

## A Case Study on Non-Visual Pen-Based Interaction with the Numerical Data

**Tatiana G. Evreinova**  
School of Information Sciences  
University of Tampere  
Tampere, 33014, Finland

*Tatiana.G.Evreinova@uta.fi*

**Grigori Evreinov**  
School of Information Sciences  
University of Tampere  
Tampere, 33014, Finland

*Grigori.Evreinov@uta.fi*

**Roope Raisamo**  
School of Information Sciences  
University of Tampere  
Tampere, 33014, Finland

*Roope.Raisamo@uta.fi*

---

### Abstract

A widespread tabular format still poses great problems for screen readers because of diversity and complexity of the cells' content. How to access the numerical data presented in tabular form in a quick and intuitive way in the absence of visual feedback? We have implemented and assessed the algorithm supporting an exploration of the tabular data in the absence of visual feedback. This algorithm helps to solve the most commonly encountered problems: retrieving the position of extreme values and the target value that can also be linked to the specific content of the virtual table. The performance of 11 blindfolded subjects was evaluated when they used the StickGrip kinesthetic display and when they relied on the Wacom pen and auditory signals. The results of the comparative study are reported.

**Keywords:** Numerical Data, Pen-based Interaction, Kinesthetic Display.

---

### 1. INTRODUCTION

Douglas and his co-authors [1] have pointed out that according to statistical data one out of every three working blind adults have to use spreadsheets and databases on a daily basis. Visually impaired people still find it very challenging to access to composite tabular information through commercially available assistive technologies. Screen reading the table on a cell-by-cell basis can be used when the number of records (cells) is limited to a reasonable amount of data. Listening to the long sequences of the monotonous data can cause a loss of concentration on the main task, overload the person with unnecessary details and such a reading leads to frustration. Recently Burton and Lilly [2] assessed the Huntington National Bank website with different screen readers for navigating the tables and forms for bill paying and transferring funds. They found that this site was well accessible with JAWS, Window-Eyes, System Access, NVDA (NonVisual Desktop Access) and Thunder. As a result, they concluded that NVDA was the winner to support different online banking procedures.

Nevertheless, some companies (Jyske Netbank of Denmark and other international units of Danske Bank in Finland, Estonia, Latvia, Lithuania, Sweden, Norway, Ireland, UK, Germany, Russia) motivated by desire to achieve better security of the Online Banking service increase the length of identifiers and the method of logging in to their system. At that, they often ignore the abilities of the customers. In particular, bank identifiers can consist of a user ID, a personal

password and an additional security card number indicated via a key number printed in the security card (Figure 1). Moreover, if the person is not able to enter the required code correctly and within a certain time interval the security number cannot be identified then the system will require repeating the identification procedure for the new key number. Of course, the card content could be scanned (captured by camera phone), converted into the tabular format and accessed with the screen reader. However, reading of the entire security table (containing ninety 10-digit numbers) might be time consuming and not safe at all. The procedure of access to the specific number should also be optimized in terms of preventing an “observation attack” including any means of possible registration of accompanying physical signals.

Beyond the online banking service and other minor problems detected with different screen-readers [2], a slow exploration of large tabular data sets and the cells with a complex content hampers everyday activities and solving practical tasks, vocational and educational issues in the absence of visual feedback. By the large data sets, we mean tens of columns and tens of thousands of records when a combination of several digits could not be a unique as a key number in the security card.

Let us look at a few more examples of the tabular data sets. A cell of the table can display different types of data including textual description, pictures and a set of different but somehow linked together features [3]. In particular, the number 38 (Strontium) can be found 9 times in interactive periodic table (<http://www.ptable.com/>), an abbreviation of “Ta” (Tantalum) – 14 times if it was retrieved by the search engine and so on. The data collected in cells of the table can be the results of measurements, statistic functions, descriptions of the products or their components, vehicles, furniture, dress, apartments, etc. When the person wishes to get an overview of the empirical data or the product portfolio, what s/he needs to know in the first step?

These are the extreme values, the most expensive and the cheapest products, the products with the specific features. Nobody wishes to scan the entire table records. Instead, sighted people try to use sorting or another way of visualization, which could facilitate even visual evaluation of data: the extreme and average values, their trend and variation/volatility.

The widespread opinion that to analyze the data the person should explore perceptually all the records by sight, by hearing, or by touch is becoming obsolete. The tables were invented as a visual means facilitating data sorting and analysis. Since then, there have been developed a variety of methods and algorithms to analyze the data and present a result of the data processing in a compact and efficient way in any modality. Non-visual navigation in tables and spreadsheets based on the direct data conversion into another accessible modality and efforts spent to develop the tools to overview the specific information and to maintain the various procedures of the data processing would become useless, as such a way is often not efficient and frustrating [4], [5], [6]. To easily infer indicative features, patterns and trends, the data records have to be grouped and sorted and these procedures should not be seen as continuous efforts of the person. The tabular data can be converted into the multidimensional array and be further processed by an appropriate software application with respect to the task assigned by the user [3]. The information-processing medium is intended and must be designed to meet the needs of the users, by reducing the perceptual and cognitive load when dealing with any type of data presented in any format.

Another great problem with the numerical data manipulation stems from devices (tablet PC, smart phones and other mobile gadgets), which are exclusively optimized for visual-based touch input. At that, the touch sensitive surface cannot provide intuitive haptic feedback as mechanical keys of the ordinary keyboard. That is the question of how to access the numerical data presented in the tabular form in a quick and intuitive manner in the absence of visual feedback remains open.

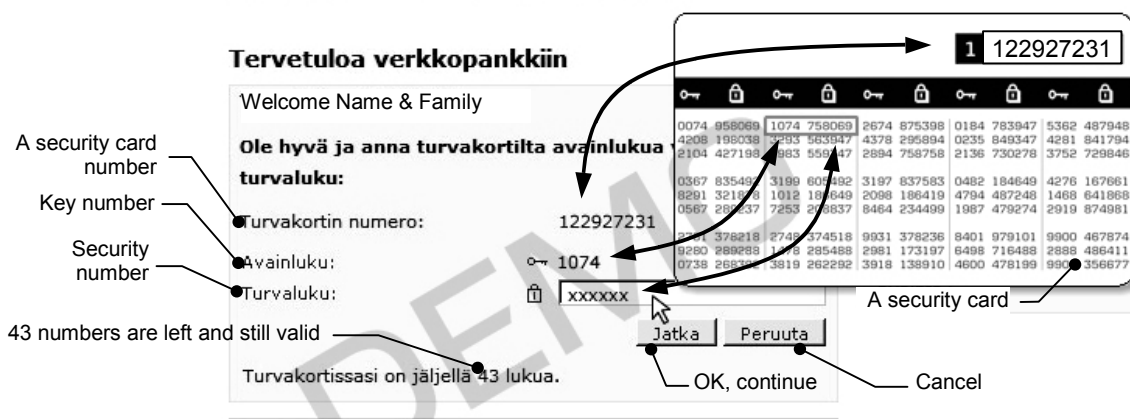


FIGURE 1: Demo shows the use of the security card when logging in to the online banking system.

## 2. BACKGROUND

The tabular data are usually grouped spatially and functionally. The spatial layout of the table can provide hyper and metaphorical information to the sighted person that often can not be directly extracted from the content of the individual cells (see e.g., periodic table of elements). Therefore, the navigation and selective reading the content of separate cells aggravate mental integration of the tabular data. The latter makes the relationships between the values of the cells unintelligible [7]. Spatially distributed tabular data require different overview and attentional mechanisms to explore and comprehend the content and the features of the layout. Many features have to be translated into accessible forms while preserving their relationships and generalization – spatial and functional grouping of the records of numerical, textual, graphical, audio, and other types of data values [5], [8]. At that, an alternative visualization of the hyper and metaphorical information has to demand a minimum of perceptual resources from an observer.

The vast majority of the previous research work was based on the cross-modal data transformation and focused on the use of sonification, including non-speech sounds, sound spatialization and musical techniques to convert visual information into auditory (phonological) representations [9], [10], [11], [12], [13], [14]. Various research projects have been undertaken to convert visual presentation of the entire table directly into another perceptual modality. Different generic properties of non-speech sounds such as the pitch, loudness, timbre and time (repetition rate) have been tested and combined with numerical data, their main statistical features and exploratory behavior of the user to facilitate browsing and access to large sets of data. Converting tabular data to series of pitches with different frequencies resulted in better perception among the visually impaired than when other perceptual properties were employed for data sonification [15].

Ramloll with co-authors [11] introduced a sonification approach to overview 2D tabular data arranged in 26 rows and 10 columns. The person could listen to any column or row as a continuous non-speech sound sequence by using keys on the numeric keypad for navigation (arrow keys and the center key), shortcut keys (PgUp/PgDn, Home, End), modal keys (to select 'value', 'pitch' or 'labels' mode) and overview keys ("+" and Enter). In this approach, the listener has to perform a sophisticated listening analysis to discover a tendency in the sound pattern and, thus, retrieve the relationships between the values of the cells within each row or column. Nevertheless, as cognitive load factors are not simply additive, when the number of records (i.e., the cells to be analyzed) is great enough, the complexity of perceptual parsing and the mental workload required to perform perceptual filtering and comprehension tasks in the absence of visual feedback is enormously increasing. The proposed sonification concept is only a reasonable way for presenting a limited number of data [8] or indicating general tendencies as a result of the statistical analysis without specifying exact values.

To overcome difficulties of exploration of the numerical data using the traditional speech-based techniques consuming time and working memory, Kildal and Brewster undertook several attempts [16], [17] to speed it up by sonifying data values at once in a selected row or column in very rapid succession. Duration of sound per value was 60 ms with an interval of 20 ms between sounds of neighboring cells. However, to avoid problems with indistinguishable sounds, the technique was limited for sonification of only integer values between 21 and 108, which have corresponded to 88 musical tones. Moreover, only the main characteristics of the relatively complex sound pattern could be detected and identified by the subjects in more detail. Then, in [17], the authors decreased the duration of sound per value to 7 ms by making quick overviews, like the cross-sectional slices over the entire table, distributing the complete range of 66 MIDI notes of piano instrument. They were also able to sonify the selected parts of the data in row or column with adjustable sound parameters. To provide the user with kinesthetic feedback, the researchers mapped the table to a Wacom graphic tablet. Thus, users could navigate with help of the pen relying on the sense of the movement and relative position of their limbs. The results of research have shown that while listeners managed well to perceive general overview of the data series, they were not able to assess in detail integrative picture about the specific features of the rapidly played back sound sequences. In the absence of spatio-temporal synchronization (compare the rhythmic structure of music notation), rapidly played back sound sequences were perceived as separate auditory events. The authors concluded that proposed accessibility technique did not facilitate the task of exploration of the numerical data for blind and visually impaired users.

Stockman with colleagues [18], [19] brought forward another approach for supporting interactive sonification of spreadsheets and diagrams. They provided automatic sonification of the rows and columns and activated the sonification cursor to provide the user with an overview by approximating parameters of the neighbor area of the arbitrary number of cells of the spreadsheet. Exact information was delivered through verbal feedbacks using the screen reader. The authors demonstrated that the system which is tailored to the user requirements and can display on demand overviews of the data from a range selected will decrease cognitive load and will speed up the process of exploring groups of cells for particular features [7]. The drawback of this approach was that the user had to navigate with the sonification cursor by controlling four additional keys (E-up/D-down, S-left/F-right) with the left hand while operating with the PC cursor by the right hand and arrow keys. To synchronize behavior of two logic cursors (activate/superpose/lock together), four modification keys were used and more keys were required in order to control parameters of auditory feedback. Abundance of controls and conditional signals needed for operating the application significantly aggravated the usability of the system implemented.

In order to facilitate manipulating any types of the tabular data, retaining a reasonable level of redundancy to indicate the trend of the variable in question and a possibility of navigation across the data, the application has to provide more intuitive and efficient way of processing and displaying the information relevant to the user's task, query and intention. Why the person should waste a time by listening to many sounds (e.g., 66 or 88 MIDI notes, [10]) which is not a straightforward task to detect the cells containing the highest/lowest values? This is also inconsistent with the Human Factors Guidelines on the use of non-speech acoustic signals and auditory tones [20] specifying that only tones having a frequency (pitch) between 300 Hz and 3000 Hz (that is, less than 50 notes or 4 octaves) can be employed to be clearly distinguished by the human ear.

Wall and Brewster [21] also undertook attempt to augment non-visual exploration of the numerical data with haptic force feedback. The entire tabular data were converted into a virtual 3D surface (plot) and explored with the PHANToM force-feedback device.

Other methods, techniques and tools have recently been implemented to facilitate exploration and manipulating numerical data in the absence of visual feedback. Nevertheless, when multiple records should be accessible in the absence of visual feedback the time to perform analysis of information presented is a critical factor. That is, the time and efforts (perceptual, cognitive and

behavioral) should be compatible to those that the blind person spent to comprehend the text block of about the same size accessed with a screen reader, including a minimum of additional knowledge and skills required to use the technique. For instance, like the reading of about 100 words, detection of the extreme value in the data set of 100 records should not exceed 20 s [4], [5], [6], [22].

By analyzing the results reported in the literature, we came to conclusion that rather than to spend additional efforts for understanding the encoding semantics of the tabular layout, any transformations of the data should reduce their redundancy while retaining some features of the spatial arrangement and a possibility to observe the trend of the variable in question, if necessary. At that, the important issues raised by Colwell and Sjöström should be taken into consideration [23], [24]. The researchers argued that it is necessary to provide visually impaired and blind user with auxiliary guiding feedback. When the users employ new technique, they have to get confident about the task and the workspace being explored, otherwise, they might easily become lost with any approach.

For instance, the task of the person is to detect the position of the numerical value that is a key to a specific (linked) data or another content in the table. Sighted person could get an exact value using find method in Excel and by filtering out perceptually irrelevant information or by creating an array formula. However regarding the skills and knowledge required, creating an array formula in the absence of visual feedback is not an easy task especially for elderly people (see e.g., Excel Magic Trick [25]) even with a suitable screen reader. Another approach consists in the formalization of the human-computer dialog when the application or the input technique (at the level of driver) could mediate the interaction between the user and the data. In such a way, the data analysis has to begin from a specification of the task to the application-mediator, e.g., by activating the appropriate functionality (mode). Then, instead of exploration of original tabular data the person would be able to supervise the data analysis performed by the application-mediator by making movements of the stylus or finger along (or parallel to) the physical borders of the study area (e.g., along the touchscreen bezel). At that, the size of the physical space of interaction could be optimized to better fit it to the touch-sensitive surface (touch pad, touchscreen or touch tablet). So, the user could systematically explore intermediate information presented in a suitable form.

The goal of the study presented was to examine how fast and accurately in the absence of visual feedback people are able to analyze the numerical data presented in a tabular form by following the proposed methodology. We approached this question by measuring both the mean time spent by the participants to complete the task (to search for the extremes or the given value in the array), and the number of errors committed when different modalities of auxiliary signals were used: an auditory and haptic signals accompanying the pen-based interaction.

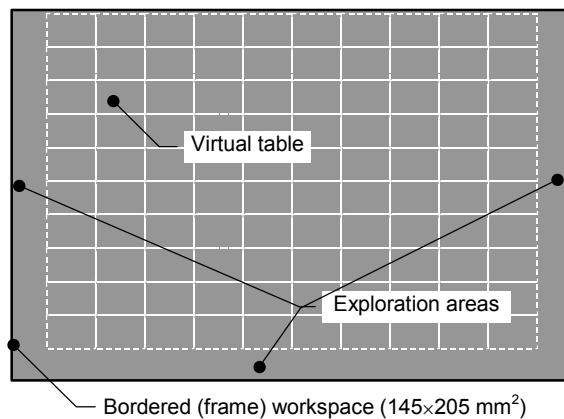
### **3. METHOD**

#### **3.1 Retrieving the Position of Extreme Values in the Data Array**

The algorithm has to be easily adjustable to the human perceptual and cognitive resources. Let us examine the particular task to detect the position of extremes in the 2D numerical data array. At the first step, the user has to specify the task to the application-mediator by activating the appropriate functionality and to select the desired modality of auxiliary guiding feedbacks. As shown in Figure 2, the 2D data array can be imagined as a virtual table (e.g., composed of 10 rows and 10 columns), size of which is maximized to fit the table to the size of the bordered frame (study area). Then the person should rely on the specific features of the workspace (the tablet), being explored in the particular sequence. For instance, the user can move a stylus across the rows within the left or the right exploration areas to allow the application to detect the extreme values (a local minimum or maximum) by rows and to display these values in the preferred modality (auditory or haptic) when the Y-coordinate of the pointer location corresponds to the current row. Thus, instead of listening to or relying on haptic signals to perceive many numerical values which would present the translated content of the cells in each row, the user can get

intermediate information – a local minimum/maximum value in the current row. Sliding across the rows will generate a number of local extremes. Still, these values retain information regarding their location (the order) within the array.

When a number of rows and local extremes are limited to ten or less, all of them could be converted into non-visual signals to present their value and relative position with respect to the virtual rows. At that, if the length of the array is also limited to a reasonable number of elements, it would allow to apply the linear model for cross-modal transformations, that is, preserving the same scale and relationships between non-visual values of the signals. However, such an approach is relatively inefficient when the task requires to focus on the identification and localization of the global extremum and if the number of elements in an array is large enough (typically 1000 or more).



**FIGURE 2:** The application-mediator augmented with auxiliary guiding feedbacks projected onto the Wacom tablet. It was viewed on the monitor of the experimenter (see subsection 3.4).

Another solution consists in alternative imaging not the local extremes (their exact values) but the distance between the current extremum and the global extremum in the array. In such a way, the distance might be presented in non-visual modality using the following algorithm for imaging the distance to the target (global extremum) through a relatively limited number of non-visual signals (see also subsection 3.4 Table 1 and Table 2):

- 1) When the distance between a local extremum and the global extremum is less than 20 units of measurement of the values stored in the array then every value of the distance can be translated into a non-visual signal;
- 2) When the distance between a local extremum and the global extremum is more than 20 units and less than 50 units of measurement of the values stored in the array then every third value of the distance can be translated into a non-visual signal;
- 3) When the distance between a local extremum and the global extremum is more than 50 units and less than 100 units of measurement of the values stored in the array then every fifths value of the distance can be translated into a non-visual signal;
- 4) When the distance between a local extremum and the global extremum is more than 100 units of measurement of the values stored in the array then further increasing the distance can be accompanied with the same specific non-visual signal.

Thus, there is no need to use a wide range of non-visual signals to convert a large amount of data stored in the array. The concept of imaging the relative position of global extremum with a limited number of non-visual signals could further facilitate detection of the row where the global

extremum is located. Then, to make navigation along the row less dependent on the absence of visual feedback, by sending the command through an accessible input, the user can lock the row number and continue exploration of the row content.

By dragging the stylus along the bottom exploration area the pointer will follow within the row locked. At that again, the distance between the current column and the column where the global extremum is located can be indicated in a similar way as it was already discussed with respect to the local extremes in each row. Moreover, any local or linked information can be read aloud by the screen reader.

### **3.2 Retrieving the Position of Any Value in the Data Array**

Processing the virtual array remains a possibility of mutually inverse operations with respect to the user task. Let us consider an array of 100 elements of integer values varying from 0 to 99. The elements of the array can be presented in the table in a random order. In the previous subsection 3.1 the algorithm of retrieving the position of extremes of the array was discussed in detail. In more general case, any element of the array can be assigned as the target having the minimum value equal to "0". Then, values of other elements of the array should be changed accordingly. That is, the value of the *i*-element (e.g., having the value 78) could be turned into the minimum and be equal to "0" by shifting all other numerical values accordingly. However, some elements, which had values less than the target, can get a negative value. The use of the distance of *i*-element from the target helps to convert the numerical values into non-visual signals regardless of their sign. This also decreases the number of non-visual signals required for translation. Thus, the task of retrieving the position of the *i*-element having the numerical value 78 can be reduced to the task to detect the position of the minimum. The algorithm provides quick search for numeric values in the data array of any length presented in a tabular form in the absence of visual feedback.

Moreover, non-visual imaging of intermediate data that helps to navigate within the array would require less than 50 out of 128 notes of the MIDI Synthesizer or about 40 clearly distinct states (levels of the penholder displacements) of the StickGrip kinesthetic display (see detailed description in subsection 3.4) to solve the task. Such an approach would be entirely consistent with the Human Factors Guidelines on the use of non-speech acoustic signals and auditory tones [14].

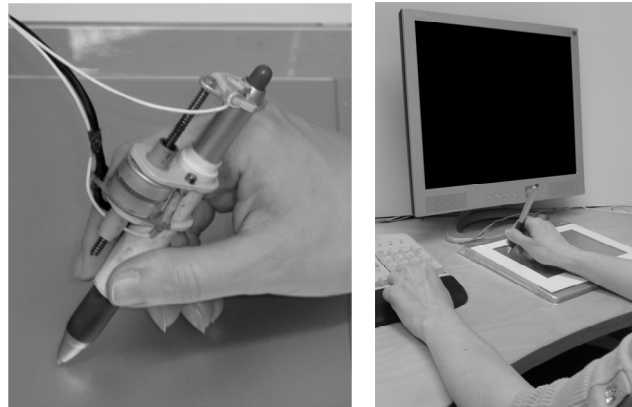
### **3.3 Participants**

Three totally blind persons successfully tested the technique, but these data were excluded from analysis. For ethical reasons, we could not involve blind people in testing when many factors were unknown. That is, only sighted people were able to evaluate the benefits and shortcomings of the new technique introduced. Eleven volunteers from the local university participated in this study. All had normal hearing abilities (but we did not make audiometric tests) and none of the participants had any extensive training in music. They did not experience issues related to skin sensitivity or movement disorders. Five of the subjects were males and six were females. The age of the subjects ranged from 29 to 58 years with a mean of 35 years. All participants were regular computer users and reported being right hand dominant or using their right hand most frequently. None of them had participated in similar experiments before.

### **3.4 Apparatus and Equipment**

A desktop PC with Intel Core Quad CPU Q9450 (2.66 GHz, 3.25GB of RAM) was used in the present experiment equipped with the PCI sound card Creative Sound Blaster Live! 5.1 and active stereo speaker system KOSS SA/30. The Wacom Graphire-4 USB tablet having an interactive surface of 280×260 mm<sup>2</sup> and including a cordless stylus was employed as an input technique. In addition, Wobbrock with co-authors have proven [26] that physical edges can be efficiently used to provide greater stability and higher accuracy in the use of stylus input techniques in the absence of visual feedback. To limit extra movements of the Wacom pen out of the workspace we used the rectangular frame of 145×205×1 mm<sup>3</sup>.

To provide the participants with auxiliary haptic feedback the motorized penholder for the Wacom pen (the StickGrip) was used [27]. The device (Figure 3) provides a range of 40 mm ( $\pm 20$  mm) of the vertical displacement of the penholder (the point of grasp) with an accuracy of ( $\pm 0.8$  mm) and with an average speed of 30 mm/s for the Wacom pen having a length of 140 mm.



**FIGURE 3:** The StickGrip kinesthetic display (on the left) and experimental setup (on the right).

Thus, when the participants explored the data distributed within the workspace, they could feel as a point of grasp is sliding up and down the shaft of the Wacom pen and their hand being displaced towards and away from the physical surface of the pen tablet. Distance and direction of the penholder displacements were coordinated with a distance to target – the results of the data processing by the application-mediator (see subsections 3.1, 3.2). The software was developed in the Microsoft Visual Basic environment under Windows XP. An embodied Microsoft Speech API was used to announce short verbal messages or any variable on demand (by clicking the arrow key down of the keyboard).

The easily distinguishable MIDI sounds of the patch 21 (Accordion instrument) were generated via the JWCS MIDI Synthesizer Control [8], [28]. Table 1 and Table 2 demonstrate how the distances to the target were encoded with a minimum number of MIDI notes to detect the position of the global extremum values. We used an original sound without any additional effects and modification by equalizer.

DISTANCE (D)	MIDI NOTE (D)	Note number // name // Frequency in Hz
0 to 19	60 + d	60, 61, 62...79 // C5, C#5...G6 // 262, 277...784
20 to 49	75 + Integer(d/3)	81, 82, 83...91 // A6, A#6...G7 // 880, 932...1568
50 to 99	83 + Integer(d/5)	93, 94, 95...102 // A7, A#7...F#8 // 1760, 1865...2960

**TABLE 1:** MIDI Notes (0-127) Used to Indicate the Distance to a Minimum Value in the Virtual Array.

Distance (d)	MIDI Note (d)	Note number // name // Frequency in Hz
99 to 80	82 + (d-80)	102, 101...82 // F#8, F8...A#6 // 2960, 2793...932
79 to 50	54 + Integer(d/3)	80, 79, 78...70 // G#6, G6...A#5 // 830, 784...466
49 to 0	60 + Integer(d/5)	69, 68, 67...60 // A5, G#5... C5 // 440, 415...262

**TABLE 2:** MIDI Notes (0-127) Used to Indicate the Distance to a Maximum Value in the Virtual Array.

### 3.5 Experimental Procedure

The experiment was conducted in the usability laboratory at the local university. The subjects were tested individually. Prior to data collection, the experimenter briefly introduced the



techniques and the goal. The detailed instructions were given for the use of kinesthetic display, the task and overall testing procedure. The formal consent of the participants to collect and use the data obtained were gathered. No private or confidential information was collected and stored.

Still the participants were given an opportunity to refuse the continuation of the experiment at any point without any explanation of the reason and they were told that such a refusal would not have any further negative effect to them or to the experiment.

The subjects were advised to anchor the hand on the desk where the Wacom pen tablet was located. They were asked to leave the hand relaxed while grasping the pen or the StickGrip. The participants could perceive MIDI sounds, speech and haptic feedback cues when they explored the numerical data relying on hearing, kinesthetic and proprioceptive senses. No visual feedback was provided during the test. The monitor was switched off during the whole experiment, however the participants were asked to keep their eyes closed in order to better focus on the feedback signals.

In the beginning of each session, the subjects were allowed to practice shortly with the sequence of actions in the absence of visual feedback. The activity of the subjects was monitored on the computer screen of the experimenter. The participants could ask questions if they felt that the testing procedure required more clarification.

### **3.5.1 The First Experimental Task**

The first task consisted in detection the position of the global extremum value in the virtual table. The experimenter did set the appropriate mode of the application according to the task. To accomplish this task, the participants used Wacom tablet bordered with a rectangular frame, as shown in Figure 4. The preferred positions of the keyboard and the tablet were adjusted beforehand.

The subjects started the trials by clicking the spacebar on the keyboard. To find out the location of the virtual cell containing the extremum (minimum or maximum value) the participants had to move the Wacom pen (Figure 3) or the StickGrip up and down along the borders of the rectangular frame, within the left or the right exploration areas (Figure 3). When the Y-coordinate of the pointer location was within the row of the virtual table, the application made analysis of the row data and generated the signal, which designated the “relative distance” between the local extremum value and the target value (the global extremum).

The participant had to perform an auditory or haptic perceptual analysis of only 10 non-visual signals and make a decision in which row the global extremum is located. The minimum value of the distance was translated into the lowest sound pitch or position of the penholder. The highest sound pitch and displacement of the penholder designated the maximum distance to the target value. After detection the row in question where the exact position of the global extremum value might be located, the subject had to press the “0” (zero) key on the numeric keypad to lock this row. After the first step, the subjects had to continue an exploration of the global extremum position within the row. The participants had to move the Wacom pen or the StickGrip along the bottom exploration area (Figure 3) of the frame. They should press the 0-key again when the target value was located. When the virtual cell (extremum value) was incorrectly pointed out, the person heard the error sound to stimulate him/her to be more attentive toward the analysis of the non-visual signals. Manipulating with the spacebar and 0-key of numeric keypad did not cause any problems in the absence of visual feedback. Nevertheless, there was a possibility to assign any other two keys on demand.

The participants had to complete two blocks of 20 trials each (10 tasks to detect the minimum and 10 tasks – the maximum value), under two conditions of feedback with no time limit. In one block, a pen input was accompanied with auditory signals, in another block the StickGrip was used as a kinesthetic display of the relative distance to target.

The numerical data were placed in the virtual table in a random order but the conditions, type and location of extremum were presented in a counterbalanced order. The type of extremum in question was announced as a verbal message (“minimum” or “maximum”). The end of the test was accompanied with the applause sound. Between trials, the participants were allowed to have a short break to ask questions and give comments.

A session of two blocks lasted for an hour and the test was performed 3 times per week, during 2 weeks, according to the ethical standards. Thus, each of eleven subjects performed 40 trials per day and 240 trials in a total. The data recorded were stored in a log file for further analysis. With a break of one week, the same group of participants performed the second task.

### 3.5.2 The Second Experimental Task

The second experimental task was to locate any value in the virtual table (data array). Described in subsection 3.2, the algorithm was used to support non-visual interaction with numerical data. The elements of the array for each task were transposed automatically when the value in question was selected and announced verbally in synthesized speech. The value, which has to be detected, was designated as a global minimum. The value of  $i$ -element ( $Val_i$ ) in the array after transposition was converted and presented as follows:

$$Val_i = Abs (Val_i - Val_{target}) + 1 \quad (1)$$

where  $Val_{target}$  was the announced value in question.

After listening to the task, the participants followed the procedure similar to the first experimental task (see subsection 3.5.1). The protocol of the testing procedure was also similar and comprised of 5 sessions completed during one week with 40 trials per day (2 blocks). All cells of the virtual table (100 values of the array) were presented in a counterbalanced order.

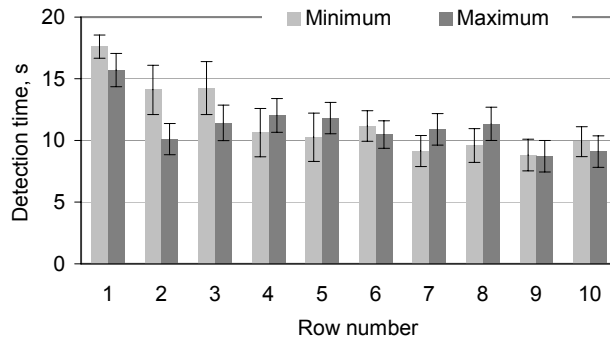
## 4. RESULTS

Within a bordered area of the tablet, the subjects demonstrated a similar navigation activity required by the task and the techniques used. At the beginning of the test, the subjects needed some practice to comprehend the way of interaction with tabular data in the absence of visual feedback to learn how to manage with the Wacom pen or the StickGrip device and how to interpret non-visual signals indicating the relative distances to the target value instead of the true values. They moved the input device very carefully and slowly back and forth along the left or right edges of the frame in order to inspect the sequences of sounds or the penholder displacements. The subjects started the test from upper row. After a series of repeated observations and a thorough perceptual analysis, they confirmed their choice according to the test protocol. With the experience gained during the first session, the subjects felt more confident in the absence of visual feedback relying on auxiliary information and they performed the task faster already during the second session. At that point, the heterogeneity in decision-making has increased and the subjects have committed a greater number of errors.

At the beginning of the test, 4 of 11 subjects tried to detect the extremum in the first (upper) or the last (bottom) two rows. They were expecting to directly find out the extremum value encoded by sound pitch, while they were told that non-visual signals indicate the relative distance to the target cell and not the true value. Still, they immediately confirmed their choice without making thorough inspection of all the rows to be sure that the decision was made correctly. On the contrary, (based on the subject's reports) the penholder displacements seemed “more obvious”, “natural”, “intuitive” and “realistic” non-visual indicator of the physical distance.

The mean times spent by the subjects during the first week to detect the position of the global extremum in the virtual table using the Wacom pen and sound signals (Condition 1) are presented in Figure 4. The results revealed that 3 of 11 participants had difficulties to identify the minimum value than the maximum value. This would be possible due to the fact, that we

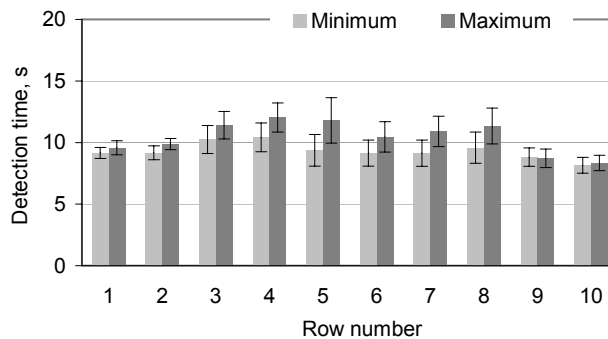
intentionally did not apply any adjustment of the audio system. The participants were the computer users and non-musicians and could have different hearing sensitivity at different frequencies (pitch and volume). However, by retrieving the position of the minimum values the subjects made a thorough inspection with fewer errors.



**FIGURE 4:** The mean time spent by the subjects to detect the global extremum (target) value in the virtual table using the Wacom pen and auditory signals. The data of the 1st week of testing were averaged over 33 trials in each row.

The mean task completion time of detection the position of global minimum was about 11.5 s with a standard deviation (SD) of about 2.9 s, varying from 9 to 17 s. To find the position of global maximum they spent in average 11.2 s (SD=1.9 s). This time varied from 8 to 15 s. The paired samples t-test revealed no statistically significant difference in this task under Condition 1:  $t(9) = 0.592$  ( $p > 0.5$ ), the correlation index of these data was 0.694 ( $p > 0.01$ ).

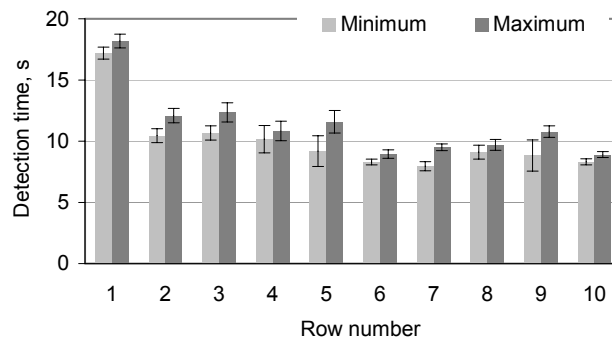
The subjects had to analyze all the local extrema and consequently all non-visual signals indicating their distances to the target (global extremum) over 10 rows. Therefore, it took more time to identify the cell position when it was located in the upper row than in the bottom row due to differences in the travel distance towards the bottom row and backwards to the target. If the global extremum value was located on the way towards the bottom row, it took less time to find the position of extremum. The similar situation was observed independently of the signals modality.



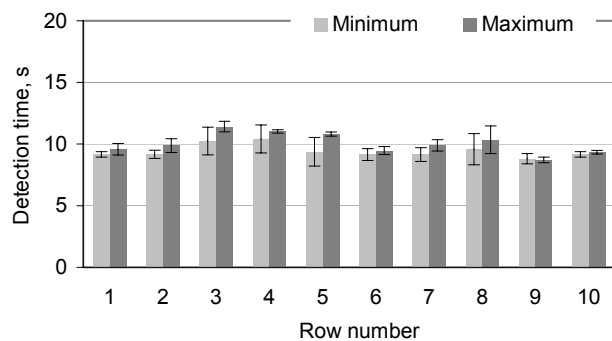
**FIGURE 5:** The mean time spent by the subjects to detect the global extremum value in the virtual table using the Wacom pen and auditory signals. The data of the 2nd week of testing were averaged over 33 trials in each row.

During the 2nd week of testing, the mean task completion time of detection the position of global minimum was about 9.3 s (SD=0.65 s), varying from 8 to 11 s (see Figure 5), while to find the position of global maximum the subjects spent in average 10.4 s (SD=1.9 s). This time varied from 8 to 14 s. The paired samples t-test revealed a statistically significant difference in this task under Condition 1:  $t(9) = 4.4$  ( $p < 0.005$ ), the correlation index was high and significant 0.85 ( $p < 0.005$ ).

The mean times spent during the 1st and 2nd weeks to detect the position of global extremum using the StickGrip (Condition 2) are presented in Figure 6 and Figure 7.



**FIGURE 6:** The mean time spent by the subjects to detect the global extremum value in the virtual table using the StickGrip device. The data of the 1st week of testing were averaged over 33 trials in each row.



**FIGURE 7:** The mean time spent by the subjects to detect the global extremum value in the virtual table using the StickGrip device. The data of the 2nd week of testing were averaged over 33 trials in each row.

At the beginning of the test (Figure 6), the subjects also spent more time to detect a position of the global maximum with an average value about 11 s (SD=2.7 s) varying from 8 to 18 s. Detection the global minimum required of about 10 s (SD=2.7s), varying from 7.5 to 17 s. As regards the task completion time the results of the paired samples t-test revealed a statistically significant difference in this task under Condition 2:  $t(9) = 6.178$  ( $p < 0.001$ ), the correlation index was high and significant 0.97 ( $p < 0.001$ ).

Still a practice with the StickGrip took a short time to learn how to distinguish and interpret haptic signals arising from vertical displacements of the penholder. To the end of the test, the subjects mastered their skills and it was required less time to find a location of extremum. To detect the global minimum the subjects spent on average about 9.4 s (SD=0.5 s) and the global maximum of about 10 s (SD=0.8 s). The paired samples t-test also revealed a statistically significant difference in this task under Condition 2:  $t(9) = 4.4$  ( $p < 0.005$ ), the correlation index was high and significant 0.885 ( $p < 0.005$ ).

The ANOVA analysis strengthened these findings and has shown that the novice-to-expert transition was significantly above chance when the participants was asked to detect the position of the global extremum in the virtual table with two different techniques. The data from the 1st and 6th test sessions were statistically assessed and compared (Table 3).

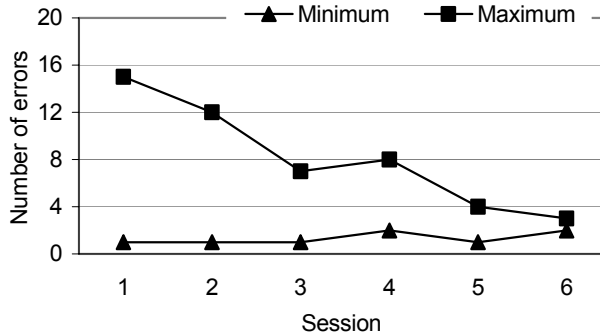
Extremum	Condition 1	Condition 2
Maximum	$F = 2.36; p < 0.001$	$F = 4.15; p < 0.001$
Minimum	$F = 1.63; p < 0.001$	$F = 3.75; p < 0.001$

**TABLE 3:** ANOVA results of detection the position of extremum values with two techniques in the 1st and 6th test sessions (10 trials, 11 participants).

As it can be seen from Figure 8, the subjects demonstrated a tendency to commit a greater number of errors when they were asked to detect the position of the global maximum under Condition 1. In particular, the average number of errors recorded from 11 subjects per session (110 trials) was about 7.4 (SD=4.3) for maximum and only 1.5 (SD=0.8) for the global minimum. The paired samples t-test revealed a statistically significant difference in this task under Condition 1:  $t(5) = 4.344$  ( $p < 0.005$ ), still the correlation index was low and not significant  $-0.4$  ( $p > 0.1$ ). The greater number of errors made in the beginning of the test was also influenced by misinterpretation of sounds indicating the distance to the target.

Relying on haptic signals (Condition 2), the subjects experienced similar difficulties to detect the position of the global extremum (Figure 9). They needed time to learn how to interpret the displacements of penholder with the distance to the target. However, when they were asked to detect the global minimum, the subjects again committed fewer mistakes.

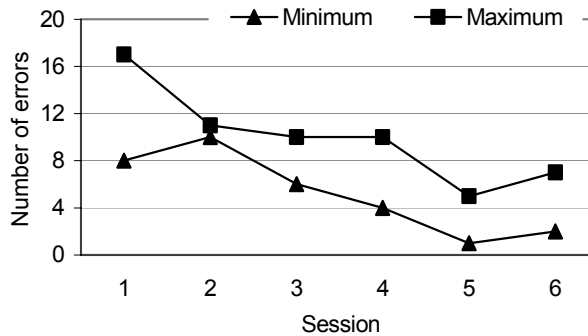
The mean number of errors recorded from 11 subjects per session (110 trials) was about 4.7 (SD=3.6) when the subjects were asked to detect the position of the global minimum and about 9.1 (SD=4.6) when global maximum was asked to locate. The paired samples t-test revealed a statistically significant difference in this task under Condition 2:  $t(5) = 6.2$  ( $p < 0.001$ ), a correlation was 0.855 and significant ( $p < 0.001$ ).



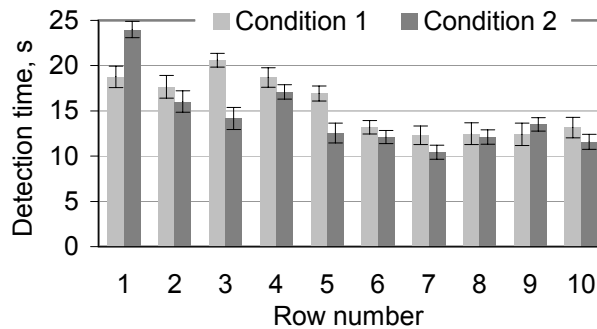
**FIGURE 8:** The average number of errors committed by 11 participants (110 trials per session) during detection of the global extremum value in the virtual table relying on the Wacom pen and auditory signals throughout six test sessions.

Figure 10 demonstrates the variations in the mean times spent by participants to detect the position of the numerical values (the second task) under two conditions. The data in each row were averaged over 110 trials.

In general, the participants spent a little less time to detect the position of numerical value being converted into the global minimum when relied on the StickGrip device: 14.3 s (SD=3.9s) vs. 15.6 s (SD=3.2 s). The paired samples t-test revealed no statistically significant difference in this task:  $t(9)=1.296$  ( $p > 0.1$ ), the correlation index was 0.646 ( $p > 0.01$ ). This result was confirmed by the ANOVA  $p < 0.001$ ,  $F = 0.97$ .



**FIGURE 9:** The average number of errors committed by 11 participants (110 trials per session) during detection of the global extremum value in the virtual table relying on the StickGrip kinesthetic display throughout six test sessions.



**FIGURE 10:** The mean time spent to detect a position of any value of 100 in the virtual table using the Wacom pen and auditory signals (Condition 1), and the StickGrip kinesthetic display (Condition 2). The data in each row were averaged over 110 trials.

Even with getting a practice in non-visual inspection of the virtual table during the first task the subjects spent greater time to detect the position of the cell when it was situated in upper part (rows 1-5) rather than in bottom rows (6-10). The mean of detection time in rows 1-5 was about 18.5 s (SD=2.4 s) vs. 17s (SD=4.5 s) varying from 13 to 23 s and from 9.6 to 26 s under Condition 1 and 2 accordingly. When the value in question was located in bottom rows 6-10 an average time was about 12.7 s (SD=2.1s) vs. 11.9s (SD=1.7s) varying from 8 to 17s and from 7.5 to 16 s under Condition 1 and 2 accordingly. As it was already discussed, it took more time due to differences in the travel distance towards the bottom row and back to the target position.

## 5. CONCLUSION

Different authors have tried to find a solution to improve accessibility to the data presented in a tabular form. However, both the data used, techniques, experimental conditions and methods of testing and analysis of the results reported in the literature vary widely. The available experimental results, which can still be compared to our approach, are presented in the Table 4.

We have implemented, described and assessed the technique and algorithm of retrieving the position of any value in the data records which are appropriate for non-visual navigation with numerical data when the number of elements in the array is large enough (typically 1000 or more). Moreover, the navigation within the virtual table was performed using the touch sensitive surface and pen-based input techniques. The number of non-visual cues required to support navigation is almost independent on the amount of data and was less than 50 signals, including the short verbal messages.

Still the goal of this study was to examine human performance and an accuracy of solving the most common tasks dealt with the virtual table composed of 100 cells. The tasks were retrieving the position of extreme values and the target value that can be linked to the specific content, as it was discussed in the scenario of access to the security number in the card (table) of online banking system.

Data format Techniques [Ref.]	Elements in array	Tasks assigned included: identification of	Auxiliary modality Auditory		Auxiliary modality Haptic	
			Time, s (SD)	Accuracy %	Time, s (SD)	Accuracy %
Tabular MIDI vs. StickGrip [new]	100	- the maximum value	10.8(1.9)	92.6	9.7(1.8)	90.9
	100	- the minimum value	10.4(1.6)	98.5	10.5(1.6)	95.3
	100	- the specific value of 100 integers	15.6(3.2)	100	14.3(3.9)	100
Tabular Speech vs. MIDI [11]	260	- the highest number of 100 integers (speech)	133(55)	4.6	--	--
	260	- the highest number of 100 integers (127 notes)	44(14)	5.3	--	--
Tabular MIDI [17]	28, 168, 744	- the highest/lowest value (66 notes)	*22(10) *26(10) *32(10)	*87(10) *90(10) *80(10)	--	--
Tabular Speech vs. Braille [6]	49	- the maximum value (browsing with speech)	42	88	--	--
	70	- the minimum value (browsing with a formula)	30	75	--	--
	21	- the specific cell (browsing with Braille)	*15	--	18	20
Graph Tactile paper vs. Novint Falcon [5]	11 (points)	- the max / min point (tactile paper)	--	--	23	42
		- the max / min point (Novint Falcon)	--	--	15	100

\* The data were not presented exactly in the paper; these values were extracted from the figures.

**TABLE 4:** Comparison of the new approach with existing methods.

The assessment of human performance revealed that the subjects committed the greater number of errors in the beginning of the test and relying on haptic signals rather than when they used the Wacom pen and auditory signals. Nevertheless, an experience with the StickGrip device took a short time to learn how to distinguish and interpret haptic signals arising from vertical displacements of the penholder.

The time required to detect the position of the virtual cell containing the extremum announced verbally was about 10.6 s (SD=1.75 s) when they used the Wacom pen and auditory signals vs. 10.1 s (SD=1.7 s) when the subjects relied on the StickGrip device. During the second task of detection any value announced verbally, the task completion time was about 15.6 s (SD=3.2 s) when the subjects relied on the Wacom pen and auditory signals vs. 14.3 s (SD=3.9 s) when they used the StickGrip device. However, the variation of the task completion time might have resulted from differences in the travel distances to the target position and the limited speed of penholder displacements of the kinesthetic display. Based on the subjective evaluation reports, we can

conclude that the StickGrip device has a potential as intuitive kinesthetic display to substitute auditory feedback in non-visual interaction with numerical data. The penholder displacements seemed to the subjects as “more obvious”, “natural”, “intuitive” and “realistic” non-visual indicator of the physical distance.

The algorithm allowing non-visual navigation in the virtual table on a touch sensitive surface and access to any value without extensive learning the numerical data converted into tactile and auditory patterns constitute a novel approach to make access and manipulating the tabular data in the absence of visual feedback.

## 6. FUTURE RESEARCH

Exploration of the large data set cannot be efficient when the cell content (e.g., numerical values) being converted into non-visual signals of any accessible modality would become redundant. Even detail overview of pieces of data does not allow to retrieve the data with a minimum number of steps. Rather than to spend additional efforts in understanding the encoding semantics of the tabular layout, the data translation methods should reduce their redundancy to a reasonable level while retaining some features of the spatial arrangement and a possibility to access to key variables indicating the trend of the parameter and descriptive statistics. The alternative approach might consist in the formalization of the human-computer dialog when the special application and the input technique could mediate the human-data interaction with respect to the user task. Then, only a limited number of navigational signals might be enough to find out a specific position in the virtual table and to display intermediate information if necessary during the data processing.

For visually impaired persons and in extreme situations, haptic channel is the most natural and perhaps the most efficient way of presenting a large amount of data (both discrete and continuous) in a limited amount of space [27], [29]. The StickGrip kinesthetic display was used in this study to provide haptic signals to support an intuitive interaction within the numerical data array: by indicating the relative distances (closer and further) between the position of the pointer and the target numerical value, providing an easy access to intermediate information and navigation in the data set. There are also many other tasks related to the interactive data visualizations.

In further research we also will focus on the issues allowing the person in the absence of visual information to interact with the data by perceiving, exploring, and understanding the data values and their relationships. The development of intuitive data visualization and exploration tools making human-data interaction comfortable and effective in the total absence of visual support remains one of the main challenges in the field of data visualization and related areas of research in human-computer interaction.

## 7. ACKNOWLEDGMENT

The authors gratefully acknowledge the support of Finnish Academy Grant 127774.

## 8. REFERENCES

- [1] G. Douglas, C. Corcoran, S.A. Pavey “The role of the WHO ICF as a framework to interpret barriers and to inclusion: visually impaired people’s views and experiences of personal computers,” *Br J of Vis Impair*, 25(1) 32–50, Jan. 2007, DOI: 10.1177/0264619607071773.
- [2] D. Burton, J. Lilly “NonVisual Desktop Access and Thunder: A Comparison of Two Free Screen Readers,” *Access World* 12(8) Aug. 2011.
- [3] Y. Wang, I.T. Phillips, R.M. Haralick “Table structure understanding and its performance evaluation,” *Pattern Recognition*, Vol. 37, No 7 Jul. 2004, pp. 1479-1497.



- [4] M. Chiousemoglou, H. Jürgensen “Setting the table for the blind,” in Proc. of PETRA, Int. Conf. on Pervasive Technologies Related to Assistive Environments, Crete, Greece, ACM Press, New York, 2011, doi>10.1145/2141622.2141624.
- [5] I.A. Doush, E. Pontelli “Non-visual navigation of spreadsheets,” UAIS, Mar. 2012, pp. 1-17, <http://dx.doi.org/10.1007/s10209-012-0272-1>.
- [6] I.A. Doush, E. Pontelli “Non-visual navigation of tables in spreadsheets,” in Proc. of ICCHP 2010, Vienna, Austria, Part I, LNCS 6179, Springer, Berlin, 2010, pp. 108–115.
- [7] C. Jay, R. Stevens, R. Hubbold, M. Glencross “Using haptic cues to aid nonvisual structure recognition,” ACM Trans Appl Percept. 5(2), Article 8, May 2008, 14 p.
- [8] T.V. Evreinova, G. Evreinov, R. Raisamo “Non-visual Gameplay: Making Board Games Easy and Fun,” in ICCHP 2008, Linz, Austria, Springer, Berlin Heidelberg, LNCS 5105, 2008, pp. 561-568.
- [9] L.M. Brown, S.A. Brewster, R. Ramloll, W. Yu, B. Riedel “Browsing Modes For Exploring Sonified Line Graphs,” in British HCI Conf, London, UK, vol. 2, 2002, pp. 2-5.
- [10] R. Ramloll, S.A. Brewster “Generic approach for augmenting tactile diagrams with spatial non-speech sounds,” in CHI Ext. Abstracts on Human factors in computing systems, ACM Press, NY, USA, 2002, pp. 770–771.
- [11] R. Ramloll, S. Brewster, W. Yu, B. Riedel “Using Non-speech Sounds to Improve Access to 2D Tabular Numerical Information for Visually Impaired Users,” in BCS IHM-HCI, Lille, France, Springer, Berlin, 2001, pp. 515–530.
- [12] S. Smith, R.D. Bergeron, G.G. Grinstein “Stereophonic and Surface Sound Generation for Exploratory Data Analysis,” in CHI SIGCHI Conf on Human factors in comp systems, ACM Press, NY, USA, 1990, pp. 125-132.
- [13] B.N. Walker, M.A. Nees “Theory of sonification,” in Hermann T, Hunt A, Neuhoff J.G., (eds), The Sonification Handbook, Logos Publishing House, Berlin, Germany, ch. 2, 2011, pp. 9-39.
- [14] B.N. Walker, J.T. Cothran “Sonification Sandbox: a graphical toolkit for auditory graphs,” in Proc. of ICAD Int. Conf. on Auditory Displays, 2003, pp. 161-163.
- [15] L.M. Brown, S.A. Brewster, R. Ramloll, M. Burton, B. Riedel “Design Guidelines For Audio Presentation of Graphs and Tables,” in Proc. of ICAD Int. Conf on Auditory Displays, Boston, MA, USA, 2003, pp. 284-297.
- [16] J. Kildal, S.A. Brewster “Explore the matrix: Browsing numerical data tables using sound,” in ICAD Proc. of Int. Conf. on Auditory Displays, Limerick, Ireland, 2005, pp 300-303.
- [17] J. Kildal, S.A. Brewster “Providing a size-independent overview of non-visual tables,” in Proc. of ICAD Int. Conf. on Auditory Displays, London, 2006, pp. 8-15.
- [18] T. Stockman “The Design and Evaluation of Auditory Access to Spreadsheets,” in Proc. of ICAD Int. Conf. on Auditory Displays, Sydney, Australia, 2004, at: <http://icad.org/Proceedings/2004/Stockman2004.pdf> [accessed on Mar. 2013].
- [19] T. Stockman, C. Frauenberger, G. Hind “Interactive Sonification of Spreadsheets,” in Proc. of ICAD Int. Conf. on Auditory Displays, Limerick, 2005, pp. 134–139.

- [20] "Human Factors Guidelines for ICT products and services," Design for All, ETSI EG 202116, V1.2.1 (2002-09), p. 146, available at: <http://www.cettico.fi.upm.es/aenor/etsi202116v010201p.pdf> [accessed on Mar. 2013].
- [21] S. Wall, S. Brewster "Providing External Memory Aids in Haptic Visualisations for Blind Computer Users," *Int. J. of Disability and Human Dev*, Vol. 4, No 4, May 2005, pp. 331-338.
- [22] I.A. Doush, E. Pontelli, D. Simon, T.C. Son, O. Ma "Making Microsoft Excel™: multimodal presentation of charts," in *Proc. of ACM SIGACCESS Conf. on Computers and Accessibility*, ACM Press, NY, USA, 2009, pp. 147-154.
- [23] C. Colwell, H. Petrie, D. Kornbrot, D. Kornbrot "Haptic virtual reality for blind computer users," in *Proc. of ASSETS Int. ACM Conf on Assistive technologies*, ACM Press, NY, USA, pp. 92-99, 1998.
- [24] C. Sjöström "Designing haptic computer interfaces for blind people," in *Proc. of ISSPA Int. Symp. on Signal Processing and its Applications, Workshop on Website Evolution*. Kuala Lumpur, Malaysia, IEEE CERTEC, Lund, Vol 1, pp. 68-71, 2001.
- [25] Excel Magic Trick 731: "Retrieve Cell Address of Minimum Value in Column," at: <http://www.youtube.com/watch?v=cpRIYaatzx4> [accessed on Mar. 2013].
- [26] J.O. Wobbrock, B.A. Myers, J.A. Kembel "EdgeWrite: A stylus-based text entry method designed for high accuracy and stability of motion," in *Proc. of UIST the ACM Symp. on User Interface Soft and Tech*, Vancouver, British Columbia, Canada, ACM Press, NY, USA, 2003, pp. 61-70.
- [27] T.V. Evreinova, G. Evreinov, R. Raisamo "An Evaluation of the Virtual Curvature with the StickGrip Haptic Device: a Case Study," *UAIS 12(2)* June 2013, pp. 161-173. DOI:10.1007/s10209-012-0273-0.
- [28] T.G. Evreinova, G. Evreinov, R. Raisamo "Camera-Based Target Acquisition Augmented with Phosphene Sensations," in *ICCHP 2010*, Vienna, Austria, Springer, Berlin Heidelberg, LNCS 6180, 2010, pp. 282-289.
- [29] A. Israr, P.H. Meckl, Ch.M. Reed, H.Z. Tan "Controller design and consonantal contrast coding using a multi-finger tactual display," *J. Acoust. Soc. Am.* 125(6), June 2009, pp. 3925-3935.