

TEXT INPUT TECHNIQUES FOR IMMERSIVE VIRTUAL ENVIRONMENTS: AN EMPIRICAL COMPARISON

Doug A. Bowman¹, Christopher J. Rhoton¹, and Marcio S. Pinho²

¹Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, Virginia, USA

²Instituto de Informática, UFRGS
Faculdade de Informática, PUCRS
Porto Alegre, RS, BRAZIL

Symbolic input, including text and numeric input, can be an important user task in applications of virtual environments (VEs). However, very little research has been performed to support this task in immersive VEs. This paper presents the results of an empirical evaluation of four text input techniques for immersive VEs. The techniques include the Pinch Keyboard (a typing emulation technique using pinch gloves), a one-hand chord keyboard, a soft keyboard using a pen & tablet, and speech. The experiment measured both task performance and usability characteristics of the four techniques. Results indicate that the speech technique is the fastest, while the pen & tablet keyboard produces the fewest errors. However, no single technique exhibited high levels of performance, usability and user satisfaction.

INTRODUCTION

Very few text/numeric input techniques been proposed for immersive virtual environments (VEs). We speculate that there are two reasons for this. First, symbolic input may seem to be an inappropriate task for immersive VEs. Many of the most successful VE applications are simple walkthroughs/flythroughs, where the user is mostly a passive observer. In this context the VE is seen simply as a visualization tool, while the “real work” is done back at the desktop. Clearly, this situation is changing, as more and more highly interactive VE applications become prominent. There are many situations where symbolic input would be useful in such applications, such as:

- leaving a brief, precise annotation for the designer in an architectural walkthrough
- entering filenames for open/save operations
- adding labels to virtual objects
- specifying numeric properties (e.g. thickness, position) for virtual objects
- setting parameters in a scientific visualization

Second, there seems to be only one “natural” technique for symbolic input in VEs. One might claim that since we perform “symbolic input tasks” in the physical world through speech, we should use that technique in the virtual world as well. Speech should certainly play an important role in VE user interfaces, but this should be based on its desirable characteristics (hands-free, efficient, descriptive) rather than simply on its naturalism. Speech does not provide the perfect solution for every symbolic input scenario, however (especially those in which precision and semi-random strings are required). Thus, other techniques must be investigated.

The objective of the work presented here was to explore some of the most promising avenues for text input in immersive VEs. We were interested in techniques that allowed for efficient and error-free task performance, high levels of

learnability and user satisfaction, and comfortable operation. These goals led to the development of the Pinch Keyboard (Bowman, Wingrave, Campbell, & Ly, 2001), a technique that simulates typing on a standard keyboard in midair, implemented using Fakespace Pinch Gloves™. In this paper we present the results of an experiment comparing the Pinch Keyboard to a “soft keyboard” implemented in a pen & tablet metaphor, a one-handed chord keyboard, and an idealized speech recognition technique.

RELATED WORK

The most significant prior research on text input for VEs is Poupyrev’s Virtual Notepad system (Poupyrev, Tomokazu, & Weghorst, 1998). In this system, users carried a stylus and a graphics tablet while wearing a head-mounted display (HMD) to view the virtual world. Users could write or draw on the surface of the tablet with the pen, and see these pen strokes on a graphical representation of the tablet in the virtual world. This allowed for some simple text or numeric input. However, the input was saved only as a series of pen strokes, and not converted into actual text or numeric data. The Virtual Notepad also suffered from large tracker latencies, making it difficult to write or draw quickly. Nevertheless, this work suggests a potential technique (recognition of pen input) for entering text and numbers in immersive VEs.

Another technique that has been investigated involves mapping hand gestures to words/phrases (Fels & Hinton, 1998). This technique uses a data glove and a neural network recognizer to allow the user to produce synthesized speech. Such systems rely on the user’s memory for the gestures and the network’s ability to perform consistent and flexible recognition.

We can also draw from the fields of mobile and wearable computing for ideas, since users of such systems are constrained in many of the same ways as users of VEs. They are often standing; they are carrying or wearing devices; and

they do not have access to traditional keyboards. Besides pen-based input, four main approaches are used for symbolic input to wearable computers: speech, miniature keyboards in a standard layout, chord keyboards (Matias, MacKenzie, & Buxton, 1993; Noyes, 1983), and “soft” keyboards (Zhai, Hunter, & Smith, 2000) whose virtual keys are pressed by selecting them with a pointing device or finger. We have adapted three of these techniques for use in an immersive VE.

There has been at least one empirical study comparing text input devices for wearable computing (Thomas, Tyerman, & Grimmer, 1998). This study found that a forearm-mounted miniature keyboard performed best both initially and after significant usage time. This is important because it is often claimed that the standard QWERTY layout is inefficient, and that other devices/layouts could surpass its performance given enough learning time. In this study, at least, other devices did not perform as well as the QWERTY keyboard even after learning time was allowed. Two of the techniques tested in our experiment use this traditional layout, although both of them involve virtual, not physical, keys.

THE PINCH KEYBOARD TECHNIQUE

We have developed a technique for text input in VEs called the Pinch Keyboard. It uses Pinch Gloves™ (figure 1), lightweight gloves with conductive cloth on each fingertip that sense when two or more fingers are touching. The gloves are comfortable to wear, and because of their discrete nature, there is no ambiguity to user actions. Our technique also uses a standard QWERTY keyboard layout, so that users can take advantage of the typing skill they already have.



Figure 1: User wearing head-mounted display and Pinch Gloves™

The basic concept of the Pinch Keyboard is that a simple pinch between a thumb and finger on the same hand represents a key press by that finger. Thus, on the “home” row of the keyboard, left pinky represents ‘a’, left ring represents ‘s’, and so on. We also need the ability to use the “inner” keys such as ‘g’ and ‘h’, and the ability to change rows of the keyboard. We accomplish this through the use of 6 DOF trackers mounted on the gloves. Inner keys are selected by rotating the hand inward. Users calibrate the location of the rows before using the system by indicating the middle of the top and bottom rows while holding the hands palm-down.

Visual feedback indicates which letters can be typed at any given time (figure 2).

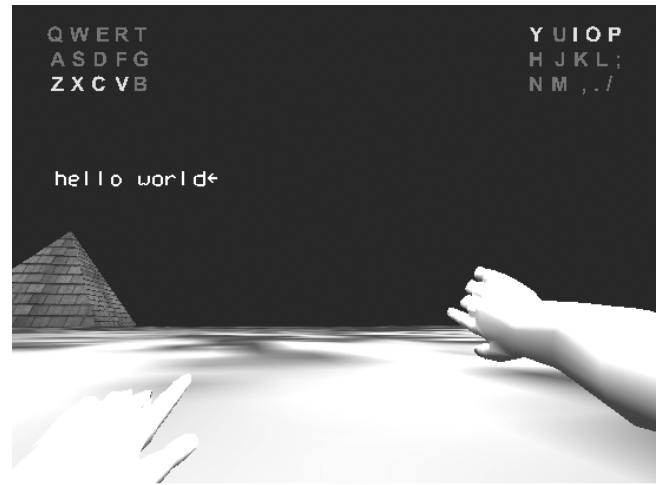


Figure 2. User’s view of the Pinch Keyboard technique

EXPERIMENT

We had previously run a small user study to demonstrate the usability of the Pinch Keyboard technique (Bowman et al., 2001). Our objective here was to do a summative evaluation, comparing the Pinch Keyboard to other candidate techniques for text input in immersive VEs. We tested four techniques in total, and measured both task performance and the usability of these techniques.

Techniques

The four techniques compared were the Pinch Keyboard, a pen & tablet keyboard, a chord keyboard, and speech. The Pinch Keyboard technique has already been described above.

The pen & tablet keyboard technique is a soft (virtual) keyboard implemented within the pen & tablet metaphor (Angus & Sowizal, 1995). The user holds a tracked physical pen (stylus) and tablet, and sees a virtual pen and tablet in the VE (figure 3). Virtual keys in the standard QWERTY layout are displayed on the surface of the virtual tablet, and users type a letter by touching it with the stylus then pressing the stylus button.

The chord keyboard technique uses a commercially available device, the Twiddler2 chord keyboard (figure 4). This is a 12-key device that can be held in either hand. Characters are produced by depressing either a single key or multiple keys (a chord). We provide a visual aid in the HMD that shows the user the layout of the keys on the device, so that even novice users can determine which keys to press.



Figure 3. Physical (left) and virtual (right) view of the pen & tablet keyboard

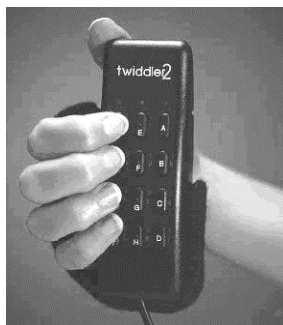


Figure 4. Twiddler2 chord keyboard

For the speech technique, we use a “wizard of oz” approach. There is no actual speech recognition software; rather, one of the evaluators listens to the user’s utterances and types on a traditional keyboard to input the text. We made this decision because speech recognition software may require training; it may have high error rates; and it may be difficult to integrate with the existing VE software. We wanted to test an idealized speech technique that would not be hindered by such recognition or implementation issues. To allow for fair comparison with the other techniques, we only allowed users to speak a single character at a time, rather than entire words. This is a reasonable assumption for many text input situations where the desired text is a non-word string (e.g. filenames).

In each of the techniques, the typed characters appear in the middle of the display. An audible click is played when the user types a letter. For the experiment, users were placed in a simple VE consisting only of a ground plane and objects specific to the technique being used (e.g. virtual hands, virtual pen & tablet, visual aids).

Apparatus

All users wore a Virtual Research V8 head-mounted display (HMD) for immersion in the virtual world. This HMD operates at 640x480 resolution. We used the HMD in biocular mode (same image presented to both eyes). A Polhemus Fastrak device was used to track the user’s head; this device also tracked the user’s hands in the Pinch Keyboard technique and the devices in the pen & tablet technique. Images were generated by a personal computer running Windows NT. The frame rate was at least 30 fps for all techniques. The applications were written using the SVE library (Kessler, Bowman, & Hodges, 2000).

Experimental Design

The experiment had a between-subjects design (each subject used only one of the techniques). Each trial consisted of a word or phrase being presented in the visual field of the subject. Words and phrases were between three and fifteen characters long. The subject was to type the word or phrase; the trial ended when the subject had successfully completed the entire word or phrase. Audio feedback told the user when the trial had been completed. There was a one second pause in between trials to allow the subject to prepare for the next trial. Each subject was to complete 72 total trials, although some subjects completed fewer trials due to technical problems during the experiment.

The independent variable in the experiment was the interaction technique used. Dependent variables were time for task completion (dependent on word/phrase length), number of correct characters typed per minute (independent of word/phrase length), number of typing errors, and subjective comfort ratings in five categories (arm strain, hand strain, neck strain, dizziness, and nausea). Time for task completion was measured automatically by the system, and the number of characters in the word/phrase was divided by this value to obtain the characters per minute measure. The number of errors was recorded manually. Comfort ratings were each on a ten-point scale, and were obtained at the beginning of the experiment and after each set of trials.

Subjects

Twenty-eight subjects participated in the experiment (seven subjects for each technique). Subjects were recruited from undergraduate computer science classes and received extra credit for their participation. There were 23 males and 5 females, and the mean age of the participants was 19.92 years.

Procedure

Subjects were given written instructions for the experiment in general and the particular technique they were to use. After this they donned the HMD and other devices for the technique being tested.

Trials were divided into three sets of 25 trials each. The first three trials in the first set were practice trials and were not counted in the statistics, resulting in 72 timed trials. After each set of trials the subject was allowed to take off the HMD and take a rest break if needed.

RESULTS

Interaction technique had a statistically significant effect on time per trial ($F=55.67$, $p < 0.001$) and on characters per minute ($F=46.00$, $p < 0.001$). We will focus on the characters per minute (cpm) metric, since it provides a more accurate representation of the speed of the technique.

Post-hoc tests revealed that the mean performance of the speech (65.99 cpm) and pen & tablet (49.68 cpm) techniques was significantly better than the mean performance

of the Pinch Keyboard (31.69 cpm) and chord keyboard (21.13 cpm) techniques. Even though there was a large absolute difference between the mean performance of the speech and pen & tablet techniques, this difference was not statistically significant due to very high variability in subjects' performance with the speech technique.

Learning appears to have been an important factor for all of the techniques except speech. As figure 5 shows, users of the other three techniques greatly improved their performance over the course of the experiment.

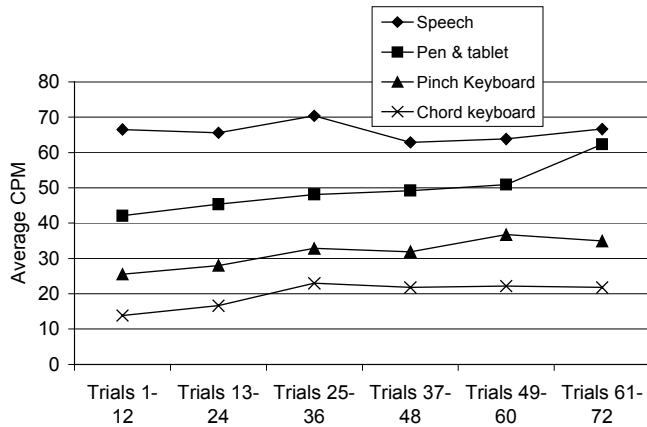


Figure 5. Learning curves for the four tested techniques

Technique also had a statistically significant effect on the total number of errors made by subjects ($F=12.49$, $p < 0.001$). The fewest errors were made with the pen & tablet technique (7.14 errors per subject), followed by speech (22.43), Pinch Keyboard (43.17), and chord keyboard (81.43).

Subjects made approximately the same number of errors, on average, during each set of trials (the error rate was not improving over time). The exception to this rule was the chord keyboard. Subjects made 40.86 errors on average with this technique during the first set of trials, but only 16.57 errors during the second set. This indicates that subjects were still learning how to produce the proper character with the chord keyboard for a significant number of trials.

The subjective comfort ratings also produced some interesting results. Figure 6 shows the average difference between the initial comfort rating and the final comfort rating for each of the four techniques and each of the five comfort categories. In other words, we used the initial rating as a baseline against which the other ratings are compared to determine the effect on comfort of using the technique. As the figure shows, the pen & tablet technique produced moderate arm strain and high levels of hand strain, while the chord keyboard produced moderate levels of hand strain and neck strain.

Observations taken by evaluators during the experiment helped to explain and expand upon the quantitative results.

Although the speech technique performed well, errors with this technique were often a problem. Errors occurred either when the subject misspoke or when the evaluator mistyped or misheard the subject. Since there was usually a

slight delay in between the utterance and the appearance of the character on the display, subjects tended to be one or more characters ahead of the evaluator. Therefore, when an error occurred, especially a “system” error, subjects would have to delete several characters to correct the error. Subjects also had more difficulty recognizing when an error had occurred with the speech technique. The other usability issue noted with the speech technique was that most subjects did not speak as fast as they would like; rather, they tried to match their speed to the perceived speed of the system.

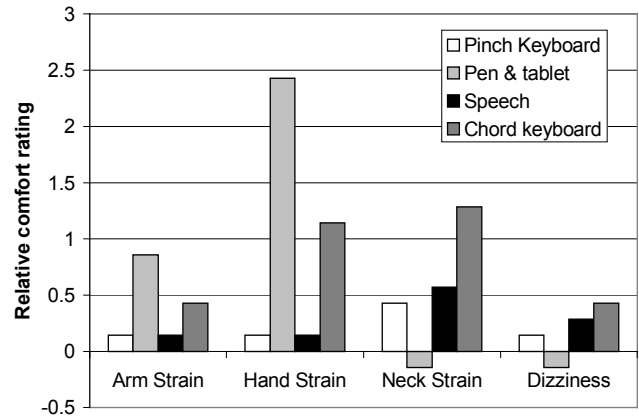


Figure 6. Differences between the initial and final subjective comfort ratings for each of the four techniques in the experiment

The Pinch Keyboard technique also had some crucial usability issues. Some subjects had trouble making contact between two fingers, forcing them to type the same character several times. In addition, no subjects were observed using the alternate technique for accessing inner keys (pinching the thumb to both the index and middle fingers rather than rotating the hand) even though this technique was explained in both the written and verbal instructions. A problem with the hand rotation technique is that the rotation often changes the hand position as well – an example of the “Heisenberg effect” (Bowman et al., 2001) – causing the active keyboard row to change. Calibration of the rows with the Pinch Keyboard technique seemed to be a critical factor in determining performance. If the distance between rows was too small or too large, trial time suffered.

We also observed differing emotional responses by users to the various techniques. In general, most subjects were visibly more interested and engaged when using the Pinch Keyboard and the pen & tablet keyboard, while subjects using the speech technique seemed bored and subjects using the chord keyboard appeared frustrated.

Finally, the post-experiment interview revealed some important information about usability and user preference. We asked subjects if the technique they had used felt “natural” to them. In almost all cases, subjects who had used a traditional keyboard layout (Pinch Keyboard or pen & tablet keyboard) responded that the technique was natural. Users of these techniques also thought they were quite easy to learn and understand. Surprisingly, most users of the speech technique

felt it was somewhat unnatural, both because they are not used to spelling words aloud and because they had to speak more slowly than they desired. We also asked subjects about their perception of their performance. Almost all subjects felt that their performance improved over the course of the experiment. Interestingly, most subjects also reported that they could continue to improve their performance with additional practice. Users of the speech technique, however, felt that their improvement was constrained by the recognition speed of the system.

DISCUSSION

The experiment showed that none of the techniques we implemented is clearly the best for text input in immersive VEs. The speech technique was the fastest, but it also produced more errors than the pen & tablet keyboard, and was found to be tedious by many of the subjects. The pen & tablet keyboard was relatively fast, had the fewest errors, and was reported to be natural and easy to learn, but it also produced high levels of arm strain. Subjects found the Pinch Keyboard technique natural and easy to learn, but its performance was sub-par due to some unresolved usability issues. The most definitive single statement that can be made from the experiment is that the chord keyboard (at least the one we used) should not be the device of choice for text input in VEs.

We can posit a few general guidelines for the use of text input techniques in actual VE applications. First, the technique used must integrate well with other interaction software and hardware used in the application. If both hands need to be free to do other tasks in the VE, then only speech and the Pinch Keyboard (of the techniques we tested) need be considered. If at least one hand must be free, the pen & tablet technique may still be used assuming that the stylus can be used for the other interaction tasks. The pen & tablet technique has the advantage that it integrates well with other tablet-based 2D interfaces, allowing the same basic interaction technique to be used for many different interaction tasks. This is true of the speech and Pinch Keyboard techniques as well, since speech may be used for system commands and pinch gestures may be used for commands, object selection, etc. In all three cases, however, text input would likely be an explicit mode of operation, which may not be desirable. Finally, our subject interviews and observations showed that if user satisfaction and engagement is an important factor, then the pen & tablet keyboard and the Pinch Keyboard should be considered.

CONCLUSIONS AND FUTURE WORK

In this paper we have explored the task of entering text into an immersive VE. We have argued that text input is and will be an important task for many highly interactive VE applications. The technique we have implemented, the Pinch Keyboard, is a novel approach using Pinch Gloves™ and the traditional keyboard layout. We have described an experiment and its results comparing four important candidates for text input in immersive VEs. Although the results of the experiment do not suggest a definitive answer to the question of which text input technique should be used in VEs, the

experiment did increase our understanding of the various techniques and the usability and performance issues related to text input. In fact, such ambiguous results are typical in empirical comparisons of 3D interaction techniques, because no single factor (task performance, user satisfaction, comfort, etc.) is clearly the most important.

There is still much work to be done in this area. In particular, pen-based techniques for text input should be designed and evaluated. A simple pen-based input mechanism such as those used on today's personal digital assistants (PDAs) could be implemented using a pen & tablet metaphor in an HMD-based VE, or using an actual PDA in a projection-based VE. Further work is also needed to understand the different categories of text input tasks for VE applications, and the performance and usability of text input techniques for each of these categories. For example, we might find that a soft keyboard is much more usable for filename entry tasks than a speech-based technique. Finally, these techniques need to be evaluated in the context of full-featured interactive VE applications, so that the integration of these techniques into a complete user interface can be studied.

ACKNOWLEDGEMENTS

The authors thank the subjects in the experiment for their time and effort. We also acknowledge Drew Kessler for his help with the SVE toolkit, and the members of the 3DUI mailing list for their discussion of text input in VEs. Marcio Pinho was supported by grant number BEX0316/01-6 from the Brazilian Foundation for the Coordination of Higher Education and Graduate Training (CAPES).

REFERENCES

- Angus, I., & Sowizral, H. (1995). *Embedding the 2D Interaction Metaphor in a Real 3D Virtual Environment*. Proceedings of SPIE, Stereoscopic Displays and Virtual Reality Systems, 282-293.
- Bowman, D., & Wingrave, C. (2001). *Design and Evaluation of Menu Systems for Immersive Virtual Environments*. Proceedings of IEEE Virtual Reality, 149-156.
- Bowman, D., Wingrave, C., Campbell, J., & Ly, V. (2001). *Using Pinch Gloves for both Natural and Abstract Interaction Techniques in Virtual Environments*. Proceedings of HCI International, New Orleans, Louisiana.
- Fels, S., & Hinton, G. (1998). Glove-TalkII: A Neural Network Interface which Maps Gestures to Parallel Formant Speech Synthesizer Controls. *IEEE Transactions on Neural Networks*, 9(1), 205-212.
- Kessler, G., Bowman, D., & Hodges, L. (2000). The Simple Virtual Environment Library: An Extensible Framework for Building VE Applications. *Presence: Teleoperators and Virtual Environments*, 9(2), 187-208.
- Matias, E., MacKenzie, I., & Buxton, W. (1993). *Half-QWERTY: A One-handed Keyboard Facilitating Skill Transfer from QWERTY*. Proceedings of ACM INTERCHI, 88-94.
- Noyes, J. (1983). Chord Keyboards. *Applied Ergonomics*, 14, 55-59.
- Poupyrev, I., Tomokazu, N., & Weghorst, S. (1998). *Virtual Notepad: handwriting in immersive VR*. Proceedings of the Virtual Reality Annual International Symposium, Atlanta, 126-132.
- Thomas, B., Tyerman, S., & Grimmer, K. (1998). Evaluation of Text Input Mechanisms for Wearable Computers. *Virtual Reality: Research, Development, and Applications*, 3, 187-199.
- Zhai, S., Hunter, M., & Smith, B. (2000). *The Metropolis Keyboard - an Exploration of Quantitative Techniques for Virtual Keyboard Design*. Proceedings of the ACM Symposium on User Interface Software and Technology, 119-128.