Cognitive Evaluation of the User Interface and Vocabulary of an Outpatient Information System

Andre Kushniruk, M.Sc.[1], Vimla Patel, Ph.D.[1], James J. Cimino, M.D.[2], and Randolph A. Barrows, M.D.[2]
[1]Cognitive Studies in Medicine: Centre for Medical Education, McGill University, Montreal, Quebec, Canada
[2]Department of Medical Informatics, Columbia University, New York, New York, U.S.A.

This paper describes an innovative approach to the evaluation of the user interface and vocabulary of a medical information system. The use of video recording for collecting usability data is detailed. The technique employed involves the collection of data consisting of transcripts of physicians as they "think aloud" while interacting with the system, along with a video record of the complete user-computer interaction. Using methods of analysis from cognitive science, the study was able to distinguish the source of physician problems in using the system's interface and in interacting with its controlled medical vocabulary. Analysis of the protocols indicated that all subjects encountered several generic problems, the most common ones indicative of a need for greater consistency in the interface design. Based on this evaluation, parts of the user interface have been re-implemented in an ongoing process of iterative system development.

INTRODUCTION

Two of the current challenges of medical computing, clinician data entry and coded data capture, appear to be in direct conflict. On the one hand, allowing clinicians complete freedom of expression suggests a user friendly approach, but such data are notoriously difficult to reuse for purposes such as automated decision support, clinical research, and case management. Natural language processing is sometimes used to generate coded data from narrative text, however the reliability of the coding is generally less than an ideal 100%. Some system developers resort to structured data entry to capture coded data; however, this approach is limited when the domain of discourse is large (such as patient problem lists).

In the Decision-supported Outpatient Practice system (DOP), under development at the Columbia-Presbyterian Medical Center (CPMC), clinicians record patient problems, allergies and medications through direct interaction with the CPMC Medical Entities Dictionary (MED).[1] Figure 1 shows a sample screen with two windows displayed. The larger window is the DOP main screen, which displays various components of the outpatient medical record, including adverse reactions, current medications, results, and active patient problems. The smaller window (upper right hand corner of Figure 1) is a term look-up function being used, in this case, to add a problem to the patient's problem list. Here, the user has selected "Problem List" from the main screen and then the function "Add Problem". When the vocabulary look-up screen (the "MED Viewer") appeared, the user then entered "travel" and three terms in the MED were retrieved. The user might then select a term, modify the selected term, or attempt another search.

As part of our evaluation of the system, we are naturally interested in knowing whether our coded data entry approach is successful in allowing clinicians to express their information properly. However, the final output of an interaction with the system will show only the presence or absence of terms. The presence of a term may not tell us if it is the desired term. The absence of terms may mean that there were no data to capture or that the data capture system failed. For example, if a user attempts to look up a term and is unsuccessful, the reason might be the user interface, the vocabulary content (term not present), or the vocabulary organization (term present but not in the place where the user was looking).

Teasing apart these reasons for failure is not straightforward. Traditional approaches to user interface evaluation are not likely to support a detailed analysis which would distinguish whether a look up problem was, for example, due to the user interface or the vocabulary organization. Such approaches, including subjective evaluations involving questionnaires and interviews, have a number of serious limitations including their imprecision and their reliance on subjects' recollection of their experience in using a system. Psychological research has shown that human recall is highly affected by long-term memory, i.e. retrospection may involve a loss of information and

0195—4210/96/$5.00 © 1996 AMIA, Inc.

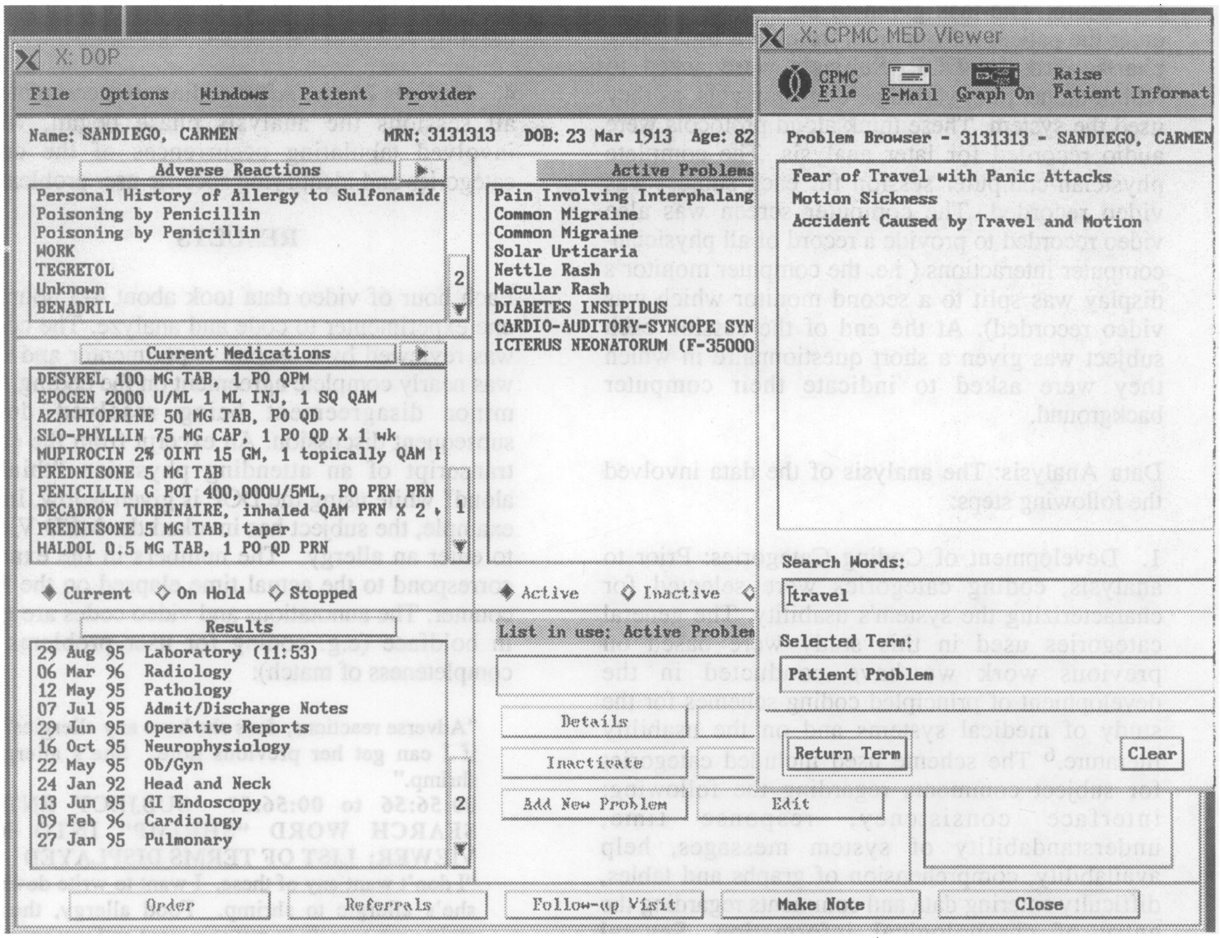

Figure 1 - DOP main screen, with MED Viewer

furthermore events are liable to be reconstructed inaccurately.[2] Thus, users may inform us of what they think they do in using a computer system, however, this can be considerably different from their actual behaviour.

In order to overcome limitations of conventional system evaluation, we have developed and refined a number of cognitive techniques for understanding human-computer interaction and the use of controlled vocabularies. One of the most useful methods is the collection of "think aloud" protocols, which allows us to focus on the process of human problem solving and reasoning. In the context of human-computer interaction, this involves the audio recording of computer users as they verbalize their thoughts while using a system to perform a task (e.g. Mack, Lewis, and Carroll's study of first time word processor users[3]). Systematic methods for analyzing verbal protocols have been developed in cognitive research, in order to provide detailed information about subjects' cognitive processing and problems.[4] More recently, work involving the use of video recording of computer users has begun to appear in the study of human-computer interaction.[5] In this paper, we describe the evaluation of the user interface and medical vocabulary of the DOP. The approach represents a novel integration of ideas from research in human-computer interaction (involving video recording) along with an analytic perspective from cognitive science. An important aspect of this evaluation is the determination of the source of problems in the user interface and physician difficulties in using the medical vocabulary. The long-term goal of this work is to feed information about user difficulties back into system design in a process of iterative system development.

METHODS

Subjects: Nine subjects participated in this study: 4 attending physicians and 5 residents. One of the subjects had briefly used the system being tested. None of the other subjects had any experience with the DOP system.

Procedure: The task given to all subjects was to enter the patient information, from their own paper charts, into the DOP. Subjects were asked to "think-aloud", i.e. verbalize their thoughts, as they used the system. These think aloud protocols were audio recorded for later analysis. The complete physician-computer session for each subject was video recorded. The computer screen was also video recorded to provide a record of all physician-computer interactions (i.e. the computer monitor's display was split to a second monitor which was video recorded). At the end of the session each subject was given a short questionnaire in which they were asked to indicate their computer background.

Data Analysis: The analysis of the data involved the following steps:

1. Development of Coding Categories: Prior to analysis, coding categories were selected for characterizing the system's usability. The general categories used in this study were based on previous work we have conducted in the development of principled coding schemes for the study of medical systems and on the usability literature.[6] The scheme used included categories for subject comments regarding the following: interface consistency, response time, understandability of system messages, help availability, comprehension of graphs and tables, difficulty entering data and comments regarding the entry of chronological information. Several additional categories, specific to the analysis of the "cognitive" interface to MED, were also derived, based on initial viewing of the videotapes. These include the characterization of MED vocabulary look-ups as (a) complete matches, (b) partial matches and (c) no matches.

2. Transcription: The audio tapes of the physicians' "thinking aloud" were transcribed verbatim.

3. Annotation Phase: This phase involved the coding for the video tapes for the presence of the categories described above. This involved the use of CVideo, a computer tool that facilitates coding by allowing the experimenter to link annotations in a computer text file to scenes on a videotape.[6] The annotation phase involved first transferring text files, containing the think aloud transcripts, to CVideo and then annotating the transcripts with the codes, i.e. "time-stamping" the coded sections of the transcripts to the corresponding video sequences. In this manner, each of the video tapes of the physician-computer interaction was coded, from start to finish, for the presence of both general usability and MED-specific categories.

4. Analysis Phase: After coding was complete for all sessions the analysis phase began, which involved tabulating occurrences of the coded categories and identifying specific user problems.

RESULTS

Each hour of video data took about 4-5 hours for one experimenter to code and analyze. The coding was reviewed by a second experimenter and there was nearly complete agreement on the coding, with minor disagreement being resolved during subsequent discussion. An excerpt from the coded transcript of an attending physician "thinking aloud" while using the DOP is given below. In this example, the subject has invoked the MED Viewer to enter an allergy. The numbers in the example correspond to the actual time elapsed on the VCR counter. The annotations and video codes are given in boldface (e.g. coding for user problems and completeness of match):

> "Adverse reactions, does she have any allergies? See if I can get her previous note. She's allergic to shrimp."
> **00:56:56 to 00:56:57 SUBJECT ENTERS SEARCH WORD "SHRIMP" INTO MED VIEWER; LIST OF TERMS DISPLAYED**
> "I don't want any of these. I want to write down that she's allergic to shrimp. Food allergy, that's it, makes me specify in my comment and my entry here will be shrimp"
> **00:57:16 to 00:57:17 SUBJECT SELECTS TERM "FOOD ALLERGY" AND TRIES TO ENTER COMMENT**
> "Oh, can't enter, try to enter again"
> **DATA ENTRY BLOCKED**
> "Alright, no big deal, it doesn't say which food allergy it is, I would like to see food allergy to shrimp, right up there"
> **PARTIAL MATCH**

As can be seen, in this example the physician has invoked the MED VIEWER to enter "shrimp allergy", followed by the selection of a term that partially matches what the physician wanted.

General Usability

The frequency of coded problems is given for all nine subjects in Table 1. As can be seen, the problem most frequently coded was that of data entry being blocked in certain conditions (with all 9 subjects having this problem). The next most frequent category dealt with comments by subjects regarding the consistency of the user interface (e.g. in some screens different types of data had to be entered in

Table 1 - Frequencies of all coded problems identified in the subjects' transcripts

Problems	Residents S1	S2	S3	S4	S5	Attendings S6	S7	S8	S9	Total Number	# of subjects with 1 or more occurrences
Data Entry Blocked	2	3	8	6	4	5	4	5	2	39	9
Comment on Consistency	0	1	1	1	0	2	8	0	0	13	5
Comment on Response Time	2	1	2	0	0	0	5	0	0	10	4
Comment on Overall Time	1	0	0	0	0	2	6	0	0	9	3
Understanding System Message	0	1	2	0	0	0	6	0	0	9	3
Comment on Operation Sequence	0	0	4	0	1	0	1	1	0	7	4
Help Accessed - not available	0	0	2	1	0	0	0	3	0	6	3
Comprehension Problem - Graph	1	1	1	0	0	1	0	0	0	4	4
Difficulty in Specifying Dates	2	0	0	0	0	0	1	0	1	4	3
Data not Displayed - Graph	0	1	1	0	0	1	0	0	0	3	3
System Crash - inadvertent exit	0	0	0	0	0	1	1	0	0	2	2
Comprehension Problem - Table	1	0	0	0	0	0	1	0	0	2	2
Data not Displayed - Table	0	0	1	0	0	1	0	0	0	2	2
Unable to Qualify Vital	0	0	1	0	0	1	0	0	0	2	2
Comment on Amount of Typing	0	0	0	0	0	1	0	0	0	1	1
Locating Program Section	0	0	0	0	0	0	1	0	0	1	1

different ways). A number of other categories had several coded occurrences, including comments regarding time, problems in understanding some system messages and difficulties in entering dates. It can also be seen that problems of the same general usability categories were present and coded for in the protocols from both groups of subjects, i.e. residents, and attending physicians.

Physician Interaction with the MED Vocabulary

Table 2 provides the results of all the searches by users to the Medical Entities Dictionary. Of the 19 cases entered, 78 attempts were made to enter controlled terms (see Table 2). A complete match was found in 48 look-ups (62%). In 11 look-ups (14%), a partial match was found. Video analysis showed the reasons for these matches to be: part of term returned (4), dosage not matched (3), too many terms returned - needed to restrict (2), and match but subject commented that he/she wanted a better term (2). In the remaining 19 look-ups (24%), no match was found. Video analysis showed the reasons to be: abbreviation not matched (5), <CR> not entered (4), inadvertent exit from VIEWER (3), term over specified (2), synonym not accepted (1), term misspelled (1), and other - term not found (3). In no cases did a match fail because the term was under specified by the user.

DISCUSSION

In this paper we have described a cognitive approach to the analysis of a user interface and medical vocabulary. The collection of video data, along with think aloud protocols allowed us to

Table 2 - Results of all look-ups using MED Viewer

	Residents S1	S2	S3	S4	S5	Attendings S6	S7	S8	S9	TOTAL
Number of Cases Entered	2	1	1	1	1	5	6	1	1	19
Total Time (minutes)	67	15	39	12	19	52	62	38	32	336
Total Number of Searches	23	0	6	2	10	22	8	4	3	78
Complete Match	11	0	2	2	6	16	5	4	2	48
Partial Match	4	0	0	0	0	3	3	0	1	11
No Match	8	0	4	0	4	3	0	0	0	19

examine the on-going cognitive processes of physicians as they used an information system. The overall approach combines aspects of the observational tradition, with rigorous and theory-based analyses that have emerged from the psychological, laboratory-based tradition.

The work we have described in this paper is part of a longer-term iterative cycle of interface evaluation and re-design. Based on the analyses, both short and long-term recommendations were made to the system design team. For example, it was recommended that the system indicate more clearly when user operations are applicable, e.g. when a field is editable. Generally, the analysis indicated that there is a need for greater consistency in a number of aspects of the interface, including data entry procedures and selection methods. Other recommendations dealt with ways of stream-lining and speeding up the entry of patient information. One suggestion, based on user comments, was to extend the data entry screen to include templates from which data could be selected. In addition, particular changes were identified in the display of graphs and tables, in order to increase their comprehensibility from a cognitive point of view. Based on these recommendations, the user interface has recently been re-programmed and the effect of resultant interface changes on physician-computer interaction is currently being evaluated. Our experience also indicates that the video analysis can be very effective in pinpointing the source of problems in using a controlled medical vocabulary. Recommendations based on these findings included the modification of the interface to provide greater consistency in the procedure for looking up a term. Longer term recommendations included the extension of the number of abbreviations, synonyms and dosages that are accepted by the system and mapped into equivalent terms that match MED terms. Thus the "terminological mapping" from user terms and abbreviations to terms that are recognized by the system could be improved and expanded. Empirical, cognitive studies could also be initiated to determine how physicians of various levels express medical problems, in order to provide a "cognitive basis" for increasing the "hit" rate of look-ups.

We have found that video data can provide much richer information for system design than retrospective reports and questionnaires.[7] In the past, analysis of video recordings was difficult due to the inherent richness and complexity of such data. However, the use of principled coding schemes, in conjunction with the application of computer-based tools for coding, have made the process of video analysis both feasible and cost effective. The approach taken can provide valuable information with very few subjects (all problem categories were identified in the protocols of the first subjects run). In addition, we have successfully applied the same general methodology for assessing a range of medical computer technologies, including physicians' use of a lap-top computerized patient record system[8] and a multimedia tutorial program.[6]

CONCLUSION

The cognitive evaluation described in this paper provided a useful data set for analyzing essential aspects of physician-computer interaction. The results of such analyses can inform the design and iterative refinement of medical systems.

Acknowledgments

This work was supported in part by a contract from the National Library of Medicine.

References

1.Cimino JJ., Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *JAMIA*, 1994:1(1):35-50.
2.Bartlett FC. Remembering: a study in experimental and social psychology. Cambridge, England: Cambridge University Press, 1932.
3.Mack RL, Lewis CH, Carroll JM. Learning to use word processors. *ACM Transactions on Office Systems*, 1983:1:254-271.
4.Ericsson KA, Simon, HA. Protocol analysis: verbal reports as data. Cambridge, MA: MIT Press, 1993.
5.Nielsen, J. Usability engineering. New York: Academic Press, 1993.
6.Kushniruk AW, Patel VL. Cognitive computer-based video analysis. In Greenes R, et al., eds. *Proc of the Eighth World Conference on Medical Informatics*, 1995:1566-1569.
7.Patel, VL, Kushniruk, AW. Human-computer interfacing in health care. To appear in van Bemmel JH, ed: Handbook of Medical Informatics.
8.Kushniruk AW, Kaufman DR, Patel VL, Levesque Y, Lottin P. Assessment of a computerized patient record system: a cognitive approach to evaluating an emerging medical technology. *MD Comp*, 1996:13(5).