Human Factors issues in multi-modal interaction in complex design tasks

Stéphane Rossignol, Louis ten Bosch, Louis Vuurpijl, Andre Neumann,
Lou Boves, Els den Os, Jan Peter de Ruiter

University of Nijmegen, NICI,
P.O. Box 9104, 6500HE Nijmegen, The Netherlands
S.Rossignol@nici.kun.nl

The European project COMIC (cf. http://www.hcrc.ed.ac.uk/comic) studies "conversational multi-modal interaction with computers". An important research issue in COMIC are the human factors aspects of computer-supported multimodal interaction in complex problem solving tasks. This article describes an experiment performed to get more insight in the human factors aspects of the specification of the ground plan of a bathroom using multimodal speech, handwriting, and pen-based gestures, as an example of a complex task.

1. Research questions

It is often said that, in general, users prefer speech to pen input, but whether that is the case in design applications is not obvious, cf. Oviatt (1999) who found that pen input is more often used to create graphics. How do users switch between within-task and about-task dialog acts? These questions are important for the design of (i) the user interface, (ii) the interaction dialogs, (iii) recognition engines for speech and pen input, and (iv) the fusion engine that merges multimodal input streams. In order to investigate the behavior of users when entering information, a pilot experiment was carried out in which 8 subjects entered the data of 9 blueprints of various complexity, using either only the pen, only speech, or pen and speech in combination. In addition, we intended to get insight in the gesture and speech repertoires used in the task.

2. Design of the experiment

The experiment comprised two parts and lasted about 60 minutes. In Part 1 the subjects had to enter information about the shape and dimensions of 9 bathrooms. Each bathroom blueprint was shown (printed on paper) to the subject, including a number of specifications (lengths, heights) in text form. We asked each subject to study the bathroom layout and enter the data into the computer: the shape of the room, the dimensions (length of walls), the position of the doors and how they open, and the width and height of the bottom of the window (if there was one). The subject was asked to "imagine" that he/she communicated with a computer that can understand his/her pen and speech input. The experimenter told the participant for each blueprint which modality (pen only, speech only or both) he/she had to apply.

In Part 2 these bathrooms had to be decorated. The sanitary may include bath, sink, etc., complemented by furniture such as mirror, cupboard, etc. Prior to the actual experiment, the participant was asked to sketch his/her own bathroom. Next, he/she had to perform Part 1 and Part 2 of the experiment. Finally, subjects completed a short questionnaire, which among others inquired about preferences for modalities.

3. Results

From the questionnaires it appeared that subjects did not like the ‘speech only’ mode. The ‘pen only’ mode was rated much more favorably, but overall subjects preferred the combination of pen

and speech. This is probably related to the fact that subjects were left totally free in the choice of the interaction mode in the 'pen plus speech' condition.

3.1 Pen gesture repertoire

In the context of this application, three pen gesture modes can be distinguished: (i) deictic gestures (e.g., tapping, encircling, arrows), (ii) handwriting, containing words, characters and digits (e.g., "The length of this wall is 3.23 m"), and (iii) drawing gestures (wall, door, window). For pen input, most dialog acts cannot be interpreted without contextual knowledge. For example, the interpretation of a line (as part of a door, window or wall), may have to be performed by considering its relative size (walls are longer than doors, doors are typically drawn with a certain angle with respect to a wall, walls are perpendicular to each other). Other kinds of pen input, in particular the specification of lengths, widths and heights, can only be interpreted if it is known what a user is specifying. This knowledge may stem from the dialog manager, or by combining the recognized specifications with recognized speech utterances. Although users tend to start with drawing all walls, then the doors and finally the windows, we observed much variation in this order. For all types of gestures we observed situations that will cause difficulties. For example, a user might first draw only a part of one object (e.g., a window) and after producing some more streams for other objects (walls, door), return to the unfinished object.

3.2 Speech repertoire

We observed large inter-subject variation in entering the shape of the room and the decoration. From the data it seems that a gesture act is supported by a speech act much more often than reverse. A gesture "icon" usually goes with quite an amount of speech. It is not straightforward to divide the incoming speech into speech acts on the basis of speech only. However, relations between a group of gestures on the one hand and a group of utterances on the other are can be distinguished. In the speech mode three types of speech acts can be distinguished. 1) Thinking loud ("I think I would place this over here"). This is not to be confused with mumbling, which did not occur frequently. 2) Requests for approval from the experimenter ("I hope this is OK"). 3) Factual input to the "interpreting computer" ("Here a bath.", and "Now I will..."). The fact that we have at least three classes of speech acts raises the question whether vocabulary and syntax are sufficient to recognize the class an utterance belongs to. Interestingly, there is a high degree of correspondence between the classes of speech acts that we observed, and the classes distinguished in the evaluation of the DARPA Communicator task: "dialog acts" (i.e., information transfer), "about dialog" speech acts (mainly asking for confirmation), and "about task" speech acts (Walker & Passonneau, 2001).

4. Conclusion

The results of our experiments show that subjects' behavior differs considerably in this design task. This variation is a serious challenge for the design of the interaction, especially with respect to the level of feedback. The question to what extent a specific feedback will limit the user in the variety of his actions will be the focus of subsequent experiments.

References

Oviatt, S. (1999) Ten myths of multimodal interaction. Communications of the ACM. Vol42, n11.

Walker, M. A., and Passonneau, R. (2001) DATE: A Dialog Act Tagging Scheme for Evaluation of Spoken Dialog Systems. *Human Language Technology Conference*. San Diego.