MultiPoint: A Case Study of Multimodal Performance for Building Presentations

Anoop K. Sinha, Michael Shilman, Niraj Shah
Group for User Interface Research
CS Division, EECS Department
University of California at Berkeley
Berkeley, CA 94720-1776 USA
+1 510 642 3437
aks@cs.berkeley.edu • http://guir.berkeley.edu

ABSTRACT

MultiPoint is a speech and pen user interface for building presentations, implemented as an add-on to Microsoft PowerPoint™. In this study, we compared users' performance between building presentations with MultiPoint and with PowerPoint. We also compared participants' performance between using Wizard of Oz (WOz) speech recognition and computer speech recognition. In aggregate, six participants with WOz recognition completed tasks in about the same time and with about the same number of errors, as they did using PowerPoint alone. Six participants with computer speech recognition took twice the time and committed four times as many errors, as they did using PowerPoint.

Keywords

Multimodal user interfaces, speech and pen input, recognition, Wizard of Oz, measurement

INTRODUCTION

Pen and speech interfaces are likely to become more important in the future with increasing use of tablet computers and other handheld devices. In an attempt to expand on the tasks where multimodal input might be more appropriate, we compared the performance of a multimodal interface for building presentations to a regular graphical user interface with keyboard and mouse.

Our interviews with professionals in industry led us to the conclusion that building presentations is a task potentially well-suited to a multimodal interface on a tablet computer. Out interviewees often draw the first draft of their slides; we designed MultiPoint as a tool for creating the first draft of a set of slides on a pen computer.

MULTIPOINT IMPLEMENTATION

Multipoint is implemented as an add-on to Microsoft PowerPoint and uses SRI's Open Agent Architecture (OAA) [3]. It communicates to PowerPoint via Visual Basic for Applications (VBA). OAA facilitates synchronous and asynchronous communication managed by a central facilitator among different recognition agents in the system. This provides the facilities for pen and speech recognition in MultiPoint.

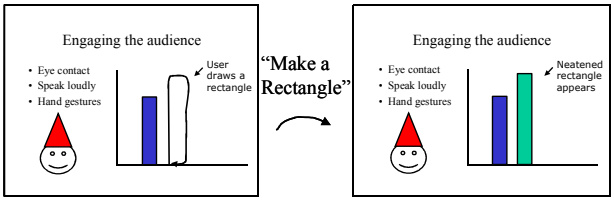

Figure 1. The MultiPoint command "make a rectangle" converts a sketched rectangle into neatened form

MultiPoint includes a subset of PowerPoint functionality; there are approximately 60 multimodal commands that can be used to draw and dictate slides. A user starts with the freeform drawing tool and draws a word or a shape on the slide. He simultaneously speaks speech commands to act on the freeform object.

As an example, the participant draws a rectangular shape, says, "make a rectangle," and the sketch is converted into recognized form (Figure 1). MultiPoint confirms the command audibly by repeating the recognized command via text to speech.

Speech commands in MultiPoint are particularly useful for

1. guiding recognition of sketched items
2. changing properties of existing objects
3. setting animations of existing objects

Some example commands used in the user study include:

- "add title"
- "add bullet"
- "make a rectangle"
- "make a triangle"
- "color shape red"
- "fly from right"
- "delete"
- "undo"

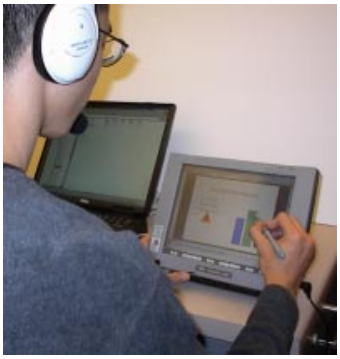

Figure 2. A participant using MultiPoint on a tablet computer

By explicitly switching into "dictation mode" a participant can also dictate text. Text can also be inputted via a handwriting recognizer agent.

METHOD

In the user study, 12 paid student participants were asked to copy four slides using MultiPoint and PowerPoint. For MultiPoint, each used a Fujitsu Stylistic 4000 pen computer and a headset microphone (Figure 2). For PowerPoint each used a laptop computer and an external mouse.

Each participant was trained on the MultiPoint commands and the PowerPoint operations that they could use to draw the slides. The copy tasks were presented in randomized order. One-half (six) performed their tasks with WOz speech recognition, and the other half (six) of the participants performed the tasks with computer speech recognition from IBM ViaVoice™.

Each task was timed, and the number of steps and errors were counted. Steps included "adding a square," "changing the color of a triangle," and were counted at the same granularity in both MultiPoint and PowerPoint. Errors included mistakes, undos, redraws, and speech misrecognitions.

RESULTS

In the WOz speech recognition tests, there were no statistically significant differences in the time (Figure 3) or the steps for completion (Figure 4) or the errors (Figure 5) between MultiPoint and PowerPoint. (There were no speech recognition errors in the WOz SR condition.)

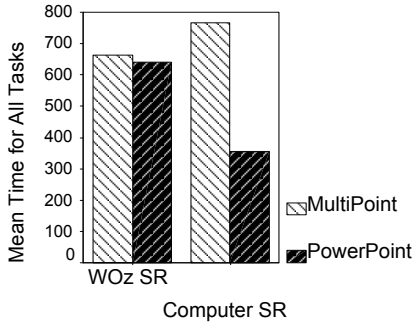

Figure 3. Time for completion in the two groups of participants

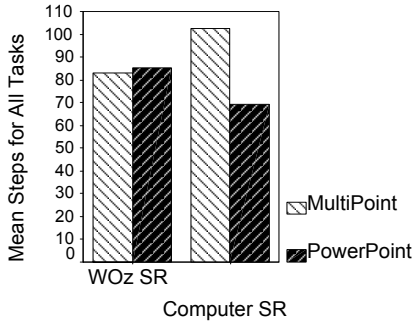

Figure 4. Steps for completion in the two groups of participants

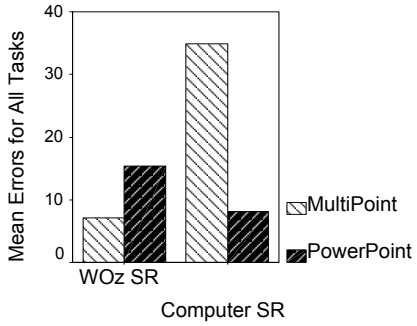

Figure 5. Errors for completion in the two groups of participants

For computer speech recognition, MultiPoint users took 215% of the time (M=409, SD=163), $t(5)=6.12$, $p<0.01$ (Figure 3), 148% of the steps (M=33, SD=13), $t(5)=6.11$, $p<0.01$ (Figure 4), and 427% of the errors (M=27, SD=13), $t(5)=5.11$, $p<0.01$ (Figure 5), versus PowerPoint. Most of the additional MultiPoint steps versus PowerPoint steps were from speech recognition errors.

For WOz speech recognition participants, there were no statistically significant differences in participant rankings of ease of use, quickness, and naturalness. All three beginners to PowerPoint ranked MultiPoint higher for ease of use, noting specifically its simplicity and pleasantness.

Participants using the computer speech recognizer were uniformly disappointed with the speech recognition performance. Those who used the computer speech recognizer perceived themselves as making many errors and waiting long times for commands to execute, though there was no statistically significant difference in the time per command versus WOz recognition. There was approximately one speech recognition error per four commands for computer speech recognition participants.

FUTURE WORK

Cohen, et al., have shown average 3.5 times performance improvement per step for multimodal input versus GUI input when mutual disambiguation was used [1]. Mutual disambiguation would no doubt help in MultiPoint, and we plan to explore it in the future, but it was not an issue in the case of WOz recognition in which there were no recognition errors.

MultiPoint performance could be improved by speeding speech recognition or excluding audio feedback. MultiPoint can also be expanded to include greater presentation building functionality.

CONCLUSIONS

Building presentations is indeed a viable task for multimodal interaction. However, novice computer users are more likely than expert computer users to be attracted to a multimodal interface for presentation building.

Recognition performance needs to be quite high to ensure good performance and user satisfaction in a multimodal application such as MultiPoint.

Participants need to feel that the multimodal application that they are using is proceeding smoothly and quickly. They do not like waiting for recognition results.

REFERENCES

[1]. Cohen, P. R., Johnston, M., McGee, D., Oviatt, S. L., Clow, J., Smith, I. "The Efficiency of Multimodal Interaction: A Case Study." *Proceedings of the International Conference on Spoken Language*, 1998.

[2] Oviatt, S., "Multimodal Interfaces for Dynamic Interactive Maps." *Proceedings of CHI'96*, pp. 95-102, 1996.

[3] Moran, D. B. and A. J. Cheyer, L. E. Julia, D. L. Martin, and S. Park, "Multimodal User Interfaces in the Open Agent Architecture," in *Proc. of the 1997 International Conference on Intelligent User Interfaces*, (Orlando, Florida), pp. 61--68, 1997.