New Interaction Techniques 2003

Grigori Evreinov
University of Tampere
Department of Computer and Information Sciences
TAUCHI Unit for Computer-Human Interaction
Pinninkatu 53B
FIN-33014 TAMPERE, FINLAND
http://www.cs.uta.fi/~grse/NIT_2003/NIT2003.htm

PREFACE

The course "New Interaction Techniques" [1] was organized by Prof. Roope Raisamo in the spring of 2000 at the Department of Computer and Information Sciences in the University of Tampere. The course is one of the advanced courses on human-computer interaction supported by TAUCHI Unit for Computer-Human Interaction.

The main goal of the course is deep analysis of novel technological achievements for augmented communication as well as user behavior in different situations, including extreme ones. Based on advanced engineering, software design and presentation of strategically-important scientific directions, the topics of lectures should stimulate creative capabilities of the students for development of new approaches to the solution of current and future tasks in computer science.

Obviously, the most of students have different background. Some of students have own experience in their use of novel devices; some students, probably, could propose own ideas to improve existing interaction techniques. Others could wish to carry out a usability exploration of a new method or to do a comparative theoretical study, for instance, to build a conceptual model of interaction based on hypothetical communicative environment. Any self-expression should be admitted and help should be done to support development of the potential innovations and innovators.

The course was designed so that in the beginning of the spring term through a series of lectures (24 hrs) on the key problems in computer-human interaction the students could be oriented in a huge information flow of the concepts, decisions, approaches and vital-important tasks. In particular, this course included the next topics:

- **Philosophy of Human-Computer Interaction**
  - Spaces for interaction (time, space, modality)
  - Semantics, symbolism, metaphors
  - Sound & voice
  - User & devices: merging and interaction of virtual reality and natural activity
  - Temporal dimension
  - Audio-haptic manipulations
- **Engineering basics for computer interaction**
  - Device capabilities and their future evolution
  - Trends in component technology
  - Display technology
  - Input devices
  - Communication with PC
    - Joystick port
    - Parallel port
    - Serial port
    - USB
- **Text entry as the model for pointing & selection**
  - Introduction to the problem and samples
- **Text input: Techniques and research tools**
  - Visiting lecturer: Poika Isokoski
- **Interactive Surfaces and Perceived Textures**
  - Human tactile sense
  - Tactile matrixes, displays and actuators
  - Textures and tactile symbols
  - Sound Pen
- **Data Sonification**
  - Auditory direct manipulation
  - Sound Simulation
  - Sound mapping and evaluation
- **Wearable Computing**
  - Head-Mounted Displays
  - Input Techniques
  - Eye movements
  - PenComputing
  - Special techniques for extreme conditions: medical, military, assistive and other applications

Whereas only theoretical considerations is a passive activation for intellectual activity and requires both engineering basics and a high self-concentration on difficult materials we used as much as possible presentations of working prototypes of devices, software or video demonstrations. To present eye and haptic and sonification interaction techniques special visits were organized by TAUCHI Multimodal Interaction Research

UNIVERSITY OF TAMPERE
DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES
SERIES OF PUBLICATIONS B
B-2003-5, June 2003

New Interaction Techniques 2003

Proceedings

Grigori Evreinov (ed.)

UNIVERSITY OF TAMPERE
DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES
P.O. BOX 607
FIN-33014 UNIVERSITY OF TAMPERE, FINLAND

ISBN 951-44-5188-0
ISBN 1457-2079

New Interaction Techniques 2003
Spring 2003

Index

Interaction Techniques papers

Technotes

Group and Gaze-based Interaction Group (6 hrs).

The practical phase of the course included individual and group activity in designing or/and usability exploration of the prototype (hardware or software) one of the novel interaction techniques, that is, carrying out the pilot project (about 8 weeks). That part was more difficult both for the students and instructor due to very short time, different directions of the research and individual skills of the students. In any case, doing the project the students should receive some experience how design the experiment, how organize usability evaluation of some technique or device, how to collect and process data, and, finally, how to present results of research project via a scientific report. After difficult choice, 11 projects were selected for investigation. Before start the project development all students wrote a research plan and did a special presentation. The goal of this stage was to clarify the tasks for researchers and to involve other students in active discussion around the proposed topics (similar to brainstorm activity). Presentations took 12 hrs.

To pass the course the students should write a research paper formatted according ACM UIST template [2] or earlier version Standard UIST'96 Conference format. The goal was to draw attention of the students to the general requirements and errors during scientific writing and presentation of own results. There were several categories of the papers in the course: short paper (2 pages), tech notes (4-6 pages) and full paper (6-8 or 10 pages). The papers have been grouped in this publication accordingly.

In total, 15 students participated in the lectures. Of them, 11 finished the course. Several papers were submitted to relevant conferences or recommended for sending after revision.

ACKNOWLEDGMENTS

We would like to thank Prof. Roope Raisamo for organizing this vital course, support and valuable advices, the researchers in the TAUCHI Unit who voluntarily help to present new techniques, visiting lecturer Poika Isokoski for actual topic and discussion. We acknowledge all students who were not daunted by the practical difficulties and successfully finished the projects and course.

REFERENCES

1. New Interaction Techniques course hope page available at: http://www.cs.uta.fi/~grse/NIT_2003/NIT2003.htm

2. ACM UIST Conference format: http://www.acm.org/sigs/pubs/proceed/pubform.doc.

Evaluating Non-Visual Feedback Cues for Touch Input Device

Selina Sharmin
New Interaction Techniques 2003
Department of Computer and Information Sciences
FIN-33014 University of Tampere, Finland
Selina.Sharmin@uta.fi

ABSTRACT
Non-visual tactile feedback improves user experience by providing user a wider area of the target and faster access to the same. Development of haptic communication and appearance of input device with tactile feedback provide scope to systematically investigate how people perceive the world indirectly through the use of various intermediate objects. Non-visual tactile feedback can be applicable in blind manipulation as well as in the situation when the vision is occupied in some other tasks. The goal of this study was to carry out a research to evaluate non-visual feedback cues, tactile and sound, during navigation in a Maze. Results of the study can be used for the development of novel pen-based input techniques, haptic interfaces and applications for people with special needs.

KEYWORDS: Haptics, non-visual feedback, tactile feedback, texture, sound feedback, force feedback.

INTRODUCTION

Touch or texture plays an important role in displaying images and forming subjective impression about natural objects and simulated surfaces. Tactile is related to the cutaneous sense but more specifically the sensation of pressure rather than temperature or pain [6]. Non-visual tactile feedback improves user experience by providing user a wider area of the target and faster access to the same. Tactile feedback is the quickest one for a target selection task comparing to different sensory feedback conditions involving normal, auditory, color, and combined [1]. The haptic technology or force feedback allows users to feel their interfaces. In this case we will be able to use our powerful sense of touch as an alternative mechanism to send and receive information in computer interfaces. In addition to visual and audio feedback, touch gives users an improved perceptual experience.

Development of haptic communication and appearance of input device with tactile feedback provide scope to systematically investigate how people perceive the world indirectly through the use of various intermediate objects such as probes or media and finger coverings and so on. Texture may be used efficiently as tactile output of textual and graphic information for people who cannot see well. Besides, blind manipulations in different situations, when vision is occupied by another task, or visual field is hidden, or the user eyes cannot leave a primary focus point, non-visual tactile feedback can be applicable.

Auditory stimuli are used moderately in human-computer interfaces, usually to signal an error or the completion of an operation. Such stimuli are simple to include since speakers are built-in on present-day systems. Graver describes in [4] a complete GUI (Graphical User Interface), a modification of the *Macintosh's* Finder, using auditory feedback to inform the user of many details of the system, such as file size or the status of the open and close operations. Numerous other examples exist in which auditory stimuli have been exploited as ancillary cues in human-computer systems [3]. Non-visual feedback is more appropriate when it needs to work in the dark.

This study evaluates non-visual such as sound and tactile feedback cues for a touch input device. An experiment was performed here aimed to measure human performance with these feedbacks. Method of the experiment with details of participants, apparatus and procedure is stated in the following section. The results and discussion of the study is described next and the last section concludes the discussion.

METHOD

Subjects

Eight volunteers from the staff and students at the University of Tampere took part in the test. All had a normal sight and tactile sensitivity but different visual and hearing acuity. None of them used hearing aid. All used computers on a daily basis, reporting four to nine hours of usage per day. Four subjects were female and four were

New Interaction Techniques '03, Spring, 2003, University of Tampere.

male. The mean age of participants was 25 years. All of them were novice for the system.

Apparatus

The experiment was carried out using a tactile pen over a maze on the ELO LCD touch screen. The software was created in Microsoft Visual Basic 6.0 under Windows 2000 that presents the maze on the screen with timer and error counter. A special frame served as the tool panel for editing MIDI parameters. A data collection system stored all the tracks of the pen from the start position to the end of the maze passing. It also stored time required to complete each trial and the number of errors occurred. The data were stored in input files for subsequent analysis. The experiment has taken place in the usability laboratory of TAUCHI Unit, the University of Tampere. The experimental setup and instruments used are described in the following subsections.

Tactile pen. A tactile pen was used to get tactile feedback during the experiment. It is a pen prototype, which currently uses of iFeel MouseMan electronics providing tactile-feedback signals through USB interface. A vibration mechanism consists of miniature DC motor with stopped rotor. Both the general view and some design features inside of the pen are shown in Figure 1. A partial-rotation motor is used inside the pen with a spindle that can only turn a few degrees (shaking). It has capability of very fast response. The motor gets an appropriate pattern of current's pulses from the mouse with required frequency, phase and other parameters.

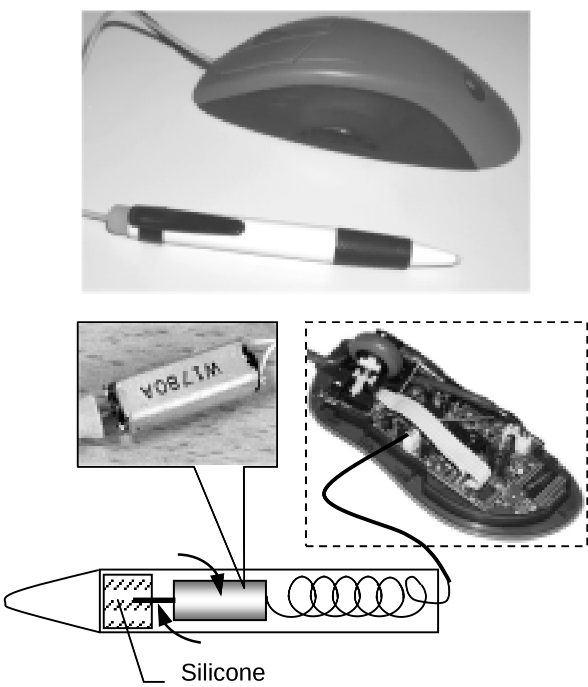


Figure 1: The general view and design features of the Tactile pen

There are only two commands to start and stop the vibration in the device. A "vibrate" command starts the vibration and an "abort" command will stop any ongoing vibrate command before it completes the running mode. However, Immersion Studio 4.0.3 graphical environment allows to edit TouchSense effects adjusting physical parameters, feeling sensations, and then saving them as "feel resources" which can be automatically loaded and executed by application through Immersion File Resources (IFR) [5].

The IFR patterns provide the pen vibrating with the specified density (0-255) and delay between pulses (0-255ms) for the specified count of pulses (1-255). It is possible to make delay longer than 255ms or more than 255 pulses by using multiple vibrate commands.

Figure 2: Experimental setup

ELO touch screen and other equipment. 15" LCD ELO Touch screen was used to detect position of tactile Pen. ASUS A7V133 was equipped with AMD Duron Processor, VIA AC'97 Audio Controller (WDM) and 256MB RAM under Windows 2000. The special driver was installed that supports Immersion's TouchSense. Figure 2 represents the touch screen used during the experiment.

Maze. The experiment screen consists of a maze as shown in Figure 3. The maze was 580 pixels high and 790 pixels wide.

There are two special areas of the maze, the starting position and the end position shown as light dark blocks in Figure 3. Earcons and feedback were added to these areas for the user to understand the entrance in to the maze as well as the completion of the trial.

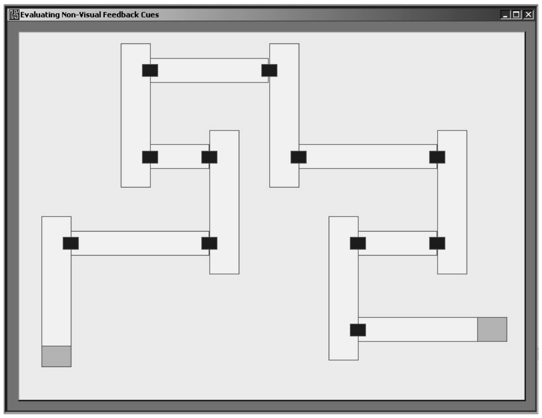


Figure 3: Maze arrangement

If the pen crosses the starting position there is an earcon indicating that the system is ready to receive all the data. Later, if the pen reached the end position of the maze, there is another sound feedback indicating that the trial has completed. The width of the path in the maze is divided in to three fields or zones as shown in Figure 4.

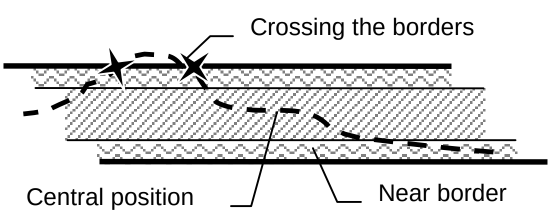


Figure 4: Path width and the fields of feedback cues

The signals for the central position of the path are different from those of the positions near the borderlines for both in sounds and tactile feedback. Crossing the borderlines is also marked by different feedback for both types. It is possible to rearrange the combination of different feedbacks for different fields of the path. There is no feedback out side the maze.

Sound feedback cues. Sonification of Tactile Pen position within the maze was based on conventional MIDI synthesis. To control by MIDI, a component of the library FASTlib.NET was used in this study. Balance was used to redistribute sound signal along X-axis like distance from start point. Frequency deviation along of Y-axis was provided by musical interval of Banjo channel consisted of 32 notes from F#6 (1480.1 Hz) in the top position up to A#3 (233.08 Hz) in the bottom position. In this study we used only a simplest mode of sonification any collisions with borders – through Agogo channel, D3 (146.87 Hz). Uninterrupted sonification between different segments of the maze supported by the special passes, please see black spots in Figure 3.

Tactile feedback cues. It has been used 1.4 magnitudes over the maximum magnitude for this type of the motor. Symmetrical offset was used to adjust the vibration. The square waveform of the current's pulses was selected as well. Three different patterns with frequency, magnitude, offset and other parameters determined in preliminary tests were saved as .ifr files. These vibration patterns were used to sign different fields of the maze path - central part, near the border and crossing the border. Envelopes of designed vibro-tactile patterns are shown in Figure 5.

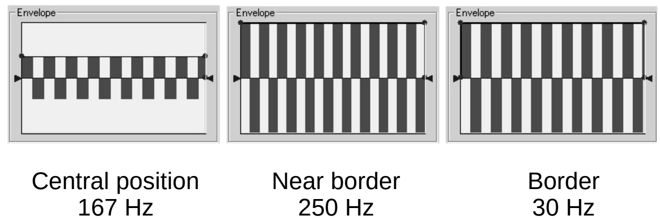


Figure: 5 Envelopes of vibro-tactile feedback cues

Procedure

The subjects were instructed to go through the maze using the tactile pen for both non-visual modes used with sound and tactile feedback. 10 trials for each mode were performed. The maze was hidden during the experiment and the subject had to move through the maze on the basis of the given feedback. During the movement along the central field of the path, near borderline or during the crosses of the border, the subject should immediately get a feedback according the test protocol described in previous section. If the pen goes outside the maze, it will not receive any feedback. Thus the subject should understand that s/he is out of the maze. Error was counted for each case when crossing the border occurred from inside position to outside. Simultaneously the time was counted for each subject required to complete each trial.

There was a 2 minutes demonstration about the system and the task for the subjects and 1-minute training. The subjects were free to choose the feedback s/he wanted to start with. The experiment took about an hour for each subject.

All the cursor activities, which are the movement of the tactile pen along the maze, were saved. These data provided the traces of all subjects' behavior during the task and thus became useful at the stage of data analysis. The 'Data' frame for editing sound feedback, switching the mode, entering special information, file's operations and other service is shown in Figure 6.

In particular, experimenter can change MIDI Patch (musical instrument), dynamical range of the frequency deviation along of Y-axis and volume. Software buttons are used in file's operations, check boxes – to select the mode of feedback cues.

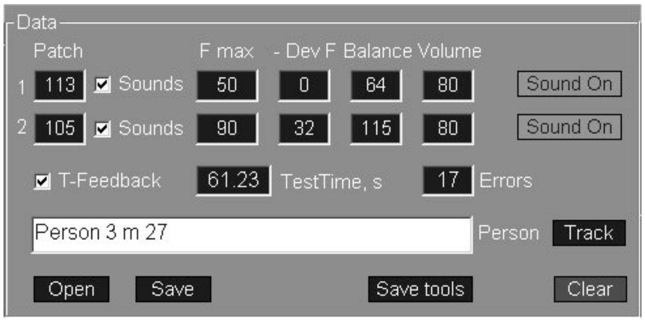


Figure 6: Service frame for editing and selecting the mode of feedback cues

It is possible to use both the feedbacks at the same time. “Clear” button is used to clear all data arrays; “Track” button can display a sequence of manipulations being performed during the test.

RESULTS AND DISCUSSION

Subjects were asked to choose a feedback either sound or tactile at the beginning of the experiment. Most of them started with sound feedback. It was told to complete the maze 10 times for each feedback. As there is learning affect, the users were asked to switch to the other feedback after four or five trials randomly. The following subsections present the findings of the observations.

Performance Analysis

Figure 7 below shows the mean time required for each subject during each mode of feedback. The line diagram indicated that tactile feedback took less time than sound feedback for all the participants. May be it was observed due to for catching or noticing feedbacks by different subjects. With sound feedback, the subject has to hear first and then move the pen to the right position accordingly. Tactile sense provides adequate feedback directly when hand moves Tactile pen. Consequently, a tactile mode took less time. Moreover, for slow movement tactile found much better than sound feedback cues. If the user moves slowly, with sound feedback they sometimes lost their position in the maze and get no feedback, which is less likely to occur with tactile feedback.

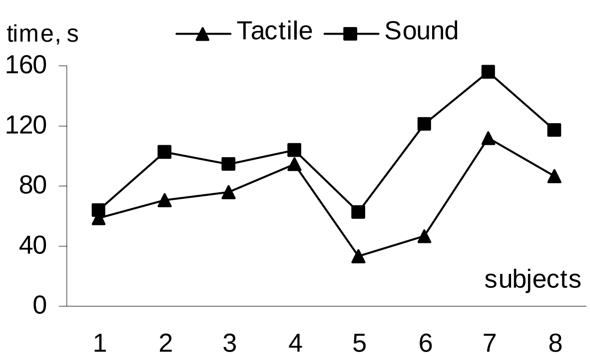


Figure 7: Mean time for tactile and sound feedback

The Figure 8 shows the mean error of each subject for both tactile and sound feedback modes. From the bar diagram, it is difficult to conclude which feedback provides less errors than the other, as half of the participants did less error with tactile feedback and half did less error with sound feedback.

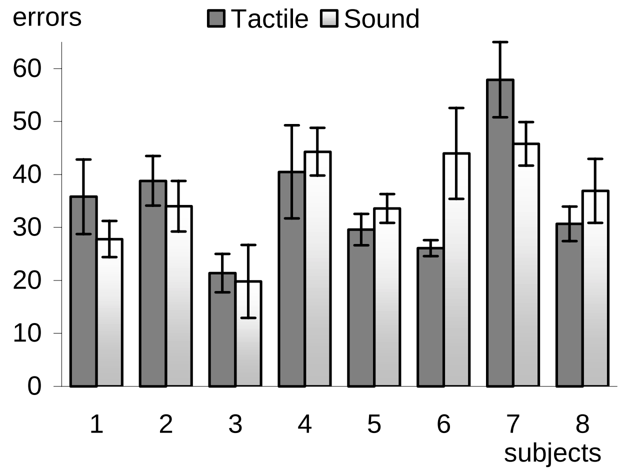


Figure 8: An average and standard deviation for errors during tactile and sound feedback modes

The reason behind this may be due to the feedback varies person-to-person. It depends on individual hearing or kinesthetic sensitivity, or preference. One may feel more comfortable with sound feedback and another one may feel comfortable with tactile feedback by perceiving sound like annoyance. The standard deviation was calculated here to measure of how widely values are dispersed from the average value, that is from the mean.

Behavioral strategy

While there was no time limitation for the task, the subjects had different performance attitudes to complete the task. Among all, few subjects took short time for each trial and few spent more time. Some of them were very careful about making mistakes. On the other, hand some subjects were just wanted to complete the task as fast as possible without considering the increased number of errors. At the beginning participants were enjoying the task with sound feedback. After switching to tactile feedback they found that tactile feedback helps more than sound feedback. The subjects were not instructed to anchor their body part. Sometimes they put the elbow on the desk and hold the wrist in air when using the tactile pen. Thus it is natural for the subjects to get tired after a while which might affect the performance of the task. Some interesting strategies of the subjects’ behavior are discussed in the following subsections.

Critical zones. Observing the collected data of all the subjects it was found that most of the errors occurred near the turning points or corners of the maze track. Going through a straight line seems much easier than turning the movements. Subjects found both the feedback insufficient

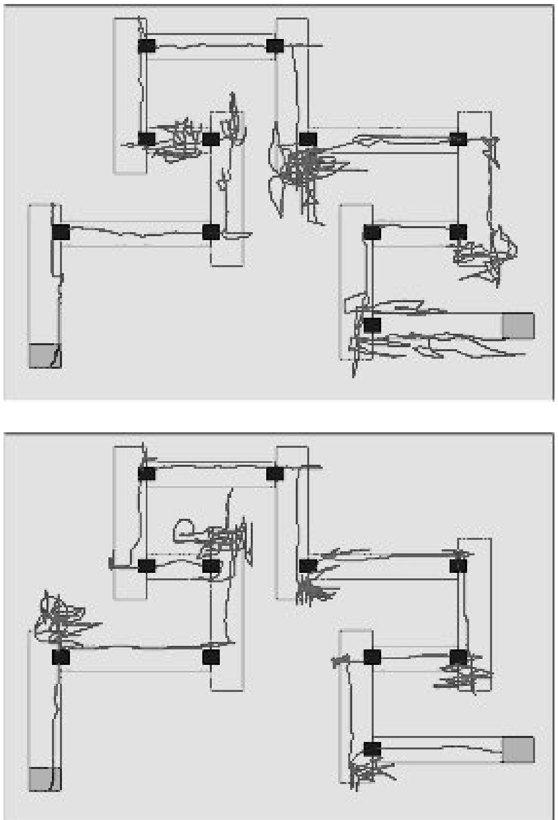


Figure 9: Critical zones

to tell them which way to follow. The Figure 9 shows the difficulties of finding appropriate track through the pass to the next segment for several subjects.

Tendency to make shortcut. Few subjects became familiar with the track after completing several trials. They got an idea about the track without watching it and tried to guess which way to go. Sometimes the subjects were moving their pen on the screen though there was no feedback. Figure 10 shows two screenshots of such cases.

Tendency to move near the borders. It was also observed that subjects were moving along the track near borderlines of the maze. It seems they preferred to follow the feedback on the borderlines rather than the central position.

The reason behind this kind of behavior can be explained by the way that they did not get sufficient feedback while going through the central part of the track. It may also cause for the type of feedback. Feedback for the borders was much stronger than that in the central part. Figure 11 presents the track of the subject who completed the task moving the pen through the borderline of the maze.

Modes Switching

The subjects were asked to switch the mode without any pre-notice. The effect of switching between sound and tactile feedback is not obvious. Of course, the learning time had effect on the result.

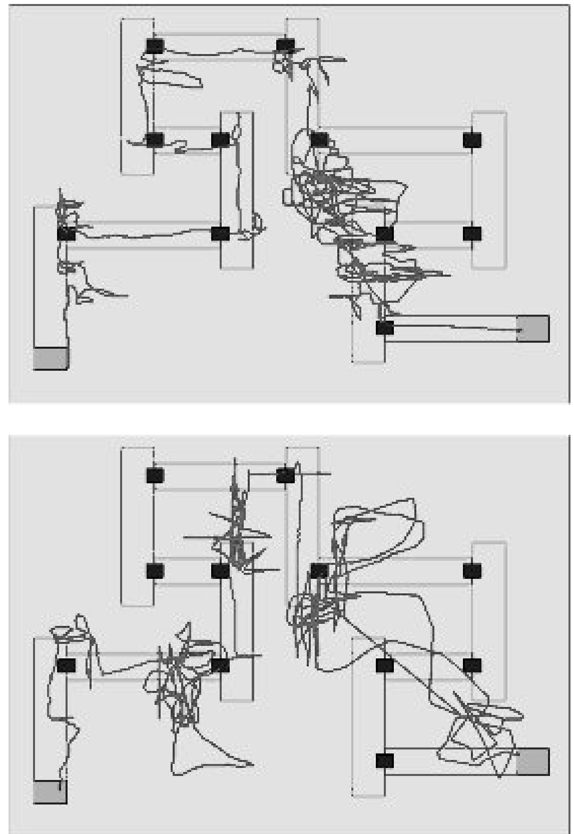


Figure 10: Moving shortcut to the destination

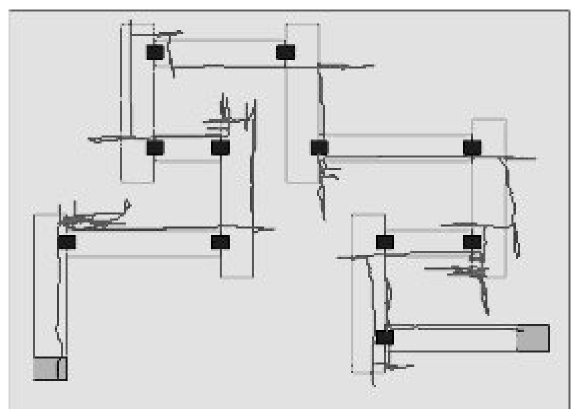


Figure 11: Tendency to move near borderlines

At the beginning during the training period, the subjects found sound feedback much easier than tactile. Starting with sound feedback at the moment they switch to tactile feedback after few trials the same subject changed the opinion on the other way. In the time being, the subjects were perhaps feeling refresh with the new feedback.

Expert Versus Novice

All the subjects were chosen novice for the system. But at the end of their task, they obviously became bit expert. A lot of differences have been observed in data. Figure 12 represents the required time for the first and tenth trials of each subject. The bar diagram shows the improvement of the task completion in shorter time by the subjects. It is also obvious from the graph that required time varies from person to person. One subject spent more than 165 seconds while the other took only about 45 seconds to complete the one trial for the first time.

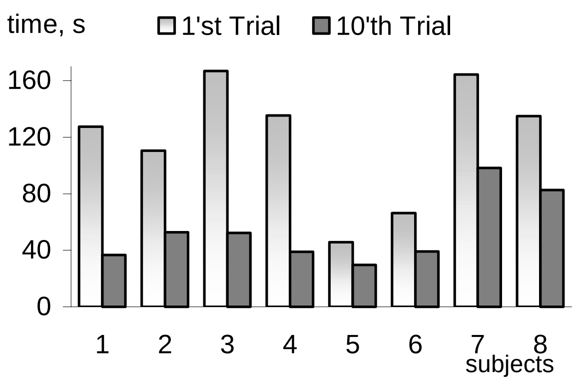


Figure 12: Time required for the novice and expert subjects

Total number of errors also reduced while the subjects getting experts. Figure 13 presents the total number of errors occurred by the subjects in their first and tenth trials. It can be observed from the graph that several subjects improved a performance while became experts and four improved moderately. It is surprising that one subject deteriorated the performance after being an expert. There were several reasons for this kind of result. The task might become boring after a few trials or the subjects felt frustrated with the feedbacks.

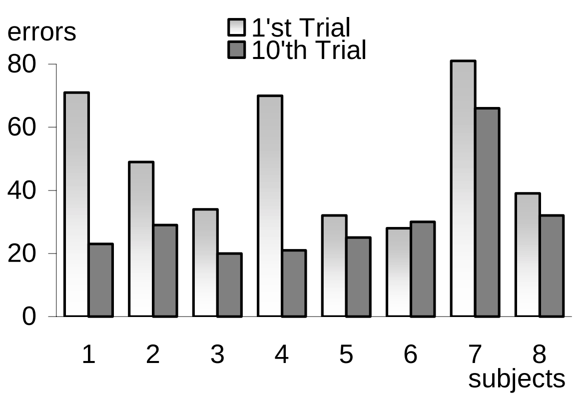


Figure 13: Error occurred by the novice and expert subjects

Limitations of the Study

There are a number of limitations to this study. Tactile feedback used in the maze was not much distinguishable between central part (167 Hz) and near the border (250 Hz) fields of the track. For crossing the border the feedback is stronger (30 Hz) than the others. Thus, the subject felt comfortable to complete the task following the maze near borderlines. It would be better to change the feedback for the central part. Different sound and tactile feedback for the joining part or the passes of the maze should provide more information to the participants. It would be more realistic to use special tactile/sound icons in this regard. It would be helpful for the participants to understand their position more precisely during the movement inside the maze towards the next or previous passes. Besides, it would be more challenging if the shape or pattern of the maze would change randomly for each trial.

Future Investigation

Though several studies had been investigated several modes of the non-visual feedback cues and their implementation, still there are a lot of things to do. The subjects' behavior for other patterns of tactile feedback might give some interesting results. Due to time limitations the investigation were not completed here. Comparing different tracks of participants and analyzing them will give more information about using non-visual feedbacks. Besides blind manipulation, it would be possible to create some interesting games for the children as well as for the adults using these feedbacks.

CONCLUSION

It has been observed from the present study that the subjects behave differently from each other in response to sound and tactile feedback cues. It was not possible to conclude which one is best than the others in a case of error occurrence. But, it is obvious from the study that tactile took less time than sound feedback to complete the task. Sound feedback sometimes may be problematic if it has lower deviation in frequency or volume when position is changed. It might make the user tired. Almost all the subjects of the test agreed at the end of the task that they preferred tactile sense than sound feedback.

Due to several limitations of the study, the number of error rate did not differ much between expert and novice users. It is necessary improving the user interface design by providing more feedback cues for different directions and the critical positions through the track.

In a complete human-machine interface, the use of non-visual feedback modalities such as sound or tactile feedback are expected to yield performance improvements in a case where visual channel is near capacity. This will occur if the operator's attention is divided among different parts of the display or among multiple tasks. The results of the study can be used for further development of information kiosks, novel pen input devices, haptic interfaces and for applications for people with special needs.

ACKNOWLEDGMENTS

I would like to acknowledge Grigori Evreinov and also the participants for their precious time.

REFERENCES

1. Akamastu, M., MacKenzie, I. S. and Hasbrouq, T. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics, 38*, 1995, 816-827.

2. Chapanis, A.. and Kinkade, R.G. Design of Controls, In H. P. Van Cott and R. G. Kinkade (Eds.), Washington, DC: U.S. Government Printing Office. *Human engineering guide to equipment design*, 1972, 345-379,

3. DiGiano, C. Program aurilization: Sound enhancements to thr programming environment, *Proceedings of Graphics Interface '92*, Toronto: Canadian Information Processing Society, 1992, 44-52.

4. Garver, W. The Sonic Finder: An interface that uses auditory icons, *Human-Computer Interaction*, 1989, 4, 67-94

5. Immersion Studio 4.0.3 available at: http://www.immersion.com

6. Oakley, I., McGee, M.R., Brewster, S. and Gray, P. Putting the Feel in 'Look and Feel'. Glasgow Interactive Systems Group, University of Glasgow, Glasgow, U.K.

7. Sellen, J.A., Kurtenbach, P.G. and Buxton, S.A.W. The Prevention of Mode Errors Through Sensory Feedback. University of Toronto. In *Human-Computer Interaction*, 1992, Volume 7, pp. 141-164.

8. Yu, W., Ramloll, R. and Brewster, S. Haptic Graphs for Blind Computer Users. Department of Computer Science, University of Glasgow, U.K.

Two Characters per Stroke - A Novel Pen-Based Text Input Technique

Natalie Jhaveri
New Interaction Techniques 2003
Department of Computer and Information Sciences
FIN-33014 University of Tampere, Finland
+358-04-833-2420
natalie.jhaveri@uta.fi

ABSTRACT

A novel pen-based text entry technique, which yields two characters with one stroke, is introduced, and the results of its investigation are presented. The 2CPS technique is implemented as a gesture-based user interface on a QWERTY soft keyboard. Eight participants completed eight text entry trials on a fully functional version of this stroke-based system on a Pocket PC. The difference between the means of the eighth trial and the first trial demonstrated a substantial improvement of 41%. In spite of this evident progress on the learning curve, the participants' mean rate for the eighth trial was a mere 5.6 wpm. The reasons behind the results are discussed and future improvements to the system are proposed.

KEYWORDS: Soft keyboard, pen-based text entry, gesture-based user interface, stroke

INTRODUCTION

Text entry for mobile systems has been popular research area for over ten years now. Mobile computing poses many challenges for designers of text entry systems. A text entry system is restricted by small portable input devices and minimal screen space. The use of a stylus to tap keys on an on-screen soft QWERTY keyboard is the widely accepted solution. Evaluations of this technique have shown that even novice users achieve good text entry rates. Though novice user rates are predicted to be 8.9 wpm with the QWERTY layout, a study found novice rates to be in fact much higher, 20.2 wpm [5]. This considerable difference is attributed to the effect of skill transfer from the users' vast experience with the QWERTY layout on a physical keyboard.

The two-character per stroke technique, 2CPS, is, in essence, an enhanced QWERTY soft keyboard. This system presents the user with the familiar QWERTY layout and adds the 2CPS functionality to it. Without any additional controls or visualizations, this technique allows the user to select a character with the traditional method of tapping a key, but also allows the user to enter a secondary vowel or space character with a directional stroke. Further details of its design will be presented in the next section.

The idea of using a stroking gesture in pen-based text entry is motivated by the fact that making a stroke gesture with a pen is a very fast and natural way of handling a pen [2]. In fact, most of us have been started learning to make directed and oriented strokes with pens, pencils, or crayons, since we were as young as two or three. Yet, most pen-based text entry systems rely on tapping as the only method of input.

Incorporating continuous gestures into stylus-based text entry is certainly not a new idea. Popular techniques that make use of character recognition technology, such as *Graffiti* [1] and *Unistrokes* [3], are the most obvious gesture-based text entry techniques for mobile systems.

Nevertheless, gesture-based user interfaces do exist outside the realm of character recognition. Some of the main novel gesture-based stylus-based text entry techniques include: Quikwriting [9], T-cube [10], and Cirrin [8]. Each of these techniques presents a novel user interface that incorporates the use of continuous gestures. Quikwriting for instance, requires the pen to enter and exit one of eight sectors within one of two different square-shaped areas, for the selection of a character; thereby allowing the user to enter text without ever lifting the pen from the screen. In contrast, T-cube, allows the user to select a character with a flick gesture starting from one of nine different positions and aimed in one of eight different directions. Cirrin presents the alphabet in a circular layout. The user can stroke the pen in and out of one of the 26 section to select a character. Therefore, through one continuous stroke, the user can enter an entire word.

Each of these techniques requires the user to overcome a high initial learning curve. In contrast, being an enhanced QWERTY keyboard, the 2CPS technique anticipates a shorter learning period, while at the same time improving on the text entry rate achieved by the basic QWERTY soft keyboard.

An empirical study was carried out to evaluate the 2CPS technique against the conventional tap-typing technique. The results of this study are presented and discussed here. Based on these results, possible improvements to the current design of the 2CSP system are proposed. First, however, the 2CSP technique is described along with the user interface of the system in which it is implemented.

TWO CHARACTERS PER STROKE

As mentioned earlier, the 2CPS technique itself does not require any visible modification to the QWERTY soft keyboard. Each key is enhanced with the same set of secondary characters that can be selected with a direction-specific stroke. For example, simply tapping the "R" key results in the entry of "R"; whereas, touching the pen down on the "R" key and stroking directly north anywhere past the top border of the key, results in the entry of two characters, "RE". This gesture is illustrated in Figure 1.

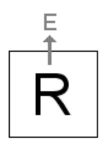


Figure 1. Stroke gesture required to enter two characters, namely R and E.

Therefore, the challenge of this technique lies in remembering the layout of the secondary keys. On the other hand, the keyboard remains simple, without any visual distractions such as secondary menus that show the stroking possibilities, at the cost of concealing and cluttering the keyboard.

Keyboard

The layout of the keyboard used in this system is a slight variation of the traditional QWERTY layout. In this layout, as illustrated in Figure 2, the keys are vertically aligned and some additional keys are included on the far right side of the keyboard. The traditional QWERTY keyboard is mainly arranged with a diagonal layout to aid typing on a physical keyboard. As this is not an issue concerning touch-typing, the variation to the layout can be justified for this system.

Figure 2. 2CPS soft keyboard: Adapted layout of the QWERTY soft keyboard.

Layout of secondary characters

The 2CPS system, which was originated, designed and developed by Grigori Evreinov, the lecturer of the New Interaction Techniques 2003 course at the University of Tampere, offers great versatility in its user interface. The secondary characters can be edited by the user of the system. It certainly makes sense to include the most frequently letters of the English language in the eight locations around they key. However, during some initial experiences with the system, it was observed that using such a layout can be very cognitively demanding for the user. Therefore, for this study, a simple arrangement that includes only the main vowels of the alphabet and the space character was used. The layout is demonstrated in Figure 3.

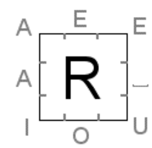


Figure 3. Arrangement of secondary characters.

The prime locations contain the most frequently used characters, "a", "e", and space. It was decided that the letters "a" and "e" would span two sections so that selecting either of them requires less precision and recall. As well, the order of arrangement, "a, e, i, o, u" is more intuitive than any other arrangement. Likewise placing the space character directly to the right of the primary character, in a prime location, is justified by the fact that the space is the most commonly entered character in text entry.

Feedback

Sound feedback is used to indicate when a key has been selected. Although, a one-character tap entry is not distinguished from two-character stroke entry. During a test situation, when a participant enters an incorrect character, a different sound is given as notification of the error. When in test mode, the participant selects the space key when he or she is ready to start entering the phrase. This start is also marked by a different sound.

METHOD

Participants

Eight voluntary participants with various educational backgrounds and of various nationalities were recruited for this study. This group, which was comprised of 3 males and 5 females, covered an age range of 21 to 30. None of the participants had any previous experience with pen-based computing and their daily computer usage ranged from low to high. Of the eight participants, three were left-handed.

Apparatus

The study was carried out on a Compac iPac, Pocket PC with its 10.5 cm stylus. The keyboard of the software application is approximately 4.5 cm X 1.75 cm, with each key measuring approximately $0.3 cm^2$ (i.e. 16x16 pixels).

The application itself collects all the text-entry data required. Therefore, key figures such as mean text entry rate (wpm), number of errors per trial, and number of keystrokes per phrase, are stored for each trial.

A paper with an illustration, as shown in Figure 4, was presented to each participant as a visual aid for understanding the system and memorizing the layout during the training period. However, it was taken away for the actual text entry trials.

A	E	E
A	R	_
I	O	U

Figure 4. Illustration of the layout for the participants

Procedure

The evaluations took place in a variety of locations, such as at the participant's residence, the university cafeteria, or the university lounge. This was done, firstly to avoid a test-type environment in which users are often not at ease, and secondly, to mimic an environment in which users might actually enter text into a mobile device, in other words one with some distraction and background noise.

Each participant was given two minutes to familiarize themselves with the Pocket PC and pen-based text-entry on the QWERTY keyboard of the software application. This was important so that the participant could find their comfortable pen-holding position, and screen tapping angle. For instance, one participant preferred entering text with the Pocket PC placed on a table. Whereas the others found that they preferred to hold the Pocket PC in their hand and sit back in their chair to type. After the two minutes of familiarization and training, the participant was instructed about the trials and told to simply "enter the phrase shown".

One trial consists of entering three phrases, which are randomly selected from a set of ten test phrases, and displayed one at a time at the top of the screen. The test phrases were 19 to 25 characters in length. As suggested by Mackenzie [6], the phrases were short, easy to remember and representative of the English language. Moreover, every letter of the alphabet was included.

To begin a phrase the participant first clicks the space bar, and at this moment, the timer is started. The timer stops upon correct entry of the last character in the phrase. After that, the phrase disappears and the next one to be entered appears. At this point the participant can rest and then, as before, he or she can start by selecting the space key.

It is important to note that erroneous characters are not accepted by the system, and therefore, of the participant enters an incorrect character, an error sound is given, and the user must attempt to enter that character again.

First, the participant completed three trials of text entry with the original tapping method on the QWERTY keyboard. After completing three trials with the basic QWERTY keyboard, the 2CPS method was introduced to the participant as discussed in the previous section. The participant was then given a five minute training period to learn this new technique. Next, the paper illustration was taken away and the participant proceeded through eight trials entering phrases, as before.

In total, the entire evaluation took approximately one hour per participant and resulted in 88 trials worth of data.

RESULTS

For all trials, and on with both techniques, unreasonably high error rates were observed. Errors per trial reached as high as 26 errors, and over the 88 trials, few demonstrated a low number of errors. There was only one trial with zero errors, only one trial with 1 error, and still only 1 trial with 2 errors. With the design of this experiment, this issue strongly affects the text entry rates and the keystrokes per characters values. Therefore, the observed text entry rates for the basic QWERTY layout fall very short of those found in past studies [5, 7]. For this reason, these two measures will be presented within the context of the 2CPS technique alone. A further discussion as to the exceptional error rates will follow in the next section.

As the software was capable of collecting an extensive amount of data, many interesting results have been extracted from this study. The main figures of interest include the text entry rates (wpm) and the keystrokes per character (KSPC) rates achieved.

Text entry rate

Over the 8 trials of entering text with the 2CPS technique, a steady improvement rate can be observed (Figure 5). Furthermore, the difference between the mean of the last trial (5.62 wpm) and the first (3.31 wpm) demonstrates an overall improvement of 41%.

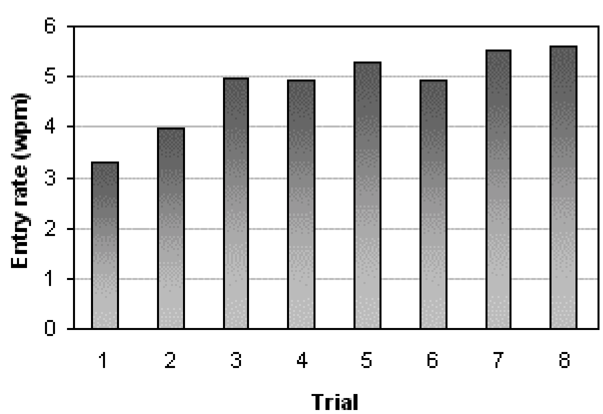


Figure 5. Mean text entry rates achieved with the 2CPS technique over eight trials.

Using text entry rates to evaluate new text entry techniques can be awkward. In cases when the rate does not include errors, it can be an unrealistic portrayal of the technique's capabilities. However, when it includes errors, it does not compare well to text entry rates of studies that disregard errors. Therefore, we next consider another metric for evaluating text entry.

Keystrokes per character (KSPC)

Keystrokes per character (KSPC) describes "the number of keystrokes required, on average, to generate a character of text for a given text entry technique in a given language". For the QWERTY soft keyboard, where one tap gives one character, KSPC is exactly 1.00, when considering only lower case characters. [4]

Determining the KPSC value for the 2CPS technique is rather problematic because the value varies depending on the sequence of characters entered. For instance, the word "elk" requires three strokes, while the word "cat" only requires two.

In order to determine a benchmark KSPC for the 2CPS technique for this study we consider it in the context of the phrases used for the study. Since this system support tapping-only, a given phrase can be entered an assortment of ways. However, being true to the technique, and the goals of text-entry, only the best combination of strokes and taps, which yields the lowest number of keystrokes, is considered. These values are presented in Table 1. Therefore, based on the values for the ten test phrases, the KSPC of this technique is 0.65.

Phrase	No. of characters	Minimum no. of keystrokes	KSPC
the quick brown fox	19	13	0.68
jumped over a lazy dog	22	13	0.59
a tree of gold and silver	25	16	0.64
email me the class notes	24	16	0.67
the best football player	24	16	0.67
software engineering	20	14	0.70
slower than a turtle	20	12	0.60
business application	20	13	0.65
a little green table	20	13	0.65
call me next thursday	21	14	0.67
		mean:	**0.65**

Table 1. KSPC values of the ten phrases used in this study

As these values depict a situation of ideal stroking technique and zero errors, it is not representative of the real KSPC result of novice users. In the case of the participant's trials, a higher KSPC result of anywhere between 0.70 KSPC to over 1.00 KSPC was observed. Again, this large disparity is caused by the differences in individual error rates, individual stroking strategies, and in the phrases themselves. Not surprisingly, the trials which resulted in a high text entry rate and less errors are also those for which the KSPC was close to the ideal values listed in Table 1. The trials that unquestionably demonstrate this characteristic are illustrated in Table 2.

Participant	Trial No.	Text Entry Rate (wpm)	Total No. of Errors	KSPC
P1	3	7.95	3	phrase 1: 0.95 phrase 2: 0.80 phrase 3: 0.75 **mean: 0.83**
P2	4	7.66	0	phrase 1: 0.70 phrase 2: 0.70 phrase 3: 0.76 **mean: 0.72**
P3	5	7.87	5	phrase 1: 0.71 phrase 2: 0.75 phrase 3: 0.85 **mean: 0.77**
P4	8	8.55	4	phrase 1: 0.75 phrase 2: 0.70 phrase 3: 0.70 **mean: 0.72**

Table 2. Trials in which close to ideal KSPC rates were achieved

Observations and Participant Feedback

Overall, participants were excited by the technique but after a while, they were frustrated by the enduring frequency of errors. At times they felt that they have definitely selected the correct key, or definitely stroked in the correct direction and location, and yet they still received an error.

As early as the first trial, participants had an easy time remembering the frequently used secondary characters "a", "e", "i", and space. However, the less common characters "o" and "u" were more difficult to remember and caused some hesitation even on the last trials.

One out of the three of the left-handed users had some initial difficulty with stroking to the right for the space character. This participant was the only participant who chose to place the Pocket PC on the table rather than hold it in hand. These slight differences could justify the fact that, out of all the participants, this participant had the highest mean number of errors (9.87) for the 2CPS technique. However, as the participant that received the lowest mean number of errors (4.37) for the 2CPS technique was also left-handed, there is no indication of a right-handed bias in the design of the technique.

DISCUSSION AND IMPLICATIONS FOR DESIGN

Perhaps more could have been concluded from this study if the large error rate hadn't interfered with the results to such as extent. In this section, the causes of the high errors are discussed, solutions are proposed, and general ideas regarding the design of the system are presented.

Limitations of the study

As described earlier, the system does not accept incorrect entries. Therefore, upon an incorrect entry, the participant must attempt the entry again. This means that all text entry rates and KSPC values found include all the repeated entries made to rectify errors. For this reason, the reported text entry rates and KSPC values are actual values, unlike other text entry studies which depict ideal measures. If the system had not shown such a high rate of errors and if the wpm and KSPC figures did not include these errors, the text entry rate would be higher for both techniques and the KSPC of the 2CPS technique would be closer to the ideal value. In view of that, we have attempted to evaluate this technique with the inclusion of all errors, which are often hidden or ignored in similar studies.

The concern now is the cause of the unreasonably high error rate. Typical sources of such a problem include: the sensitivity and resolution of the touch screen, the parallax effect of holding the PDA on an angle [5], the small size of the keys [5], or in the design of the software application itself. However, as the built-in keyboard of the device does not exhibit an error-prone behavior, the first three possibilities can be discarded. Many participants often commented that "Sometimes it just refuses to take." To further investigate this issue, a five day longitudinal study was performed, with myself as the sole participant. As a relatively novice user of PDA's, pen-based text entry, and this 2CPS technique, using myself as a participant for this quick study is somewhat justified. For five successive days, five trials of the 2CPS technique were completed with the same software used in the main study. Though over the five days a smooth increase in text entry rate was observed (Figure 6), the number of errors was very sporadic in its downward trend (Figure 7).

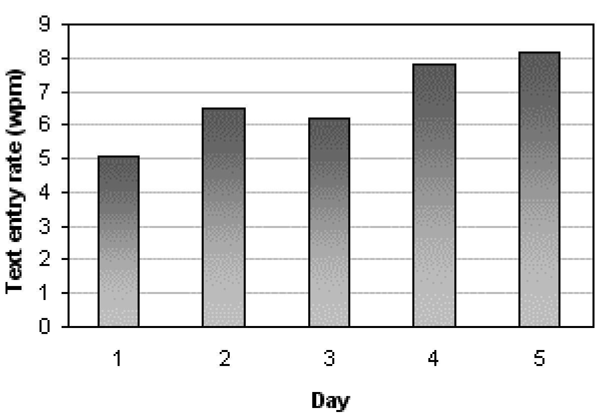


Figure 6. Mean text entry rates achieved over a five-day period with the 2CPS technique.

Even by the third day, the mean number of errors per trial was as high as that of the first trial. Certainly, this is not the expected trend of number of errors over a longitudinal evaluation. Furthermore, as in the previous study, low error numbers for a trial were rare. In this study there was never a trial of 0 errors or only 1 error. Only three trials were completed with 2 errors. Although the mean number of errors of the fifth day was quite moderate at 4.8, that final day still contained a trial with nine errors. Therefore, the high errors were not simply an issue of the eight participants of the main study; they were certainly unavoidable with this system.

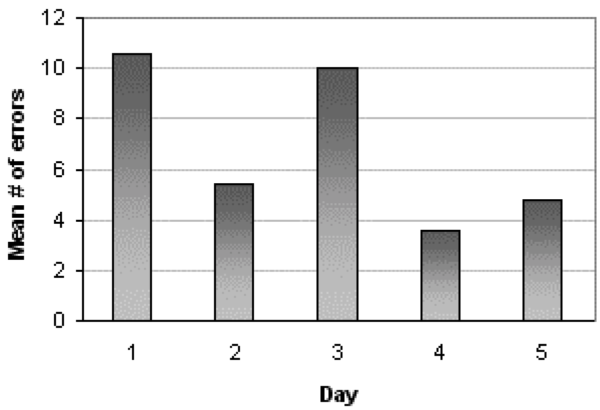


Figure 7. Mean error count observed over a five-day period with the 2CPS technique.

Before further evaluations can be carried out on this technique, the source of the high error rate must be identified and solved in the design of the system. Once this issue is solved in the system, further evaluations of the technique will certainly produce more clear and concrete results.

Feedback Improvements

As mentioned earlier, the system does not give any visual feedback for key selection. It uses sound to indicate when a key is selected. Also, the system uses sound feedback to indicate whether or not the correct key was entered. Based on the feedback of the participants, having this feedback was useful. However, this type of feedback is only useful in test situations. In actual text-entry circumstances, when the system does not know what text the user wishes to enter, error feedback not possible. In reality, the user only notices a text entry mistake by glancing at the text field.

Moreover, sound feedback may not be ideal for all occasions of mobile system use. It can be difficult to receive in noisy environments and inappropriate in certain circumstances, such as during meetings.

In contrast, visual feedback, which is typically used in most soft keyboards, would likely be a very practical solution. Text entry on soft keyboards already inherently requires the user's visual attention. Therefore, incorporating feedback into a visual element of the keyboard would limit the attention required by the user to only one sense, vision.

One possible solution for future examination would be to incorporate the visual feedback into keys themselves. For instance, the side or corner of the key that is selected by the stroke could be shadowed or colored (Figure 8). In this way, the user would at least see what he or she has selected. This is especially important as soft keyboards on mobile system screens are typically very small. Moreover, in contrast to the sound feedback, which is given after the key selection, this system provides the feedback during the selection. This would give the user the chance to at least notice an incorrect selection. Furthermore, if such an action could be supported in a future version of the system, a novice user could perhaps re-direct an incorrect gesture during mid-stroke.

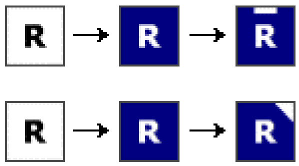


Figure 8. Proposed visual feedback: Sequence of an unselected key, a selected key, and a selected and top-stroked key (upper picture) or top-right-stroked key

Perhaps including such a visual feedback approach into a future version of the system would enhance its usability.

CONCLUSION

This initial evaluation of the 2CPS technique proved to be very informative and instructive. The technique offered many clear advantages including:

- decreased visual scan time [5]
- decreased average movement time [5]
- extended use of the stylus

However, the technique also bears some obvious disadvantages when compared to the QWERTY keyboard tapping technique. These include:

- increased cognitive load (to consider the correct stroking direction)
- increased precision required for aiming the stroke

These initial findings demonstrate the 2CPS technique's potential for decreasing the KSPC value, and thereby perhaps decreasing the text entry rate currently achieved with the QWERTY soft keyboard. While the QWERTY keyboard alone achieves a KSPC of 1.00, the 2CPS technique has the potential to decrease this value to as low as 0.65 KSPC.

Adapting this technique into the next generation on mobile devices is certainly more feasible than with other novel text entry techniques because it does not dramatically change the familiar layout of the soft keyboard. Furthermore, it can easily be included as an advanced feature that can be enabled as a personal preference.

For these reasons, it is definitely worth improving the current implementation based on the finding of this study and then re-evaluating the technique against the tap-typing technique.

ACKNOWLEDGMENTS

Sincere appreciation is extended to Grigori Evreinov for initiating the design of the 2CPS technique, developing the system, and proposing its evaluation as the topic of this study. His guidance with this project was very much appreciated.

REFERENCES

1. Blickenstorfer, C. H. (1995, January). Graffiti: Wow! *Pen Computing Magazine*, pp. 30-31.

2. Geissler, J. Gedrics: the next generation of icons. *Proceedings of the 5th International Conference on Human–Computer Interaction (INTERACT'95)*, Lillehammer, Norway, June 27–29, 1995, pp. 73–78.

3. Goldberg, D., & Richardson, C. (1993). Touch-typing with a stylus. *Proceedings of the INTERCHI '93 Conference on Human Factors in Computing Systems,* 80-87. New York: ACM.

4. MacKenzie, I. S. (2002). KSPC (keystrokes per character) as a characteristic of text entry techniques. *Proceedings of the Fourth International Symposium on Human Computer Interaction with Mobile Devices*, pp. 195-210. Heidelberg, Germany: Springer-Verlag.

5. MacKenzie, I. S., Zhang, S. X., & Soukoreff, R. W. (1999). Text entry using soft keyboards. *Behaviour & Information Technology*, 18, 235-244.

6. MacKenzie, I. S., & Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI 2003*, pp. 754-755. New York: ACM.

7. MacKenzie, I. S., and Zhang, S. X. (2001). An empirical investigation of the novice experience with soft keyboards. *Behaviour & Information Technology, 20*, 411-418.

8. Mankoff, J., & Abowd, G. A. (1998). *Cirrin*: A word-level *Unistroke* keyboard for pen input, *Proceedings of the ACM Symposium on User Interface Software and Technology - UIST '98* (pp. 213-214): New York: ACM.

9. Perlin, K. (1998). Quikwriting: Continuous stylus-based text entry. *Proceedings of the ACM Symposium on User Interface Software and Technology - UIST '98*, 215-216. New York: ACM.

10. Venolia, D., & Neiberg, F. (1994). *T-Cube*: A fast, self-disclosing pen-based alphabet. *Proceedings of ACM Conference on Human Factors in Computing Systems - CHI '94*, 265-270. New York: ACM.

F-Pointer: Finger-Manipulated Device and Usability Evaluation

Remberto Martinez Gonzalez
New Interaction Techniques 2003
Department of Computer and Information Sciences
FIN-33014 University of Tampere, Finland
358-400-784273
remberto.martinez.gonzalez@uta.fi

Abstract
Human-computer interaction techniques include sensor technology and strategy of user behaviour as joint parts of the interface. Miniaturisation and economized space constraints require simple design to perform pointing and selecting tasks, while keeping efficiency as high as possible. Having finger motion detection along a surface, usability of a single finger manipulated device is assessed within a text entry scenario. The experiment was designed to measure novice user performance by screen typing where speed, accuracy, and response time were registered. The results showed a throughput of 1.66 bps and velocity of 10 wpm. Future applications should take advantage of the integration capability of this device.

KEYWORDS: finger manipulation, pointing and selecting task, text entry technique.

INTRODUCTION

Input devices merit important attention regarding to computer interaction techniques. For a device is always desirable to keep its primary function of interaction with user only when needed as well as when used to accommodate motor and mechanical capabilities and limitation of humans. Not only ergonomic comfort and safety should have serious impact on design but also usage strategy and mental effort. Taking a look at successful devices manipulated by hands and fingers, like touch pad, track balls, mouse, rubber stick, joystick, the common relevant aspect is the adaptation to the anatomy of hands and fingers. Mice took the leading rate since its first appearance in desktop interface after light pens due among others factors to its re-acquisition and hand-finger muscles operation abilities [5].

In wearable computing and mobiles phones where miniaturization or economized space is a serious constraint, new requirements have to be fulfilled by input devices: portable, user-independent, wearable and non-obtrusive, but also integrated and efficient. At this point the solutions tends to register finger motion from ranges of a simple mini joystick to expensive virtual keyboard (for current virtual keyboards see http://www.alpern.org/weblog/stories/2003/01/09/projectionKeyboards.html).

Previous research in finger motion can be found in [3] where a virtual keyboard is projected over a desktop and a CCD camera senses fingertips movements. Capacitance and light sensors have been used in [6] and [1] respectively but no relevant results in device integration have being proposed until [2] where an unobtrusive device uses capacitive sensing of wrist movements and finger positioning in a tactile pad.

F-pointer is a simple finger manipulated device developed by Grigori Evreinov, PhD at the Computer Sciences Department of Tampere University. The design comprises unobtrusiveness and transparency, simple and easy to learn when other computer input devices have been used but still has its own input method that needs to be acquired.

For this reason we carried out a comparative study where we could evaluate the performance of the device, study the features and strategy of the input process, test compliance with existing standards and make a forecast of possible applications. In order to assess user performance we designed an experiment where text entry is accomplished by pointing and selecting in a screen keyboard (2D). By means of that we were able to record values like movement time, accuracy and user response.

The purpose of this project was to evaluate a new non-expensive pointing device for recording finger motion and position without direct contact or special fixation to the finger.

METHOD

The design of the experiment included usability elements. Performance was measured with a piece of software developed in Microsoft Visual Basic 6.0 for Windows 2000 by Grigori Evreinov. This software allowed subjects to be

New Interaction Techniques '03, Spring, 2003, University of Tampere.

immersed in the task of typing in a screen QWERTY keyboard while measuring the time needed to point a character and select it by dwelling in the position a certain time. Comfort was evaluated by asking questions to participants during and after the test interval. The questionnaire in Annexed 2 showed participants rating of usability aspects like operation, fatigue, easy to learn and others.

Testing used multidirectional pointing (2D Fitts' discrete task) with F-pointer device.

The experimental design includes some elements about information theory discussed in [7]. The rate of task execution can be interpreted as the human rate of information processing considering the analogy between task difficulty and information [4].

It is recommended when evaluating user response and throughput of a pointing device to include some index of difficulty as follows:

$$ID = \log2\ (D / Wef + 1), \quad (1)$$

Throughput (TP) is then computed as

$$TP = ID / MT \text{ (bits per second)}, \quad (2)$$

and it captures both the speed and accuracy in performance. MT is the mean movement time (in seconds) and D is the distance to target (in pixels). W_{ef} is the effective width of target after affecting the width W by a correction factor:

$$W_{ef} = W \times Corr \quad (3)$$

This factor is intended to include the user perception as once "over" the target no correct selection is made due to the shape of target and/or the angle of direction of movement respect to target.

After several trials and provided that no discrete error is recorded, we assumed the correction to be equals 1 and so Wef equals W. In terms of evaluation and comparison this assumption is valid since the correction is constant value. For details in using other coefficients see [7] where the unit-normal curve is used to normalize the effective width of target.

The use of a text entry technique allows measuring other features related directly to index of difficulty. Human performance is also a function of the time consumed to accomplish certain task and the error rate related [4].

The number of keystrokes per character (KSPC) is particularly important in characterizing and comparing text entry methods. KSPC is the number of keystrokes, on average, to generate each character of text in a given language using a given text entry technique. In our case this value give also information about the efficiency of the device used. In general, speed of text entry is measured in words per minute (wpm) and is proportionally directed to the number or characters per second (CPS).

For characterising the accuracy of the device, the error rate was defined as the percentage errors from the total amount of input characters. In this case as well this parameter is related to the text entry technique.

Besides the visual stimulus, in task completion, sound was provided for the beginning and end of input sequence, error or bad character, next word and end of trial. Other functionality of the software tool used was provided for the experimenter as well.

Participants

5 volunteers (4 female and 1 male) were asked for cooperation in this study. All the subjects were right handed but one – left handed person, all used computers with graphical user interfaces on a daily basis. According to their experience in input device they were all novice with F-pointer, two of them had previous experience with touch pad. All used the F-pointer with the right hand and index finger (Figure 1b).

Apparatus

A desktop PC (ASUS A7V133) was equipped with AMD Duron Processor, 256MB RAM, VIA AC'97 Audio Controller (WDM), game port and speakers.

The tested device, F-pointer, was attached to the game port (Figure 1). The device driver maps the circular area of the surface directly to the respective area of screen. The sensibility can be adjusted in the calibration procedure provided by the operating system from the experimental software tool.

Figure 1: F-pointer, use mode.

The software tool was written in Microsoft Visual Basic 6.0 SP5 developed by Grigori Evreinov for running in Windows 2000 operating system. This tool for Text Entry testing (Figure 2) used a component Joystk32.ocx developed by Mabry Software Inc. (www.mabry.com)

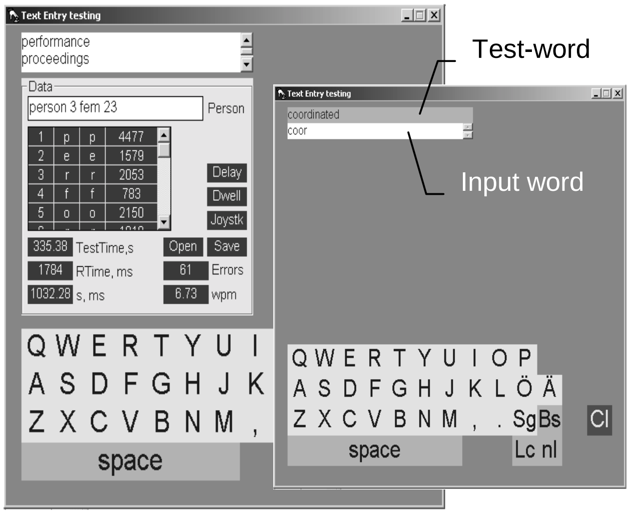


Figure 2. Software tool for text entry

Procedure

Participants were instructed to move the pointer by moving a finger inside the surface of F-pointer. Dwelling the pointer for a certain time over the desired letter did the selection. Normally, the used finger was the index when the device was resting on the table. Other positions were allowed like lifting the device and operate it through thumb finger similar to a mobile phone position, but participants preferred the conventional one.

We asked subjects to retype test-words as fast as possible in the input text window after demonstration. Each time the target letter was selected in the on-screen keyboard; the equivalent character appears so as to complete entry of the requested word below the displayed test-word. After 5 minutes of familiarisation with the device the control of pointer was good enough to start data collection. Taking into account the subjects were novice, the software settings were constant to provide a dwelling time ('Dwell') of 400 ms and repetition ('Delay') 300 ms in all trials. These values can be changed in the software by clicking in the respective label and moving arrow keys to increment or decrement them (Figure 2).

The block of trial test starts after selecting with F-pointer the space bar. This is the first character typed (not included in measurements of speed in CPS nor wpm), the origin of distance to any other character in the QWERTY layout, and the starting point in time measurement. (See Annex 1 for a relative approximate distance between each letter in the QWERTY layout displayed on the screen).

Each trial consists of 20 words displayed one by one. The words from english language were selected randomly from a text containing 150 words. The Table 1 shows some characteristics of this word set.

After starting the test, the first word appears and an audible icon informs participant to start the input by selecting the space bar. If a select operation occurred in a wrong character sequence respect to the input word, a beep signaled an error condition and the wrong character was recorded by the software but not displayed.

File	Twords.txt
words	150
unique words	150
minimum length	6
maximum length	14
average word length	9.0
words containing non-letters	0
letters	1350
correlation with english	0.8012

Table 1. Word set used in the experiment

Participants were instructed to take small pause between words since the starting time is only measured after selecting the space bar. In a case of several errors or character time-out exceeded 10 seconds the program instructs to continue with the next symbol by audible feedback.

Before gathering data, the task and the selection technique were explained and demonstrated to subjects. The total time spent by each subject was approximately one hour and 15 minutes while the performance test range from 45 to 50 minutes.

Design

Pointing Performance

The design of the experiment used F-pointer device and text entry task as a between-subjects factor. The controlled variables used in experiment were set in the software like follows:

Target Width: 40 pixels (16 mm)

Target Distance: from 20 to 369 pixels (Annex 1)

Trials: from 1 to 20

Blocks: from 1 to 7

The target sizes was of 40×40 pixels that corresponds to on-screen width and height about 16 mm respectively. The dimensions of on-screen characters drown inside the target were about 20 x 20 pixels in the Microsoft Windows environment. The distances between target letters in the QWERTY keyboard layout displayed on the screen (Annex 1) represent and Index of Difficulty from 0 to 3.47 bits.

A list of example words and relative distances needed for retyping is presented in Table 2. Note that typing starts when the cursor is over the space bar.

Word	Distance from key to key in pixels
theory	100, 28, 126, 220, 180, 60.
approach	90, 362, 0, 220, 180, 322, 90, 126.
comparison	20, 253, 113, 144, 362, 126, 140, 322, 282, 144.
application	90, 362, 0, 28, 28, 215, 90, 165, 100, 20, 144.

Table 2. Word set used in the experiment

Variables measured in each trial were total time to input a character (Tt) and number of errors (Nr). In Figure 3 the total time Tt needed to input the first character "p" in the word "application" is composed by the several elements. The software records this magnitude in ms as:

$$Tt = Rt + Mt + Dt \quad (4)$$

where, Rt is the response time measured after subject received stimulus, process the action mentally and start moving the cursor from "a" to "p", Mt is the movement time from "a" to "p", and Dt is the dwelling time needed to detect "p" as desired selected char. This constant value Dt is closely related to the entry technique used so it is not compulsory in throughput calculations.

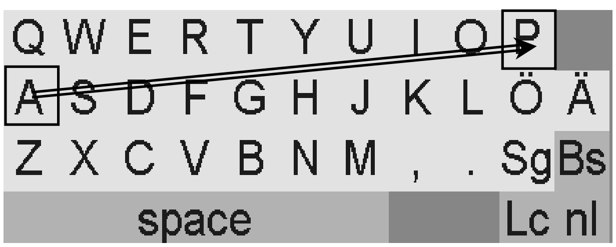


Figure 3. Total time to input "P" after "A"

The number of errors is measured when a retyped character is another one than required in test-word. We found this parameter valuable for the assessment of device since it allows us to calculate KSPC and error rate (ER).

Device Assessment Questionnaire

After test completion subjects were request to fill a device assessment questionnaire consisted of ten questions taken from common usability testing experience. Participants were asked to give a response to each question as a rating on a three points subjective scale.

The questions were related to aspects like consistency of system, simplicity of device, easy to learn, control by user, visual presentation, matching between the cursor and finger movement, error feedback, transfer of skills, and memory load.

ANALYSIS

The software tool used in the experiments collected the data for total movement time (MT) in character entry. Notice that from now on we will refer MT as Tt – Dt (see equation 4), to include both movement time and response time. The program computed automatically also the average and standard deviation of MT as well as the number of errors and word per minutes. These data were used later for analysis by calculating values of Throughput, CPS, efficiency and speed.

For throughput calculation we used the distances between characters in the screen keyboard (Annex 1), computed as the Euclidean distance between the starting and ending character. Notice that for non-diagonal adjacent letters on the QWERTY layout we ponder at this distance by half of real magnitude (20 pixels instead of 1) in order to differentiate the case of selecting consecutively the same letter by moving slightly the cursor (less than 10 pixels normally) in a target area of 40×40 pixels. In the rest of the cases the distance was calculated from centre to centre of targets. With these values the ID is computed and then throughput.

CPS was calculated as the number of input character divided by the input time in seconds. Efficiency (η) is the rate between the amount of test-word characters and the retyped word characters in each entry trial. KSPC is the reciprocal of efficiency. With these parameters we defined the speed rate as

$$Speed = \eta \times CPS \quad (5)$$

This definition includes not only the performance of the pointing device but also the text entry technique used.

RESULTS

Pointing Performance

The mean movement time MT was 1.6 seconds with standard deviation of 0.8 s. Error rate average was 26.8% with standard deviation of 6.7%. The computed throughput of all trials was 1.38 bps with deviation of 0.26.

The following graph (Figure 4) shows the best performance of the subject achieved after second trial.

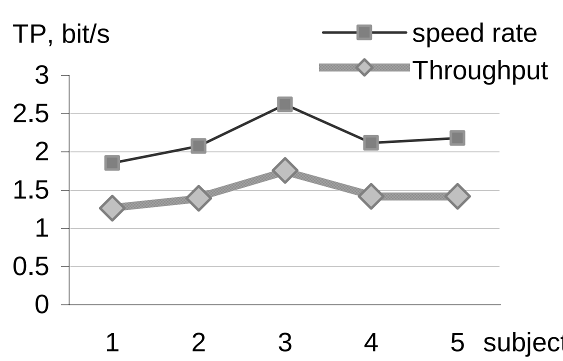


Figure 4: F-pointer: Pointing performance.

The line of speed rate is normalized in the Throughput scale to show the correlation between the pointing performance and the speed rate (5) in the text entry technique.

Text entry

The average efficiency of the device after eight trials was 73.4 %, which in turns affects the overall speed of 0.66 cps. It is worthy to signify that the mean velocity in characters per seconds without considering the efficiency rate reach the value of 0.91 cps.

The related graph (Figure 5) shows the best performance achieved by the subjects after second trial.

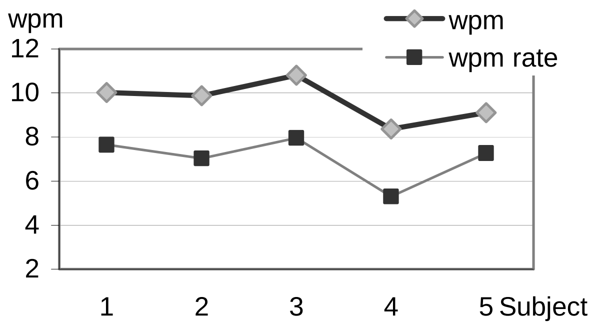


Figure 5: F-pointer: Text entry performance.

The distance between lines is the effect of the efficiency factor and the error rate.

Comfort

The stress of arm, wrist, hand or finger was not included in the device questionnaire in this experiment since we allowed taking small pause before starting each word entry. However participants complained about stress in arm muscles in some positions during the test. The force to operate F-pointer device was minimum. Subjects commented that device was too sensitive, and some effort was needed to keep the cursor stable once selected a target. One of the subjects was requested to rotate the position of the device and repeat the experiment, no difference was remarkable in results but interesting was the reversed mental mapping to accomplish task and the imperceptible time to mentally adjust to the new conditions. Another subject was requested to lift the device and use it with thumb finger simulating a mobile phone. No faster input was detected. Finally, we wanted to explore in one of the volunteers if changing the QWERTY layout for mobile phone-like layout would improve performance. No significant improvement was registered in pointing performance but a lower error rate.

DISCUSSIONS

During the study of the F-Pointer device we found useful the ISO standard for evaluating pointing performance. Throughout measurements include speed and accuracy as confirmed in the experiment results. The multi-directional selection on the screen keyboard made possible to include in this assessment accuracy and user response while immersed in a typical task. The changes in the keyboard layout redistribute the index of difficulty but do not affect the performance on the device.

All subjects agreed in the easy-to-learn quality of the device. We recommend increasing the amount of participants and redesigning the questionnaire to include descriptions about stressful conditions, movement mapping between device and cursor and degree of rejection in novice users.

The sensitivity of the device is attributed to interference of fluorescents lamps in the infrared detector and insufficient calibration. This effect added to the novice use of the interface (technique-device) and incorrect hand positions while executing the experiment might be the cause of stress in wrist muscles detected by some participants.

Possible Applications

It is noteworthy that this device could be used as stand alone (like in the experiment carried out) or integrated into another device. In the second case, assume that it could be inserted into a conventional keyboard and substitute the mouse, for instance, in the space between the alphanumerical and numerical keyboards. In a case of portable computer it could use the space dedicated to touch pad.

For those more familiar with mouse experience it is also possible to embed this unit to the left/side button to enlarge/replace the function of the wheel adding a fine control feature. This also could solve the problem of "clutching" in most of mice. The F-pointer device if integrated with analogue button (also developed by Grigori Evreinov), could support 3D manipulation using a ring metaphor. The ring allows interaction along X and Y while the analogue pressure detection could provide the Z-dimension.

Since pen is a common object in office environment, sophisticated ideas like the computer mouse pen [9], claims that muscles in the forearm and upper arm should be in resting when holding a pen device in a natural position. A normal pen can be also used to interact with F-pointer device rather than the index finger if desired. F-pointer is also suitable for mobile phones by replacing the joystick currently integrated in models with graphical screen. More ambitious will be to replace the T9 keyboard used in all models since the position of the finger will determine the number selected.

REFERENCES

[1] Mascaro, S and Asada, H.H. Finger Posture and Shear Force Measurement using Fingernail Sensors: Initial Experimentation. In *Proceedings of the 2001 IEEE International Conference on Robotics and Automation*, Vol. 2, 2001, pp. 1857-1862.

[2] Rekimoto, J. Gesture Rist and Gesture Pad: unobtrusive wearable Interaction Devices. *In Proc. ISWC2001*. Available at: http://www.csl.sony.co.jp/person/rekimoto//iswc01.pdf

[3] Matsui, N. & Yamamoto, Y. A New Input Method of Computers with One CCD Camera: Virtual Keyboard. Keio University, 2000. Available at: http://www.yy.ics.keio.ac.jp/~nozomu/research/vk/

[4] Douglas, S. A., Kirkpatrick, A. E., and Mackenzie, I.S. Testing pointing device performance and User assessment with the ISO 9241, Part 9 standard. In *Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI 99*, New York: ACM, 1999, pp 215-222.

[5] Zhai, S., Mackenzie, I.S. Teaching old mice new tricks: Innovations in Computer Mouse Design. In *Proceeding of Ergon-Axia '98*, 1998, pp 80-83.

[6] Havlik, J. Capacitance Sensor for Finger Motion Sensing. In *Proceedings of Biosignal 2002*, Czech Technical University, 2002.

[7] MacKenzie, I. S. Motor behaviour models for human-computer interaction. In J. M. Carroll (Ed.) Toward a multidisciplinary science of human-computer interaction, San Francisco: Morgan Kaufmann, 2003, pp. 27-54.

ANNEXES

Table 1: Distances in pixels between letters in the QWERTY screen keyboard layout

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	Ö	A	SP
A	**0**	165	90	60	90	100	140	180	282	220	260	300	243	204	322	362	20	126	20	165	243	126	28	28	204	20	340	380	90
B	165	**0**	60	90	113	28	20	28	144	90	126	165	60	20	179	215	179	90	126	60	113	20	144	100	90	140	204	243	20
C	90	60	**0**	20	60	28	90	126	215	165	204	243	140	100	253	291	113	90	28	113	179	20	90	20	144	60	282	322	20
D	60	90	20	**0**	20	20	60	100	204	140	180	220	165	126	243	282	90	28	20	90	165	28	28	28	126	90	260	300	60
E	90	113	60	20	**0**	28	90	126	180	165	204	243	179	144	220	260	60	20	28	60	140	90	20	90	100	113	282	322	100
F	100	28	28	20	28	**0**	20	60	165	100	140	180	126	90	204	243	126	20	60	28	126	20	90	90	90	126	220	260	60
G	140	20	90	60	90	20	**0**	20	126	60	100	140	90	28	165	204	165	28	100	20	90	28	126	126	28	165	180	220	60
H	180	28	126	100	126	60	20	**0**	90	20	60	100	28	20	126	165	204	90	140	28	28	90	165	165	20	204	140	180	90
I	282	144	215	204	180	165	126	90	**0**	28	20	28	90	113	20	60	260	140	322	100	20	179	220	253	60	291	90	126	170
J	220	90	165	140	165	100	60	20	28	**0**	20	60	20	28	90	126	243	126	180	90	20	126	204	204	28	243	100	140	113
K	260	126	204	180	204	140	100	60	20	20	**0**	20	28	90	28	90	282	165	220	126	28	165	243	243	90	282	60	100	144
L	300	165	243	220	243	180	140	100	28	60	20	**0**	90	126	20	28	322	204	260	165	90	204	282	282	126	322	20	60	179
M	243	60	140	165	179	126	90	28	90	20	28	90	**0**	20	113	144	253	144	204	113	60	100	215	180	90	220	126	165	90
N	204	20	100	126	144	90	28	20	113	28	90	126	20	**0**	144	179	215	113	165	90	90	60	179	140	60	180	165	204	28
O	322	179	253	243	220	204	165	126	20	90	28	20	113	144	**0**	20	300	180	282	140	60	215	260	291	100	330	28	90	200
P	362	215	291	282	260	243	204	165	60	126	90	28	144	179	20	**0**	340	220	322	180	100	253	300	330	140	369	20	28	233
Q	20	179	113	90	60	126	165	204	260	243	282	322	253	215	300	340	**0**	100	28	140	220	144	20	90	180	60	362	402	253
R	126	90	90	28	20	20	28	90	140	126	165	204	144	113	180	220	100	**0**	90	20	100	60	60	90	60	144	243	282	100
S	20	126	28	20	28	60	100	140	322	180	220	260	204	165	282	322	28	90	**0**	126	204	90	20	20	165	28	300	340	90
T	165	60	113	90	60	28	20	28	100	90	126	165	113	90	140	180	140	20	126	**0**	60	90	100	144	20	179	204	243	100
U	243	113	179	165	140	126	90	28	20	20	28	90	60	90	60	100	220	100	204	60	**0**	144	180	215	20	253	126	165	144
V	126	20	20	28	90	20	28	90	179	126	165	204	100	60	215	253	144	60	90	90	144	**0**	113	60	113	100	243	282	20
W	28	144	90	28	20	90	126	165	220	204	243	282	215	179	260	300	20	60	20	100	180	113	**0**	60	140	90	322	362	126
X	28	100	20	28	90	90	126	165	253	204	243	282	180	140	291	330	90	90	20	144	215	60	60	**0**	179	20	322	362	28
Y	204	90	144	126	100	90	28	20	60	28	90	126	90	60	100	140	180	60	165	20	20	113	140	179	**0**	215	165	204	126
Z	20	140	60	90	113	126	165	204	291	243	282	322	220	180	330	369	60	144	28	179	253	100	90	20	215	**0**	362	402	90
Ö	340	204	282	260	282	220	180	140	90	100	60	20	126	165	28	20	362	243	300	204	126	243	322	322	165	362	**0**	20	215
Ä	380	243	322	300	322	260	220	180	126	140	100	60	165	204	90	28	402	282	340	243	165	282	362	362	204	402	20	**0**	291
SP	90	20	20	60	100	60	60	90	170	113	144	179	90	28	200	233	253	100	90	100	144	20	126	28	126	90	215	291	**0**

Table 2: Device assessment questionnaire

Question	Disagree	Middle	Agree
I think I would like to use this device frequently	0.4	0.6	0
I found the system complex	0.4	0.4	0.2
I thought the device was easy to use	0.2	0.4	0.4
I think I would need train to use this device	0	0.6	0.4
I found the functions of the system well integrated	0	0.6	0.4
I thought there was too much inconsistency in the system	0.8	0.2	0
I imagine most people would learn this device quickly	0.4	0.2	0.4
I found the system very cumbersome to use	0.6	0.4	0
I felt very confident using the system	0.6	0.2	0.2
I need to learn a lot of things before getting started	0.8	0.2	0

Target Selection under Time Pressure Conditions

Zhang Weiwei
Department of Computer and Information Sciences
University of Tampere, Finland
+358 405 737 798
Wei-wei.zhang@uta.fi

ABSTRACT
Many factors can be involved and negatively influence on a decision-making. This project presents an exploration of the new technique for analysis of human behavior in simulated stress-conditions: two-hand coordination during the task of pointing and selection the targets under time pressure. As mentioned by S. Keele [4], planning goal-directed actions involves two important components: first, the intention to attain an action goal or to experience a desired event after executing an action; and, second, the selection of a motor program for a movement that elicits precisely this event. What would happen if different hands would control X and Y coordinates separately? What strategy will be chosen in critical situation? A new input device, the analog buttons, and designed for this purpose software allow carrying out the objective investigation of human performance and analyze data concerning individual behavior patterns. Handedness and performance are two main competitive factors, which determine behavioral dynamics of the subjects. Under time pressure conditions we have observed a variety of techniques with different accuracy-speed trade-off. However, we can say that there are, at least, three fields along track to the target and three temporal intervals when dominant hand and strategy can change, they are: long-distance behavior 0-600 ms; tuning for capture - two-radius of the target and 300-600 ms duration; approach – middle-distance behavior, the most variable tracking part.

KEYWORDS: Time pressure, target selection, decision making, human behavior, leading hand.

INTRODUCTION

Both input devices and interaction techniques are usually evaluated with a set of standard tasks. A measurement of the user performance on task completion time and error rate of results in the performance analysis serves as the basis for refinement and redesign of the devices and techniques. However, observations at the performance level often overlook important information on how users actually accomplish the task, which may offer additional insights toward a better understanding of the interaction process and design solutions [2].

To understand the underlying processes of interaction there is a challenge to carry out process-oriented studies. In human motor control research, the study of the "micro-structure" has served similar purposes [7]. For instance, Zhai and Milgram studied both the performance and the trajectory of 6 DOF manipulations of 3D objects. The trajectory analysis revealed critical differences between devices in terms of coordination, which would not be found in time performance data alone [9]

In general, people often face problems that force them to make decisions when working, studying or doing in real life. Certainly, they could take many pressures from different aspects to complete that determine. For instance, extremely nature conditions (open space, deep-water research and so on), workload or studies under time pressure and others.

In this paper a detailed analysis will be done for subject behavior strategy in simulation of time pressure conditions with redistributed haptic sensation during button manipulation in 2D visual space, capture of the lateral targets. Haptic sensations should be perceived through decrement or increment of spatial coordinates and force applied to the finger in bimanual task.

Human handedness – the consistent preference for one hand in skilled manipulative tasks – is often said to be a defining trait. While other primates may demonstrate individual preferences for the right or left forelimb in reaching and in manipulating objects, only in human populations is there a consistent tendency for the right hand to be the preferred hand. Yet there is overwhelming evidence that where such cultural pressures are relaxed, a natural preference for the left hand in skilled tasks develops in as many as one person in six. [1, 5]

In a preliminary study involving a target acquisition task through the use of two analog buttons the dominant left hand could be explicitly identified despite the reports by some of the participants that they were right-handed. But it is necessary to investigate further the influence of the extreme conditions, which could clearly recognize leading hand and subject behavior strategy, for instance in the coordination task under time pressure.

Thus, the study was aimed to investigate:

– The features of subject performance in strongly restricted temporal intervals for decision making and adequate action (capturing targets), while a target exposition is progressively decreasing;

– Individual behavioral patterns being accomplished in a short time frame when controlled parameters (cursor coordinates) are redistributed between two hands.

In general, conventional mouse or joystick could be used to control direction and tracking targets. But these devices, as a rule, are hand oriented due to designing. Therefore special-designed analog buttons were used to provide the overall separation of the controlled parameters in both the hands (a finger pressure along X- and Y-axes).

METHOD

Participants

Eight voluntary participants with various educational backgrounds and of various nationalities were served in this study. This group, which was comprised of 4 males and 4 females, cover a range of 21 to 30. None of the participants had any previous experience with analog button input device. And most of them used their computer often but only with conventional mouse, touchpad or joystick. Of the eight participants, one was left–handed, two of them have not leading hand and others were right–handed persons.

Apparatus

Analog buttons were designed in TAUCHI Unit several months before this experiment and showed high stability of parameters, that is, an additional calibration was not required throughout the study. The using of silicon tube provided stability of mechanical parameters; the mode of current amplifier compensated nonlinear characteristic of optical force-displacement transducer. A construction of the buttons is schematically shown in Figure 1. There are only two controls both having the same design (Figure 2). One controls left-right motion (X-coordinate) of the cursor and the other controls its up-down motion (Y-coordinate).

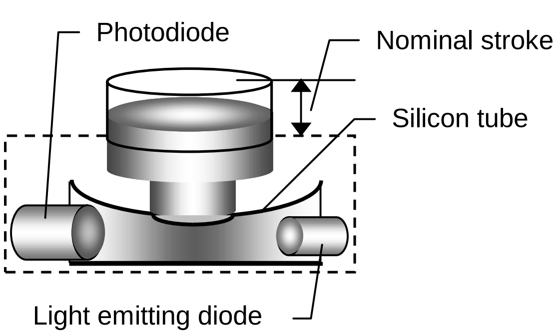


Figure 1: Design of the analog button

Figure 2: Overall view of the input device

Two software applications were written in Microsoft Visual Basic 6.0 for Windows 2000 (by Grigori Evreinov) and used in this project. One program was intended to collect data during experimental sessions. The behavior analysis was implemented through arrays of recorded tracks by other software.

Procedure

The test subjects were told that the goal of the experiment was to evaluate their behavior strategy in the target-capturing task and performance with new input device. In general, specific abilities of the subjects in these experimental conditions should depend on the leading eye or hand and motor coordination. Therefore, each subject was identified with a number of personal information.

The instruction before experiment contained the following message:

“There are 49 hidden targets. Please, use the buttons to capture on-screen targets. The targets will disappear after some time and appear again after 1 s. Do it as fast as possible and don’t stop if a mistake occurs. When the time is over, you will receive sound feedback or instructions as speech cues. If you wish, you can have several trials before testing.”

The complete testing duration was not more than 60 minutes per person 2 session × 5 trials × 49 targets, including training time, breaks and instructions.

RESULTS AND DISCUSSION

Balakrishnan and Hinckley [3, 8] presented a comprehensive overview of two-handed interaction techniques. They have considered many important factors such as focus of visual and kinesthetic attention and came to conclusion that “increasing task difficulty, divided attention, and lack of visual integration can all cause the user to adopt a more sequential style of interaction”.

Individual skills and experience can erase congenital status of the dominant hand, influence on performance of different types of movements. Length and direction of traces are also essential restricting factors of the speed in two-handed target-capturing task. Therefore we did not use any controls (rotary potentiometers, cranks, joysticks) which involve in motion most of the hand muscles. Some subjects have used forefinger others – thumb, while a dynamic range was only 4×4 mm along X and Y-coordinate accordingly.

As there were not restrictions on using analog buttons, we observed two ways of manipulation by them. Six subjects held buttons in their hands and used two thumbs to press buttons that are easier to control the tension. Two persons used forefinger to press buttons fixed on the table because it was more comfortable for them. In both cases the participants asked a pause when they have done half of test due to a tension in their fingers. That is a really hard work for controlling pressure with one finger on each button.

During test procedure with increasing task difficulty – decreasing targets exposition time – we have observed that a notion of the leading hand is very dynamical and can vary within one attempt. First, we tried to analyze data proceeding from the assumption that differences between ambidextrous technique and sequential actions with dominant and non-dominant hand should be essential throughout the test and will be more and more obvious after changing time of target exposition. Statistical analysis did not show significant differences within errors range for right-handed and left-handed groups F = 0.95, for left-handed and ambidextrous F = 0.88 and for right-handed and ambidextrous F =0.84.

Another finding was that under time pressure the subjects used different strategy. Some of them still tried to capture all targets although it seemed that it was impossible. Others only had time to choose some areas near the target to make sure that they can catch some of targets.

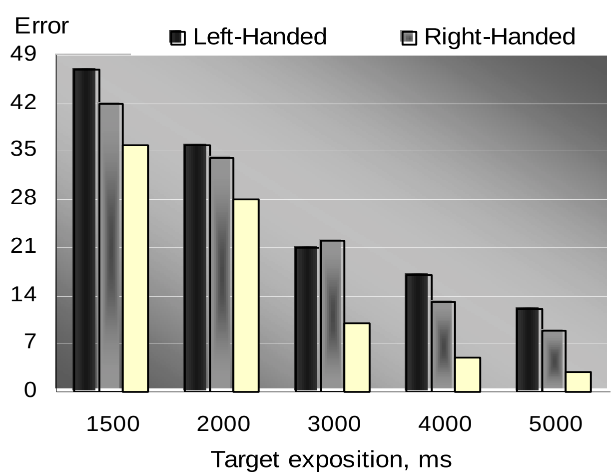


Figure 3: An average of errors for left-handed, right-handed and ambidextrous subjects

Figure 3 shows an average of errors for three groups of subjects. The ambidextrous subjects were done well in this case. At least, a less adjust time was required when time pressure condition had changed.

When target exposition time decrease, a number of missing targets is increase immediately. However, relative number of missed targets between groups of the subjects is almost the same when the target exposition time was 4 or 5 s. When target exposition time was changed to 3 s the difference between ambidextrous subjects and other groups was increased, but it still more than half targets were captured. When exposition was successively decreased a number of missed targets were progressively increased. Although 2 s or 1.5 s are quite short time for catching targets, the subjects had different behavior within these intervals.

Impact Time Pressure on Task Performance

Noteworthy, some of participants always used their right hand since the target appeared, to redirect the cursor. That is, as the X-button (coordinate) was always in their right hand, they will press it at first to track target. When they were asked to change X- and Y-buttons position in hands, it was difficult for subjects to change a strategy without readaptation. Training took some time "to adjust hands" and to select right strategy (Figure 4). Similarly, other subjects started tracking by the left hand. But this did not mean that right or left hand would have leading (dominant) position up to the moment when target will be captured. Figure 4 shows performance when asked this object to X-Y position. Although he is a righted – hand people, when he changed X button to right hand, the result is not satisfied with exposited time, even it is more than 4s. But after adjusting, it would do better.

Figure 4: Changing hands to manage by coordinates, one subject.

If we will consider times or lengths of the tracks throughout the test it cannot help to evaluate performance or specific features in behavior of the subjects. Figure 5 shows that 94% targets were captured within the interval 600-3000 ms by the subject of ambidextrous group. Meanwhile, if target exposition was decreased to 3 s an error rate was changed from 5% to 11% in ambidextrous group and up to 22% in the

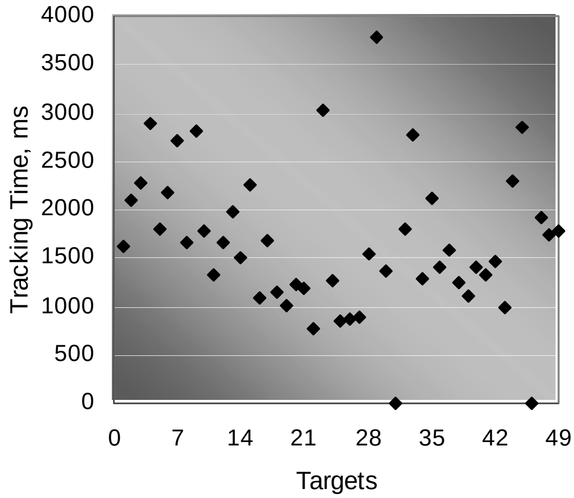


Figure 5: Times of the capture at exposition 4 s per target

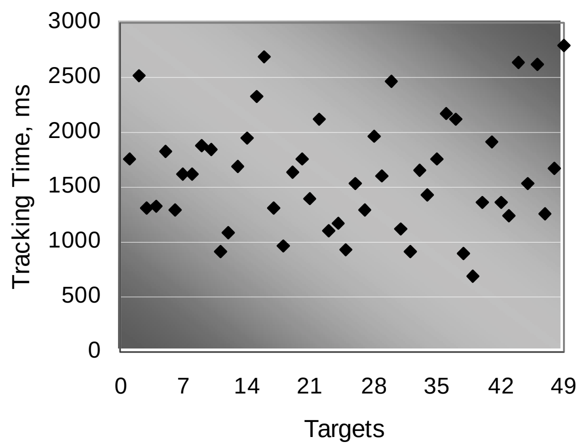


Figure 6: Times of the capture at exposition 3 s per target

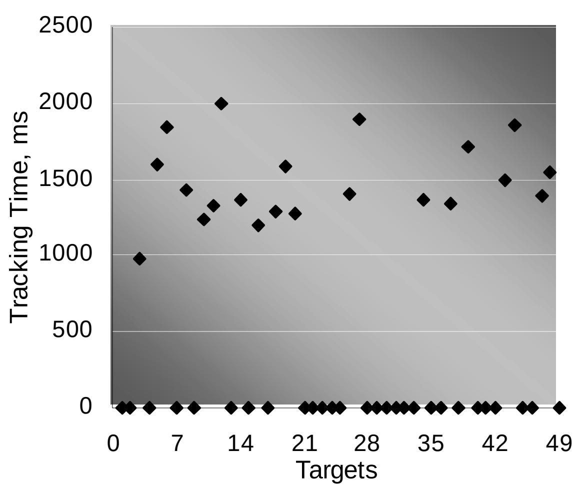


Figure 7: Times of the capture at exposition 2 s per target

groups with leading hand (Figure 3). That is, under time pressure a tension is increased and the subjects lost the time to re-coordinate or to synchronize goal-directed activity of the hands.

Figure 6 shows that all of targets were captured within the interval 500-3000 ms by the subject of ambidextrous group. When exposition time was shorter (2 s) only 57% targets were captured (Figure 7) and reply time has been lengthened (1000-2000 ms).

Individual Patterns of Behavior

Because different people have different experiences and capability, using the same event they would have different decision and behaviors. The following diagrams (Figure 8) present typical behavioral patterns of three groups of the participants.

The target position is zero and cursor can be from the left or from the right of the target, it depends on the state of the buttons. If the subject prefers to release the buttons after the target was captured, the patterns will have asymmetrical view (Figure 8, 3^{d} line and, partly, 1st line) and the lengths of the tracks will be longer. The third line shows analysis results from righted – handed people. According to these two diagrams, there were not recorded tracks below 0. All trails were begun from the left top corner of the screen. By other words, patterns show the person set free the button and cursor was returned back to this position every time, and the next target will mainly appear from the same side.

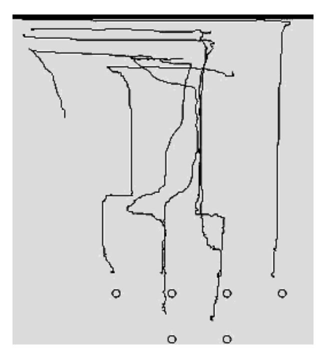


Figure 9: Buttons are free after capture of the target

For instance, six points of random sequence were selected from all tracks and presented at Figure 9. The most of tracks were begun from the left top corner, while these six points are quite near; the subject still used the same strategy.

This approach is easier to control cursor position with the help of two buttons when controlled parameters are redistributed between two hands. But this strategy takes more time. As can be seen in diagram, there is a long distance between the start cursor position and capturing the target. So, when the target exposition time changes to 2 s or 1.5 s, to prevent errors a strategy should be changed.

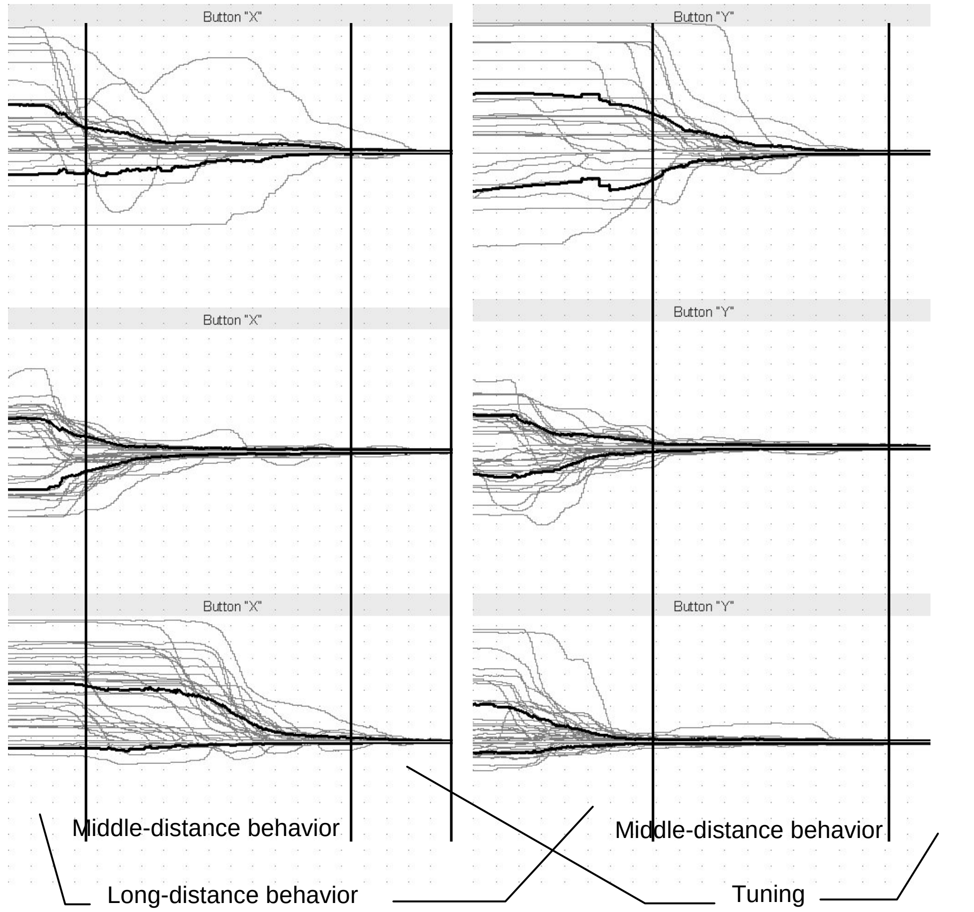


Figure 8: Dynamics of individual behavior strategies for left-handed (1st line), ambidextrous (2d line) and right-handed (3d line) groups of the subjects. Length of records and targets exposition 3000 ms

Figure 8, 2d line shows another situation (average track) with the same temporal window (target exposition 3000 ms). The pattern of subject behavior is symmetrical; any dominant start positions or movements are not presented. These tracks just depend on appearance of targets from any side of the current cursor position.

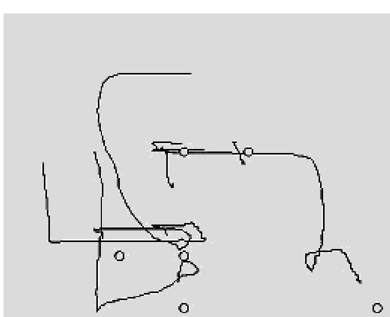


Figure 10: Cursor is near the captured target; buttons are pressed down

That is the subject is holding the buttons in non-zero state. This statement could be demonstrated through 2D-tracks. In Figure 10, although the targets are quite closely, there are not any dominant directions or similar tracks.

This strategy is quite random; capturing the targets is difficult and inefficient. But it could be the same time in similar cases, as the targets have appeared on the same distances or quite closely.

Self-Imposed Behavior Strategy

In the next section we are going to compare a behavior of the subjects under time pressure conditions, in particular, when target exposition was 2 s and the subject used only self-imposed strategy. While, we observed two evidently different strategies with capturing the targets (Figure 11 and 12), there were not found essential statistical differences in task performance under time pressure conditions among the groups. The subject who used an approach to targets from one corner was equally successful as the person moving the cursor by two hands simultaneously and hold buttons in pressed state. And the number of errors was proportionally greater when exposition time was shorter, as he had time only to catch closest targets in a short distance. He tried to track targets in other parts of the screen, but fail with that. Other subjects were more accurate and fortunate, in capturing the targets with short exposition time.

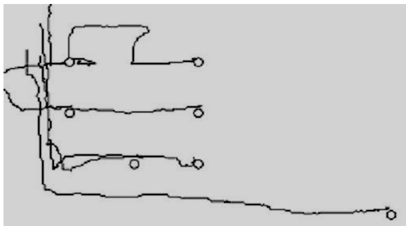


Figure 11: Self-imposed strategy – the left button is pressed, the right button free after capture the targets

The Figure 12 shows another – "freestyle" technique when the subject caught more targets He gave up some targets which are in a corner or far away but that allows to choose one suitable place to keep a cursor on average distance to the most of other targets. It is more handy approach. While being in the same conditions different people could have different strategies, one of strategies could be more suitable and have benefits.

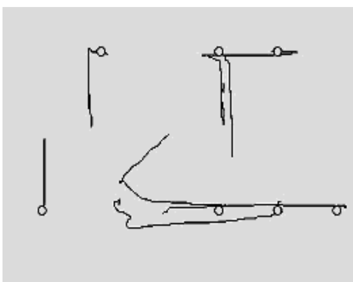


Figure 12: Freestyle – both buttons are pressed after capture the targets

Compare with Common Tools

We have tested conventional mouse and joystick to control cursor under time pressure conditions and tracking on-screen targets. The tendency of the results was similar, but mouse manipulation was slower in the first phase (long-distance behavior) and more accurate near the target (tuning).

CONCLUSIONS

In comparison to performance analysis, process-oriented study, especially points tracking trajectory analysis, is much less mature and more complex. The traditional techniques in studying performance, such as statistical variance analysis of means, did not produce informative results in our study. When we averaged button-tracking data in order to perform variance analysis, we lost much of the information contained in the data. Detailed, individual behavior analysis proved to me much more information about dynamics of strategy of the subjects. Overall, we found that participants used a variety of combinations of hand coordination patterns in various segments of tracks. This means that the design of input device algorithm should take all of these patterns into account. As mentioned by Zhai, designers cannot assume the one fixed hand coordination pattern found in direct hand pointing [8, 9].

Handedness and performance are two main competitive factors, which determine behavioral dynamics of the subjects. Under time pressure conditions we have observed a variety of techniques with different accuracy-speed trade-off. However, we can say that there are, at least, three fields along track to the target and three temporal intervals when dominant hand and strategy can change, they are: long-distance behavior 0-600 ms; tuning for capture - two-radius of the target and 300-600 ms duration; approach – middle-distance behavior, the most variable tracking part.

REFERENCE

1. Annett, M. and Kilshaw, D. Right-hand and left-hand skill II: estimating the parameters of the distribution of L-R differences in males and females. *British Journal of Psychology* 74 (1983), 269-283.

2. Card, S. K., English, W. K., and Burr, B. J. Attention on: Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics* 21 (1978), 601-613.

3. Hinckley, K., Czerwinski, M., and Sinclair, M. Interaction and modeling techniques for desktop two-handed input. In *Proceedings of UIST'98 Symposium*, ACM/SIGCHI, N.Y., 1998, pp 49-58.

4. Keele, S. W. Attention on: Transfer of learning among motor patterns with different relative timing. *Journal of Experimental Psychology: Human Perception and Performance*, 14 (1986), 241-252.

5. Kirsh D. Complementary Strategies: Why we use our hands when we think. *In Proceedings of the 7th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum, 1995.

6. Leganchuk, A., Zhai, S., Buxton, W. Manual and cognitive benefits of two-handed input. *ACM Transactions on CHI*, 5, 4 (1999), 326-359.

7. Mann, L. and Tan, Ch. The Hassled Decision Maker: The Effects of Perceived Time Pressure on Information Processing in Decision Making. *Australian Journal of Management*, 18, 2 (Dec., 1993), the University of New South Wales.

8. Grossman, T., Balakrishnan, R., Singh, K. An interface for creating and manipulating curves using a high degree-of-freedom input device. *ACM CHI Letters*, 5 (Jan. 2003), 185-192.

9. Zhai, S., Mailgram, P. Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. In *Proc. of CHI'98, ACM Conf. on Human Factors in Computing Systems*. (1998), pp. 320-327.

Usability Exploration of the Visual Textual Imaging

Tatiana Evreinova
TAUCHI Computer-Human Interaction Unit
Department of Computer and Information Sciences
FIN-33014 Tampere, Finland
Tel: +358 3 215 8569
e_tg@cs.uta.fi

Oleg Spakov
TAUCHI Computer-Human Interaction Unit
Department of Computer and Information Sciences
FIN-33014 Tampere, Finland
Tel: +358 3 215 8556
oleg@cs.uta.fi

ABSTRACT
The lack of access to verbal communication with non-deaf individuals is a major problem for the profoundly deaf people. The goal of our work was to develop communication techniques through graphical imaging of textual information for the profoundly deaf and hard-of-hearing people. Our empirical research was focused on an exploration the visibility of proposed pseudo-graphic typeface in comparison with five conventional phonetic typefaces. Our results show that Impact and Styled typefaces were perceived easier and seemed to be rather legible than Courier, Arial, Comic and Times New Roman. The tachistocopic analysis of the amount of recognized token (target stimulus) among distractors showed that subjects had fewer recognition difficulties with target stimulus among distractors for Arial and Times New Roman typefaces. For font attractiveness, Comic was perceived as being more attractive than Arial and Courier, while Styled and Impact were perceived as more attractive than Times New Roman. Of the fonts studied, Impact and Styled appear to be the more visible. Besides, being the most preferred they could be perceived fairly preattentively. We suppose that the proposed Styled typeface may have wider applications such as public display systems for dynamic imaging of current financial events in stock exchanges or present environments with different constraints.

KEYWORDS: hearing impaired, pseudo-graphic tokens, syllabic tokens

INTRODUCTION
The lack of access to verbal communication with non-deaf individuals is a major problem for the profoundly deaf people. The hearing impaired has no way to access phonetic languages, and vice versa, the non-deaf individuals have difficulties to understand the specific languages of disabled people. As a rule, they are deprived of possibilities to listen speech including radio and television programs, and possibilities to communicate on the phone without special adaptation of the unit [4].

There is a range of communication methods open to these people. Some can manage lip-reading, some sign language, Blissymbolics or other reading codes, and some can manage to use the telephone if it is amplified or accompanied with an inductive coupler to a hearing aid. However, not all existing systems for such people are so fortunate.

Lip-reading requires a great concentration of visual attention and can be very stressful. Some of phonemes are indistinguishable and unpredictable words are not properly understood. This makes the task of decoding mouth movements extremely difficult [7].

Thus, for the majority of profoundly deaf people lip-reading is quite impossible. The small number of people who use such a method mainly succeed by relying on guessing.

The visibility of hands still remains one of the most problematic aspects in using the sign language, especially for those people who have grammatical disorders and aphasia. The issue of qualified sign language interpreters has been always crucial for the deaf community. Some of deaf individuals are not aware of the amount of information which many interpreters filter out. They cannot interpret everything in group situations, for instance: cross-talk, in which people interrupt each other and/or talk over each other and specialized terms, for which no signs exist yet in Sign Language [7, 8]. Thus, the interpreter has to spend time in finger spelling them and/or use a longer sequence of signs. Some of interpreters totally ignore this kind of information. Furthermore, the amount of students with hearing impairments in the majority of universities essentially exceeds the number of interpreters and causes additional difficulties for teaching [11].

New Interaction Techniques '03, Spring, 2003, University of Tampere.

Still the usage of subtitling signing for profoundly deaf people remains as an inefficient and extremely expensive method of information imaging (it costs on average 400 dollars per hour to subtitle a TV program). Feedback from these people indicates that there is a high level of transmission difficulties, including missing words, spelling errors, breakdowns and reliability problems across all types of programs. Moreover, they read and assimilate information at different speeds and there are particular problems in relation to children because of both their age and connected reading proficiency and degree of deafness [9]. Thus, hard-of-hearing people require unifying the knowledge between the phonetic and sign language through combining visual and strengthen linguistic feedback that accompanies spoken language comprehension.

The goal of our work is to develop communication techniques through graphical imaging of textual information for the profoundly deaf and hard-of-hearing people. The display should require minimal resources of the visual perception and the use of specific pseudo-graphic symbols may reduce a cognitive load if they can be perceived preattentively.

With the introduction of legal obligation to provide subtitling signing for deaf viewers within digital television environment, there have been many suggestions concerning of the improvement of signing-specific aspects such as the visibility of hands and the visibility of the oral components of signs. However, the preferred method of achieving high visibility of subtitling signing is expensive in terms of the cost of skilled signers and the duplication of television channels [8]. Therefore an essential amount of research is being done on less expensive methods which could be acceptable to the deaf viewers.

Another well-known method to provide a language transfer created by television program to the deaf viewers is so-called close captioning (textual subtitling without signing) or the written translation of the spoken language (source language) of a television program or film into the language of the viewing audience (the target language) [12]. Nevertheless, there are many contradictions such as technical and usability problems regarding of implementing typeface design for close captioning. On the one side, the space available for subtitles is up to 30% of the screen height to present the best balance. The text is displayed at less than 140 words per minute, which can mean that the dialogue is essentially shortened [11]. That is, reducing dialogue significantly aggravates both phonological awareness and logical clarity of a language transfer. On the other side, the problem of displaying typefaces in specific format which should be available to deaf viewers has to be considered, too.

For instance, one of the common approaches is to add black borders. This does not affect the subtitling typeface but increases information readability. Another approach is to stretch the picture; this might involve stretching the subtitling typeface, which would enlarge legibility [11].

The important factor such a typeface visibility has to be taken into account. Good visibility of typeface is especially important both for people with reduced vision and for ordinary users in unfavorable light conditions [8]. This feature was rarely considered as typefaces are developed.

Every token must be acceptable, recognizable and distinguishable. Such tokens need to have as smooth shape as it is possible. On the other hand, displayed data must be coded as economically as possible that means not to use more information elements than it is strictly necessary. However, if digital interface includes graphical objects then readability problems begin to arise. Reading dynamically displayed text is a rather different task from recognizing symbols only. The latter task neither has the benefit of contextual information nor it requires intellectual processing [7].

To facilitate the apparent difficulty in recognition and perception of the tokens we see a need in adaptation of existing typeface to something rather more appropriate. The characters other than English nevertheless supporting by phonetic awareness and similar to conventional syllabic tokens should be included.

DESIGN AND IMPLEMENTATION

While many authors recognize that legibility of typeface plays an important role during the language transfer process [13], the important factor as visual processing of textual information still remains beyond designing. Some time ago Dr. Grigori Evreinov had created a pseudo-graphic typeface called as Styled. The goal was to essentially simplify visual perception of the typeface and to strengthen legibility. The Styled typeface includes 26 pseudo-graphic tokens that are very similar to standard typeface and could be perceived relying on previous user experience.

Visual perception critically depends on orientation and directions that arise early in visual processing. Behavioral measurements reveal that the human visual system is more sensitive to horizontal and vertical orientations in the visual stimuli [6]. Some research found the striking correlation between neural activity and behavior, which demonstrates that during the visual processing of the exposed stimuli humans can also easily perceive another oblique directions [14]. That is why to facilitate associative interpretation new tokens were mapped to eight directions: -135°, -90°, -45°, 0°, 45°, 90°, 135°and 180°, from the exposed coherent direction, where positive and negative values represent clockwise and anticlockwise rotations. A variant of the Styled typeface is shown in Figure 1.

While there are many studies comparing readability of on-line and printed typefaces, there is no objective research on the visibility of typefaces, which are currently used for close captioning and are available in the selection of various styles.

In the present study specific methods for visibility evaluation of dynamically displayed graphic tokens are introduced. An attempt is also made to evaluate which of tested typefaces is the easiest to perceive during imaging textual information. Before the test it was supposed that pseudo-graphic typeface would be easier to perceive

preattentively than the other ones, as there is a similarity to habitual syllabic tokens and simplified form of visual stimuli. In addition, it was supposed that at a shorter exposition time the recognition of proposed typeface would be better than the recognition of traditionally used typefaces.

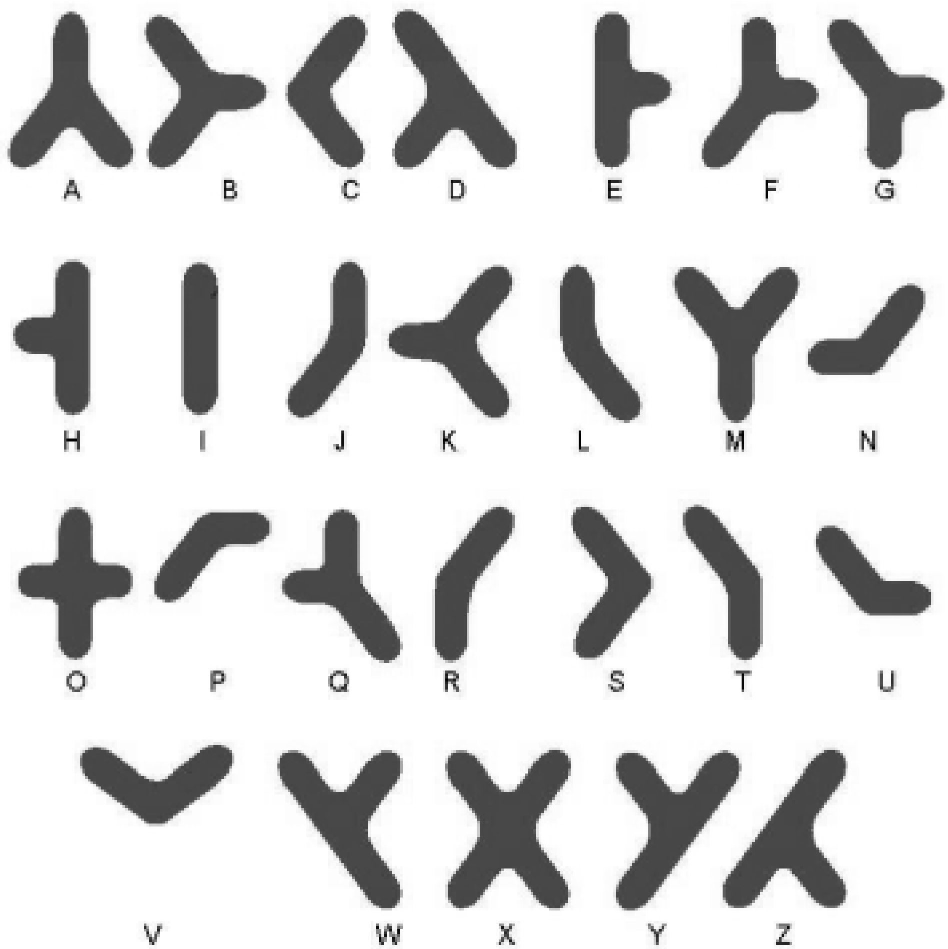


Figure 1: Styled typeface for the deaf and hard-of-hearing people.

EXPERIMENTAL EVALUATION

Participants

20 volunteers (13 males and 7 females) from staff and students at the University of Tampere participated in the study. The ages of the subjects ranged from 22 to 50 years with a mean age of 32. All had a normal color sight but different visual and hearing acuity. Seventy-three percent wore prescription glasses. None used hearing aid. The average computer experience of all participants was 8.2 years. All used computers on a daily basis, reporting 7 to 12 hours of usage per day. Ninety-five percent of the participants reported to have regularly read documents on computer screens, at least, a few times per week.

Apparatus

The experiments have taken place in the usability laboratory of Tampere Unit for Computer-Human Interaction (TAUCHI). Our testing was mostly based on using eye-tracking technology. Since eye tracking is tedious, we precisely planned the experimental procedure before starting. That is, we aimed to reduce testing time as much as possible to avoid subject over fatigue.

We used the SMI EyeLink tracking system to measure subjects' eye movements, fixations and scanpaths. SMI EyeLink system consists of headband-mounted camera assembly. Two custom-built ultra-miniature high-speed cameras provide binocular eye tracking. A third camera tracks 4 IR markers mounted on the visual stimulus display (Subject PC) for head motion compensation and true gaze position tracking. A processing system (EyeLink Operator PC) analyzes synchronously the images from all 3 cameras in real time at 250 Hz sampling rate to determine pupil position of both eyes and marker position. An Ethernet link connects the eye-tracking computer to an experimental-display computer, with support for real-time data transfer and control [5]. The tracking system operates with a very small delay of typically less than 12 ms between the eye motion and the time when data is available on the Subject PC.

Some time before this exploration Oleg Spakov wrote iComponent software, which provides a possibility to use SMI EyeLink data for other applications through a simple interface. The experimental software Font Reader was also written by Spakov in C++. Figure 2 shows a snapshot of the program Font Reader during an editorial mode for parameters tuning. Figure 3 shows a snapshot of the program during a performance of the test.

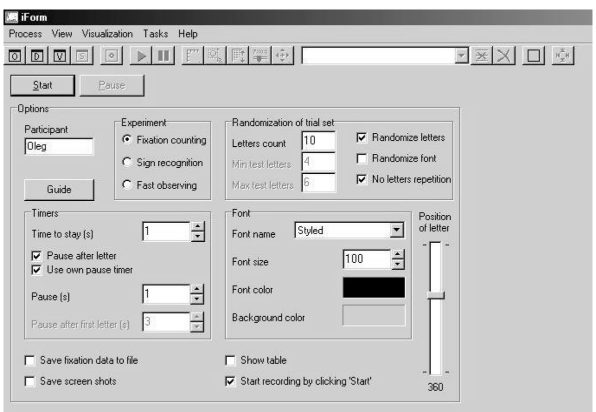


Figure 2: Snapshot of the Font Reader program in editorial mode.

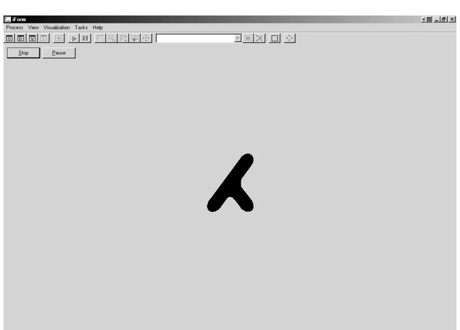


Figure 3: Snapshot of the Font Reader program during the testing.

Adjustments were made to maintain the subjects' eyes at approximately 45 cm from the 17-inch monitor. Each eye-tracking experiment was started with the calibration procedure according to the requirements of SMI EyeLink system. A spot was presented sequentially and randomly at nine different positions on the monitor screen of the Subject PC. The subject was asked to fix the gaze on the spots which appear on the screen.

During an experiment, the subject was wearing the headset of the eye tracker as shown in Figure 4. The stimuli were shown through the Subject monitor and the subject's current gaze position was indicated by two moving cursors (left and right eye cursors) on the Operator PC monitor (Figure 5) as the experiment was running. The recording of eye movements was carried out when external lightening was switched off to minimize influence of visual distracters.

Figure 4: Experimental setup (Subject PC) during the testing.

The standard iView 3.01 package provides interactive analysis functions for image-based stimuli. Recorded data and results are available for further post processing.

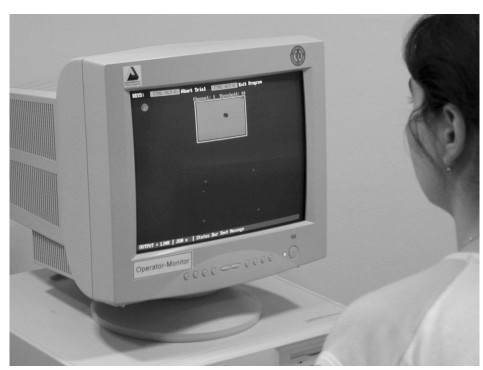


Figure 5. EyeLink Operator PC and data monitoring.

Motivation, experimental sessions and materials

Motivation

Our empirical research was focused on an exploration of the visibility of proposed pseudo-graphic typeface in comparison with conventional phonetic typefaces. The goal of this study was not to determine whether the designed typeface is better than those used habitually, but rather to define perceptive differences between the explored typeface and others, currently used.

Experimental sessions

The study consisted of three parts.

Experiment 1. We examined eye movements, fixations and scanpaths during viewing typefaces selected for testing based on objective recording of visual attention (gaze-position) through the SMI EyeLink system.

Experiment 2. An evaluation of selected fonts based on recording subjective and objective viewing time was carried out. To find out differences between the time when the subject had completed a viewing on his own judgment (subjective decision making) and when eye-tracking system had recorded fixation independently of the subjects' action, we measured visual scan time for typefaces.

Experiment 3. We investigated recognition of the tokens (target stimuli) selected from tested typefaces among orthographically similar tokens (distractors). It was suggested that pseudo-graphic typeface would be easier to perceive preattentively, because their shape was essentially simplified.

At the beginning of each experimental session the subjects were given written instructions explaining the task.

Materials

The following typefaces were compared throughout three experiments:

Selected typefaces	Motivation to select the typeface
Arial	The most commonly used sans serif font [1, 3, 2]
Comic	The mimic print facilitates syllabic tokens perception [1]
Courier	The most commonly used mono-spaced font [3]
Impact	Easy to perceive on screen, a large x-height factor [2, 9]
Times New	High legibility with economy of screen space [2]
Styled	8-directional simplified layout

Experiment 1: Tracking of the eye movements

Procedure

The subjects were shown samples of six typefaces in order of their appearance in the experiment. They were told that six consecutive passages taken of five phonetic typefaces and one of explored pseudo-graphic typeface would appear. One passage consisted of 10 visual stimuli, the stimulus exposed during the seven seconds. The order, in which visual stimuli were presented inside six passages, was randomized for each subject. The subjects were asked to gaze at visual stimuli and try to notice specific features they include. Presentation of a passage was started with pressing the button in experimental software. The whole experiment took 10 minutes in average. Both background and visual stimuli had a gray color. A stimulus-background brightness relation was 0.02. This condition was used to remove edge effects of involuntary gaze captioning.

Results

In this study about 1200 records (10 tokens per one passage, 1passages per tested typeface, 6 typefaces, 20 subjects) were gathered with SMI EyeLink system. An example of fixations and scan paths recorded by SMI iView 3.01 program for token "A" is shown in Figure 6.

The number of visual fixations was analyzed for each of typefaces. Our experiments demonstrated that Styled typeface has a lower number of visual fixations (14) than other typefaces (17-21).

The amount of areas of visual interest (AVI) was the next parameter used for a comparison of explored typefaces. AVI are defined as the areas where the most densely packed points of visual fixation are grouped (see Figure 6 and Figure 7). Before testing it was supposed that the novel typeface would evoke arousal reaction and strengthen visual interest than other ones. The results demonstrated that a higher number of AVI was observed for phonetic typefaces (5-7) than on Styled (only 4).

When asking subjects which of exposed typefaces could be preferred as easiest to perceive, they mostly selected Times New Roman, Arial and Comic. Subjects reported that the Times New Roman and Arial typefaces are "easiest to perceive" and "the most habitual", while those who selected Comic typeface commented that this typeface is "amusing" and "very interesting to gaze" (quoted from the subjects' comments). 73% of subjects identified Impact as "worst". 85% of subjects hesitated over the choice of the "best" typeface.

The subject who disliked Impact typeface gave "thickest" as the reason. Although, the subjects somehow characterized each of the phonetic typefaces, they were quite suspiciously related to the pseudo-graphic typeface. The overall impression regarding to this typeface was similarity of pseudo-graphic tokens (e.g., A, C, I, X, Y and Z) to syllabic tokens.

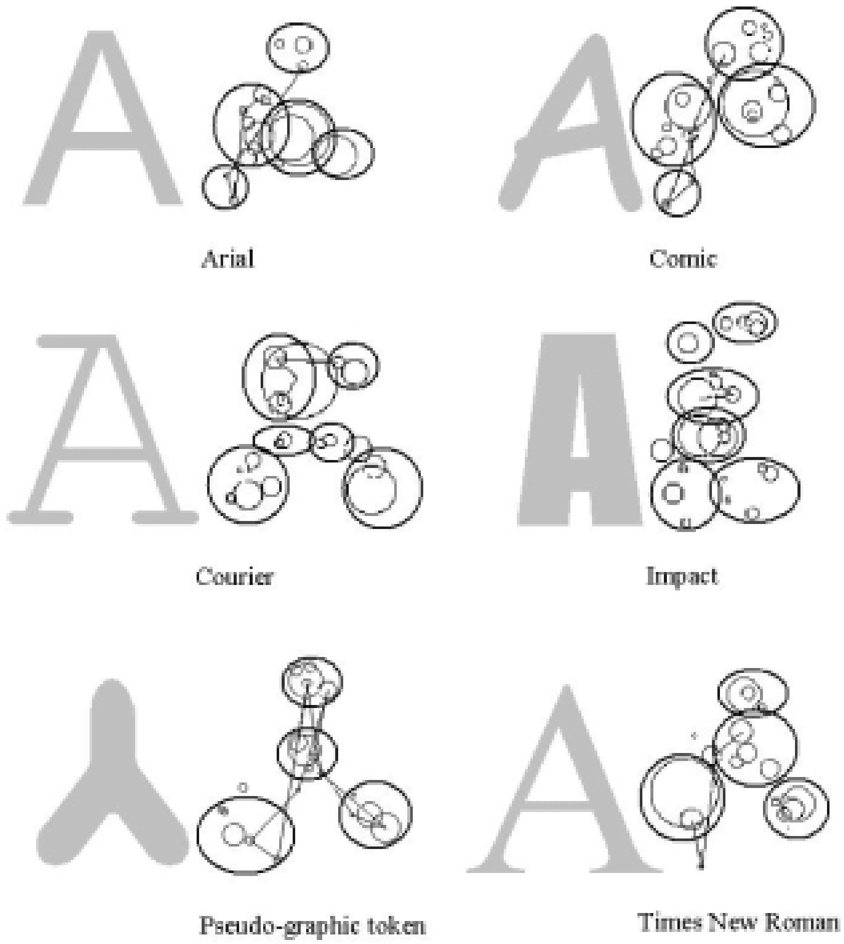


Figure 6: Areas of visual interest (AVI) recorded by SMI iView 3.01 program for the token “A”. AVI have been grouped automatically, thick black circles show groups.

A coefficient of difference (K_d) between the amount of AVI distributed across the alphabet in Styled typeface (St), in relation to an amount of AVI distributed across the alphabet in phonetic typefaces (Ph), was estimated as follows:

$$K_d = \frac{St - Ph}{Ph} \times 100\% .$$

Figure 7: AVI distributed across the Styled typeface in relation to AVI of phonetic typefaces. The coefficient of difference (K_d) is shown along Y-axis.

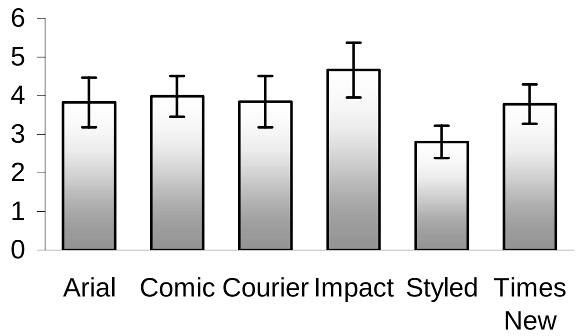


Figure 8: The average value of AVI (Y-axis) during an exposition of visual stimuli (X-axis).

That is, the amount of AVI for pseudo-graphic tokens, which were orthographically similar to the habitual syllabic ones, was essentially lower than the amount of AVI for conventional syllabic tokens (see Figure 8).

With these results in mind, it was hypothesized that Styled typeface has to be significantly more easily perceived than phonetic typefaces, since both the smaller amount of visual fixations and a lower number of AVI were recorded. However, we would not make any final decision concerning exceptional advantages of the Styled typeface.

Experiment 2: Visual scan time measurementSubjective evaluation

Procedure

The procedure was the same as in the first experiment but a passage consisted of 26 visual stimuli (alphabet). When the subject had completed a viewing on his own judgment (subjective decision making) he pressed down a space key. The session took 10 minutes in average.

Results

We have not found any statistically significant differences between the number of fixations for phonetic typefaces and pseudo-graphic typeface regarding to our previous experiment. Such a result was fully expected since the viewing of exposed visual stimuli was controlled more by the subject than by the interface.

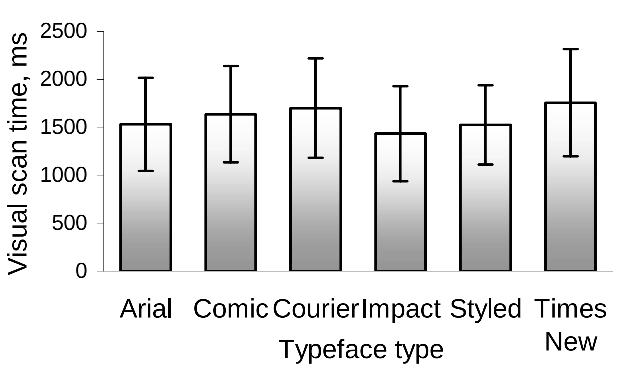


Figure 9: Visual scan time and standard deviation (vertical lines) of visual stimuli exposition in self-control mode.

However, when we examined the visual scan time of selected typefaces, we found the differences. While our subjects described the Times New Roman as “easiest to perceive” and “the most habitual” typeface during the first experiment, the viewing of Times New Roman typeface required more time to be perceived and was about 1755 ms. The pattern of results (Figure 9) suggests that there is a small advantage in the visual scan time of Impact (1434 ms) and Styled (1523 ms) typefaces over viewing of Arial, Comic and Courier typefaces (1529, 1635 and 1698 ms correspondingly). Probably the serif typefaces require greater visual scan time due to their smoothing effects which tend to strengthen depending on lowering stimulus-background contrast. It aggravates visual perception of these typefaces. We did learn that at the beginning of our testing the participants stated they had Times New Roman as their default typeface but chose Impact as the most legible and visible typeface after this study.

Thus, such a factor as a stimulus-background brightness relation can crucially affect both visual perception of typeface and the subjects preferences. Moreover, we noticed that the serif typefaces as well as pseudo-graphic typeface seemed to be more legible in this case than sans-serifs ones.

Objective evaluation

Procedure

The subjects were also wearing the headset of the eye tracker. Six consecutive passages composed of five phonetics typefaces and the pseudo-graphic typeface. One passage consisted of 5 visual stimuli. Visual stimuli were exposed in the same order (non-randomized) for each subject. The subjects were asked to gaze at exposed stimuli and try to notice their specific features. The session took 10 minutes in average. In this case, background had a gray color and visual stimuli had a black color. A stimulus-background brightness relation was -0.8.

Results

Overall examining the subjects’ AVI for five tokens (A, D, F, J, U) showed that the tokens taken from Styled typeface were considered to be the most legible (the average value of AVI was 2.8). The average values of AVI for phonetic typefaces were: 3.7 - Times New Roman, 3.8 - Arial, 3.84 - Courier. However, a significant difference was observed in perception of tokens taken from Comic (the average value of AVI = 3.98) and Impact typefaces (the average value of AVI = 4.66), see Figure 10. The tokens U and J were defined as the most easily perceived throughout all the testing typefaces. The average value of AVI was about 2.45 for U and 3.2 for J. The tokens F, A and D were defined as the most hardly perceived tokens in Styled typeface. The average value of AVI was about 4.27 (F), 4.6 (A) and 4.2 (D). The majority of subjects indicated that perception difficulties of such symbols were due to their disparateness to habitual syllabic tokens selected from phonetic typefaces.

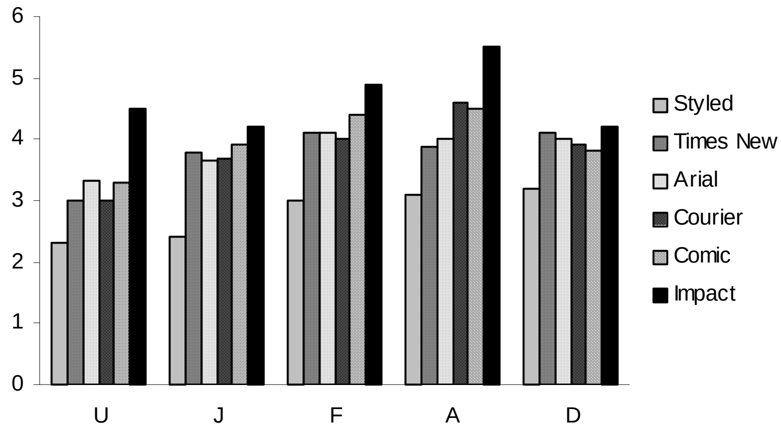


Figure 10: The average value of AVI (Y-axis) recorded by SMI iView 3.01 program for five tokens.

Experiment 3: Tachistoscopic evaluation of tokens

Procedure

Tachistocopic presentation of visual stimuli means that the stimuli should be exposed in the same place for extremely short periods of time using limits of unconscious visual perception to define how easily or hard the presented stimuli might be preattentively perceived by subject.

At the beginning of the experiment, subjects were told to fix the gaze on a spot located in the center of the screen to avoid additional errors due to an incorrect gaze position. During the session, the task of subject consisted of remembering of the first token (target stimulus) and accounting of an amount of this token appearing within presented sequence. Characters that are orthographically similar to the target token were served as distracters, for instance, if the token "A" was served as target stimulus then the token "M" was used as distracter for the token "A". The subjects were informed through the following message:

> "You will see a sequence of different tokens.
> At the beginning of each trial a token that you must remember will be shown.
> Then you should recognize this token in the following sequence of some amount of other tokens.
> Please, count up how many times you found out the first exposed token.
> Fill textual window after trial (the amount of recognized tokens excluding first exposed token)
> Press down the Enter key."

All sequences were composed equally of ten target stimuli and ten distractors. An exposition time for each token was 20 ms throughout the testing. Previous studies have shown that this duration is above the visual perception threshold [15]. However, interstimulus interval had been changed and was the following: 150, 300 and 600 ms. The test block included 3 trials for each exposition time, for a total 54 trials per subject, 9 trials per typeface. The session took 10 minutes in average.

Results

The tachistocopic analysis of the amount of recognized target stimulus among distractors obtained under the investigation of pairs was carried out at three interstimulus intervals.

The results of the recognition of target stimulus "A" are presented in Figure 11. As it was expected, the probability of an error depends on orthographic similarity of target stimulus to distractor and of the interstimulus interval. The recognition of a token is decreased in proportion to growing orthographic similarity of target stimulus to distractor and the error rate is also increased.

At the interstimulus interval of 600 ms the error rate was about 3% for all typefaces. When lowering the interstimulus interval to 300 ms the error rate grew up to 8%. At the interstimulus interval of 150 ms the amount of counted target stimuli became even greater than was actually presented within tested passage (the error rate was about 17%). The higher error rate for small interval was likely due to insufficient adaptation of subject to the first exposed token. Subjects indicated that they had fewer recognition difficulties with Arial, Impact and Times New Roman typefaces.

The full analysis of recognition of all the test sequences showed that the average value of recognized target stimulus among distractors was 62.27%. The better interstimulus time was 600 ms, overall error rate was 2.97%.

In general, the recognition of target stimuli (typefaces) depends on individual features of the person, such as the latent period of visual perception, concentration of attention on the task (tokens counting) rather than on the exposed target stimulus, or simply fatigue.

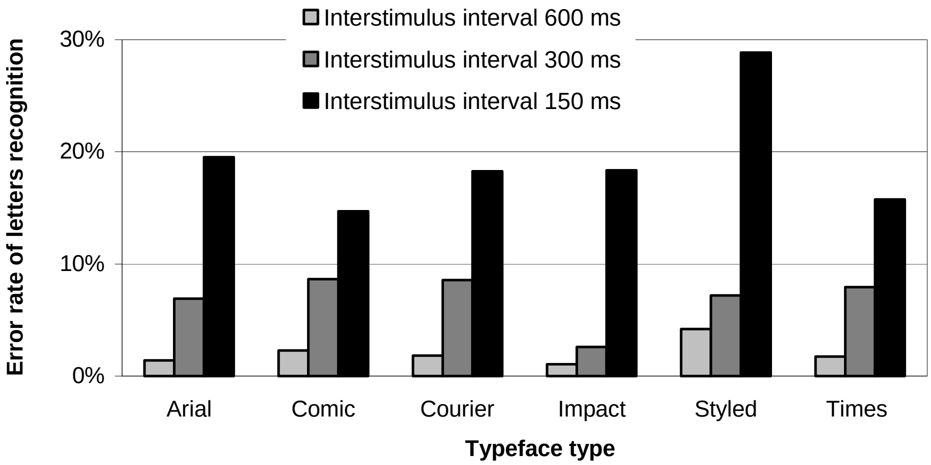


Figure 11: Recognition sample of target stimulus "A" at three interstimulus intervals

DISCUSSION

This study was focused on an exploration on the visibility of proposed pseudo-graphic typeface in comparison with conventional phonetic typefaces. Several observations can be made regarding the examined typefaces.

First of all, no significant differences were found in the number of fixations and visual scan time when subjects gazed at exposed stimuli. However, distinct differences were detected in the amount of AVI and subjective preferences of typefaces. Impact and Styled typefaces were perceived easier and seemed to be more legible than Courier, Arial, Comic and Times New Roman. The tachistocopic analysis of the amount of recognized tokens (target stimulus) among distractors showed that subjects had fewer recognition difficulties with target stimulus among distractors for Arial and Times New Roman typefaces.

For font attractiveness, Comic was perceived as being more attractive than Arial and Courier, while Styled and Impact were perceived as more attractive than Times New Roman. Of the fonts studied, Impact and Styled appear to be the more visible. Besides, being the most preferred they could be perceived fairly preattentively.

It is interesting to note that all subjects sat at the “ideal” computer range since SMI EyeLink system can be properly used only when the distance between eye and monitor is near 45 cm. That is, received data suggest that many people over 40 years old would need a spectacle correction for optimum viewing.

Considering the time scale within which we have had to work, this time spent was the optimal during of this testing could be managed. It is freely conceded that this testing is far from ideal usability investigation: all the subjects were to some extent self selected, and they were in no way stratified or subjected to any of the research criteria required for this kind of test. But we believe that comments we received from our subjects would help to indicate further areas where more work needs to be done. Although many helpful comments were made before data collection was started, some of the improvements concerning the experimental setup had already been made after observation of the preliminary results by our team. In fact, we came to terms that the new typeface, at the very least, represents a considerable improvement.

There have been several criticisms. One of the subjects reported that it is very hard to perceive preattentively the changing pseudo-graphic tokens when they were dynamically displayed in the same place. Another subject felt that it would be “irritating to perceive such simplified graphic tokens instead of habitual conventional letters”. Therefore the typeface might be essentially improved in the light of experience and further constructive criticism. Probably, it would be a great step in further development of Styled typeface if the most linguistically oriented framework could be build up to identify specific orthographic features that contribute to visual perception differences we observed. In a view of the absence of such a framework, we cannot properly suggest which of the following key factors, such as larger x-height, token smoothness, the issue of alternative size, distinguishability or simplicity, play crucial role in considering how visible or legible our explored typeface is.

We suppose that the proposed Styled typeface may have wider applications than use for subtitling only; for example, public display systems for dynamic imaging of current financial events in stock exchanges or present environments with different constraints.

ACKNOWLEDGMENTS

The research reported here was conducted at the Usability Laboratory of the Tampere Unit for Computer-Human Interaction at the Department of Computer and Information

Sciences of the University of Tampere. We wish to thank every volunteer test subject at TAUCHI for making our testing possible. A thanks is also extended to the students of the "New Interaction Techniques" course for participating in the experiment. We gratefully acknowledge Grigori Evreinov for giving in our disposal Styled typeface invented by him and his guidance in the early days of the testing and helping to us begin the initial data collection.

REFERENCES

1. Bernard, M., Mills, M., Frank, T., McKown J. (2001). *Which font do children prefer to read online? Usability News 3.1.*Available at: http://psychology.wichita.edu/surl/usabilitynews/3W/fontJR.htm

2. Boyardski, D., Neuwirth, C., Forlizzi, J., Regli, (1998). *A study of fonts designed for screen display*. In proceedings of CHI' 98, (April 18-23, Los Angeles), pp. 87-94.

3. Brookes A. *A guide to teaching and supporting students with disabilities.* Available at: http://www.aston.ac.uk/staffdev/public/tssd.doc.

4. Evreinova, T., Raisamo, R. *Multimodal communication system for the people who are deaf or have low vision.* Proceedings of the HANDICAP2002 2ème Conférence pour l'essor des technologies d'assistance. Paris - Porte de Versailles, June 2002, pp. 137-143

5. EyeLink Gaze Tracking. Project information available at: http://www.smi.de

6. Furmanski C. S., Engel Stephen A. *An oblique effect in human primary visual cortex*. Nature Neuroscience 3 (2000), pp. 535 – 536. Available at: http://www.nature.com/

7. Kilborn, R. *Speak my Language: Current Attitudes to Television Subtitling and Dubbing. Media, Culture and Society* (London), 1993.

8. Luyken, G. M. *Overcoming Language Barriers in Television: Dubbing and Subtitling for the European Audience.* Manchester: The European Institute for the Media, 1991.

9. Mercinelli M. *Guidelines-Accessibility requirements for new telecommunication equipment.* Available at: http://www.stakes.fi/cost219/smartphones.htm

10. Mills, C. B. & Weldon, L. J. (1987). *Reading text from computer screens.* ACM Computing Surveys, 19(4), pp. 329-358.

11. Quality of Sign Language Interpreters by Omer Zak. Available at: http://www.zak.co.il/deaf-info/old/interp_quality.html

12. *Television Closed Captioning.* On H.R. 4267. U.S. House of Representatives, One Hundred First Congress, Second Session, May 1990.

13. Voice of listener and viewer. Available at: http://www.culture.gov.uk/PDF/subtitle_vlv.PDF

14. Watanabe T., Nanez J. E., Sasaki Y. *Perceptual learning without perception.* Nature Neuroscience 413 (2001), pp. 844 - 848. Available at: http://www.nature.com/

15. Watt, R.J. and D.P. Andrews (1981) "APE: Adaptive Probit Estimates of Psychometric Functions," Curr. Psychol. Rev. 1: pp. 205-214.

Two Low-Footprint Approaches to Text Input by Physically Challenged Users

Oleg Spakov & Darius Miniotas
Department of Computer and Information Sciences
FIN-33014 University of Tampere, Tampere, Finland
+358 3 215 8556
{oleg, dm}@cs.uta.fi

ABSTRACT
Virtual keyboards used by physically impaired users for text entry usually display all characters simultaneously on the screen. Applications where saving physical space is critical, however, require another approach. A common option is grouping several symbols under one key and employing a hierarchical selection mechanism. We present an alternative technique (*Symbol Creator*) based on assembling characters using segments that resemble basic elements of Latin cursive. This technique was empirically evaluated along with another technique developed as an eye-gaze analogue to mobile phones' multi-tap system. In terms of accuracy and entry speed, both techniques performed equally well. Subjectively, however, Symbol Creator was the preferred method, since it was perceived as being more fun to use and less frustrating than the other technique. These findings suggest that designers of text entry systems for disabled users should not stick to menu tree-based systems – there are also other solutions to on-screen space saving problem.

KEYWORDS: Eye tracking, text entry, virtual keyboards, physically challenged users

INTRODUCTION

Eye gaze is probably the most frequently used input method in text entry systems intended primarily for physically challenged people. This is due to the fact that even severely impaired users generally retain good ocular motor control. In a standard application, the user enters text through a virtual keyboard present on the screen.

A traditional approach in designing such virtual keyboards has been to display simultaneously all the characters needed for text and numeric entry. This is the most straightforward approach since it exploits the direct analogy with a standard manually operated keyboard. Visual keyboards may have the standard QWERTY layout, or be arranged in some other order (e.g., alphabetical or optimized according to the statistically derived frequency of the letters for a particular language; see, e.g., [6] for a review).

New Interaction Techniques '03, Spring, 2003, University of Tampere.

Saving Screen Space

Taking into account the requirement for increased size of virtual keys due to limited accuracy of eye-gaze control, a keyboard with all the available symbols in view inevitably occupies a significant portion of the screen space. This narrows down the capacity and functionality of a GUI merely because much less screen space is left for other elements.

In applications where it is crucial to have as much screen space as possible to support interaction other than text entry, immediate accessibility of all ASCII symbols can be tradeoff against savings in physical space. Using this approach, it has already become a standard practice to group several symbols under one key and employ some hierarchical selection method to access individual symbols. Several early eye-typing systems were built according to this principle (e.g., HandiWriter in [8]; ERICA in [3] – see, e.g., [4] for a review), as well as some of the more recent ones (e.g., EagleEyes in [1]).

Grouping several symbols under one key, however, is not the only possible solution to the screen space saving problem. Perhaps a completely different approach would facilitate text entry by improving speed-accuracy tradeoff and/or making the process more convenient/enjoyable for the user? What if the user were given a task to assemble the character to be entered using a combination of, say, two items out of a limited set of such "assembly parts" instead of having to select a special key for that character stepping trough a menu tree? As with any human-computer interaction technique, however, one has to bear in mind that it should be easy for the user to learn in order to be seriously considered as an alternative input method.

Obviously, one is more likely to succeed in achieving this objective if one is able to utilize already existing knowledge and skills of intended users.

Skill Transfer from Elementary School

Statistics on spinal cord injuries (SCI) suggests that, for instance, in the USA 61% of the cases occur among people in the 16 to 30 years age group, whereas the share of the age group of 0 to 15 years is just 5% (http://www.cureparalysis.org/statistics/). This means that more than 95% of the people had already finished elementary school by the time they were affected by paralysis. It turn, this implies that the absolute majority of the people with SCI have the skills of writing in cursive.

Although there are several styles of Latin cursive script currently in use for teaching children, one can easily distinguish a limited number of basic elements most of the characters can be decomposed into. These basic elements can then be unified in a systematic way to yield a relatively small set of segments to be used for assembling letters and other symbols in text entry tasks. One of the strengths of such an approach is that it makes use of human skills acquired during several years of learning at school. This way, the user should be able to master the new technique in a relatively short time.

The current study compares two text entry techniques using eye tracker as the input device. User performance is evaluated empirically by measuring the speed and accuracy of text entry. One of the techniques is a variation of the well-known multi-tap technique originally developed for mobile phones and adapted now by us for eye-gaze interaction. (For simplicity, we will refer to this technique as *multi-tap* even though it has little to do with physical tapping.) The other technique called *Symbol Creator* uses a completely different approach for character entry as described below.

Symbol Creator

This technique uses seven letter segments implemented as eye-gaze activated on-screen keys (Figure 1). In addition, the visual keyboard accommodates a key labeled "End" that emulates pressing the spacebar. The "End" key has also a supplementary function of signaling the completion of the entry needed by some of the characters (see below for details).

For instance, to enter letter "i", the user is first to select segment #5 (Figure 1a, counting from the left), and then complete the entry by "striking" with their gaze on "End". Meanwhile, dwelling further on segment #5 would result in entry of either "u" or "w", depending on the duration of the gaze. Similarly, segment #3 can be used for entry of either "n" (double "eyestroke") or "m" (triple "eyestroke").

Intuitively, one can guess that letters "c" and "o" are selected by gazing on the segments having exact appearance as these letters (segments #1 and #4, respectively). As in the case of letter "i", entry of both these letters must be confirmed by "End" key; otherwise the software assumes a second segment will follow to complete selection of another character. For example, if the user selects segment #4 followed by segment #5, letter "a" appears in the text field.

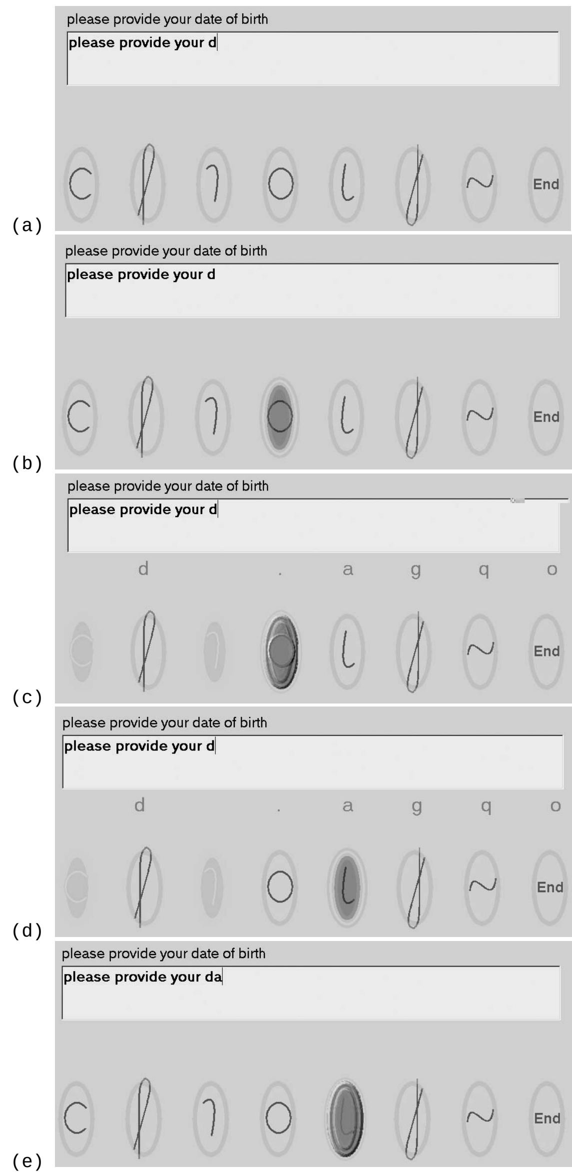


Figure 1. Entry of letter "a" using Symbol Creator

To illustrate the process of entering a letter using Symbol Creator, Figure 1 depicts the whole sequence of screenshots derived during input of letter "a". In the example, the user is asked to enter the phrase "please provide your date of birth". As the starting-point, we consider the situation where the user has already typed, "please provide your d", and now they have to enter letter "a" (Figure 1a). To do this, the user has first to look at segment #4. To give visual feedback to the user, the segment is highlighted as soon the gaze falls upon it (Figure 1b). The segment is not selected, however, before a dwell time has elapsed (set at 400 ms in our current experiment). This enables the user to verify correctness of the choice and shift their gaze to another segment if desired.

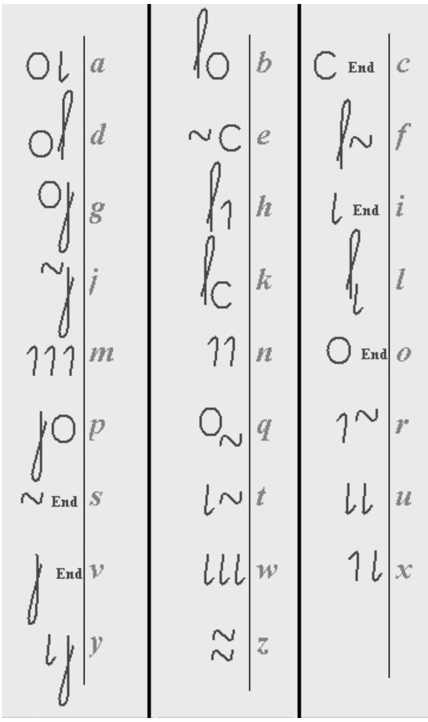


Figure 2. Assembly sequences for letters in Symbol Creator condition

When the dwell time is over, the selected segment is highlighted in a different way (Figure 1c). At the same time, a series of lower-case letters appear above respective segments that can be combined with the selected segment to form a letter. (This serves as a hint to the user.) Meanwhile, those segments that cannot be used in conjunction with the currently selected segment are disabled (turned gray).

To complete selection of letter "a", the user has now to gaze at segment #5 (Figure 1d). Upon expiry of the 400-ms dwell time, the segment gets the selection confirmation highlight (Figure 1e). Immediately after this, letter "a" appears in the text field next to the previously entered letter "d". The entry of letter "a" is thus completed, and the user can proceed to entering the next letter.

Figure 2 displays all combinations of the segments needed to enter any letter. For absolute majority of the characters, two steps are required to complete their selection. For the practical task of text entry, however, our method allows to save some keystrokes by skipping the confirmation of letters "n" and "u" by "End" key, which is optional unless these letters are followed by a space or another symbol starting from segment #3 or #5, respectively.

Multi-tap

For this technique, the on-screen keyboard layout is depicted in Figure 3. There are 8 keys for text entry accommodating the 26 letters in exactly the same fashion as the one used in standard mobile phones. In addition, there are two more keys. The leftmost key serves as the spacebar. It can also be used to enter full stops and commas, but those two symbols were not needed in the current experiment. The rightmost key is a functional key for switching between the different keyboards (not used in this study). As with Symbol Creator, all the keys are aligned in a single row at the bottom of the screen.

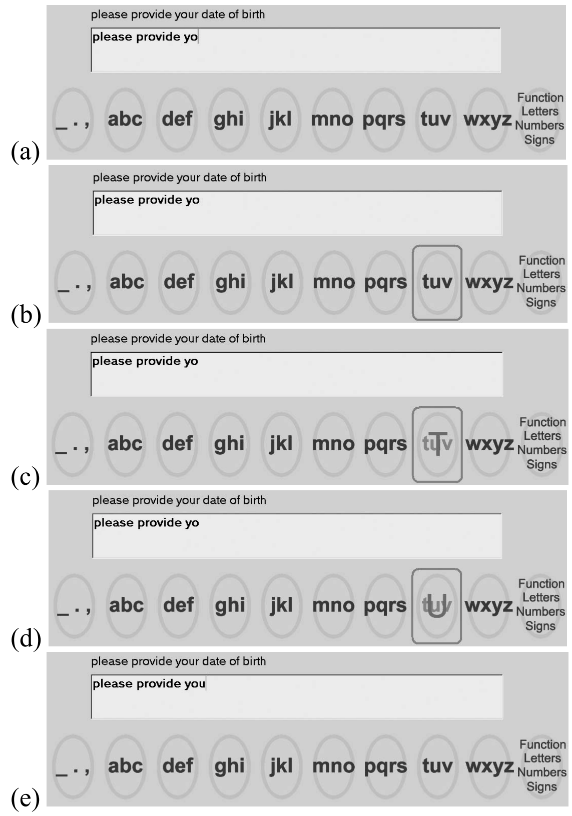


Figure 3. Entry of letter "u" using multi-tap

To select a letter, the user has first to gaze at the key containing that letter. After the gaze has been kept steadily on the key for a certain amount of time reaching the predefined threshold (set at 400 ms in the current experiment to keep full consistency with the Symbol Creator condition), the first letter contained by the key is highlighted (expands in size and changes color at the same time). If this is the desired letter for entry, the user is to confirm its selection by simply shifting their gaze to some other area on the screen. As a rule, the user shifts their gaze towards the key containing the letter to be entered next.

If the user's gaze, however, does not leave the current key's area upon highlighting of the first letter, the second letter on that key is highlighted after another 400 ms. Once again, the user has two options here: simply hold their gaze further on the key to proceed to the next letter, or shift their gaze away if the current letter is the required one. The software interprets the event of shifting the gaze away as the command to enter the letter highlighted last. That letter then appears in the text entry field.

As an illustration, Figure 3 depicts step by step the sequence of entering letter "u" using the multi-tap technique. Here the user is asked to enter the same phrase as in the previous example with Symbol Creator. Only now we have the situation where the user has already entered, "please provide yo", so they are about to type "u" next (Figure 3a). For this, the user first looks at "tuv" key, on which the required letter resides. The key is highlighted by rectangular of different color to indicate that the gaze points now at this key (Figure 3b). After the user has gazed for 400 ms at the key, enlarged letter "t" appears in the foreground of "tuv" key (Figure 3c). Since the user needs letter "u", they fixate further upon the key for another 400 ms until "t" is replaced by "u" (Figure 3d). The user then gazes away from the key, and letter "u" is entered in the text field (Figure 3e).

METHOD

Participants

Six subjects aged 23 to 46 (three male and three female) took part in the study. They all had varying degrees of computer experience, but just two of them had prior experience with eye tracking. All participants were able-bodied with normal or corrected vision.

Apparatus

The experimental software was developed using Borland C++ Builder 5.0 and ran under Microsoft Windows 98. The hardware for the experiment consisted of two PCs and a head-mounted eye tracking system EyeLink™ from Sensomotoric Instruments, Inc. A Pentium III 500 MHz was used as the participant PC. It was connected to the experimenter PC (Celeron 466 MHz) used for analysis of the captured eye images.

Procedure

The task consisted of entering phrases provided by the experimental software using one of two conditions. The conditions were: (a) Symbol Creator and (b) multi-tap. The phrases were retrieved by the experimental software randomly from the set and presented to participants one by one to enter. A group of five phrases was called a block.

In our study we used the phrase set compiled by MacKenzie [5]. This set comprises 500 phrases that are moderate in length (from 16 to 43 characters, with the mean being equal to 28.61), easy to remember, and with letter frequencies typical of the English language.

Each participant took part in five sessions of text entry. Each session lasted on average 30-35 minutes, including short breaks for rest and recalibration. The order of conditions was alternated for every session. Sessions were completed on five consecutive days, one session per day. Participants had to complete three blocks during each session. In total, each participant had thus to enter $5 \times 3 \times 5 = 75$ phrases for each condition.

Prior to the first session, participants were introduced to each text entry technique. This introduction included two stages of training. First of all, participants were showed a table listing all the symbols to be used in the text entry experiment. Although this stage of introduction was somewhat redundant for the multi-tap condition due to its self-explanatory nature, it was rather important for the other technique to have the participants study the new alphabet one by one first. The exposure was limited to 3 minutes for all participants (in the case of multi-tap condition, however, participants were free to decide whether they wanted to proceed to the next stage of training after a shorter period of the initial exposure).

After the first 3 minutes spent on studying the characters, another 12 minutes were given for the participants to try out each of the two techniques on the eye tracker. The relatively long training time could be explained by the fact that for most participants it was their first acquaintance with an eye tracker. Thus they needed to get familiar not only with each of the gaze-based text entry techniques, but also with the technology in general.

No data was recorded at this stage – any actions performed by the participants were treated as training only. The whole training procedure altogether lasted thus 15 minutes or less.

Subjects were instructed to aim for both speed and accuracy when entering the words. In addition, subjects were told if a mistake was made, they were to ignore it and continue with the phrase. Text entry speed was derived from timing values recorded for each letter. Error rates were evaluated using the Levenshtein minimum string distance statistic [7].

To help motivate subjects, summary data for accuracy and speed were displayed at the end of each block. An audible feedback click was produced upon the recording of a letter.

RESULTS

Learning Effects

The error rate fell from 9.5% in Session 1 to 3.3% in Session 5, and the main effect for session was significant ($F_{4,20} = 24.93$, $p < .0001$). The entry time per letter also improved significantly ($F_{4,20} = 137.97$, $p < .0001$). The mean fell from 2022 ms in Session 1 to 1371 ms in Session 5. The session × entry method interaction was significant both for error rate ($F_{4,20} = 13.02$, $p < .0001$) and entry time ($F_{4,20} = 12.71$, $p < .0001$). These effects are clearly seen in Figure 4. As can be seen from Figure 4a, there is a substantial reduction in error rate from Session 1 to Session 3 in the Symbol Creator condition followed by a slight and non-consistent improvement after that. For multi-tap, however, the error rates are not visibly different throughout the experiment.

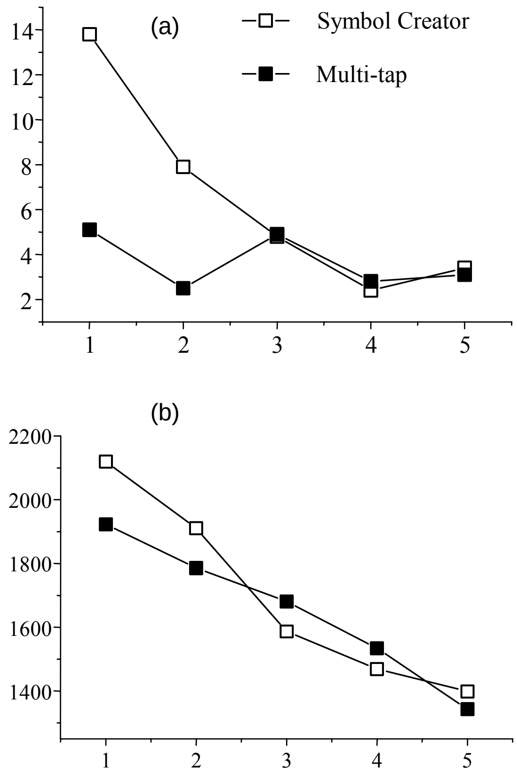


Figure 4. Performance over sessions: (a) error rate (%), (b) entry time (ms)

In Figure 4b, the entry time improvement over the 5 sessions is very evident for both the entry methods, but there is a noticeable difference in the learning patterns. In the multi-tap condition, entry time decreased almost monotonically, whereas Symbol Creator exhibited a more irregular course of improvement. In the latter condition, as many as 50% of the total reduction in entry time occurred from Session 2 to Session 3, making Symbol Creator superior in terms of text entry speed. Following Session 3, however, entry time decreased at a much slower pace ultimately reaching the performance level of multi-tap at Session 5.

Condition Effects

To test for condition effects independent of learning effects, a final analysis of variance was undertaken with the data from the last session. There was neither main effect for error rate ($F_{1,5}$ = .12, ns), nor for entry time ($F_{1,5}$ = 6.37, ns). The performance is compared in Table 1.

Condition	Error Rate (%)	Entry Time (ms)	Speed (wpm)
Symbol Creator	3.4	1399	8.58
Multi-tap	3.1	1343	8.94

Table 1. Performance comparison of the two techniques

The right-hand column converts entry time, in ms, to speed, in words per minute (wpm), for comparison with other studies and other entry techniques. (In keeping with the typists' definition, a "word" equals five characters.)

Performance by Letter

Looking at performance on a letter-by-letter basis reveals a key difference between the two conditions. In the Symbol Creator condition, for most of the letters it took on the average almost the same amount of time to enter them (Figure 5a). The exceptions here were only letters “m”, “w”, “n”, “u”, and “z”. The marked prolongation of entry time for letters “m” and “w” is the outcome of triple dwell times, whereas reduction in entry time for letters “n”, “u”, and “z” can be attributed to simplified entry procedure requiring only a single segment for letter assembly instead of two different ones (see subsection 1.3). Entry of space characters made a separate case, because it required only a single action of gazing at “End” key – hence the dramatic reduction in entry time (the leftmost column in Figure 5a).

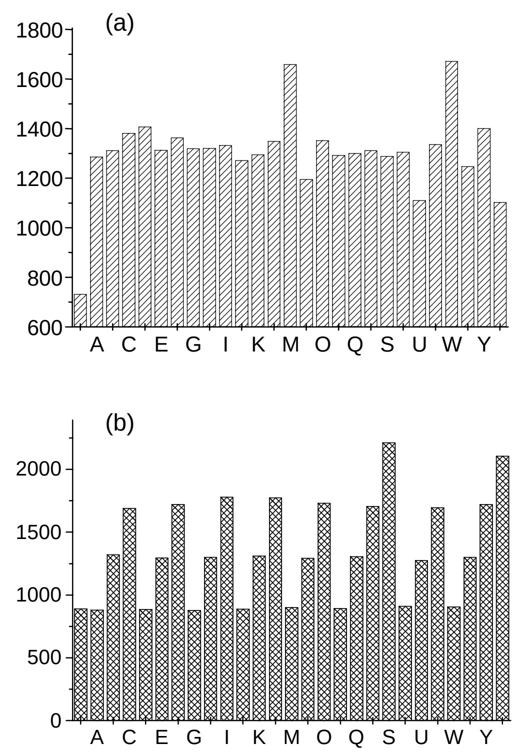


Figure 5. Entry time (ms) by letter: (a) Symbol Creator, (b) Multi-tap. The leftmost columns display entry times for space character

Meanwhile, in the multi-tap condition, entry time per letter depended heavily on the letter’s position relative to the first letter on the key (Figure 5b). This was largely determined by the nature of the multi-tap technique itself. Here, entry time is a quasi-linear function of a letter’s location on the key it belongs to. That is, for any of the front letters (e.g., “a”, “d”, “g”, etc.), entry time is minimal. It is composed of scanning time (the time needed to locate the letter on the keyboard) and the 400-ms dwell time for selection confirmation (constant component). Our data show that scanning time varied somewhat for different letters, but only slightly. For the letters located second on the keys, such as ”b”, ”e”, “h”, etc., total entry time will always be 800 ms plus varying scanning time. Accordingly, for the two letters that come as the fourth (i.e., “s” and “z”), total entry time will always exceed 1600 ms. (In fact, in our experiment the measures obtained for both letters were slightly above 2000 ms.)

Participant Comments

Participants were surveyed for their impressions and their perceived performance. During the initial introduction to the text entry techniques, when participants were shown the Symbol Creator assembly table, they all believed multi-tap would be much easier to use than the novel technique. After 15 minutes of training, however, five of the six indicated that Symbol Creator was just as easy as multi-tap, and only one participant indicated that the latter was easier.

In the questionnaire provided, all participants rated Symbol Creator as both highly guessable and learnable. Multi-tap received also high scores for these characteristics, but it was natural and quite obvious. On the other hand, four of the six participants described Symbol Creator as more fun to use, whereas multi-tap was characterized as more frustrating. A typical explanation for their frustration was that text entry with multi-tap seemed time-consuming as most of the time they could do nothing but just sit watching the delayed output of the right letter because of the sequential scan through all the letters on the key. The situation was particularly apparent during entry of words involving frequently used letters such as “i” or “o” that come as the last on the keys. On the contrary, in the case of Symbol Creator, participants had an impression of being able to exercise full control over the text entry process resulting in better satisfaction.

CONCLUSIONS

In terms of accuracy and speed of entry, both Symbol Creator and multi-tap performed equally well. For both text entry techniques, the mean entry speeds were just below 9 words per minute, whereas the error rates were as low as 3%. This is consistent with the study on eye-based text entry [2], where speeds up to 10 words per minute were reported. Moreover, their system used word prediction to facilitate text entry, whereas our system in its current implementation did not employ any support of this kind. That is, participants had to actually enter every letter and still were able to achieve comparable text entry speed keeping at the same time relatively low error rate.

Even though both the techniques exhibited very similar performance characteristics, participants preferred using Symbol Creator to multi-tap. This might be explained, at least partly, by the differences in entry time per letter. For Symbol Creator, entry time was distributed more or less evenly across the letters with just few exceptions. Meanwhile, with multi-tap the distribution was very uneven, since entry time per letter was directly influenced by the arrangement of letters within the keys.

Further work is required to investigate how word prediction might improve the performance. Also, in our future experiments we intend to recruit physically challenged people to serve as participants, since they represent the target population for the techniques described here.

REFERENCES

1. Gips, J., Dimattia, P., Curran, F.X. & Olivieri, P. (1996), Using EagleEyes – an electrodes based device for controlling the computer with your eyes – to help people with special needs, *in* J. Klaus, E. Auff, W. Kremser & W. Zagler (eds.), *Interdisciplinary Aspects on Computers Helping People with Special Needs,* Oldenburg, 77-84.

2. Hansen, J.P., Hansen, D.W. & Johansen, A.S. (2001), Bringing gaze-based interaction back to basics, *in* C. Stephanidis (ed.), *Universal Access in HCI*, Lawrence Erlbaum Associates, 325-328.

3. Hutchinson, T.E., White, K.P., Martin, W.N., Reichert, K.C. & Frey, L.A. (1989), Human-computer interaction using eye-gaze input, *IEEE Transactions on Systems, Man & Cybernetics,* 19, 1527-1534.

4. Istance, H.O., Spinner, C. & Howarth, P.A. (1996), Providing motor impaired users with access to standard Graphical User Interface (GUI) software via eye-based interaction, *Proceedings of the 1st European Conference on Disability, Virtual Reality and Associated Technologies* (*ECDVRAT ’96*), Maidenhead, UK, 109-116.

5. MacKenzie, I.S. (2001), A note on phrase sets for evaluating text entry techniques. Available at http://www.yorku.ca/mack/RN-PhraseSet.html

6. Majaranta, P. & Räihä, K.-J. (2002), Twenty years of eye typing: Systems and design issues, *Proceedings of Eye Tracking Research and Applications Symposium (ETRA 2002)*, New Orleans, USA, ACM Press, 15-22.

7. Soukoreff, R.W. & MacKenzie, I.S. (2001), Measuring errors in text entry tasks: An application of the Levenshtein string distance statistic, *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems (CHI 2001)*, ACM Press, 319-320.

8. Ten Kate, J.H., Frietman, E.E.E., Stoel, F.J.M.L. & Willems, W. (1980), Eye-controlled communication aids, *Medical Progress through Technology*, 8, 1-21.

Sonification of Facial Expressions

Ioulia Guizatdinova
TAUCHI Computer-Human Interaction Unit
Department of Computer and Information Sciences
University of Tampere
FIN-33014 Tampere, Finland
+358 3 215 8556
ioulia_gizatdinova@pochtamt.ru

Zhiguo Guo
Department of Computer and Information Sciences
University of Tampere
FIN-33014 Tampere, Finland
+358 3 041 440 9668
Zhiguo.Guo@uta.fi

ABSTRACT

Visual attributes specifically describing human facial expression can be inaccessible when visual perception is blocked or an image is hidden for observer. Current computer programs can process facial expressions with speech output or alternatively by adding tactile feedback cues. However, facial traits and expressions are still not sonificated. The goal of this project was development and usability evaluation of the sonification system for alternative access to facial expressions through eARmoticons. eARmoticons should evoke the emotional feelings of a listener like those a visual image could produce. The results of our study showed that after some training time the auditory imaging facial expressions could also be accessible for a listener. The proposed technique, which can briefly display an array of related attributes like facial traits, could facilitate communication and interpretation of visual images for people with special needs.

KEYWORDS: emotional expression, facial traits, sonification, earcons, eARmoticons.

INTRODUCTION

The lack of visual information can become a great problem when visual perception is temporary blocked or even unavailable. For example, for blind people an access to visual information has always been an obstacle. Nowadays some computer programs with text-to-speech and video processing can recognize simple pictures and sonificate some attributes with auditory output or alternatively by adding tactile feedback cues. Both an emotional content of the images and facial expressions are still not accessible for the blind or when an image is hidden for observer.

New Interaction Techniques '03, Spring, 2003, University of Tampere.

In the development of alternative communication and access to visual information, sonification, as a method, has been used to map the communication between two modalities. Sonification is the use of non-speech audio to convey information; more specifically sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation [3].

Some sonification methods suppose to set only quantitative conformity of sound parameters to the visual ones. It is supposed too that a person may learn any sound code (like Morse code) and substitute visual mental notions (percepts) by hearing complicated sound patterns [8]. Other investigators are relying on cognitive transfer, like as synesthesia phenomenon, or intermodal sensations [2].

P. Meijer [6] proposed the sonification method of arbitrary pictures registered by a video camera. His approach is based on agreement that amplitude of a sinusoidal oscillator is proportional to pixel gray level, frequency is dependent on a vertical pixel position, and a horizontal pixel position is translated into time-after-click. After that image is scanned through one column at a time, and the associated sinusoidal oscillator outputs are presented as a sum (chord) followed by a click before presentation of a new sequence of columns. The example of smiling face which is sonificated using “The vOICe Java Applet” [9] and its spectrogram are shown in Figure 1. The method requires a lot of training to recognize the complicated earcons.

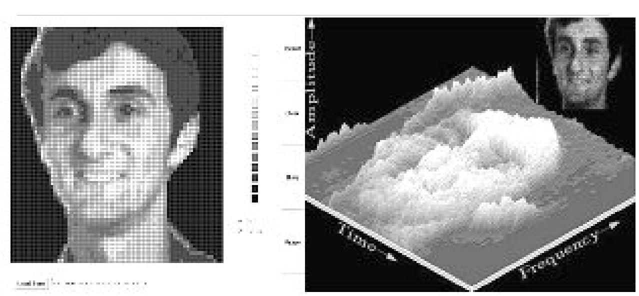


Figure 1: An original image (on the left) and image-to-sound mapping (spectrogram) on the right. http://www.visualprosthesis.com/javoice.htm [6].

However, in contrast to emoticons, present methods of transformation cannot be used to communicate meaningful visual attributes such as facial traits or expression without long training, based on previous experience of a listener. And the question remains how to convey the facial traits with emotional tinge of some person presented like any visual image into the sounds, for instance, earcons or music chords.

In a paper on auditory emotional access to information [8], Harald Schwende writes that sonification system enables an emotional access using the similarity of feeling, by watching a picture or hearing music and transforming visual information in music compositions.

The goal of our work was designing and usability evaluation of the sonification system for facial expressions, which should allow a transformation of the visual information (related attributes specifically describing emotional expression/state) the into sound patterns. This kind of sounds ("eARmoticons") should evoke the similar emotional feelings of a listener, that is, to provide an emotional access through "audio snapshot" when visual perception is blocked or an image is hidden for observer.

SOUND MAPPING

There are two main blindness categories, they are: congenital and adventitious. The adventitious blind people have experience with real images, whereas the congenital blind people have not any associations to visual image [5], but extensive tactile and auditory experience with surround objects.

There are also some facial features, which are hardly transformed by sounds, for instance, congenital or non-classified attributes.

Let us assume that a visual image of facial expression (video image, picture) or related attributes specifically describing facial expression were processed and classified. Then, the problem is how to convey an array of known visual traits into brief and informative sounds – stimuli, which will provoke similar behavioral patterns and feelings (or emotions), or evoke a mental association with emotional content.

First, two approaches to image-to-sound mapping could be defined. One way is a transformation of each facial expression as a whole. We called this mode in the future system – direct sonification (DS). The second technique could be based on sonification of the remarkably sparse set of facial landmarks like forehead, eye, nose bridge and mouth defining a human face, called as indirect sonification (IS). The conceptual model of both approaches is shown schematically in the Figure 2. The model is obvious but it is a good start point for evaluating sound parameters, which could be adopted or rejected through empirical investigation.

Six basic emotions, which could reliably be identified through corresponding six universal expressions proposed by Ekman and Friesen, were selected and marked by the following labels: *surprise, fear, anger, disgust, sadness, and happiness* [1]. According to this classification, there are generally three types of emotions and facial expressions: positive, neutral and negative. *Anger, disgust, fear* and *sadness* are classified into the negative type; *happiness* and *surprise* are classified into the positive type; and additionally there is one neutral type. Then, the simplest approach to sound mapping is to divide a frequency range according emotional state (Figure 3).

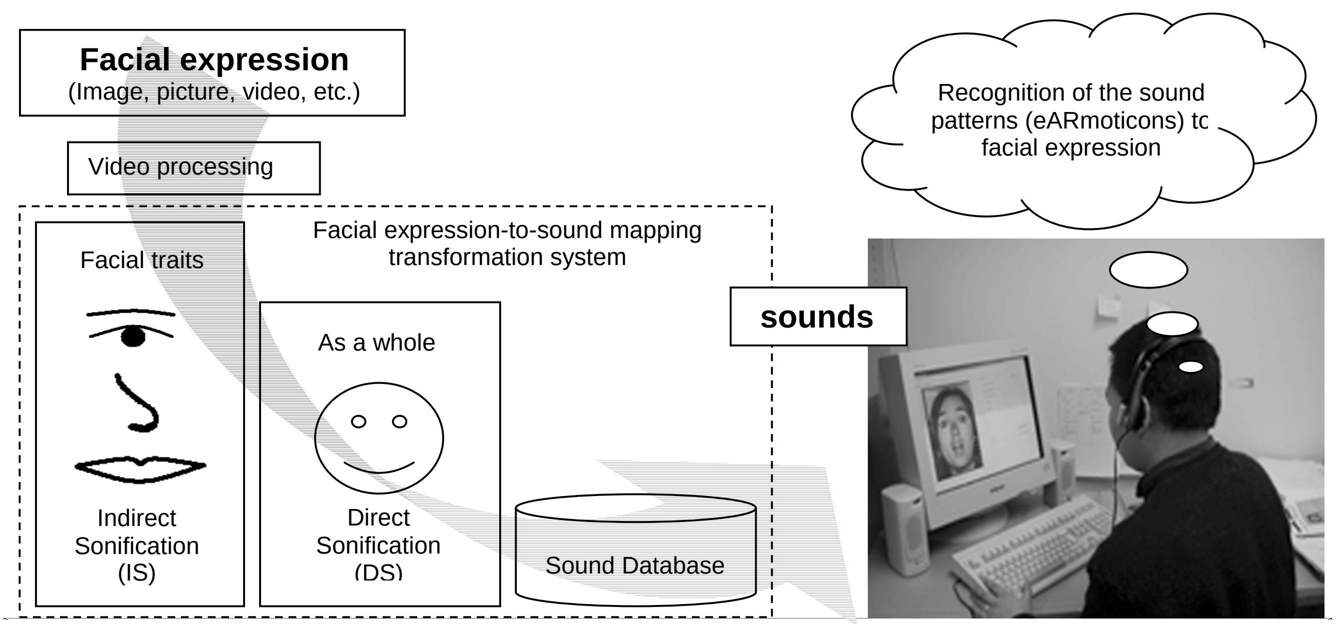


Figure 2: Conceptual model of the image-to-sound-mapping.

The positive emotions (happiness and surprise) could be sonificated through high-frequency sounds, the negative emotional states (sadness, anger and disgust) could be conveyed into sounds with low frequencies, and the neutral emotional state could be presented by sounds of middle-frequency range. The expressive performance of sounds includes different spectral shapes of the signals. Probably, the timbre is more correct sound parameter describing eARmoticons by taking into account spectrum of MIDI tools used in the synthesis. We will discuss this parameter in the next sections. Meanwhile the fear belongs to the negative type, but in our study we considered the fear to be the positive type regarding sound frequency.

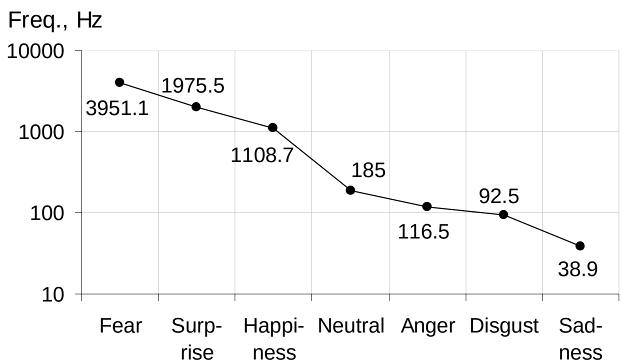


Figure 3: MIDI Frequency (Hz) for sonification of the facial expressions for both IS and DS modes.

To design eARmoticons for DS and IS modes we have used different methods and MIDI tools by creating the sounds [4]. With IS technique it is supposed that facial landmarks should have well-recognized timbres which are determined by emotional state. The tools, used to change sound envelope and timbre, were organ, piano, guitar and trumpet (Appendix 1). When we used DS mode to sonificate the emotion as a whole, we used Orchestra Hit (Appendix 2). The instruments for both sonification modes are shown in Figure 4.

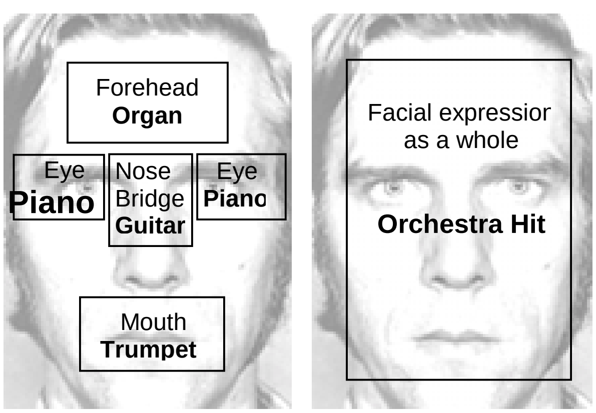


Figure 4: Modes and instruments: Indirect Sonification (the left image); Direct Sonification (the right image).

EXPERIMENTAL EVALUATION

Participants

13 volunteers (7 males and 6 females) from staff and students at the University of Tampere participated in the study. The ages of the subjects ranged from 21 to 33 years with a mean age of 27. All of the participants had a normal vision and hearing. None used hearing aid.

Apparatus and Procedure

The experiments have taken place in the usability laboratory of Tampere Unit for Computer-Human Interaction (TAUCHI) during two weeks.

A desktop PC (ASUS A7V133) was equipped with AMD Duron Processor, 256MB RAM, VIA AC'97 Audio Controller (WDM) and stereo system. The 15" LG LCD Flatron Monitor, 575 LM Multimedia Speakers were used.

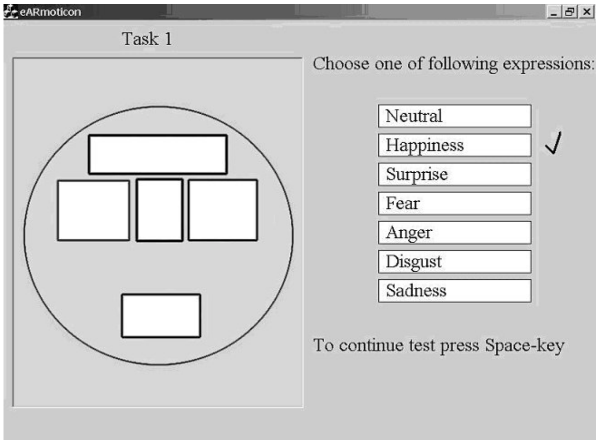
The experimental software was written by Ioulia Guizatdinova in VC++ running in Windows 2000 operating system.

In order to test both techniques through DS and IS, we carried out two experimental sessions. Each experimental session consisted of two parts, training and the test. During training images were displayed in the following order: neutral, happiness, surprise, fear, anger, disgust and sadness. The subjects could manipulate sounds by clicking on the whole image or on the different facial landmarks. After presentation of the facial expressions and sounds the pictures were masked. The subject could make a decision by clicking on the mask (the same layout of the facial landmarks) or just by listening sounds of the whole face. During the test, hidden emotional expressions were presented in a random order. Figure 5 shows a snapshot of the experimental software during IS and DS modes correspondingly.

Both the size and optical parameters (brightness and contrast) of images and sound volume were not changed throughout testing. There was a 5 minutes demonstration about the system for the subjects, they also were instructed to click on the image, listen to the sounds, and try to memorize them. After that, eight trials were performed for each subject.

While test time was not limited, each trial lasted about 30 minutes, 15 minutes per session. A five-minute break was allowed between experimental sessions. The test took a total time of 4 hours for each participant. To motivate a performance of the subjects after each trial, spent time and recognition rate per image were displayed. At the end of the test subjects were asked to fill a questionnaire.

During training and the test we investigated learning and retention of the sounds. In particular, we recorded a number of clicks and the time needed to associate and memorize eARmoticons regarding the image.



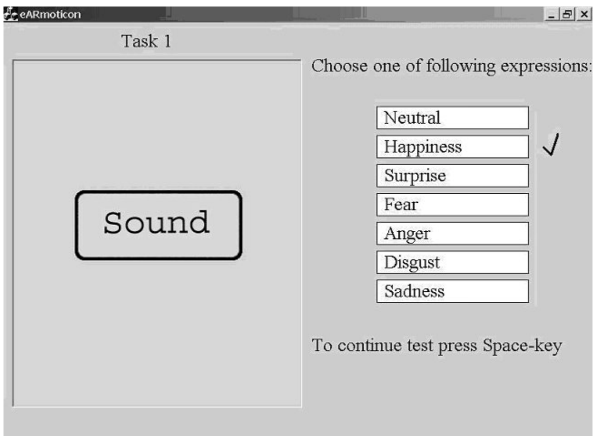


Figure 5: Snapshots of the program in the test mode: Indirect Sonification (the top image), Direct Sonification (the bottom image).

RESULTS

Data of two experimental sessions were analyzed separately.

Indirect Sonification

The time required to memorize and recognize eARmoticons was analyzed regarding each facial expression. Before experiment, it was supposed that the subjects could be confused whether eARmoticons with high or low frequencies belong to positive or negative type, or when their frequency or spectrum are similar (Appendix 1). Nevertheless, experimental results showed that the subjects spent more time to investigate neutral eARmoticons, which were sonified in middle-frequency range (Figure 6).

A number of clicks was used to estimate learning and retention of the eARmoticons. A number of clicks was averaged over facial expressions and facial traits. The results demonstrated that neutral expression required a greatest number of clicks in training mode. The measure of ANOVA (analysis of variance) revealed the statistically significant differences between neutral expression and others ones.

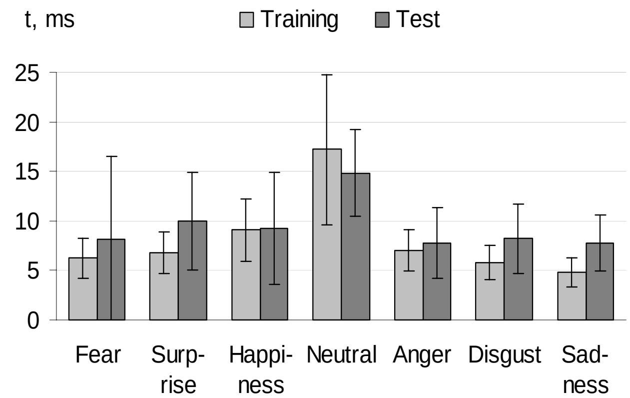


Figure 6: The elapsed time to memorize (Training) and recognize (Test) emotion through eARmoticons.

Figure 7 shows that the greater values of differences were detected for sadness in relation to neutral eARmoticons when the subjects tried to memorize them. For example, $F_{forehead}$ = 10.96; F_{eyes} = 7.19; $F_{nose\ bridge}$ = 9.35d; F_{mouth} = 8.34, α = .05.

Eyes' landmarks required the greatest number of clicks in training mode when visual image was available (Figure 7). Moreover, five participants chose strategy to remember only eARmoticons of the eyes and ignored other facial landmarks. Seven subjects noted that eARmoticons well designed with piano timbre and these sounds were more familiar and pleasant than organ, guitar or trumpet ones. They told also that it was hard to understand differences between guitar and trumpet regarding their timbre. They assumed that these eARmoticons might be easily distinguishable if duration would be longer.

While all the participants reported that they have a normal hearing, recognition of eARmoticons depend on individual sensitivity to sounds of high and low frequency. The overall recognition picture was good for neutral, anger and sadness eARmoticons, which had low frequencies (Figure 10). Recognition of eARmoticons, which had high frequency, was worse. However, some of the subjects showed very good recognition rate of high-frequency eARmoticons (Figure 11), in particular, women.

Direct Sonification

No facial features were investigated in this experiment. Sounds described facial expression as a whole. The results for a number of clicks and elapsed time were averaged over type of expressions (Figure 8 and Figure 9). We did not find of specific features in behavior of the subjects, for instance, through a number of clicks in both training and the test.

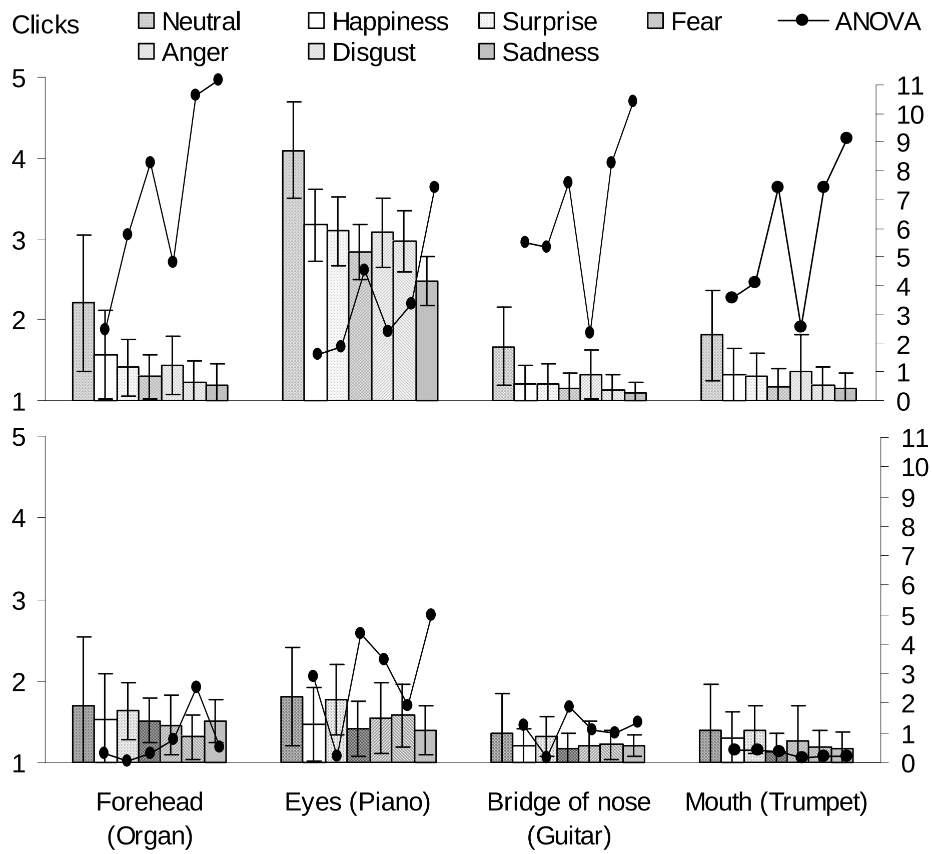


Figure 7: The mean number of clicks (and ANOVA) needed to memorize and recognize facial expression or facial traits through eARmoticons; the top image – in training mode, the bottom image – during the test.

Figure 8: The mean number of clicks required to memorize and recognize facial expression through eARmoticons.

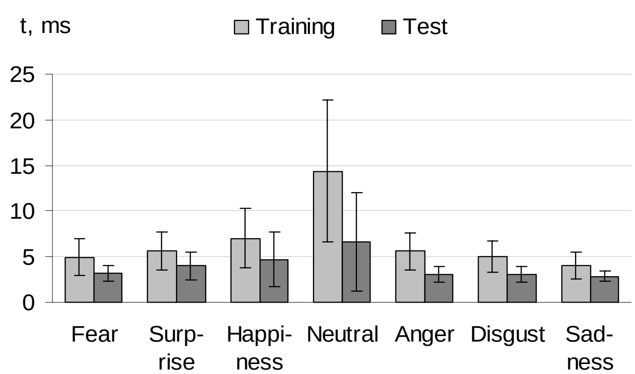


Figure 9: The elapsed time to memorize and recognize emotion through eARmoticon.

Only elapsed time to recognize neutral expression revealed significant difference. The maximum difference was observed between a repetition of eARmoticons that symbolized sadness and neutral expression: in training mode $F = 13.49$, during the test $F = 10.60$. The results of this experimental session showed that recognition of low-frequency eARmoticons was significantly better (Figure 10).

eARmoticons that symbolized anger and disgust were recognized correctly in 100% of cases. The overall recognition performance of high-frequency eARmoticons remained low, with an exception of the fear eARmoticon. As in a case of IS, some of the subjects have demonstrated an excellent recognition of high-frequency sounds (Figure 11). But we did not take into account musical experience of some participants, as the sonification should satisfy needs of all people independently of individual skills.

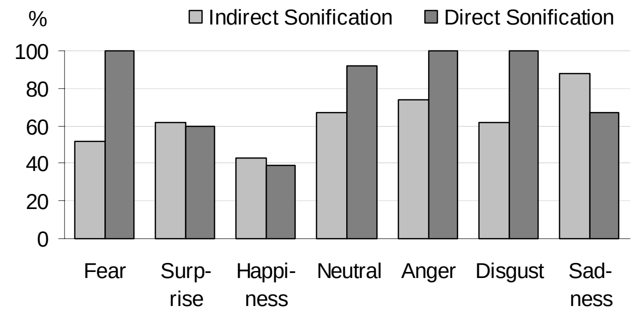


Figure 10: Recognition rate of the expression type.

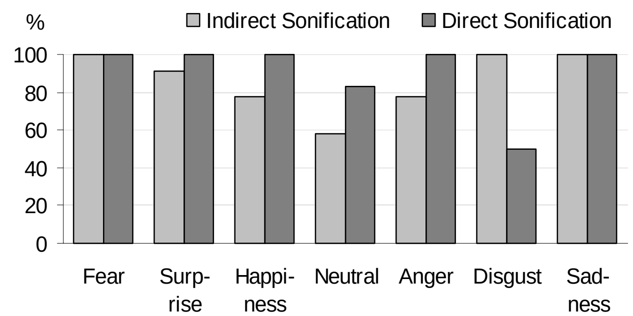


Figure 11: Example of the good recognition with high-frequency eARmoticons.

DISCUSSION

There is incontestable evidence about vision domination under other attention mechanisms of human beings [10]. On the other hand, it is found that eyes are the most informative regions on the face [11]. As the test consisted of not only of audio stimuli (eARmoticons) but also visual presentation of the pictures of verified emotional expressions, forming of mental associations took place under dominant influence of visual analyzer. Thus, the subjects might subconsciously give preference for the investigation of the eyes only.

The received data gave us information that sonification of eyes should be made more accurately as this trait captures more attention of an observer. Other facial features had similar recognition rate through eARmoticons when visual expression was hidden.

Data of both IS and DS sessions showed that in comparison to the neutral expression, sonification of other ones had required fewer repetitions of eARmoticons for memorizing. We have considered several reasons for these phenomena. First, it should be noted that the concept of a neutral face is a problematic one. Some psychologists state that there are no perfectly neutral faces. A neutrality of the exposed face may depend on the individual experience of the observer. It has been shown [7] that reactions to expressive displays can be a joint function of the type of the display and the observer's attitudes. Perception of the neutral face may also depend on the context of other (previously) presented facial expressions. For example, in the context of smiling faces a neutral face may be classified as sad. Thus, ambiguity in visual perception of the facial neutrality can essentially influence onto the auditory perception of eARmoticons. Second, due to some similarity of spectrums between neutral, disgust and anger eARmoticons (Appendix 2) the subjects have felt confusion during recognition in IS mode. Subjects could be confused whether eARmoticons with middle-frequencies belong to the negative type. The same distraction was observed for eARmoticons that symbolized surprise and happiness both with IS and DS techniques. Third, short duration of eARmoticons (<600 ms) might also present some difficulties in recognition of sounds.

CONCLUSION

Overall, the results of our experiments showed that facial expressions could be briefly presented through sounds with emotional content called eARmoticons. In presented project there were used two techniques. The first method was sonification of facial expression as a whole through one sound. The second technique was sonification of emotional expression with the help of the set of eARmoticons that symbolized traits like forehead, eye, nose bridge and mouth. In spite of good overall recognition picture through eARmoticons and both techniques, several remarks could be formulated.

The further work could concentrate on designing more complicate brief patterns. But, similar to conventional menu alternatives, sound parameters of eARmoticons should be unambiguous, mutually exhaustive, exhaustive, and non-overlapping. The results demonstrated that designers of sonification system should take into account destination of symbolic sound system, whether sonification system is intended for the visually impaired or users with normal vision.

ACKNOWLEDGMENTS

The research presented here was performed at TAUCHI Computer-Human Interaction Unit of Department of Computer Science of University of Tampere. We would like to thank every volunteer participating in our usability test. We greatly acknowledge Grigori Evreinov who gave us this interesting topic and guidance during our work.

REFERENCES

1. Ekman, P. and Friesen, W.V. *Facial Action Coding System (FACS): A Technique for the Measurement of Facial Action*. Palo Alto. CA.: Consulting Psychologists Press, 1978.
2. Galeyev, B.M. What is Synaesthesia: Myths and Reality. *Leonardo Electronic Almanac 7*, 6 (August 1999). Available at: http://prometheus.kai.ru/mif_e.htm.
3. Glossary. ICAD International Community for Auditory Display. Available at: http://www.icad.org/

4. GoldWave. Product information is available at: http://www.goldwave.com.

5. Kurze, M. TDraw: a computer-based tactile drawing tool for blind people. In *Proceedings of 2nd Annual ACM Conference on Assistive technologies,* (April 11-12, Vancouver, British Columbia), ACM Press, Canada, 1996, pp.131-138.

6. Meijer, P. An experimental system for auditory image representation. *IEEE Transactions on Biomedical Engineering 39*, 2 (Feb. 1992), 112-121.

7. Schmidt, K.L. and Cohn, J.F. Dynamics of facial expression: normative characteristics and individual differences. In *Proceedings of ICME International Conference on Multimedia and Expo*, (August 22-25, Tokio), Jupan, 2001, pp. 728-731.

8. Schwende, H. Auditory Emotional Access to Visual Information. In *Proceedings of Computers Helping People with Special Needs, 8th International Conference, ICCHP* (July 15-20, Linz), 2002, Austria, pp. 445-447.

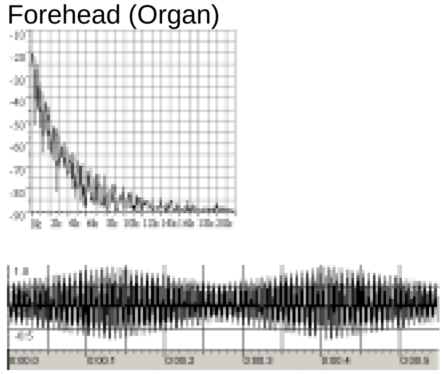
9. The vOICe Home Page. Available at: http://www.visualprosthesis.com/.

10. Noton, D. and Stark, L.W. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vis. Res. 11*, (1971), 929–942.

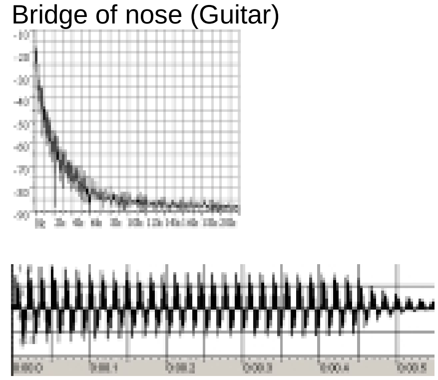
11. Yarbus, A.L., *Eye movements and vision*. Plenum Press, New York, 1967.

APPENDIX 1.

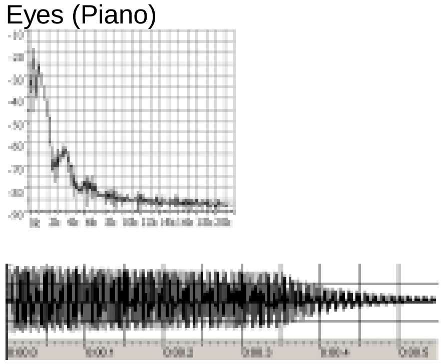
Forehead (Organ)



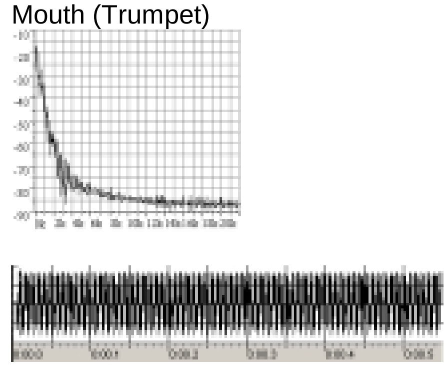
Bridge of nose (Guitar)



Eyes (Piano)

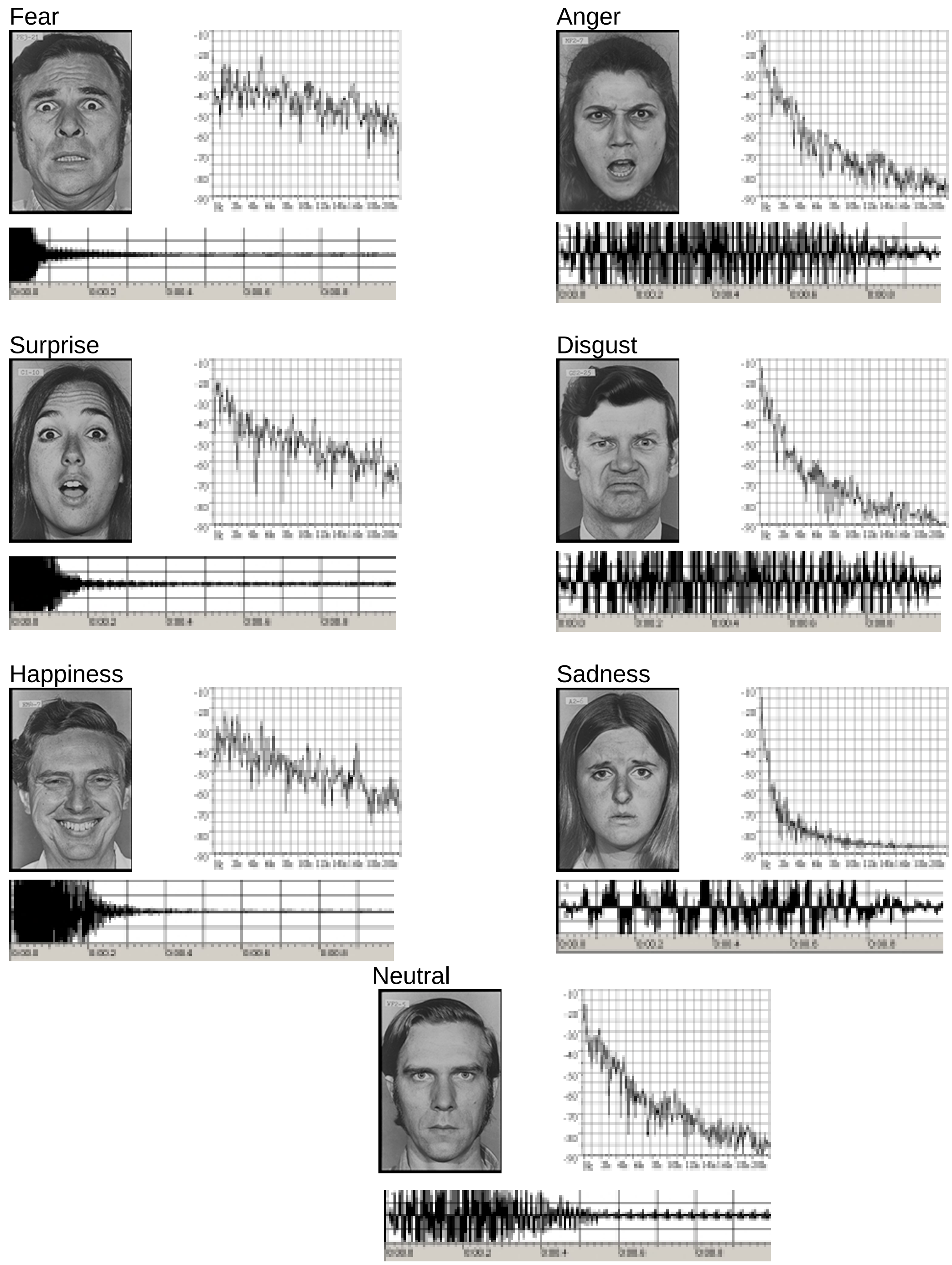


Mouth (Trumpet)



Sound signals (wav) and spectrograms (frequency spectrum) of neutral eARmoticons (185 Hz) for imaging facial traits with Indirect Sonification technique.
Sound processing was performed by GoldWave digital audio editor [http://www.goldwave.com].

APPENDIX 2



Sound signals (wav) and spectrograms (frequency spectrum) of eARmoticons
for imaging facial expressions with Direct Sonification technique.
Sound processing was performed by GoldWave digital audio editor [http://www.goldwave.com].

Symbol Creator: Usability Evaluation of the Novel Pen-Based Text Input Technique

Melinda Luoma
New Interaction Techniques 2003
Department of Computer and Information Sciences
FIN-33014 University of Tampere, Finland
Tel: 358-44-522-0476
melinda.luoma@uta.fi

ABSTRACT
The goal of this project was a usability study of Symbol Creator, a pen-based text input technique. Symbol Creator (SC) is a new technique, which aims at faster and more intuitively usable text entry. It is based on assembling characters using symbols that resemble basic elements of Latin cursive. Since keyboards are getting smaller and smaller, and the number of text messages has increased in the last decade, it is important to study ways to make multitap input technique more fast and easy to learn. Usability was being evaluated by text entry speed, number of errors made, and subjective user experiences. Results showed multi-tap input technique to be faster than Symbol Creator technique. However, the number of errors was less with Symbol Creator technique than with multi-tap technique. Also, when measuring text entry rate with keystrokes per character, the two text input techniques reached the same rate of 2 KSPC. Subjective experiences were negative regarding Symbol Creator before testing, but changed to more positive during testing the technique.

KEYWORDS: Pen-based text entry, handwriting, KSPC, software keyboard, Symbol Creator, multi-tap.

INTRODUCTION

Text entry is used more and more in the modern world. The number of text messages sent has been increasing exponentially during the last years and is now approaching 1 billion SMSs per day [2]. As mobile devices get smaller, new text entry methods are needed - methods that require less space and are fast and convenient enough to satisfy user needs. These methods need to be easy to learn, since our world is already full of stressful and memory-consuming issues. New devices and methods should thus be ones that help users manage their everyday life, not bringing new tasks to learn.

But why do we need these new techniques? Are not the old methods good enough? The reasons are simple: consumers want more manageable and usable devices. Previous text entry methods are good, but maybe we could find something even better. Traditionally, the one problem with handheld devices is that text entry is slow when being compared to e.g. full sized QUERTY-keyboard. The reason for this is the use of multi-tap and pen-based methods [12]. Before going deeper into Symbol Creator technique, let us briefly outline the existing text entry methods.

TEXT ENTRY METHODS

Text entry methods can be divided into key-based and stylus-based text entry [3] through techniques direct finger pointing or pen-shaped device. Among key-based entry we can find various onscreen keyboards both for desktop and wearable small-size touchscreens like PDA. A number of software keys can vary from 3 up to 104 with different layouts [11]. Most of these methods are based on a single tap or multi-tap entry techniques similar to manipulating with physical keys and used in mobile phones [3, 4].

Meanwhile, stylus-based text input can use gestures to select characters or words [12] through different optimised layouts or traditional handwriting recognition. However, the user needs to have experience in order to reach fast text input speed.

Both of these text input methods have their advantages and drawbacks. Among the key-based text input, QWERTY has the disadvantage of a large software keyboard which makes it hard to fit in small space. The telephone keypad and five-key method need less space but require some learning of multi-tap typing. Key-based text entry does not require any additional equipment, whereas stylus-based input is done with a pen or some other pointing device.

When comparing the speed of text entry between the two methods, keyboard exceeds the stylus-based methods. With a miniature QWERTY keyboard, it is easy to reach the speed of 20 words per minute, and experts are able to type even 40-60 words per minute [4]. Table 1 demonstrates text entry times in different text input methods.

Scheme	**WPM** (finger)	**WPM** (thumb)
Iconic	19.8	17.6
Two-key	25.0	22.2
Multi-tap	27.2	24.5

Table 1. Text entry speed of iconic, two-key and multitap text input techniques (adopted from Jannotti, 2002 [5].)

Symbol Creator, developed by Grigori Evreinov at the University of Tampere, is a novel text entry method, which may be considered as a combination of key-based and stylus-based methods. It uses a pen for input and is based on a set of symbols just like the stylus-based methods, but at the same time, has software keys, which are being tapped with the pen. Symbol Creator can also be seen as a multi-tap method since every character is created with at least two taps. Figure 1 illustrates the set of symbols of this method. As we can see, these seven symbols resemble parts of Latin cursive, and each character is constructed by combining these symbols in different sequence [1].

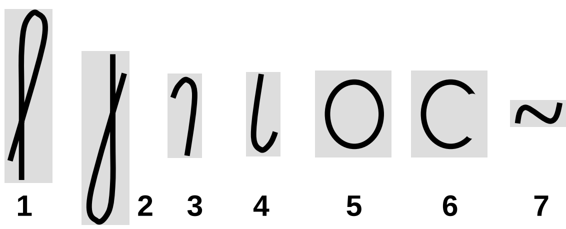


Figure 1. Symbols of Symbol Creator (adopted from Evreinov, 2003 [1])

Symbol Creator can also be considered as an iconic text entry system while it is based on icons which when combined, resemble our letters. (For a more extensive report of iconic text entry systems, see [5].)

Continuing the classification of Symbol Creator, we can see it fitting well to the group of coders, instead of pickers. Coder is a text entry method that uses sequences of buttons to create letters; picker is an interface in which the user needs to pick the letter from a list. Coders have the advantage of higher speed and less need for visual feedback; however, they have the disadvantage of longer learning time [5].

METHOD

The primary aim of this project was to investigate the usability of a novel pen-based text entry method, Symbol Creator. Symbol Creator was compared with the conventional 12-key phone entry.

Experimental Setup

The software for the testing was made in Microsoft Visual Basic 6.0 for Windows 2000 by Grigori Evreinov. The handheld device used for running the software was a Jornada HPC 720. The tests were performed in Tauchi Unit of the University of Tampere. The screens in the tests are illustrated in Figure 2 and Figure 3.

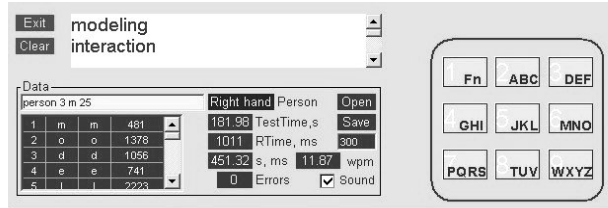


Figure 2. Screenshot of the test for multitap text input technique. Data frame is hidden during testing.

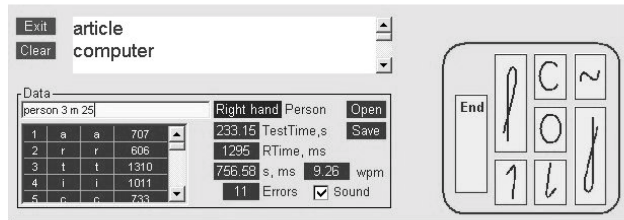


Figure 3. Screenshot of the test for Symbol Creator text input technique. Data frame is hidden during testing.

Usability was measured by speed of text entry (wpm), clicks per word, number of errors, and subjective user experience. Novice performance was measured in order to see how intuitive this method is.

Subjects

8 voluntary subjects took part in the study. The subjects were students of Computer Science at the University of Tampere. The average age of the subjects was 27.3 years, ages varying from 23 to 32 years. All of them, 1 left-handed and 7 right-handed subjects had normal vision. All subjects were experienced computer users and seven of them had previous experience with pen-based text entry and handheld devices. All subjects had a mobile phone with 12-key entry, so they were considered experts in the setting of multitap input technique. However, they did not know Symbol Creator in advance, thus being classified novice users.

Procedure

The study was carried out in two sessions of experiments, testing Symbol Creator and multitap in different days. This was done because one session took approximately one hour and we did not want to tire the subjects out. Four of the subjects completed the test of Symbol Creator first, the other four started with multitap input technique, thus controlling the possible learning from affecting the results.

One test session consisted of seven trials. Each trial had twenty words appearing in a random order from the list of 150 words. An average length of one word was eight characters, the longest words being 11 characters and the shortest ones 7 characters long. Before the test, subjects were given five minutes practice time for multitap input technique and fifteen minutes for Symbol Creator technique. Subjects were handed out a paper of the symbols and their combinations for creating characters (Figure 3) as practice for using Symbol Creator. Subjects were told the logic of the method and advised to memorise the symbols. When the test started, the paper was taken away.

The delay time between taps was set to 300 ms. Several delay times were tried out before the tests, and this one seemed the most suitable for our purposes. The problem with setting the delay time was that it affected both the speed of moving from one key to another and the speed of tapping. So, if it was convenient for tapping, it was slightly too slow when moving from one key to another; on the other hand, if moving between keys was fast enough, then the tapping speed needed would exceed the capability of average subjects. That is why we made a compromise of 300 ms. It allowed the tapping speed to be convenient, not too fast, but made a little delay between the keystrokes. The setting of delay time could be a reason why we did not reach the often mentioned 20 wpm with the multitap text input technique.

In both experimental settings the subjects had audio feedback related to each tap. When they hit the right key, there was a short clicking sound indicating that the subject could move to the next symbol. When they hit the wrong key, there was a lower sound saying "wait" to indicate the subject should try again. The words appeared in the text box one at a time, and the subject had to press the button "End" or "Fn" to start retyping the test word.

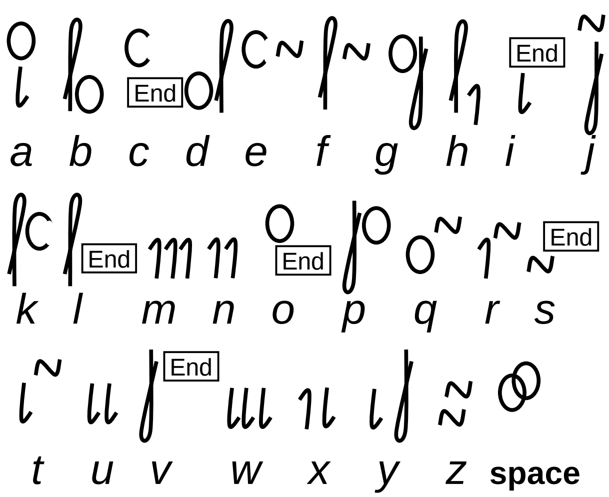


Figure 4. Alphabet and entry technique through Symbol Creator. “End”- an additional click. Segment “End” can also be used instead of the next symbol.

This procedure allowed subjects to rest between the words. There was no possibility to delete a character but the subject had to try so long that he either got the right character or the time of 10 s to input one character ran out. Every missed symbol was counted as an error.

RESULTS

Speed of text entry

Text entry speed was measured by the number of words produced per minute. In novice users, the multitap input technique seems to be faster than Symbol Creator. Both methods witnessed learning, which is increase in speed by approximately three letters per minute - multitap input technique 3.6 wpm, Symbol Creator technique 3.1 wpm). Average speeds in the first trial were 11.9 for multitap input technique and 8.0 wpm for Symbol Creator. Speeds in the last trial reached 15.5 for phone entry and 11.1 for Symbol Creator. These results are illustrated in Figure 5. Average speeds of all trials were 14.2 wpm for multitap input technique and 9.8 wpm for Symbol Creator technique. Standard deviations were 3.6 for multitap input technique and 3.1 for Symbol Creator. There was a significant difference between the two methods ($F(55) = 0.60$, $p<.05$).

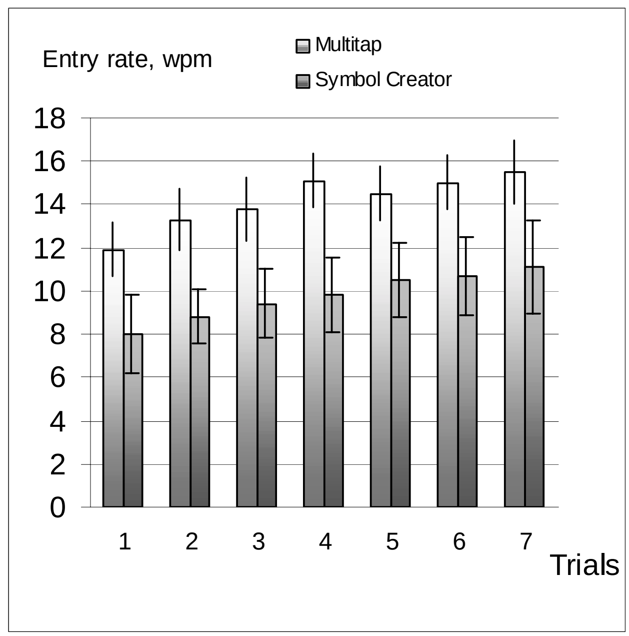


Figure 5. Text entry rate for multitap text input technique and Symbol Creator, measured by words per minute.

Another way to evaluate performance and measure the efficiency of text entry is counting the necessary clicks per word (cpw). The comparison between the two text input methods showed a difference here, too: multitap input technique resulted in an average of 24 clicks per word, whereas with Symbol Creator the subjects reached an average of 20.7 clicks per word. These results can be seen in Figure 6. The number of clicks varied from a minimum of 14 cpw to a maximum of 50 cpw with PhEntry, and 12-36 cpw with Symbol Creator. Standard deviations were 6.8 for multitap text input technique and 10.1 for Symbol Creator. Again, we found a significant difference between the two methods ($F(6) = 16.44$, $p<.01$). The number of clicks did not change significantly from one trial to another, indicating that learning was happening rather in the reaction times and motor skills (speed of word entry), than in reducing the number of clicks.

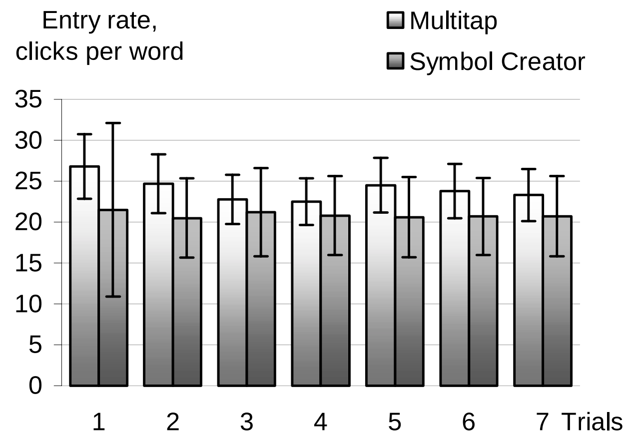


Figure 6. Text entry rate for multitap text input technique and Symbol Creator, measured by clicks per word.

These results should not be taken for granted, however. Evaluating new text entry techniques with text entry rates can result in awkward results. When the evaluation does not take errors into account, the full potential of these techniques can be hidden. When the evaluation includes errors, the comparison to methods that disregard errors becomes difficult [6].

Evaluation of the above results regarding text entry rate, can be better understood when using a different kind of approach, taking into account KSPC, that is keystrokes per character. Keystrokes per character is "the number of keystrokes required, on average, to generate a character of text for a given text entry technique in a given language" [8]. KSPC is calculated with the following equation:

$$KSPC = \Sigma (Kc \times Fc) / \Sigma (Cc \times Fc),$$

where Kc is the number of keystrokes required for entering a character, Cc is the size of a character (always 1), and Fc is the frequency of a given character in a corpus. When converting the text entry rates of multitap text input technique and Symbol Creator from words per minute to keystrokes per character, we get essentially the same text entry rate for both techniques. Thus, both techniques require very similar values of KSPC (multitap text entry technique 2.034, Symbol Creator 2.074).

Number of errors

The learnability and usability of Symbol Creator was studied by comparing the number of errors made with Symbol Creator text entry method and multitap input technique. In this comparison, we received significantly less errors with Symbol Creator than with the other method. The average number of errors per trial, that is false clicks, in the first trial with Symbol Creator was 32.4. By the last trial the number of errors were reduced to an average of 20.4. The corresponding values for multitap input technique were 60.1 errors in the first trial, and 31.4 errors in the last one. There was a significan difference between the two methods ($F(55) = 1.95$, $p<.01$). Hence, it seems that multitap input technique is much more prone to errors than Symbol Creator. These results are illustrated in Figure 7.

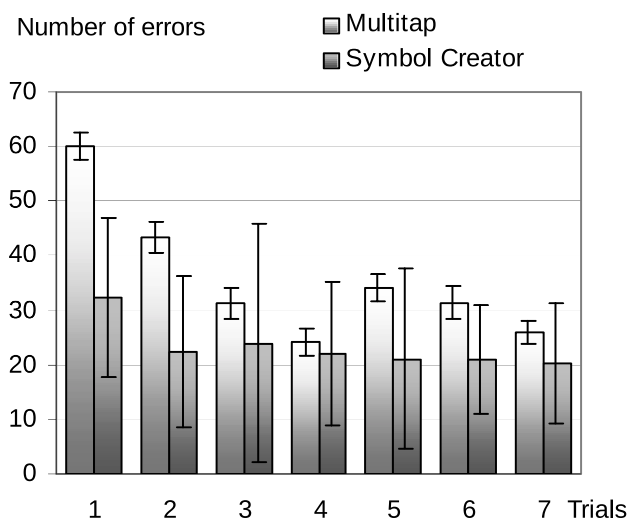


Figure 7. Number of errors for multitap text input technique and Symbol Creator.

Subjective experience

The subjects were given a short interview both before and after the tests. In the initial interview they were asked which of the two text entry methods they think would be easier to use, how easy would the learning of Symbol Creator be, how fun it would be, and which method would they prefer to use in the future. All subjects said that conventional method used in mobile phone would be the easier one, since that is the one they already have experience of using. Six out of eight subjects thought Symbol Creator would be difficult to learn and seven out of eight thought they would prefer the conventional multi-tap input technique.

After completing both tests, subjects were asked the same questions again. Still, all of the subjects said that multitap input technique was easier. This is understandable since

seven trials is quite short practice time. However, six out of eight subjects said Symbol Creator was much easier to learn than they thought it would. Also, five out of eight said it was actually quite fun method, and three of the subjects could consider switching to Symbol Creator if it was possible in the future.

DISCUSSION

Text entry speeds were low in both phone entry and Symbol Creator entry experiments, compared to previous studies. One reason for this might be the already mentioned slow delay time. It has been suggested that for one press, even an approximately 200 ms delay would be enough [3], but in our pilot experiments, we found 300 ms more comfortable for the users.

Symbol Creator reached only an average of 9.8 wpm, which was surprising. Maybe the practice time, 15 minutes, was too short to understand the logic of the method and memorise the combinations of symbols. Nevertheless, it is not surprising that regular multitap input technique reached faster entry times, since it is far more practiced than Symbol Creator. However, when these results were converted into a measure of keystrokes per character, we found a similarity in the text entry rate, an average of 2 KSPC for both techniques.

Evaluating new text entry techniques with text entry rates can result in awkward results. Text entry speed might have been slow but the number of errors produced with Symbol Creator was smaller than with multitap input technique. One reason for this could be that usually, when speed increases, so do errors. Maybe we are so used to multitap input technique that we have become too confident and even sloppy. Symbol Creator, being a new method, demanded more attention from the subjects, thus reducing the number of errors, but also effecting text entry speed. Maybe subjects were too careful not to make mistakes with Symbol Creator, thus resulting in slower speeds.

Considering subjective experiences, the results are promising for Symbol Creator. Although initial attitudes toward this new method were negative, after completing the experiment, more than half of the subjects thought that Symbol Creator was after all fairly easy and fun to use.

Text entry techniques require research and development in the future. One important issue will be the personalization of text input style. That is, the program would learn the way the user is doing things and adjusts its settings accordingly. The problem of the delay time slowing performance could be overcome by this kind of self-learning devices. Once the user gets to know the text input method, e.g. Symbol Creator, he becomes faster and the system could adjust the delay times and other settings correspondingly.

Future work should consider not just how people could learn to use new systems, but how these systems could learn the way users are using them!

REFERENCES

1. Evreinov, G., Raisamo, R. An exploration of vibration during interaction with polarized texture. Acoustics 2003, 16-18 June, Cadiz, Spain, WIT Press, UK (in press).
2. Miniotas, D., Spakov, O., Evreinov, G. Symbol Creator: an alternative Eye-based Text Entry Technique with Low Demand for Screen Space. In Proceedings of INTERACT 2003, September 2003, Zurich (in Press).
3. Edwards, A. & Yanyu, Li. How many ways can you use one button? Timing data for button presses. Unpublished paper. 2002. Available at: http://www-users.cs.york.ac.uk/~alistair/wip/Button.pdf
4. Hashimoto, M. & Togasi, M. A virtual oval keyboard and a vector input method for pen-based character input. CHI ´95 Proceedings (1995).
5. Jannotti, J. Iconic Text Entry Using a Numeric Keypad. 2002. Available at: http://www.pdos.lcs.mit.edu/~jj/jannotti.com/papers/iconic-uist02.pdf
6. Jhaveri, N. Two Characters per Stroke - A Novel Pen-Based Text Input Technique. NIT 2003. Project report.
7. MacKenzie, I.S. Mobile Text Entry using Three Keys. Proceedings of the Second Nordic Conference on Human-Computer Interaction NordiCHI (Oct 2002), 27-34.
8. MacKenzie, I. S. KSPC (keystrokes per character) as a characteristic of text entry techniques. Proceedings of the Fourth International Symposium on Human Computer Interaction with Mobile Devices (2002), 195-210. Heidelberg, Germany: Springer-Verlag.
9. MacKenzie, I. S., & Soukoreff, R. W. Text entry for mobile computing: Models and methods, theory and practice. Human-Computer Interaction, 17 (2002), 147-198.
10. MacKenzie, I. S., & Soukoreff, R. W. A model of Two-Thumb Text Entry. Proceedings of Graphics Interface (2002), 117-124.
11. Tiainen, S. Designing a Keyboard for Small Mobile Phones. October 30, 2001. Available at: http://kotisivu.mtv3.fi/seppo/article.html
12. Toshiyuki, M. An Efficient Text Input Method for Handheld and Ubiquitous Computers. Lecture Notes in Computer Science (1999), 289-300. Handheld and Ubiquitous Computing, Springer Verlag.

Text Entry through a Single Button and Auditory Feedback

Yang Xiaoqing
New Interaction Techniques 2003
Department of Computer and Information Sciences
FIN-33014 University of Tampere, Finland
+358 40 5737226
Xiaoqing.Yang@uta.fi

ABSTRACT
The main objective of this project was to study the features of user perception of spatial-temporal mapping through sonification of the spatial patterns. An alternative text entry with a single button was considered as a model for a menu selection task. A traditional seven-segment display element was used as a spatial layout for symbol input and imaging. Empirical research was carried out for the rhythmic musical sequences coordinated to the spatial seven-segment layout. Seven notes were assigned to each segment in a temporal sequence, which the user had chosen by pressing a button. The segments were activated by cyclically after the first click. When all segments have been cycled, the result was interpreted as a character according to a set of rules and depending on the character set used. The research was focused on examining temporal components, user behavior strategy and decision-making taking place under time pressure. The rationale for the test design and the results of a preliminary evaluation are presented.

KEYWORDS: Spatial-temporal pattern, sonification, user behavior strategy, decision-making under time pressure

INTRODUCTION

In the document prepared for the National Science Foundation [8] there is a precise definition of sonification. Sonification is defined as “the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication and interpretation”. In the Auditory Display Community (ICAD) there has been a strong focus on the auditory side of this transformation, emphasizing psychoacoustics, parameterization of sound, spatial sound and sound design [1, 5, 9, 12, 14]. However, Sigurd Saue writes, “sonification is considered more as a technique than as a display, and that a thorough discussion of how to interact with such a display is lacking.

This becomes even more important when exploring spatially defined data. Sound is inherently a temporal medium, and its benefits depend crucially on changes over time” [10]. Hence, there is a challenge to investigate mapping of spatial-temporal and perceptive features and relations.

SPATIAL-TEMPORAL ATTRIBUTES & RELATIONS

How could spatial and temporal attributes be superposed and perceived as similar qualities or relations? Several solutions have been suggested, for instance, sonic probes controlled by mouse [13], virtual microphones [6] and control paths [2]. These approaches set up a model for the mapping process and are based on music or audio metaphors (Table.1).

Sound Entities	Sound Attributes	Spatial Entities	Spatial Attributes
Virtual source	interferential maximum or spot	Dot	(x, y, z) coordinates
Track	graph, stream, rhythmic structure	Line Contour	∞ dots (x_n, y_n, z_n)
Texture	wavelet, noise	Surface	Texture Texton
Shapes	envelope	2D-shape	shadows or projections foreshortening or linear prospect
3D-shell [Hollander, A.J., 1994]	∞ virtual sources	3D-shape	Volume

Table 1: Entities, their attributes and perceived qualities and relations.

There is interesting approach visualizing the temporal structure of musical waveform [4]. Figure 1 shows the spatial imaging for sounds with first three bars of the score, that is the quite normal way of people used, but here the repetitive nature of the piece should be clear even to those unfamiliar with musical notation. The acoustic similarity between any two instants of an audio recording can be displayed in a static 2D plane, which makes structural and rhythmic characteristics like visible pattern (Figure 2).

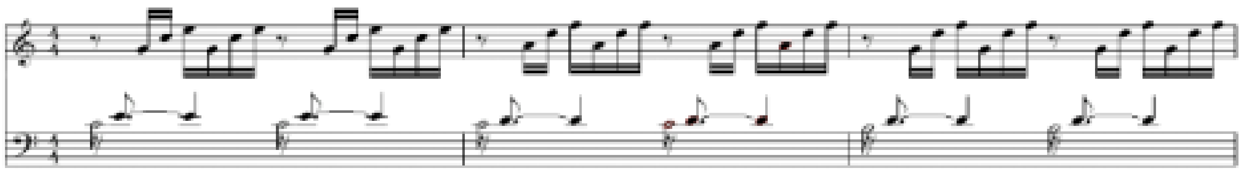


Figure 1: Spatial imaging for sounds in music sheets, first bars of Bach's Prelude No. 1 in C Major [4]

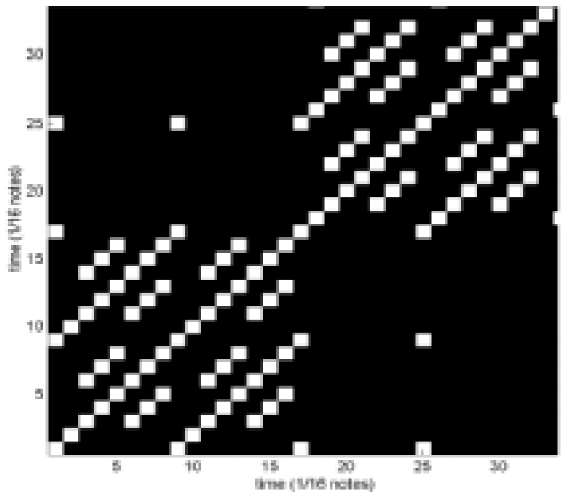


Figure 2: Self-similarity of Prelude No. 1: computed from MIDI note events [4]

Figures 1 and 2 show the same sequence of temporal events - the first two bars of Bach's Prelude No. 1 in C Major. But, Figure 2 shows the similarity image of this music, derived directly from the MIDI data. Matrix entries *(i,j)* were colored white if note *i* was the same pitch as note *j*, and left black otherwise [4]. Foote and Cooper have also presented other examples for classical and popular music.

BACKGROUND

Text input can be considered as temporal processes or stimulus-dependent goal-directed behavior of the user (Figure 3)

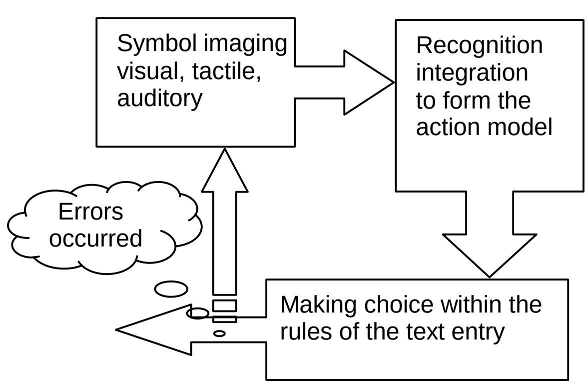


Figure 3: The forming of behavior model during text input

The goal of the system, described in this paragraph, was to provide an experimental platform with low-level control of text entry parameters for carrying out an exploration of spatial-temporal mapping. A traditional seven-segment display element was used as the layout for spatial imaging of the sequence of temporal intervals. Each segment was activated (lighted) during a definite time and the subject could choose the segment by pressing a button. At the same time, visual imaging of the segment state was substituted by sounds. The duration of the cycle (temporal pattern) in this case depends on the speed on which the user can act. In any case, before the start-click, the user must imagine a spatial layout and accomplish a sequence of pressings in own rhythm or in a definite speed [3]. If temporal pattern can be perceived like a simple musical phrase, its spatial projection should have a known equivalent like alphabet token.

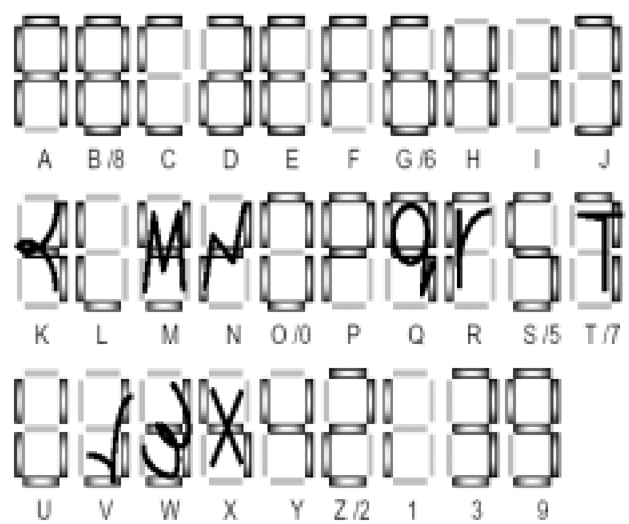


Figure 4: Cyclic synthesis of alphabet tokens and numerals

The seven-segment display element allows imaging 128 diverse combinations of lighted and dark segments. But only 16 uppercase characters and 10 numerals coincide with conventional imaging, i.e. A, B, C, 1, 2, 3 have the same graphics as alphabet tokens. The synthesis of some characters uses simple graphic analogies of handwriting and lowercase characters imaging. There are 29 test symbols presented in Figure 4.

TEST

Test Design

The experiment was carried out using ASUS A7V133 was equipped with AMD Duron Processor, VIA AC'97 Audio Controller (WDM) and 256MB RAM under Windows 2000. The software was created in Microsoft Visual Basic 6.0. A special frame (Figure 5) served as the tool panel for editing MIDI and other parameters. A data collection system stored all temporal intervals required to complete each trial and also the number of errors occurred. The data were stored in input files for subsequent analysis. The experiment has taken place in the usability laboratory.

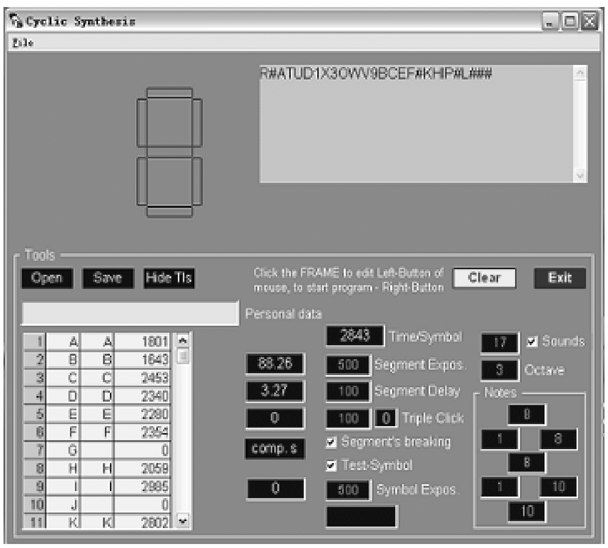


Figure 5: Snapshot of the program with opened frame of the tools

Three tonal schemes were tested as shown in Figure 6. Three temporal patterns were formed by series of notes generated with MIDI synthesizer. These patterns could be played like simple harmonic sequences, two bars of the birthday song and children's song.

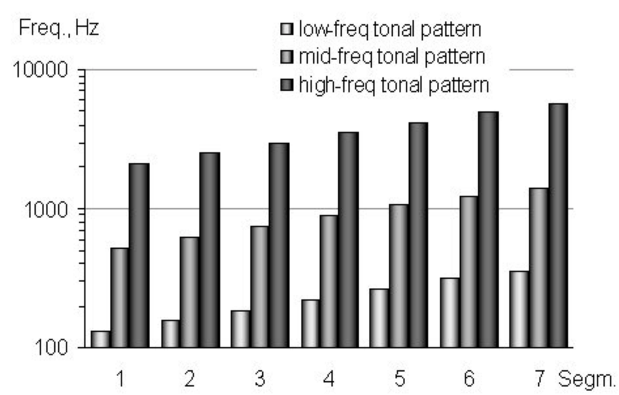


Figure 6: Tested key feedback sound schemes

When tonal patterns were played with middle frequency, all components were perceived and recognized clearly, the subjects easily remembered patterns. According to preliminary tests the low and high frequencies were difficult for perception and remembering. In addition, five durations for sonification of each segment were presented: 300, 400, 500, 600 and 700 ms (T1-T7 exposition time per segment).

Five sessions were arranged and each of them was composed of 3 trials with middle-frequency tonal patterns. The trial was based on the text entry technique through a single button manipulation. In the preliminary test the schemes were compared to each other in an undistracted usage context. There was a 3 minutes demonstration about the text entry technique, the task for the participants and 10-minute training. After training period another test was arranged, consisting of two blocks. In the first block learning of the concept was tested in an undistracted usage context (help file), familiar with the cyclic synthesis alphabet tokens and numerals, particular notes. The second block tested a performance of the subjects in a usage context requiring testing 5 segment expositions with a single button of the keyboard (a space key).

Subjects

Nine university students 21 to 30 years of age served in this experiment. All participants had normal or corrected-to-normal vision with no known hearing deficit. This group was comprised of 6 males and 3 females. As the reason for the test was to study the auditory feedback, it was decided to utilize users already familiar with the other aspects of the music, but really not all the users were fit for this requirement.

Procedure

The subjects were presented with a computer screen with playback the sound (speech synthesis) of the letters to be entered one by one without seeing the visual images when the user pressing the button at a time. The letters array was reduced to 27 characters to have unique combinations of segments. Each trial consisted of 27 test symbols presented in a random sequence. After each trial experimenter changed exposition time of the segments. The order of auditory feedback schemes presented was varied so that as many permutations as possible were used. That was done to avoid lower performance due to both learning effect and subject's fatigue during monotonous work. To simplify following analysis, and to avoid having to use even bigger number of permutations, the texts were normalized.

Before each test the subjects were given a chance to familiarize themselves with the test by entering text with the auditory feedback. The subjects were instructed not to see on the screen when doing the test. And during the test computer screen was shut down. They were also instructed to pay more attention on any errors. This was to avoid spending time missing the next sound letter. It is very natural to intuitively correct the error right on the second time after making it that the subjects were not forced to avoid their natural behavior.

Figure 7 shows a model of transforming (mapping) temporal events (tonal sequences) into spatial seven-segment layout and conversely. We have suggested that simple actions with temporal events could integrate an imagination and a perceptive experience of the person.

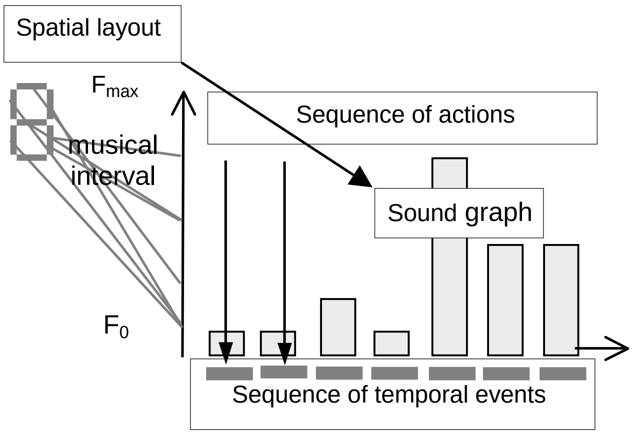


Figure 7: Conceptual model of the spatial–temporal transformation and sound mapping

Efficiency is the success rate of typing, divided by the time spent typing [11]. The success rate can be calculated by subtracting the percentage of typing errors from 100%. Amount of the errors remaining in the final result were taken in account in calculating the error percentage. Most of statistics was provided by used software.

RESULTS AND DISCUSSION

The First Test Session

In the first session, three different sound feedbacks were tested against each other. The purpose was to find if there were any significant differences in recognition of the patterns. It was found in the test that 5 out of 9 subjects had preferred the normal rhythm of the simple harmonic sequences. For them the typing speed should not be affected but only the error rates. The errors recorded through 3 trials within 3 different tonal sequences are shown in Figure 8.

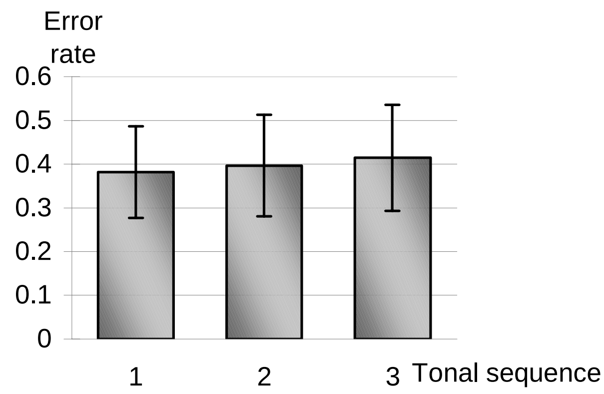


Figure 8: Rate of error occurrences with different feedbacks

Each result of the subjects was analyzed and then overall statistics was processed. The errors were compared to each other using F-test. It can be said that the concept of the first tonal pattern was better than two other ones. This simple mapping has produced fewer errors. The sequences with own rhythm have evoked confusion and additional distraction.

In general, we can conclude that the sonification of the simple harmonic sequences was the most acceptable and efficient among three proposed tonal sequences in a particular case of the mapping temporal events onto seven-segment spatial layout.

The Second Test Session

After the first session, each subject was familiar with the sonification mapping and technique to compose the characters. The same 9 subjects were served again. In the first session the subjects only found that the first tonal sequence is more efficient but did not reveal any significant differences when they were composing the letters by listening for a sound feedback. From previous investigation, the tonal patterns depended on a speed on which the person can act. Even if the subject had experience, s/he could not predict what next symbol will be presented. Therefore the first segment is the most difficult, at least, during the test [3]. According the results of the first session in the following experiments only the first tonal pattern was used. I analyzed the process of input for each symbol in with the same condition (segment exposition 500 ms) and found out that time of the decision making brings an essential component into the duration of user manipulation.

From the data collection, we found that first, there are some letters, those difficult to entry with sound feedback only (Figure 9). By the segments decreasing, the validity was getting higher and higher. If there is the empty segment among the cyclic synthesis, it is easy to getting mistake and difficult to make decision, here like “D, G, Q”. Otherwise, regardless of settings, if the subjects made a wrong decision in the first step, the result would be incorrect.

Second, Decreasing exposition time of segments lead to the short sounds which would be not clearly recognized. This unfinished process of segment recognition disturbs and breaks a sequence of actions (Figure 10) and perception integrity of the sound pattern.

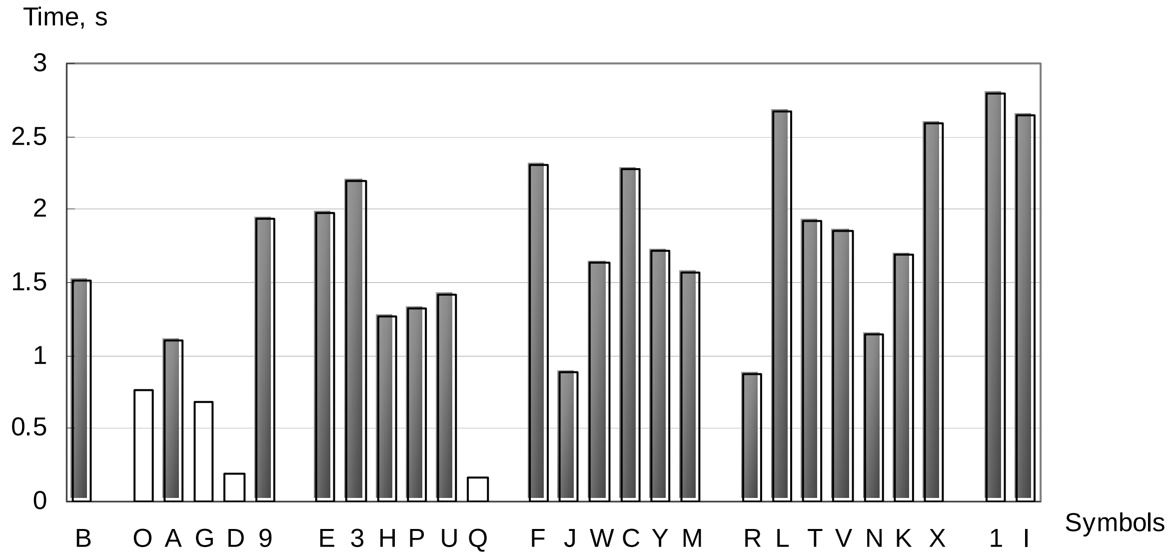


Figure 9: Relative difficulty in dependence on an amount and position of symbol's segments

In Figure 10 it is obviously that the correct rate of decision making is decreasing by the segment exposition getting shorter.

As for key tones, in general most of the participants said they would not require changing the tested notes. Three of nine subjects said spontaneously that they would like to input text without seeing screen but listening tones only, i.e. like blind people.

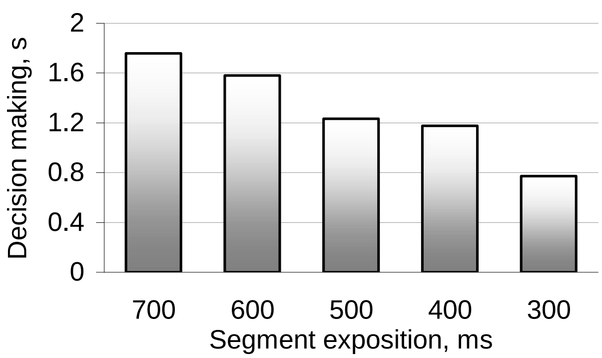


Figure 10: Decision making with different expositions time of segments

Conclusion

In this paper, the conceptual modeling for spatial-temporal structures was considered and constructive research was carried out to simulate the temporal components and relationships. Usability evaluation of user behavior in different time conditions and communication with computer without visual feedback was done. The rhythmic musical sequences were coordinated to spatial seven-segment layout of the software indicator. I suppose that the sonification should integrate in visual display a quality comparable to what the potential users normally heard. Auditory feedback can strengthen and supplement the visual interpretation of data. It should support pointing and fixation of attention to provide engagement between spatial and temporal which can be easy to learn and to use.

Through a temporal pattern it is necessary not only a high accuracy and repeatability of manipulations, but a good perception of the rhythm. As shown, there were some measurable differences between the different sound feedbacks (Figure 8). A simple spatial layout could decode the state of a process or simplify perception, facilitating navigation through sequence of actions [7]. Also for the error situation, more understandable spatial layout lowers error rate. To manipulate by temporal events by relying on internal temporal perception only it might be difficult. If a design of temporal structure will include certain "demands" like musical rules, it could have a good prospect.

REFERENCES

1. Barrass, S. Auditory information design. Ph. D. Thesis. Australian National University, Canberra, Australia, 1998.

2. Choi, I. & Bargar, R. Interfacing sound synthesis to movement for exploring high-dimensional systems in a virtual environment. In *Intelligent Systems for the 21st Century, IEEE International Conference on Systems, Man and Cybernetics* (vol. 3), 1995: 2772-2777.

3. Evreinov, G. & Raisamo, R. Cyclic Input of Characters through a Single Button Manipulation. *In Proceedings of the 8th Int. Conference on Computers Helping People with Special Needs*, ICCHP 2002 (15-20 July, Linz, Austria), LNCS Vol. 2398, Springer-Verlag Berlin Heidelberg, 2002, pp. 259-266.

4. Foote, J. & Cooper, M. Visualizing Musical Structure and Rhythm via Self-Similarity. *In the ICMC 2001 Conference Proceedings.* Havana, Cuba. Foote,

Jonathan. Music Visualization and Analysis. http://www.fxpal.com/people/foote/musicseg/

5. Gaver, W.W. Auditory interfaces. In *Handbook of Human-Computer Interaction* (ed. Helander, M.G., Landauer, T.K. & Prabhu, P.), 2nd Edition, Elsevier Science, Amsterdam, 1997.

6. Gröhn, M. & Takala, T. MagicMikes – method for spatial sonification. In *Proceedings of SPIE* 2410 (Visual Data Exploration and Analysis II), 1995: 294-301

7. Kirsh D.: Complementary Strategies: Why we use our hands when we think. In Proc. of the 17th Annual Conference of the Cognitive Science Society. Hillsdale, NJ, Lawrence Erlbaum, 1995. Available at: http://cogsci.ucsd.edu/~kirsh/Cogsci95/cogsci95.html

8. Kramer, G. & al. *NSF Sonification Report: Status of the field and research agenda*, 1999. Available at: http://www.icad.org/websiteV2.0/References/nsf.html.

9. Kramer, G. Some organizing principles for representing data with sound. In *Auditory Display: Sonification, Audification and Auditory Interfaces* (ed. Kramer, G.), SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley, Reading, Mass., 1994: 185-222.

10. Kramer, G. An introduction to auditory display. In *Auditory Display: Sonification, Audification and Auditory Interfaces* (ed. Kramer, G.), SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley, Reading, Mass., 1994: 1-77.

11. Ronkainen, S. & Marila, J. Effects of Auditory Feedback on Multitap Text Input Using Standard Telephone Keypad. *Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan,* July 2-5, 2002. Available at: http://www.icad.org/websiteV2.0/Conferences/ICAD2002/proceedings/57_ronkainen-marila.pdf

12. Scaletti, C. & Craig, A.B. Using sound to extract meaning from complex data. In *Proceedings of the SPIE* vol. 1459, San Jose, 1991: 207-219.

13. Smith, S., Grinstein, G.G., & Bergeron, R.D. Stereophonic and surface sound generation for exploratory data analysis. In *Proceedings of CHI'90*, ACM Conference on Human Factors of Computing Systems, 1990: 125-132.

14. Wenzel, E.M. Spatial sound and sonification. In *Auditory Display: Sonification, Audification and Auditory Interfaces* (ed.Kramer, G.), SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley, Reading, Mass., 1994: 127-150.

Text Input Through One-Dimensional Head Tracking

Juha Pieviläinen
Department of Computer and Information Sciences
University of Tampere
juha.pievilainen@uta.fi

ABSTRACT
In everyday life we encounter many situations where we need to manipulate some machine when using hands is not possible. Nowadays there are commercial solutions that allow hands-free remote control by the stationary device such as computer, TV-set or similar one. But they do not allow mobility for user as these devices are based on interaction between external unit (IR/RF transmitter and receiver or receiver) and marker (mirror or transmitter) attached to user head, forehead, finger, glasses etc. This paper describes a new technique for head tracking that allows the user to be mobile by using users torso as a reference point. While usability testing for prototype was carried out in stationary conditions, our input device allows mobility of the user and could be applied for wearable device like mobile phone, PDA or alternative input device for people with special needs.

KEYWORDS: hands-free technique, head tracking, wearable device, SymbolCreator

INTRODUCTION
In everyday life we encounter many situations where we need to manipulate some machine when using hands is not possible. Computer users who cannot use a conventional hand-operated computer mouse and/or keyboard due to some disability or an injury might use a head operated mouse or joystick in order to control their computer through on-screen keyboard, to type any text, draw graphics etc. [1].

Head tracking can be based on the following techniques: mechanical, magnetic, non-inertial, optical, acoustic and inertial [2]. No of these techniques are for mobile use. This is because each of these techniques requires an external stationary source or coordinates system regarding which head position could be measured. In this paper a new approach for designing a head tracker was proposed and usability evaluation was carried out. The tracker was based on radio waves and directional antennas. The goal was to check the resolution, sensitivity and reliability for pointing and selection of software controls, by using a new technique of text entry for mobile devices.

DEVICE DESIGN
The device uses radio waves and directional antennae both in transmitter and receiver. It allows detecting changes in frequency or magnitude depending on relative position of two antennae. The device consists of generator, receiver, antennae and connection cables. The generator and receiver build in a small box, which is connected to the antenna by thin screened cable. This assembly makes up the main unit (Figure 1). In addition there is a connection cable connecting the box to the joystick port of the PC. The prototype is powered by two standard AA 1.5 V batteries, which are placed in the same unit. Changing of radio signal, due to a head rotation in horizontal dimension, is transformed into the input current of X-coordinate of the joystick port. A potentiometer was used instead of Y-coordinate.

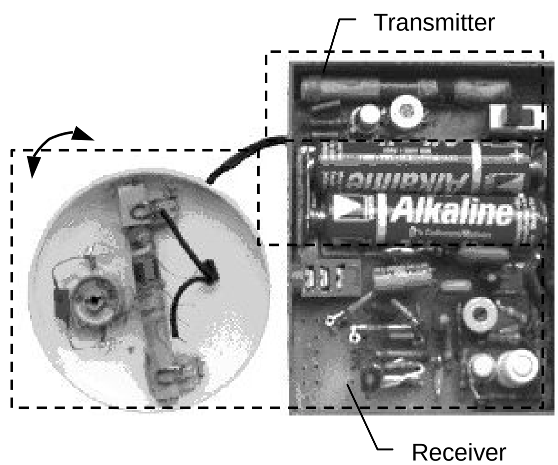


Figure 1. Directional antenna and design of the unit inside

The antenna was placed on top of the subject's head and held still by a thin metallic band. The main unit was placed on the back near the neck (Figure 2).

USABILITY EVALUATION
The usability evaluation was carried out by doing a comparative empirical research of one-dimensional head tracking and conventional input techniques through mouse and joystick with text entry technique for mobile devices called SymbolCreator. The decisive factor for comparing the entry methods was percentage of missed letters.

Software

For usability testing of the head tracker software (Figure 2), designed for mobile application, was modified and adapted in Microsoft Visual Basic 6.0. The software called SymbolCreator is a new text input technique based on cursive handwriting elements to compose alphabet characters.

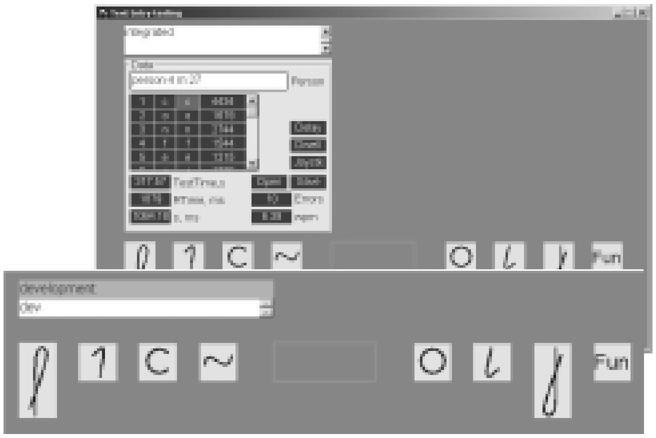


Figure 2. Snapshot of SymbolCreator interface and fragment in the testing mode

Figure 2 demonstrates that the SymbolCreator uses seven basic segments and their combinations creating letters and one button (Fun) for choosing special options for editing. The software randomly chooses 20 test-words from the list of 150 words and sequentially presents them for retyping.

Subjects and Procedure

6 volunteer experienced computer users were hired for usability study. Subjects were 21 to 30 years old. Each subject implemented the test three times with 5 minute break between sessions. Each time they used a different control method, a cordless mouse, a conventional joystick and the head tracker device. The subjects were given 10 minutes to get used to the SymbolCreator, at this point they were using mouse and were given SymbolCreator alphabet on paper. The paper was removed before testing. Every subject had at least 30 minutes experience with the head tracker device. Dwelling time was set at 530 ms. Dwelling time was used only with head tracker.

RESULTS

In preliminary testing it was noted, that optimal dynamic range for head movements could be defined during calibration. Therefore a special care was taken when calibrating the device for each subject. Several pre-processing steps were performed in order to reduce noise in radio channel occurred due to amplitude modulation to remove most of the variability of the cursor position. The preliminary results predicted that the head tracking would have more errors in comparison with conventional pointing methods.

As the total amount of letters in the test word was not constant the results were compared through percentage of missed letters. The data showed that with average error rate of 10.4% the mouse is the less error-prone input device. With mouse the error rate ranged from 1.1% to 34.2%. The second most accurate input device was the joystick, which had average error rate of 10.6% and error rate ranged from 1.1% to 34.1%. The head tracker was proven to be the most error-prone input device with average error rate of 63.4% and error rates ranging from 45.7% to 98.3%. There were some differences between test subjects (Figure 3). Some subjects seemed to adapt to the SymbolCreator much easier than others.

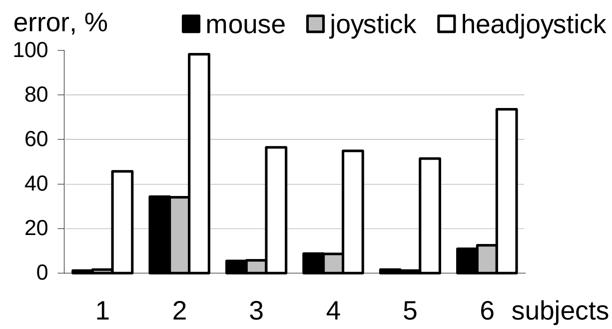


Figure 3. Missed letters' rate with different input devices

CONCLUSION

According to the results received during usability evaluation of the head tracker, it was found out that proposed design to be too sensitive for wearable application it to be useful. Also because of the nature of the construction the vibration of the cursor is really an issue that should be dealt with for this technique for constructing a head tracker to be usable in any situation. The sensitivity problem and increasing dynamical rage could be solved if relative movements (mouse input) would be recorded instead of absolute positioning technique like joystick input.

For a side note, SymbolCreator was found to be a quite good method for text-entry.

REFERENCES

1. Evans, G. and Blenkhorn, P. *A Head Operated Joystick – Experience with use*, *CSU*N 99, available at http://www.csun.edu/cod/conf/1999/proceedings/ session0019.htm

2. Youngblut, C., Johnson, R.E. et al. *Review of Virtual Environment Interface Technology, Institute for Defence Analyses*, available at http://www.hitl.washington.edu/scivw/scivw-ftp/publications/IDA-pdf/