Integrating Data Sonification With Synthetic Speech For Unified Data Access

Dr Tony Stockman

***Abstract*—Data sonification holds considerable promise as a means by which visually impaired computer users can obtain improved access to data. The ability to overview large, possibly multi-dimensional data spaces rapidly, plus the possibility of using sonification to pinpoint segments of data for detailed examination, should lead to considerably reduced times for data exploration tasks. For reasons of cost and speed of access, the most popular medium for screen reader output is speech. This means that careful design is required when integrating data sonification into a screen reader environment, to avoid overloading of auditory channels and allow flexible control of both the sonified and speech outputs. This paper will examine a number of issues in the effective integration of data sonification with a speech-based screen reader. We shall explore the nature of the controls required in order for sonification to complement and enhance data exploration with speech, and discuss the requirements for an effective and accessible interactive sonification tool.**

***Index Terms*—Data sonification, synthetic speech, screen reader, multi-modal interaction.**

I. Introduction

SCREEN readers [1] are the primary mechanism used by visually impaired computer users to access computer systems. Although screen readers exist for other systems such as the Macintosh and Unix, by far the most widely used screen readers, Jaws for Windows (JFW) [2] and Window-Eyes (WE) [3] run under the Microsoft Windows operating system. Both JFW and WE support output in both Braille and speech, but for a number of reasons, most users of screen readers either use speech output mode alone, or use Braille only in combination with speech. The reasons for the general preference for speech output can be briefly summarised as follows:

1) The additional hardware required for Braille output is relatively expensive, and is an additional component to be carried in mobile use

2) A significant proportion of visually impaired computer users are not fluent in reading Braille

Tony Stockman is with Queen Mary, University of London, Mile End Road London E1 4NS, United Kingdom (phone: +44 (0)20-7882-5202; fax: +44 (0)20-8980-6533; e-mail: tonys@dcs.qmul.ac.uk).

In contrast, synthetic speech represents a relatively inexpensive option, which can often be provided via a built-in sound card, and which can deliver output at easily variable speeds, which are only limited by the users perceptive ability [4].

Research into the use of non-speech sound to convey information, known as sonification, has grown rapidly since the early 1990s [5, 6]. There has however, so far, been little take up of this research in screen reader technology. The release of JFW version 5, in September 2003, represented the first significant use of non-speech sound in a commercial screen reader. In this paper we shall examine the potential for integrating sonification techniques in screen reader systems, describe specific experiences of developing interactive sonification support for spreadsheet access, and discuss issues involved in integrating sonification with speech-based access technology.

II. Screen Readers and Non-speech Sound

The recent release of JFW version 5 [7] contains limited use of non-speech sound to customise the feedback provided to users about the current state of the interface. These customisations, called “behaviors” in JFW, are defined in speech manager files, known as schemes. Examples of the ways in which non-speech sounds may be user defined include: to identify when focus has moved to a particular type of interface widget, to signal the different states of a checkbox, to indicate upper/lower case or degree of indentation, or identify the values of HTML attributes. The inclusion of this limited support for the use of non-speech sound is intended to reduce the time and improve the accuracy of user tasks such as navigating between objects in the Windows GUI, and the screen-based proof reading of lengthy documents.

Data sonification, i.e. the use of non-speech sound to render data characteristics of interest, has considerable potential to enable visually impaired people to have improved access to data in both work and leisure situations [8, 9]. Data sonification is potentially useful to visually impaired people in at leased the 2 following data analysis tasks:

1) In providing a rapid overview of a data set. Most people can listen to the sonification of a set of data very much faster than they can browse it using speech or Braille.

2) Identification and navigation to points or areas of interest. It is always likely that speech or Braille will be required at some point during the examination of data, to obtain specific values. However, a well designed sonification system can reduce the amount of data to be navigated in speech/Braille, both by helping identify the area of interest in the first instance, and possibly additionally by providing a second, more localised sonification of the immediate area of interest.

The recent inclusion of non-speech sound in JFW does not directly support these overview and data navigation tasks, although the potential to provide the support existed in releases of the system prior to JFW 5 through the medium of JFW's scripting language. The following section describes how support for the above tasks has been prototyped within the Microsoft Excel spreadsheet environment. Issues arising from the development of this prototype, and more generally relating to the integration of sonification with speech technology, are discussed in subsequent sections of the paper.

III. Prototype Development

The objective of the project was to explore the issues in developing support for the overview and data navigation tasks described in section II above, and to test the value of sonification in supporting these tasks within the context of use of a speech-based screen reader, in realistic, work-based situations. Informal discussions with potential users determined that by far the most common source of data they encountered requiring support for these tasks are Excel spreadsheet data. The prototype to be developed therefore needed to produce an interactive sonification enabling rapid overview of spreadsheet cell values, and rapid navigation and exploration of cells of both global and local interest.

A. Software Environment

The system was built and tested on a Toshiba laptop running Windows XP. Interactivity was initially encoded in Visual Basic for Applications macros [10], whilst Jaws version 4.51 was used as the screen reader. Csound version 4.23 [11] was used as the sonification engine. Csound uses two files in order to synthesise sound: an orchestra file, which defines the instruments to be used in the sonification, and a score file, defining the notes to be played by each instrument (note that in recent versions of Csound, these two files have been merged into one, integrated file format).

Within the prototype, A pre-written orchestra file is used to specify the instruments to be "played" during sonifications. An enhanced version of the system could include the functionality to change the orchestra file dynamically during spreadsheet interaction. Each line of the score file corresponds to a note to be sonified. In simple sonifications, the parameters on each line of the score file can be specified as follows:

-- Instrument to be used.

-- Start time of the note to be played in seconds (0 corresponds to immediately).

-- The duration of the note to be played in seconds.

-- The amplitude of the note.

-- The frequency in Hz.

B. Operation of the Prototype

The prototype operates by using VBA code fragments invoked by the user, which prepare data to be sonified by csound. An initialisation macro performs a pass through the data in the worksheet, identifying global maximum and minimum numeric values, which are used to scale the raw data to be within an appropriate frequency range. When a user requires to sonify a range of cells, the required macro is invoked using the corresponding "hot-key" combination. The called macro steps through the set of cells to be sonified, for each cell writing a line to the score file, the frequency value of which corresponds to the scaled value of the cell. The prototype provides the following functions:

1) Sonify a selected range of cells with end of row/column indicators, with optional move to global max or min

2) Sonify: current column, current row, rest of column, rest of row, previous part of row or column, all with optional move to max or min

IV. Evaluation

The evaluation of the system, which is still in progress, has so far included 4 visually impaired users, using 3 different spreadsheet applications. User U1 has evaluated the system using application A1, which is a spreadsheet of children and families social work data, drawn from this users's real work context. Similarly user U2 has used spreadsheet application A2, a university admissions spreadsheet, again drawn from that users work domain. Unfortunately, the fact that applications A1 and A2 contain "live" data, makes them inappropriate for evaluation by other users. Therefore, a further application, A3, has been employed for general evaluation, and this has been tested by users U3 and U4. Application A3 comprises a spreadsheet of module results for a group of 134 2nd year undergraduate students, from which all identifying attributes such as student names and numbers have been removed. All the users involved so far possess a similar level of knowledge of Excel, in that they have used Excel occasionally in a basic way, to review spreadsheet data and navigate to specific cells.

A. Evaluation Method

Evaluations were conducted using the development environment described in section III. All users involved so far

have a good level of familiarity with the Jaws screen reader. An initial preparation session of about 30 minutes was used to do the following:

1) Ascertain each user's previous experience of spreadsheet applications in general, and Excel in particular.

2) Explain the built-in keystrokes used in Excel for selecting multiple rows and columns

3) Explain and demonstrate the basic idea of sonification, and show users how to invoke the macros to perform the available sonification functions described in section III.

Users were then given an amount of time determined by themselves to explore the spreadsheet used for their evaluation, employing a combination of sonification and spreadsheet information spoken by Jaws. During this time they were free to ask questions about operation of the system and make comments about its use. The times for these exploratory sessions varied from 15 to 25 minutes. The evaluation session concluded with an informal discussion of each users experience of using the system.

B. Evaluation Results

Background: all users possessed a similar level of prior experience with spreadsheet applications, in that Excel was the only spreadsheet application they had used, all to a similar level of being able to review data in cells and columns, and navigate to specific cells. 3 of the 4 expressed a lack of confidence in using the application efficiently to extract the type of information they would normally require.

The main points arising from the free form spreadsheet exploration sessions can be summarised as follows:

1) All users seemed to develop an understanding of the idea of using sonification to overview data, and appreciate how it could compliment spreadsheet exploration with a screen reader.

2) The up-up polarity mapping of data values to frequency seemed to be a natural one for the variables in the 3 applications employed here. In the U2A2 evaluation however, the user was frequently interested in the difference between two variables, for example between the numbers of applicants to specific university courses and the numbers actually enrolled. For this user it would have been extremely useful to be able to sonify the difference between selected variables, either by interleaving sonifications of the two variables of interest, or by directly sonifying the difference between them.

3) In general, the more values involved, the more substantial was the advantage obtained from using sonification. For example, in the evaluations using application A3, sonifications of all 134 students results, or of the results of specific examination questions saved substantial time in screen reader navigation. A precise quantification of this gain has not been made, and will of course vary depending on the pace of interaction between any given user and screen reader. From observations during these experiments however, experienced users would typically review cell data values at a rate of approximately 3 per second, whereas sonifications were delivered at the rate of 5 values per second.

4) Users commented that the sonifications helped to provide an overview perspective of the data in a way that screen-reader-based review does not, i.e. it seems easier, at leased for a short time, to remember the "shape" of the data as conveyed by the sonification, than it is to remember the "shape" of the data as conveyed by the list of numbers spoken by a screen reader. The ability to comprehend the sonified overview was illustrated by the following examples:

-- User U1 being able subjectively to compare the numbers of referrals in different months of the year in application A1.

-- User U2 being able easily to perceive the domination of male over female recruits to university courses in Computer Science in application A2.

-- Users U3 and U4 being able to compare the relative difficulty students found between different examination questions in application A3.

5) Two users remarked it would be useful to be able to navigate the spreadsheet from cell to cell, sonifying the values as they went.

6) Although users employed both, the macros involving no movement of the active cursor to global or local maximum or minimum values, were used more often than the macros that included movement.

7) The use of global maximum and minimum values to scale the sonified frequency values was important in facilitating comparisons across sonifications of rows and columns. It was felt though that more could be done to help anchor the understanding of the values being sonified, either through speech and/or additional non-speech sounds to annotate sonifications.

8) It was felt by user U1 that the sonification of a moving average to facilitate monthly comparisons between data may be helpful in providing a smoother comparison.

9) Some confusion was seen, particularly in the early stages of the exploratory sessions, between the keystrokes required to operate Excel, Jaws and the VBA macros.

V. Issues in Integrating Sonification with Speech-based Screen Reader Software

A. Use of the Audio Channel

The most obvious single problem in combining sonification and speech-based screen reader technology is that both use the same sensory channel for communication. It is important to have the ability to synchronise sonifications with speech from the screen reader. It is likely that Most of the time, most users will not want to put up with the cognitive load of simultaneously processing both speech and non speech sound, however, whether simultaneity is sometimes permissible or not, the key issue is that overlap must be both controllable and avoidable.

B. Use of Hot-Keys

The keyboard actions used to initiate events must have a consistent form across both speech and non-speech events.

C. It Should be Easy to Synchronise Speech Data Output with Sonification

Have specific values from a sonification or ranges of sonified values rendered in speech. This will help to anchor the sonification to the absolute values underlying the sonified data.

D. Scrolling

It should be possible to scroll backwards and forwards through the sonification, and where required drop into speech mode to examine specific values.

E. Out of this Last Point Arises the Notion of a Sonification Cursor

That is notion of a sonification cursor and speech cursor for pointing. The default is likely to be that they are synchronised and move together, however during data analysis, it may sometimes be necessary to use them separately, in which case it becomes desirable to route them back together again when required. A simple example of this arises when using a speech cursor to move between specific monthly values, e.g. “July”, “August” etc. and then use the sonification to reveal an overview of the details of monthly data for the month currently indicated.

F. Sonification During Navigation

The evaluation identified a requirement to sonify individual spreadsheet cells during cell to cell navigation. The implementation of this feature raises at leased two issues for integration with speech. Firstly, initiating the inter-cell navigation will normally cause a speech event to speak the value of the target cell, which users may prefer to be switched off, or at leased control. Secondly, users are likely to wish to perform the inter-cell navigation rapidly, and so any latency between the keyboard event initiating the navigation and the resulting sonification must be kept to a minimum. For both these reasons, it becomes increasingly important that the software used to invoke sonifications is more tightly integrated with the speech component.

VI. Summary and Future Work

The basic prototype and early evaluations demonstrate that there is value in using sonification in real-world applications to convey overviews of data. The advantage that this brings increases with the amount of data to be reviewed. Further functionality is required to be able to sonify user-selected compound values of data, such as the sums or differences between rows and/or columns. Considerably improved sonifications may be produced by providing more sophisticated orchestra files for the csound-based sonifications, and where possible putting the control of these in the hands of the user. To improve the responsiveness of the system overall, and support some required functionality such as sonification of inter-cell navigationh, it is desirable to integrate the sonification software as tightly as possible with the speech-based screen reader software, for example through the use of Jaws scripts and sound manager.

References

[1] http://www.rnib.org.uk/technology
[2] http://www.freedom-scientific.com
[3] http://www.gwmicro.com
[4] C. Asakawa, H. Takagi, S. Ino, and T. Ifukube, “Maximum listening speeds for the blind (Published Proceedings style),” in *Proc. of the 9th International Conference on Auditory Display*, Boston, Ma, 2003.
[5] G. Kramer, “Auditory Display, Sonification, Audification and Auditory Interfaces (Book style),” Addison Wesley 1994.
[6] K. Gregory. (Year, month, day). The Sonification Report (edition) [Online] Available: http://www.icad.org/websiteV2.0/References/nsf.html.
[7] http://www.freedom-scientific.com/fs_products/software_jaws5intro.asp.
[8] L.M. Brown, S.A. Brewster, R. Ramloll, M. Burton, and B. Riedel, “Design guidelines for audio presentation of graphs and tables (Published Proceedings style),” in *Proc of the 9th International Conference on Auditory Display*, Boston, Ma, 2003.
[9] T. Stockman, “The Sonification of SQL constructs in multi-modal database interaction (Published Proceedings style),” in *Proc of the 9th International Conference on Auditory Display*, Boston, Ma, 2003.
[10] http://www.msdn.com.
[11] http://www.csounds.com.

Dr Tony Stockman ( ) was born in Blackburn, UK in 1960. He obtained a First Class Honours degree in Computing and Data Processing from Bradford University, where he remained while studying for his PhD in the Signal Analysis of Physiological Control.

He is a member of the Interaction, Media and Communication research group, within the Department of Computer Science at Queen Mary College, University of London. His main research interests are in data sonification and auditory display design. He is a board member of the International Community for Auditory Display (ICAD).