Interactive Generation of Overview Information Using Speech

Johan Kildal and Stephen A. Brewster
Glasgow Interactive Systems Group
Department of Computing Science
University of Glasgow
Glasgow, G12 8QQ, UK
{johank, stephen}@dcs.gla.ac.uk
www.multivis.org

CHI 2007, April 28 – May 3, 2007, San Jose, USA
ACM 1-xxxxxxxxxxxxxxxxxx.

Abstract
In non-visual interfaces, using non-speech audio can be a more effective and efficient way of obtaining overview information than using speech. However, users who are blind regularly use speech-based tools to access information in computers, and often prefer this technology over others that pose steeper learning curves. This paper proposes a technique to explore numerical data tables interactively in order to extract overview information by optimising the use of speech.

Keywords
Speech, tablet, data tables, overview, accessibility

ACM Classification Keywords
H.5.2. [Information Interfaces and Presentation]: User Interfaces---Voice I/O.

Introduction
When encountering a new data set or collection of information that is going to be analysed, obtaining overview information is the first task that needs to be completed [4]. In most cases, information is presented in visual forms (printed on paper or laid out on a visual display, sometimes in the form of specifically constructed information visualisations), and in most of those cases people accomplish this task using vision, a

sense that provides a very synoptic view of the information. There are many cases, however, in which users cannot analyse information visually, either because the visual channel is already occupied with other tasks or if users have visual impairments. The latter is the target group of users for the work presented here.

Blind and visually impaired (VI) computer users often utilise screen reading software that converts information (commonly text) into speech. While speech is a very natural way of obtaining information, it poses serious limitations for browsing information quickly to obtain an initial overview. In particular, information is accessed sequentially and in full detail, providing poor contextual information, which rapidly saturates working memory before an overview of the complete set of information is constructed. Other electronic information accessibility technologies, like refreshable Braille displays, present similar limitations to those discussed for screen readers in this paper.

When the information being analysed is numerical, the use of non-speech audio to represent the data has been shown to be a successful approach [5]. Kildal and Brewster developed TableVis, an interface implementing techniques that permit users to obtain overview information from large numerical data tables quickly and easily [3]. They also showed that using non-speech sounds to obtain overview information from numerical data tables is more effective and efficient than using speech generated with a typical screen reader [2]. The same research work also showed that VI users who regularly use screen readers tend to favour speech-based techniques over more novel ones, mainly due to familiarity. This is a classic example of the performance vs. preference paradox [1]. Taking this qualitative information into account, while ultimately the solutions offering better performance and good usability are expected to be favoured by more expert users, it was considered worth exploring further the possibilities that speech can offer to obtain overview information from numerical data sets. This paper reports the approach taken and some initial results from a pilot evaluation with blind users.

Design principles

The interface setup and interaction technique were adapted from the “cells” exploratory mode of TableVis [3]. Contextual information, which is normally gathered visually and that is necessary to construct a coherent image of the complete data set (a data table in this case), is obtained using a graphics tablet augmented with a tangible frame delimiting the active area on which data are presented. Any data table that is going to be explored is scaled to fill the complete working area of the tablet (see figure 1). Thus, combining the fixed reference provided by the physical frame (indicating the boundaries of the data set) with the proprioceptive information provided by the hand that explores the table with the tablet’s electronic pen, users can maintain the context of the information retrieved. With this layout, pointing at a certain location on the graphics tablet is equivalent to pointing at a cell on the table that has been mapped on the surface of the tablet, and in TableVis this produces the sonification of that cell, representing with non-speech sound the numerical value stored in that cell. The equivalent action in the speech version of TableVis would be to have a synthetic voice read the numerical value stored in the cell that the pen is pointing at. In this way the limitation of having to access information sequentially is overcome, and

the focus+context paradigm is implemented for non-visual speech-based information browsing.

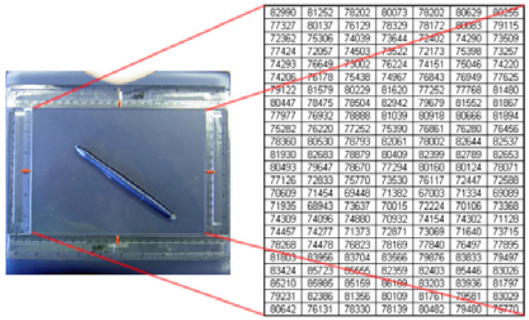

figure 1. The data table to be explored is presented on the active area of the tablet, scaled to fill it completely, so that spatial layout of data becomes an invariant [3].

As mentioned above, reading numerical values in speech conveys more detail than is necessary at the stage of browsing for overview information. All this information soon saturates the user's working memory, making it impossible to perform enough comparisons to identify trends or patterns on the table. Additionally, it takes a long time to read numbers, particularly those representing large values or with many decimal places, aggravating the problem of decay of the information in working memory.

To compensate for these limitations of speech, the concept of most significant digit (MSD) was introduced. The MSD in a number is the leftmost, non-zero digit in that number. It is the digit with the greatest value in the number. For example, in the number 4,326,436.394, the most significant digit is the 4 at the left end. If a user is exploring a data set with numbers that are as long as in the example above, a first approach to gaining an overview could be by restricting speech presentation to only the information contained in a few of the most significant digits. The number of significant digits that are required to obtain overview information at a particular level of detail, however, will vary depending on the data in the table.

3,425,263	4,579,389	5,320,147
6,568.453	7,679,128	9,474,596
9,141,853	10,002,730	11,256,026

table 1. Example of a data table in which the information in the MSD is enough to identify the trend that the values in the table follow.

3,125,263	3,279,389	3,320,147
3,468.453	3,579,128	3,674,596
3,741,853	3,802,730	3,956,026

table 2. Example of a data table in which the first two MSD's are required to extract information about the trends that the values in the table follow.

In the example shown in table 1, simplifying the numbers in the table to only the MSD transmits enough information to identify that values grow both going to the right and down. In the example in table 2, the MSD only would show a table with the same value – three million – in all the cells. This would already be valid information as a first stage of overview, informing that

all the values in the table are in the range from 3,000,000 to 4,000,000. Presenting also the second MSD, however, would reveal that the values in this example follow a similar trend as the previous example. This simple principle was implemented in two different ways that are described in the following sections.

First implementation

In the first implementation, an additional control (a programmable rotary knob) is provided to select the number of significant digits that the application is going to present in speech. By default, only the MSD is spoken when pointing at a cell. The user can increase or decrease the number of MSD's spoken, in increments of one. All the digit descriptors (million, thousand and hundred) are spoken for all the MSD's considered. For instance, taking the value 3,425,562, the speech generated for different numbers of MSD's considered would be as shown in table 3.

MSD's	Speech
1	Three million
2	Three million, four hundred thousand
3	Three million, four hundred and twenty thousand
4	Three million, four hundred and twenty five thousand
5	Three million, four hundred and twenty five thousand, five hundred
6	Three million, four hundred and twenty five thousand, five hundred and sixty
7	Three million, four hundred and twenty five thousand, five hundred and sixty two

table 3. In the first implementation, speech generated for the number 3,425,562 considering different numbers of MSD's.

Moving to a new cell before the speech corresponding to the previously selected cell had been fully spoken truncates the speech, produces a brief "tick" sound to signify that the boundary between two cells has been crossed (and to add an audible division between two different speech messages) and starts the production of speech corresponding to the newly selected cell. Each user can select his/her preferred rate of speech.

Second implementation

In the second implementation, there is no external control to select the number of MSD's that are used to generate the speech. This setting is fixed to speak the numbers in full, with all their digits. Like in the first implementation, moving to a new cell truncates the speech from the previous cell and initiates the speech corresponding to the newly selected cell, with an audible "tick" in between. As before, the user can set the preferred speech rate.

In this implementation, it is the speed at which the pen is moved around the table that defines how much speech is generated. In other words, the user stays on each cell only as long as it takes for the speech to communicate enough information for the level of detail required to construct a particular overview. If the pen is moved around very fast, only the "tick" between the cells is heard, and there is no time for any speech to be spoken. With English as the language for the speech synthesiser, as the user slows down the speed of movement of the pen, speech corresponding to the MSD is heard. Slowing the speed further, the descriptor of that digit will appear (if present), followed by the speech corresponding to the second MSD, and so on.

Pilot evaluation

As proof of concept, both implementations were evaluated qualitatively in a pilot study with three visually-impaired users. They were all experienced computer users who had at least completed secondary education and who were expert regular users of screen reading applications to access information in their computers.

Four data tables were used to present both implementations to each participant, though practical examples. Then, tasks were set requiring users to extract overview information from up to 8 numerical data tables. Each tables had always 7 columns and 24 rows, and the values contained in them ranged between one million and ten million (the top value of the range was never present, all values had 7 digits). The overview tasks required the participants to describe any trends present in the data or to find an area containing cells with the highest or lowest values in the table. While no quantitative information was formally collected, the participants could complete the tasks accurately, in exploration times that were comparable to those needed to explore similar tables with non-speech sound in TableVis [3], and hence faster than with conventional speech-based techniques [2]. Participants accessed only a subset of the cells before inferring the answer, implicitly assuming interpolated values in the cells that were not accessed. This assumption is valid for smoothly changing data, which was the case in the examples presented in the evaluation, but would not necessarily have to be with more general data sets. In this sense, using non-speech sounds in TableVis to scan all the cells in a table, constitutes a more general technique.

After trying the tasks with each implementation, a semi-structured interview was conducted. All three participants agreed in saying that those tasks would have been more difficult to complete with a conventional screen reader, mainly due to the need to travel sequentially across the table, without being able to jump and point at different areas of the table that could be found easily inside the tangible frame. They were all also coincident at saying that the second implementation was better than the first one. One of the participants found the functionality to define the number of MSD's in the first implementation to be quite confusing, and soon realised that setting this control to maximum accuracy the exploration could be used exactly as in the second implementation, controlling the amount of information received by modifying the speed of the hand. One participant mentioned that with the second implementation it was very quick to get an impression of the size of the table by moving the hand fast and hearing the ticks produced between cells, and then to get an idea about how the values were distributed in the table by simply waiting for the beginning of the speech to be spoken.

Additional considerations

There are some aspects about the technique described in this paper that need to be developed before it can be used with more general data sets. If values in different cells do not contain the same number of digits, there is not an equivalent MSD for all of them. Strictly speaking, the MSD would be the leftmost digit of the value in the table with the largest number of digits to the left of the decimal point. In that case, the speech for the values with less digits would be "zero", possibly followed by the descriptor of the MSD, if there was one (for example "zero million") to avoid conveying the impression that the value was an absolute zero). Alternatively, a distinctive sound could be used where there was no digit corresponding to the current setting of MSD's.

Another aspect affecting to both implementations is the sign of the numbers stored in the cells. One solution would be that if there are both positive and negative values in the table, the word “plus” or “minus” was said at the beginning of the speech. Simply waiting for this information to be spoken would already provide a first level of very gross overview information that could be useful in many cases. Additionally, it is a convenient way of speaking values, as in English the sign of the number would be spoken at the beginning anyway.

This brings us to a very important consideration for the universality of this technique, which is the way in which numbers are constructed in different languages. While the design and piloting of this technique has been done exclusively for English speech, the authors do not ignore the fact that the construction of spoken numbers may be different in other languages. For example, some European languages (French and Danish are two examples) use a vigesimal system (base-20) to construct some numbers in speech. Even in English, numbers 11-19 are irregular in the system described here.

Conclusions

This paper has introduced a new technique for using speech to browse tabular numerical information in search for overview information. This technique attempts to compromise between the familiarity of speech among expert users of speech-based accessibility tools and the limitations of gaining overview information that are intrinsic to the use of speech. This compromise is attained by interactively optimising the use of speech, applying the properties of MSD’s and the way in which speech representation of numbers is constructed in English. Among the two implementations of this technique evaluated quantitatively, the one in which the speed of movement of the hand controlled the amount of detail retrieved was preferred by the participants. With this implementation, their performance was observed to be comparable to the performance obtained with non-speech sounds in earlier studies with TableVis, and faster than with traditional speech-based techniques. A full empirical study is required to quantify these observations. Future work should focus on extending this technique to more general data sets (for example containing values with varying numbers of significant digits). Attention to languages other than English will inform about the universality of this technique.

Acknowledgements

This work was supported by the EPSRC funded MultiVis II project (GR/S86150).

References

[1] Bailey, R.W. Performance Vs. Preference. In *Human Factors and Ergonomics Society 37th Annual Meeting*. 1993, pp. 282-286

[2] Kildal, J. and S.A. Brewster. Explore the Matrix: Browsing Numerical Data Tables Using Sound. In *ICAD2006*. 2005. London, Ireland, pp. 300-303

[3] Kildal, J. and S.A. Brewster. Providing a Size-Independent Overview of Non-Visual Tables. In *12th International Conference on Auditory Display (ICAD2006)*. 2006. Queen Mary, University of London, pp. 8-15

[4] Shneiderman, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *IEEE Symposium on Visual Languages*. 1996. Boulder, CO, USA: IEEE Comput. Soc. Press, pp. 336-343

[5] Walker, B.N. and J.T. Cothran. Sonification Sandbox: A Graphical Toolkit for Auditory Graphs. In *International Conference on Auditory Display (ICAD)*. 2003. Boston, MA, USA, pp. 161-163