

EMA-Tactons: Vibrotactile External Memory Aids in an Auditory Display

Johan Kildal¹, Stephen A. Brewster¹

¹ Glasgow Interactive Systems Group, Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ, UK
{johank, stephen}@dcs.gla.ac.uk - www.multivis.org

Abstract. Exploring any new data set always starts with gathering overview information. When this process is done non-visually, interactive sonification techniques have proved to be effective and efficient ways of getting overview information, particularly for users who are blind or visually impaired. Under certain conditions, however, the process of data analysis cannot be completed due to saturation of the user's working memory. This paper introduces EMA-Tactons, vibrotactile external memory aids that are intended to support working memory during the process of data analysis, combining vibrotactile and audio stimuli in a multimodal interface. An iterative process led to a design that significantly improves the performance (in terms of effectiveness) of users solving complex data explorations. The results provide information about the convenience of using EMA-Tactons with other auditory displays, and the iterative design process illustrates the challenges of designing multimodal interaction techniques.

Keywords: vibrotactile, external memory aid, overview, visual impairment, high-density sonification

1 Motivation and description of the problem

Data explorations are performed at many different levels of detail, in a continuum that ranges from very general overview information (including size and structure of the data set, nature and meaning of the data), through global description of the relations in the data set (general trends in the data), to more detailed descriptions in particular areas of interest or even to the retrieval of each piece of information in full detail. Every data exploration should start by obtaining overview information, as Shneiderman expresses in his visual information-seeking mantra, "overview first, zoom and filter, then details on demand" [1], which was later extended to non-visual modalities [2].

Previous work by the authors focused on the problem of obtaining overview information non-visually, concentrating in particular on users who are blind or visually impaired (VI), who generally suffer great difficulties retrieving overview information using current accessibility tools. For the common problem of exploring complex tabular numerical data sets (spreadsheets are a typical example), the authors developed TableVis, an interface intended to explore numerical data tables by generating sonifications of the data interactively, with particular focus on obtaining overview informa-

tion at the beginning of the exploration of a data set [3]. In brief, TableVis uses a pen-based tangible input device (a graphics tablet) onto which a data table is scaled to fill the complete active area (Figure 1, left), providing a number of invariants for the user to rely on during the exploration. The first invariant is that the complete data set is always on display, and the tangible borders of the tablet correspond to the boundaries of the data set (Figure 1, right). The second invariant is that all of the data are directly accessible by pointing at a particular location on the tablet, which is a constant location. The third invariant is that the active area on the tablet has a fixed size. These invariants provide enough context information for users to explore data tables in search for overview information at various levels of detail. Using a sonification strategy widely tested in creating auditory graphs [4], information is transformed into sound, mapping each numerical value to a particular pitch of sound within a predefined range, in such a way that the lowest value in the data set corresponds to the lowest pitch and the highest value corresponds to the highest pitch, with all of the intermediate numerical values being mapped proportionally to pitches in that range. By default, a continuum pitch-space approximately ranging 60-2600Hz is used to map all the values in any table. Information can be accessed by listening to one cell at a time (*cells* mode) or by listening to all the cells in a complete row or column (*rows* and *columns* modes) in a single sound event. The latter modes are particularly appropriate to obtain overview information. In them, a complete row or column is sonified by playing each one of the cells in that row or column in so fast an arpeggio that it is perceived as a chord. Thus, a single complex sound (a combination of all the frequencies to which the values in that row or column map) is heard for each row or column, with a particular perceived overall pitch. This sonification technique is called *High-Density Sonification* (HDS). Comparing the relative perceived pitches of adjacent rows and columns is a very fast and effective way of scanning a large table in only a few seconds and obtaining a good overview representation of the whole data table [5].

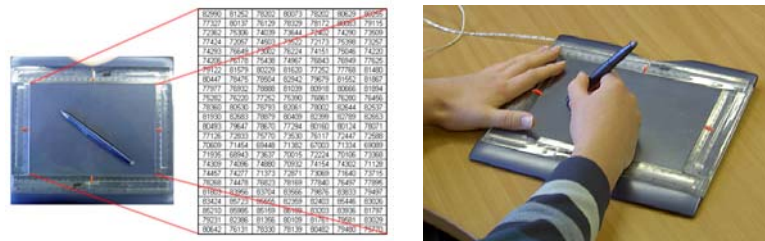


Fig. 1. The data table to be explored is presented on the active area of the tablet, scaled to fill it completely (left). A user explores a data table creating an interactive sonification with the pen, while the left hand feels the boundaries of the data set to provide contextual information (right).

Kwasnik [6] proposed the following components of browsing: orientation, place marking, identification, resolution of anomalies, comparison and transitions. TableVis was designed to provide support for these functional components by maintaining permanently a focus+context metaphor while chunking information with HDS to minimise the number of comparisons that have to be performed. During experimental

evaluation studies conducted with blind and sighted blindfolded participants, it was observed that, under certain conditions, the working memory of some users reached saturation. While the actual circumstances in which this happened will be described in detail later, they involved performing large numbers of comparisons between specific areas in a data table as intermediate steps towards the completion of an exploratory task. Qualitative data from those studies revealed that some form of external memory aid could support performing those intermediate steps, preventing saturation.

This paper introduces EMA-Tactons, vibrotactile external memory aids (EMA's) that are combined with interactive sonification techniques for the exploration of data. EMA's are used to mark interesting areas in a data set where the user may want to go back to. By explicitly marking them, the user's working memory can be freed, preventing saturation of this kind of memory before an exploratory task is completed. An iterative design process is described in detail for the illustrative case of TableVis.

2 Requirements capture and the first design iteration

During experimental evaluations of TableVis, some participants had difficulties to complete certain exploratory tasks that required performing multiple comparisons as intermediate steps, due to working memory saturation. Those tasks involved exploring numerical data tables with 7 rows and 24 columns finding overview information in terms of the meaning of the data in those tables (see [5] for a detailed description of the study). The task was completed exploring the data using HDS, comparing the 7-note chords corresponding to all 24 columns and then choosing the column with the pitch that was perceived to be the highest among the 24. This process required comparing all the chords against each other, and remembering both pitches and spatial locations. These problems arose mainly with data sets in which there were no apparent patterns and where peaks in the data were randomly located, without smooth variations that led towards them. From those observations, the characteristics of the tasks and data sets that lead to such situations were derived:

- Data tables with a moderately large number of rows and/or columns (it was observed that 24 was big enough);
- Data sets that do not contain smooth patterns. In other words, data sets where data are distributed (apparently) randomly;
- Tasks that require obtaining information with an intermediate level of detail.

Those characteristics in task and data set require a user to perform large numbers of comparisons and to remember a lot of intermediate information temporarily. In the studies described above, users had to remember where the columns with the largest numbers were (spatial memory) and what each one of them did sound like (pitch memory). A list of columns candidate to producing the highest overall perceived pitch was constructed by comparing all the columns against each other and adding them to that list or rejecting them. All that temporary information had to be held in the very limited storage capacity of working memory [7]. In situations like those, some kind of EMA could significantly improve the chances to complete the task by preventing working memory saturation. Some of the participants tried to mark the exact locations

of the isolated peaks with their fingers, so that once all the candidate columns were marked they could go back to those positions on the tablet and compare them, choosing the highest one. This technique posed several difficulties. Firstly, marking positions on the tablet with a finger was quite inaccurate. Fingers moved accidentally and references were often lost. Additionally, it was often very difficult to mark three or more positions distributed across the tablet. Rearranging the fingers to mark an additional position often resulted in accidentally losing all the references. Finally, the non-dominant hand could not assist the dominant hand that held the pen by providing good references to known positions on the data set (corners, middle points of sides etc), used in maintaining the focus+context metaphor through proprioception.

2.1 Design

A list of characteristics for EMA marks (using an analogy with marks often created with pencil on printed documents) was derived from the observations above:

- Marks should be easily added and removed;
- They should not be limited in number;
- Each mark must remain in the same position, unless explicitly moved by the user;
- Marks must be easy to find;
- Adding a mark must not alter the information in the data set;
- A mark should not obstruct the access to the information in that position;
- Marking should combine with other techniques and tools for data exploration and analysis available in the interface, to support the process of information seeking.

Using the fingers from one hand clearly does not comply with some of the characteristics in the list. An initial solution that we considered was using tangible physical objects that could be placed on the tablet. One example was to utilise reusable putty-like adhesive material (commercially available products like BluTack, Pritt-tack or others). This design would comply with most of the characteristics in the list, in addition to having many other advantages such as being cheap and disposable. There was, however, an important limitation, as the markers have to be recognised and supported by a computer in order to also comply with the last point in the list. In a realistic exploratory task with TableVis, a user needs to be able to explore complementary views of the same data set (using *rows* and *columns* modes) that, when combined, help to build an understanding of the whole data set. Changing the exploratory modality should present the marks corresponding to the new modality only, and all the marks corresponding to other views of the data would be stored and preserved. In the case of other interfaces for interactive data sonification, EMA's should combine with the functionality available in those interfaces to support information seeking, which in many cases will require that EMA's are recognised and managed by a computer.

2.1.1 Computer-supported EMA's

In TableVis, the auditory channel is used intensively to maximise the transmission of information to the user. Designing EMA's in the form of non-speech sounds, although appropriate in principle, would have increased the amount of auditory information

transmitted through this channel, potentially creating problems of masking and overloading. On the contrary, information is less intensively transmitted through the somatic senses in most interactive sonification interfaces. In the case of TableVis, proprioception and kinesthesia are used to maintain the context of the information that is in focus, but very little information is perceived cutaneously apart from feeling the tangible borders of the exploration area with the non-dominant hand. Wall and Brewster [8] used force-feedback to provide EMA's in similar situations. Incorporating force-feedback devices to TableVis would have meant removing one of the dominant criteria for its design, which was to make use of inexpensive off-the-shelf technology that resulted in affordable systems, easily scaleable and where components could be replaced flexibly. Vibrotactile actuators are much more common devices, which are already used in multiple applications (the most popular of them being vibration in mobile phones), and they can be very small and even wireless. These actuators can generate vibrotactile messages that are easy to control, and that can be perceived subtly on the skin. Vibrotactile stimuli can be used to generate Tactons, which are structured, abstract tactile messages that can communicate complex concepts to users non-visually [9]. Using Tactons as EMA's (thus, EMA-Tactons), the potential to transmit information to users via cutaneous stimulation can transcend a mere binary indication of whether certain piece of data has been marked or not. The information conveyed could be richer, potentially including the type of the annotation (a marked cell, or complete row or column in the case of a data table) and ranking the information according to its importance in a particular search. Ideally, this richness of information would approximate that of simple annotations that sighted users make while exploring a data set presented in the visual medium.

2.2 Implementation

A Tactaid VBW32 transducer (Figure 2, left) was selected to generate the vibrotactile stimuli (www.tactaid.com). The nominal vibration frequency of a Tactaid (where amplitude of frequency is highest) is 250Hz. A Tactaid transducer was mounted laterally on the rear end of the tablet's electronic pen (Figure 2, right), using adhesive tape that ensured hard contact between the transducer and the pen. First tests showed that it was important that this contact was not loose otherwise the external shell of the transducer rattled against the surface of the pen, which could be heard by the user. Lee *et al.* [10] mounted a solenoid axially on their Haptic Pen, to accurately simulate a physical event. EMA-Tactons are abstract information with no physical tangible equivalent. Thus, the lateral mounting was used instead, for ease and reliability. The user does not touch the transducer directly during the exploration; the vibration is transmitted and felt on the surface of the pen while holding it normally (Figure 2, right). The pen+transducer assembly was informally tested to observe the frequency at which the whole combined item vibrated with the highest amplitude, which would provide a good transmission of vibration to the skin without causing movement on the pen that could affect accuracy of pointing. It was observed that the vibration was most noticeable at 270Hz. Therefore, a sine wave with frequency 270Hz was chosen to generate vibrotactile stimuli. Fine-tuning the intensity of the vibration was left to the discretion of the user, for reasons of comfort.

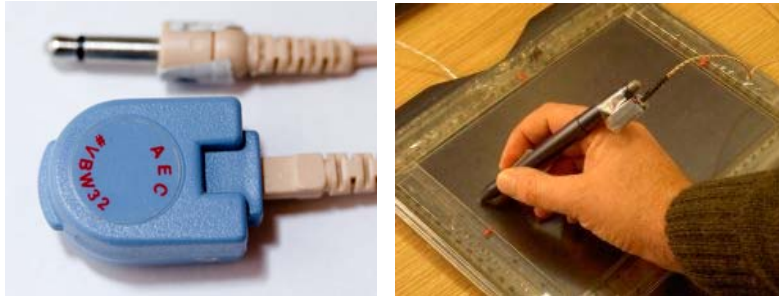


Fig. 2. Tactaid VBW32 transducer (left). Tactaid mounted on the pen of the tablet (right).

During the exploration of a data table, a user could mark any position that the pen was pointing at by pressing a button. Depending on the selected navigation mode, only the cell or the complete row or column being pointed at would be marked. The vibration would be felt as long as the pen remained on the selected cell, row or column. The vibration would be felt every time the pen re-entered a selected area during the exploration. An EMA-Tacton could be removed by pressing the same button while the pen was on the marked area. Adding or removing a mark was confirmed by a different, easily distinguishable, percussion sound. Data could be marked in different navigation modes, and switching between modes would not delete any marks, but make them selectively accessible (e.g. in *columns* mode, only marks affecting complete columns would be accessible). This model could easily be extended to selecting a cell by intersection of a selected row and a selected column. In this study, EMA-Tactons convey only binary information, i.e. whether data has been marked or not.

2.4 Pilot evaluation

Five participants took part in the pilot evaluation of EMA-Tactons in TableVis, with the implementation described in the previous section. All participants were visually impaired and used screen reading software to access information in computers. The structure of the evaluation session was as follows: introduction to the interface and to the concepts involved in exploring data through interactive sonification; set several tasks for data exploration and observe the participant perform the task, encouraging think-aloud explorations; finish with a semi-structured interview. Only the *columns* navigation mode was used, where complete columns were compared by using HDS in TableVis.

The data sets used for the evaluation were tables with 7 rows and 24 columns, the same size as some of the tables used in earlier evaluations of TableVis and with which working memory saturation problems had been observed [5]. When navigating by columns, one such table is sonified into an array of 24 chords (with 7 piano notes each) arranged side-by-side horizontally on the tablet. Participants could access the chords by moving the electronic pen from one column to another. Tapping repeatedly on the same column would replay the chord representing that column. The data presented in the tables were such that each set of 7 values in each column had a different

arithmetic mean (all means were approximately equidistant from the nearest higher and lower ones) and all the 24 columns had approximately the same standard deviation. In each data table, columns were placed in random order, so that means did not follow any discernible progression. The task set was to find the column with the highest perceived overall pitch.

For the first two participants, the task of finding the highest single column happened to be simple enough to allow them to complete it without the need for EMA-Tactons. While both participants agreed that EMA-Tactons offered help to complete the task, they completed it making little or no use of them. One of the participants explained that a first scan of all the columns showed where the highest few columns were. Then, this participant claimed to be able to remember one of those pitches and compare it against every other pitch until the highest was found. Only the pitch and position of the current highest sound had to be remembered, and after finding a higher one the previous pitch and position were replaced with the new ones, never overloading working memory. This participant was consistently correct. The authors concluded that, while the task was probably challenging enough for some users, it did not always saturate working memory and cause interaction problems. It is interesting to observe that Wall and Brewster reached a similar conclusion in their study [8].

To further challenge working memory, the task set to the remaining three participants required that the 5 columns with the highest pitched sounds were selected, instead of the single absolute highest. The number of comparisons was much larger, as was the number of intermediate results to be temporarily remembered (positions, pitches associated to those positions and number of positions selected). The procedure to be followed was to explore the data with the pen and when one of those five sounds was identified to select it by pressing a push-button, while the pen was still on the position of that column. Columns could be selected and deselected by pressing the same button. Two experimental conditions were defined: i) selecting a column would add an EMA-Tacton to that location, which would be felt when going over it with the pen; ii) selecting a column would not add an EMA-Tacton, although the selection/deselection confirmation sounds would still be heard (which would alert about trying to select the same column twice, as the user would know from the confirmation sound that a mark had been removed from an already selected column). Thus, the only difference between both conditions was that in the second one the user would not be able to find marked areas easily, having to rely on his/her memory more heavily. In the first condition, the user would not have to consciously decide to use the EMA-Tactons, as they would simply appear when columns were selected. Then, participants would be able to keep track of the number of columns that had already been selected (by counting the number of vibrating columns) and they would also be able to check more easily if each one of the selected columns did actually belong to the group of the 5 columns with the highest pitch, resulting in a better performance at solving the task.

The new task was observed to overload the participants' working memory very quickly during the pilot study, and participants reported that it was easier to complete the task when the EMA-Tactons were available. In addition to providing qualitative feedback, two of the participants completed the exploration of the 12 data sets (which were described earlier). The 12 data sets were presented in random order in each condition, and the order of the conditions was counterbalanced. Participants had up to 120 seconds to perform an exploration and select the columns. The subjective work-

load experience was assessed after each condition using NASA-TLX [11] questionnaires, followed by a semi-structured interview. The quantitative results from this experimental test are presented in the next section, together with quantitative data from the same experiment that was run with a bigger group of sighted blindfolded participants, and both are compared. From a qualitative point of view, it was concluded from the pilot study that the EMA-Tactons had good acceptance once the task was challenging enough.

In the studies presented here, data ordering was randomised to maximise the chances of users' working memory getting saturated. This was a major difference to the setup used in previous evaluations of TableVis, where data always followed more or less smoothly-changing patterns. Exploring random data sets using HDS has limitations. This technique is appropriate for obtaining overview information (general description of trends and patterns) and for finding areas in the data set with high or low values when data change smoothly (leading towards them), or if extremes are obviously high or low. In the case of very random data, like in this study, HDS can help pick and group areas by ranges of values (as in the task where the five highest sounds have to be identified), but there is no guarantee that the absolute highest pitch can be singled out reliably, or that the sixth highest sound will be thought to have a lower pitch than the fifth highest. To compensate for any confounding effects introduced by this limitation in the data discrimination technique used, the final version of the design is also evaluated using single tones instead of chords, which, although further from the scenario being replicated, provides an unequivocal criterion to judge the correctness of the answers. A parallel line of research from the authors is investigating how relative pitch is perceived in complex dissonant chords

3 Experimental evaluation of the first design iteration

A group of 8 sighted persons was recruited to take part in the experiment designed during the pilot study (due to our limited supply of visually-impaired people we often have to test with sighted blindfolded participants. The approach we take is to scope out the problem with our target users and then test sighted participants to gain more data. The performance of the two groups is commonly very similar). The setup, data sets and procedure were exactly the same as those described in the previous section and used in the pilot study. To assess quantitatively the effectiveness of the EMA-Tactons, the correctness of the results at solving the task was divided in two parts, each providing a metric of effectiveness. A third metric was obtained considering the task as a whole:

- *Sub-task 1 (number of selections)*. Correctness in selecting exactly 5 positions on the tablet. 100% correct obtained only when exactly 5 positions are selected. This metric is calculated with the formula:

$$\text{Correctness sub-task 1 (\%)} = 100 \cdot (1 - |S_s - 5| / 5) . \quad (1)$$

- *Sub-task 2 (pitch of the selected sounds)*. Correctness in having selected the positions with the highest pitch sounds. 100% correct obtained only if all the positions

selected correspond to the group of the same number of sounds with the highest pitch. For example, if 7 positions are selected and they are the 7 sounds with the highest pitch in the whole set of 24 sounds then sub-task 2 is 100% correct.

$$\text{Correctness sub-task 2 (\%)} = 100 \cdot (S_c / S_s) . \quad (2)$$

- *Overall task (Combination of sub-tasks 1 and 2)*. Metric to assess the correctness of the overall task, as the product of both sub-tasks. 100% correctness is only obtained if exactly 5 positions were selected *and* they correspond to the 5 highest pitch sounds in the set. This metric is calculated with the following formula:

$$\text{Overall correctness (\%)} = 100 \cdot (1 - |S_s - 5| / 5) \cdot S_c / S_s . \quad (3)$$

In all formulae, S_s is the number of sounds selected and S_c is the number of sounds from the selection that are in the group of the S_s sounds with the highest pitch.

Results from the evaluation with sighted blindfolded participants (Figure 3, left) show that the effect of using EMA-Tactons is small, differences not being significant for any of the metrics, according to two-tailed t-tests (paired two sample for means): sub-task 1 ($T_7=1.609$; $P=0.152$); sub-task 2 ($T_7=-0.378$; $P=0.717$); overall task ($T_7=1.27$; $P=0.245$). The results by the two VI participants (Figure 3, centre and right) are approximately within the ranges obtained in the experiment with the group of sighted blindfolded participants (although performance in sub-task 2 was slightly lower for the first VI participant).

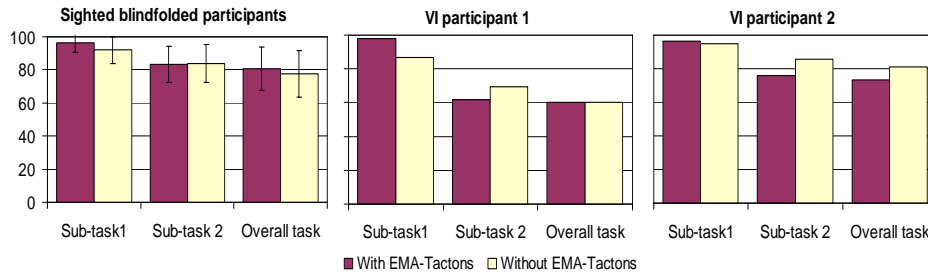


Fig. 3. Left: results (percentage of task completion) from experimental evaluation of the first design iteration (unsynchronised sound and vibration). Centre and right: results from the pilot study, by participants with visual impairments. Error-bars represent 95% confidence interval.

The hypothesis that effectiveness would be higher with EMA-Tactons could not be proved, according to these results. Among the qualitative feedback provided by the participants, many of them agreed in saying that the vibrotactile information could both help and interfere in the process of solving the task. EMA-Tactons were helpful to keep count of how many locations had already been selected (thus the slight improvements in sub-task 1). Several participants, however, reported that sometimes the vibration on the pen could be distracting, stating that it could even get in the way when the user was trying to listen to the sounds. Others said that they found EMA-Tactons helpful in general but that it was very difficult to get information simultaneously from sound and from vibration and that they concentrated on the source of information they needed at each time, ignoring the other. One participant also reported that vibration and sound sometimes seemed to be two unrelated events.

An explanation for these results and comments can be found in the differences between endogenous and exogenous spatial attention, and in aspects of crossmodal spatial attention. When a participant wanted to count how many sounds were already selected, attention was endogenously (voluntarily) diverted to the hand holding the pen, monitoring for vibrotactile cues. If, on the contrary the user was trying to listen to the sounds and unexpectedly a vibration was produced in the pen, attention got diverted to the hand exogenously (involuntarily, stimulus driven), thus potentially interfering with the listening. Multiple sensory inputs are processed selectively, and some stimuli get processed more thoroughly than others, which can be ignored more easily. There are very complex interactions between crossmodal attention and multisensory integration and much research is being carried out in that field that will inform the designers of multimodal interfaces (see chapters 8, 9 and 11 in Spence and Driver [12]).

4 Second design iteration and experimental evaluations

Results from the first design iteration suggested that presenting a vibrotactile cue simultaneously with the onset of the sound did not bind them enough to create a single multimodal event, where users could perceive both sensory cues to be related to a common event. A conscious binding of both events was required, what could increase the subjective overall mental workload despite the support that was being provided to working memory, which should reduce it, resulting in an overall increase in this metric of the subjective experience (see Figure 6, later). To improve the integration between audio and vibrotactile information so that they were more easily identified as being generated at a common multimodal event, the EMA-Tactons were redesigned to be synchronised with the audio signal, not only on their onset, but also in their decay and end. In the first design, the vibrotactile stimulus was felt for as long as the pen remained on a position that had been marked, well beyond the duration of the sound. In the second design iteration, the vibration, instead of being produced as long as the pen remained in a marked area, had similar duration (200ms) and envelope as the sound. A sharp attack was followed by a 120ms sustain period (so that the presence of the vibration was clearly perceived), and then the amplitude of the vibration decayed during the last 80ms. As an extension of sound-grouping principles from auditory scene analysis, which suggest that sounds that are likely to have originated in the same event in the physical world are grouped together [13], the authors hypothesised that two stimuli in different sensory modalities that were synchronised and equally shaped could be more easily perceived as having been generated at the same event (as when in the physical world some mechanical action generates decaying vibration and sound that are perfectly synchronised throughout).

Having made this change, the same experimental evaluation setup was conducted with another 12 sighted blindfolded participants. As it was effectiveness and not efficiency the aspect that was being targeted with this study, up to 180 seconds were allowed in this case to explore each data set in order to permit extended, thorough data explorations. The results are shown in Figure 4. Performance in sub-task 1 (accuracy in the number of selections) was statistically significantly better with EMA-Tactons ($T_{11}=3.008$; $P=0.012$). The improvement in performance for sub-task 2 (selecting the

highest-pitched sounds) was still not significant ($T_{11}=1.379$; $P=0.195$). The performance with EMA-Tactons for the overall task showed a statistically significant improvement ($T_{11}=2.89$; $P=0.015$).

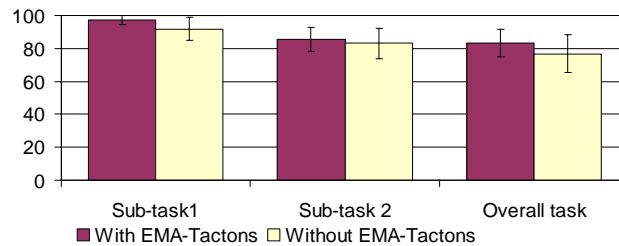


Fig. 4. Results (percentage of task completion) from first experimental evaluation (chords) with the second design iteration (synchronised sound and vibration). Error-bars represent 95% confidence interval.

The possible existence of two confounding factors was identified in this study. Many participants started their exploration taking a quick overview of all the data, looking for the areas on the tablet with the highest-pitched chords, and then they tried to concentrate their search in those areas. In every data set, the five highest-pitched chords were in a similar range of frequencies. It was therefore possible that there was a learning process during which target frequencies were learnt, reducing the initial scan to identifying those pitches only and ignoring any other sounds from the beginning. Relying on pitch memory in this way would reduce the number of comparisons that had to be performed, increasing performance in both conditions and reducing the potential benefit EMA-Tactons could offer. Another possible confounding factor was comparing the perceived overall pitches of two similar chords, as it has been discussed in 2.4, earlier. These two possible factors were addressed by creating new data sets in which sounds were single piano notes instead of chords. Each data set was a succession of 24 notes in a chromatic scale (one semitone distance between any two consecutive notes), arranged in random order. It was expected that any possible ambiguity in the judgement of relative pitches would disappear for the majority of participants. To prevent participants from remembering target pitches between data sets, each one of the 12 data sets would cover a different range of 23 consecutive semitones. Since data sets were presented in random order, it was not possible to predict what the highest-pitched sounds in a new data set would be like before every position had been examined, thus preserving the need to perform a full set of comparisons.

Having created new data sets in the way that has just been described, a new group of 12 participants was recruited to test again the effect of EMA-Tactons in their second design iteration (with audio and vibrotactile stimuli synchronised). In this case (Figure 5), the improvement in performance for sub-task 1 (number of selections) was not significant ($T_{11}=1.892$; $P=0.085$). In contrast, the performance in sub-task 2 (selecting highest pitches) improved significantly with EMA-Tactons ($T_{11}=2.216$; $P=0.049$). The performance considering the overall task saw, again, a significant improvement when EMA-Tactons were used ($T_{11}=2.490$; $P=0.030$).

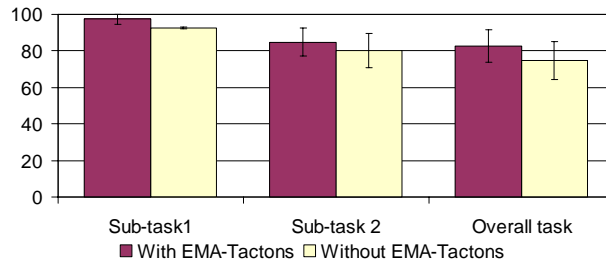


Fig. 5. Results (percentage of task completion) from the second experimental evaluation (single tones) with the second design iteration (synchronised sound and vibration). Error-bars: 95% confidence interval.

The increase in significance for sub-task 2 (selecting the highest pitches) could well be due to having removed both confounding factors. In particular, it is believed that participants could have been obtaining benefit from pitch memory in the previous setup, hence facing less working memory saturation problems and obtaining less benefit from EMA-Tactons. Other results presented in the next section support this idea. The reasons for the loss of significance for sub-task 1 need to be investigated further, but there is no reason to think that it is due to using tones instead of chords.

5 Other Results

In all three experiments conducted, the time to complete the task (an aspect not targeted by this research work) was in average longer when EMA-Tactons were used than when they were not. This was also true for the two visually-impaired participants, who required on average 90.9 and 75.9 seconds respectively to complete the task with EMA-Tactons, while it took only 67.3 and 57.8 seconds respectively to complete it without them. Based on qualitative feedback, this is attributed to the fact that with EMA-Tactons, participants could be more thorough in their search without reaching saturation of working memory, resulting in more focused, and thus longer, data explorations. The difference in time to complete task was only significant in the second design (synchronised sound and vibration) with chords ($T_{11}=2.789$; $P=0.018$). It is interesting to observe that, comparing the conditions without EMA-Tactons from both experiments in the second design (synchronised sound and vibration), the average time to complete the task was longer in the second experiment (single tones) than in the first one (chords). This supports the hypothesis that pitch memory was being used in the first case to simplify the task.

The overall subjective workload (derived from the NASA-TLX ratings) was perceived to be significantly lower when EMA-Tactons in their second design iteration (synchronised sound and vibration) were used ($T_{11}=-2.970$; $P=0.012$ for chords and $T_{11}=-3.546$; $P=0.005$ for single notes). Again, the difference was bigger in the last experiment, when saturation of working memory was higher and with more room for improvement. With the first design of EMA-Tactons, the difference was not signifi-

cant ($T_7=0.558$; $P=0.594$). The task was exactly the same in both cases, as it was the amount of information provided by both prototypes of EMA-Tactons. Therefore, the fact that the overall subjective workload using EMA-Tactons was significantly lower with the second design while showing no significant difference with the first design must be attributed to the actual design, suggesting that synchronising sound and vibration to integrate sensory channels was the correct approach.

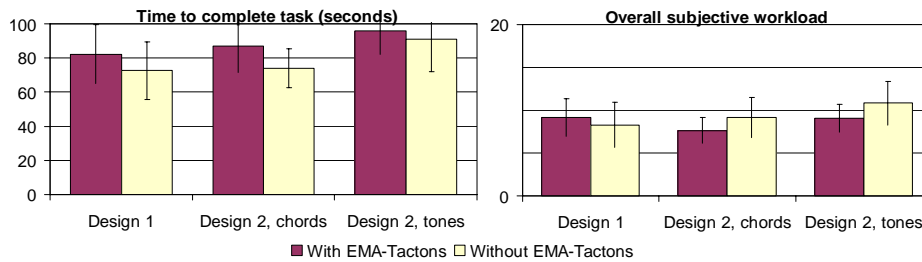


Fig. 6. Time to complete task, in seconds (left) and overall subjective workload, derived from NASA-TLX ratings (right). Error-bars represent 95% confidence interval.

6 Conclusions

This paper has introduced EMA-Tactons as a way of enhancing interactive data sonification interfaces with vibrotactile external memory aids, to tackle common problems of working memory saturation in non-visual environments. An iterative design process produced two prototype designs that were tested quantitatively and qualitatively. This process showed that designing multimodal interfaces for good integration of sensory channels is difficult and complex. Subtle changes can make a big difference in perceiving a single multimodal event instead of unrelated events in different sensory channels. EMA-Tactons were tested in TableVis, an interactive data sonification interface designed to explore tabular numerical data non-visually. In the first design iteration, enhancing the interaction with external memory aids in the form of vibrotactile stimuli to avoid saturation of the users' working memory did not produce any significant improvements in the performance in terms of accuracy of retrieved information. Careful redesign of the vibrotactile stimuli following principles of ecological perception produced a better integration of multisensory information, which led to significant improvements in performance. Consistently saturating the participants' working memory in order to test the prototypes also proved to be difficult. Even with very demanding tasks, resourceful participants were believed to have used pitch memory to simplify those tasks, so that the need for any external memory aids was reduced at the expense of very small loss in accuracy. This illustrates the difficulty of replicating scenarios where working memory saturation problems had been observed, and which would produce the same effect on the whole population of participants in a study. Despite the human resourcefulness observed, the second prototype of EMA-Tactons produced significant improvements in effective task completion. In future design iterations, using more than one bit of information from the EMA-Tactons can

permit adding richer annotations. Additionally, the combination of rich annotations in different exploratory modalities (*rows*, *columns* and *cells*) has the potential to offer support for complex exploratory tasks that today can only be done visually.

Acknowledgements

We want to acknowledge the contribution of all the participants in this study, and in particular the committed support received from the RNC in Hereford. This research is funded by EPSRC grant GR/S86150/01.

References

1. Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. IEEE Symposium on Visual Languages. Boulder, CO, USA: IEEE Comp. Soc. Press, (1996) 336-343
2. Zhao, H., Plaisant, C., Shneiderman, B., and Duraiswami, R.: Sonification of Geo-Referenced Data for Auditory Information Seeking: Design Principle and Pilot Study. Int Conf Auditory Display. Sydney, Australia (2004)
3. Kildal, J. and Brewster, S.: Exploratory Strategies and Procedures to Obtain Non-Visual Overviews Using Tablevis. Int J Disabil Human Dev, Vol. 5 (3), (2006) 285-294
4. Flowers, J.H.: Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New Directions. Int Symposium Auditory Graphs. Int Conf Auditory Display. Limerick, Ireland (2005)
5. Kildal, J. and Brewster, S.: Providing a Size-Independent Overview of Non-Visual Tables. Int Conf Auditory Display. Queen Mary, University of London (2006) 8-15
6. Kwasnik, B.H.: A Descriptive Study of the Functional Components of Browsing. Ifip Tc2/Wg2.7 Working Conference on Engineering for Human-Computer Interaction. North-Holland (1992) 191-203.
7. Miller, G.A.: The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review, Vol. 63 (1956) 81-97.
8. Wall, S. and Brewster, S.: Providing External Memory Aids in Haptic Visualisations for Blind Computer Users. Int J Disabil Human Dev, Vol. 4(3) (2006) 285-294
9. Brewster, S. and Brown, L.: Tactons: Structured Tactile Messages for Non-Visual Information Display. Australasian User Interface Conf. Dunedin, New Zealand: Australian Comp. Soc., (2004) 15 - 23
10. Lee, J.C., Dietz, P.H., Leigh, D., Yerazunis, W.S., and Hudson, S.E.: Haptic Pen: A Tactile Feedback Stylus for Touch Screens. Annual ACM Symposium on User Interface Software and Technology. ACM Press. (2004) 291-294.
11. Hart, S. and Wickens, C.: Workload Assessment and Prediction. Manprint, an Approach to Systems Integration. Booher, H.R., Editor. Van Nostrand Reinhold. (1990) 257-296.
12. Spence, C. and Driver, J.: Crossmodal Space and Crossmodal Attention. Oxford University Press (2004)
13. Bregman, A.: Auditory Scene Analysis: The Perceptual Organization of Sound. The MIT Press (1994)