the overview.

Key	Function	Labels	Value	Pitch
↑	Move to cell above	Y_{C-1}	(X_C, Y_{C-1})	$[X_C, Y_{C-1}]$
←	Move to left cell	X_{C-1}	(X_{C-1}, Y_C)	$[X_{C-1}, Y_C]$
→	Move to right cell	X_{C+1}	(X_{C+1}, Y_C)	$[X_{C+1}, Y_C]$
↓	Move to cell below	Y_{C+1}	(X_C, Y_{C+1})	$[X_C, Y_{C+1}]$
5	Reveal info about current cell without moving elsewhere	X_C, Y_C	(X_C, Y_C)	$[X_C, Y_C]$

Table 4: Navigation keys allowing user to move from cell to cell.

Key	Function	Labels	Value	Pitch
Home	Jump to top left corner of table	X_1, Y_1	(X_1, Y_1)	$[X_1, Y_1]$
PgUp	Jump to beginning of current column	X_C, Y_1	(X_C, Y_1)	$[X_C, Y_1]$
End	Jump to beginning of current row	X_1, Y_C	(X_1, Y_C)	$[X_1, Y_C]$
PgDn	Jump to bottom right corner of table	X_M, Y_N	(X_M, Y_N)	$[X_M, Y_N]$

Table 5: Shortcut keys for moving quickly to strategic positions in table.

4 Comparing the Effect of Non-speech Sounds on the Accessibility of 2D Tabular Numeric Information

The purpose of this experiment was to evaluate how successfully visually impaired and blind participants tackled data comprehension tasks under two auditory conditions namely:

1. speech only referred in this paper as (Speech); and

2. speech and pitch referred to as (Pitch).

In the speech only condition, the 'Enter pitch mode' key was disabled. Thus, users had access to the contents of the table only in speech, i.e. they could navigate the table in the 'value' or 'labels' modes only. In the Pitch condition, the user was able to navigate the table in all three modes, namely the 'labels' mode, the 'value' mode and the 'pitch' mode.

4.1 Hypotheses

Access to numerical information in tables is improved when opportunities to access numerical information both in speech and pitch are provided:

H.1. The overall subjective workload is reduced under the pitch condition.

H.2. There is a decrease in the time participants take to tackle set data comprehension tasks.

H.3. There is a decrease in the number of errors under the pitch condition.

Key	Function	Labels	Value	Pitch
+	Overview of current column without moving elsewhere	$Y_1 \dots Y_N$	$(X_C, Y_1) \dots$ (X_C, Y_N)	$[X_C, Y_1] \dots$ $[X_C, Y_N]$
Enter	Overview of current row without moving elsewhere	$X_1 \dots X_M$	$(X_1, Y_C) \dots$ (X_M, Y_C)	$[X_1, Y_C] \dots$ $[X_M, Y_C]$

Table 6: Overview keys allowing user to trigger an overview output with a single key press.

	Demonstration and Training	Session I	Session II	NASA TLX Explanation	NASA TLX Administration	Debriefing
A.1		(Speech) Crime Statistics	(Pitch) Students Statistics			
A.2		(Speech) Student Statistics	(Pitch) Crime Statistics			
B.1		(Pitch) Crime Statistics	(Speech) Students Statistics			
B.2		(Pitch) Student Statistics	(Speech) Crime Statistics			
	$\longleftarrow$ .... 1 hour 30 .... $\longrightarrow$					

Table 7: Experiment schedule.

4.2 Experimental Design

16 participants were subjected to the speech and pitch conditions as described in Table 7. The order in which the tasks and auditory conditions were presented to the participants ensured that any effects due to inherent task difficulty levels and practice (i.e. increase in task tackling efficiency due to increase in familiarity) were minimised.

4.3 Experimental Procedure

The data analysis and inspection tasks that participants had to tackle were designed according to a number of requirements. Firstly, the data analysis should not be so complicated that the focus of the task was on the interpretation of the question rather than on how the tabular data is perceptualised or accessed. Secondly, the questions should not favour a single class of data comprehension task e.g. overview tasks over data discovery tasks.

The two sets of questions that participants face in this experiment relate to Crime and Student statistics. During the training phase, participants had our non-speech mapping strategy explained to them. Through examples, we made sure they had a reasonable understanding of what we meant by pitch, *this concept cannot be assumed to be known by participants*, how the values were mapped to pitch and the spatialisation of the non-speech sounds to give them a cue about their current position in the current row or column. Participants were then allowed to experiment with a data set different from those used in the experiment. This provided them with the

opportunity to familiarise themselves with the interface and to ask the experimenter for any clarifications they needed. Also any major hearing impairment would have become apparent during this exercise.

During the next two sessions, participants were administered the relevant tasks in the required order. The time taken for tackling each question was noted. They were free to ask clarifications about questions they felt were not clear. They were encouraged to make a best guess of the solution whenever they felt any question was too hard to tackle. Participants were free to leave during the experiments whenever they wanted.

In the next stage, the experimenter explained the various factors contributing to the workload variable in the NASA task load index (Hart & Wickens, 1990). After the meanings of the various variables were clarified, the pair wise comparison step of the NASA TLX was performed to determine the weight of each contributing factor for each participant. This step allowed us to compare workload values obtained with this experiment to other values obtained form other independent experiments (e.g. a similar experiment but administered to sighted users). The participants were then asked to rate each factor for every task under the relevant condition.

4.4   Construction of Queries

The data comprehension tasks constituted of queries constructed in the same spirit as exercises used to test the basic graph or table analytic skills of late primary or early secondary school students in the UK. We adopted this approach firstly because we wanted to avoid tasks, which were so difficult to understand that participants would find it hard to construct a meaningful goal before interacting with the application, and secondly to increase the size of our pool of participants who had a wide variety of education backgrounds. Our aim was not to make an exhaustive evaluation of data comprehension tasks. They were used mainly for motivating meaningful interactions with our application. We reproduce here the two sets of questions in our experiments regarding two tables of 260 cells containing fictitious numerical data.

4.4.1   Student Statistics: Student Performance Analysis

This dataset is about the performance of a number of students in various subjects. Answer the following questions based on the information available in the table.

- Name the student or students scoring the highest marks for Biology?

- The list of student is presented in ascending order of marks for a given subject. Name the subject?

- Name the subject most likely to have the highest number of passes, note that the pass mark is greater than 40?

- Name the student or students scoring the lowest marks for the Assembly course?

- In which course is performance particularly poor, that is, most marks are less than 40?

- Name the student most likely to have the highest total marks?

4.4.2 Crime Statistics: London Crime Statistics

This table is about the type of crime and the number of cases reported in London from 1974 to 2000. Answer the following questions based on the information available in the table.

- State the year or years in which the highest number of murder cases was or were reported?

- State the year or years in which the highest number of robbery cases was or were reported?

- Which type or types of crime had a consistently high number, consider this number to be greater than 50, of cases reported?

- Which type or types of crime shows or show a consistent increasing trend?

- State the year or years in which the numbers of hate crime cases was or were lowest.

- Which type or types of crime shows or show a consistent decreasing trend?

5 Results

We now compare the various scores obtained for the individual workload categories as defined by the NASA Task Load Index.

5.1 Participants

The age of the participants ranged from 23 to 57. There were 7 women and 9 men in the experiment. The subjects had diverse visual impairment conditions (e.g. aniridia, optic atrophy, congenital rubella, colaboma etc. . . .) but none used any residual vision for tasks relevant to our experiment. Most participants were comfortable with the notion of tables and what they are used for. The majority did not have special musical skills with the exception of three subjects who were musicians (one claimed to have perfect pitch). The notion of pitch however was new for some subjects and had to be explained during the training session with examples. The majority of our subjects were familiar with screen reader applications; some also had access to Braille displays. Few were expert users spreadsheet applications and some 'played around' with spreadsheet without really using them. Two of the participants were programmers and appeared to be very comfortable with the numeric keypad.

5.2 Comparing Workload Categories

Figure 1 illustrates the average values for the various workload categories for the NASA TLX for visually impaired users. For completion, we have also included the results obtained for an earlier similar experiment with sight users who could access information in the tables only in the auditory medium.

Table 8 shows the results of a t-test (paired two sample for means) on the workload categories. Our results show a highly significant decrease in the mental ($T_{15} = 4.46$, $p < 0.01$), temporal ($T_{15} = 4.00$, $p < 0.01$), performance ($T_{15} = 3.70$,
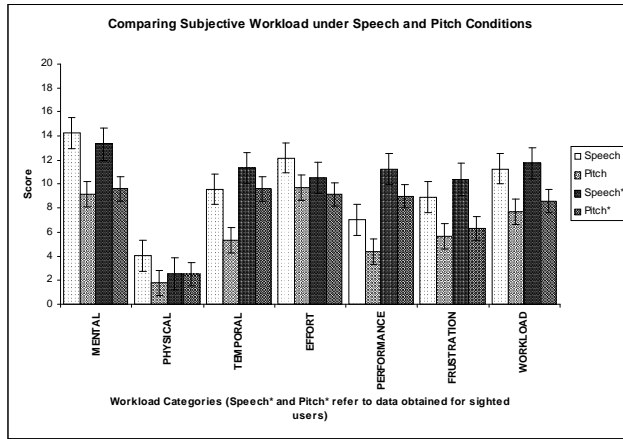
Figure 1: Results showing the impact of pitch on workload categories (standard error bars are shown).

	Mental	Physical	Temporal	Effort	Perform--ance	Frustr--ation	Workload
p	0.0002	0.0237	0.0006	0.0238	0.0011	0.0087	0.0003
T_{15}	4.4636	2.1594	4.0026	2.1568	3.6958	2.6709	4.2836
(Sighted) *p*	0.0044	0.5000	0.0725	0.0574	0.0939	0.0032	0.0082
(Sighted) T_{14}	3.0389	0.0000	1.5432	1.6813	1.3844	3.1916	3.5475

Table 8: Summary of t-test results (the cells containing $p < 0.01$ are highlighted.

$p < 0.01$), frustration ($T_{15} = 2.67$, $p < 0.01$), and overall subjective workload ($T_{15} = 4.28$, $p < 0.01$) in the pitch condition. Hypothesis H.1 is thus confirmed.

The reader should note that the conditions for these two groups of participants were not identical. In the experiment with sighted users, data was obtained from only 15 participants. Also, an earlier version of the application was used: there was no key descriptor facility and participants had access only to one speech and one pitch mode. While direct comparisons cannot strictly be made, the results does provide us insights about how visually impaired users compare to their sighted counterparts and also perhaps about the effect the interface design has on the workload experienced by users. For example, the bar chart shows that sighted users experienced more temporal and frustration workload as compared to their visually impaired counterparts. This may be due to the fact the version of the application the sighted participants used had only one speech mode which made it over verbose. This might have slowed down their interaction.

5.3   Analysis of the Average Time Taken per Question

Visually impaired users on the average took 133 seconds to tackle a question in speech (standard deviation = 55.033) and on the average took only 44 seconds to tackle a question in pitch (standard deviation = 14), this difference was statistically significant ($T_{15} = 6.2$, $p < 0.01$). In the case of the experiment with sighted users, no temporal information was captured. On the average participants got 4.6 correct in the speech condition and got 5.3 in the pitch. This difference was statistically significant ($T_{15} = 2.2$, $p = 0.0219 < 0.05$). Our quantitative evaluation strongly supports hypotheses H.2 and H.3.

5.4   Participant Feedback and General Discussions

The majority of participants found that commands and modes of interaction were easy to understand and remember. The ability to learn the interface quickly and get on with the task was a definite plus.

We have collected here the most positive and the most negative of user's comments.

Positive comments:

 'thought it was another gimmick, once I started using it to tackle tasks,
 it is quite useful ... '
 'the pitch facility is absolutely fantastic ... '
 '... pitch was easier on the memory ... '
 'was not convinced with the idea but it works ... '
 'it is so clear it is easier in pitch'
 'in pitch, I see a picture ... '

Negative comments:

 'not really useful for low values ... '
 'difficult to distinguish between low notes'
 '... pitch makes it slightly easier? can't say ... '

Our quantitative results reveal that when low values were involved in a data comprehension task, the benefits of the pitch mode decreased. This observation ties well with the feedback we got from some participants who stated that low pitches were difficult to distinguish. In our data comprehension tasks, the lowest frequency presented to participants by our application was 8.18Hz (Midi pitch parameter 0). Earlier guidelines for the design of earcons suggest that the pitch used should be no lower than 125–150Hz (the octave of C4) so that the sounds are not easily masked and are within the hearing range of most listeners (Brewster et al., 1995). Our results confirm that pitch discriminations at lower frequencies is problematic and we plan to use an improved number to pitch mapping in forth coming experiments to tackle this issue.

The majority of participants were pleased with the panning effects of the pitches during table navigation and none found the mapping confusing. One participant was oblivious to the panning effects and failed to detect spatialisation; we did

not test spatial hearing ability during this experiment. However, she was still able to complete the data comprehension tasks using pitch information alone. In our opinion, this mapping is an effective way of introducing spatial cues in the representation without having recourse to expensive sound rendering hardware to produce realistic out of the head spatialisations.

Our data reveals that, in rare cases, there is little difference on average between the time taken to tackle a question under the speech condition and the pitch condition. This may be due to the fact that introducing a number of modes in for browsing the table may impose extra strategy building load on participants attempting a data comprehension task. Most participants found choosing the right mode at the right time during the task solving process straightforward. The evidence is that there were a very small number of participants who found meaningful mode transitions difficult and were often engaged in inefficient transitions.

6　Conclusions

In summary, we investigated whether there are grounds to introduce pitch to make the experience of data browsing more effective. Arguably, the obvious way to represent a table in sound is to use speech feedback to inform the user where she is in the table and what is available there. We have demonstrated through rigorous evaluation of our application that non-speech sounds if introduced with care in a data visualisation application can have a significant impact on making it accessible to the visually impaired user. Many participants were eager to find our approach integrated in commercial products, more so, because it is simple and should not be difficult to implement in spreadsheet or screen reader applications. We hope that our work will provide some useful pointers to commercial software developers engaged in making data visualisations accessible to the visually impaired.

7　Acknowledgements

References

Alty, J. & Rigas, D. (1998), Communicating Graphical Information to Blind Users Using Music: The Role of Context, *in* C.-M. Karat, A. Lund, J. Coutaz & J. Karat (eds.), *Proceedings of CHI'98: Human Factors in Computing Systems*, ACM Press, pp.574–81.

Brewster, S. A., Wright, P. C. & Edwards, A. D. N. (1993), An Evaluation of Earcons for Use in Auditory Human–Computer Interfaces, *in* S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel & T. White (eds.), *Proceedings of INTERCHI'93*, ACM Press/IOS Press, pp.222–7.

Brewster, S. A., Wright, P. C. & Edwards, A. D. N. (1995), Experimentally Derived Guidelines for the Creation of Earcons, *in* G. Allen, J. Wilkinson & P. Wright (eds.), *Adjunct Proceedings of HCI'95*, BCS, pp.155–9.

Gardner, J. A., Lundquist, R. & Sahyun, S. (1998), TRIANGLE: A Tri-modal Access Program for Reading, Writing and Doing Math, *in Proceedings of CSUN International Conference on Technology and Persons with Disabilities*, Center on Disabilities, California State University, Northridge, California, USA. http://www.dinf.org/csun_98/csun98_104.htm.

Gomez, C., Shebilske, W. & Regian, J. (1994), The Effects of Training on Cognitive Capacity Demands for Synthetic Speech, *in Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, Human Factors and Ergonomics Society, pp.1229–39.

Hart, S. & Wickens, C. (1990), Workload Assessment and Prediction, *in* H. R. Booher (ed.), *MANPRINT, An Approach to Systems Integration*, Van Nostrand Reinhold, pp.257–296.

Kurze, M. (1998), TGuide: A Guidance System for Tactile Image Exploration, *in Third Annual ACM Conference on Assistive Technologies*, ACM Press, pp.85–91.

Mansur, D. (1975), Graphs in Sound: A Numerical Data Analysis Method for the Blind, Master's thesis, University of California.

Moore, B. (1982), *An Introduction to the Psychology of Hearing*, Academic Press.

Pollack, I. & Ficks, L. (1954), "Information of Elementary Multi-dimensional Auditory Displays", *Journal of the Acoustical Society of America* **26**, 155–8.

Roffler, S. & Butler, R. (1968), "Localisation of Tonal Stimuli in the Vertical Plane", *Journal of the Acoustical Society of America* **43**, 1260–6.

Shneiderman, B. (ed.) (1994), *Sparks of Innovation in Human–Computer Interaction*, Ablex.

Stevens, S., Volkmann, J. & Newman, E. (1937), "A Scale for the Measurement of the Psychological Magnitude of Pitch", *Journal of the Acoustical Society of America* **8**, 185–90.

Author Index

Keyword Index