VIDEO IN APPLIED COGNITIVE RESEARCH FOR HUMAN-CENTERED DESIGN

RENATE J. ROSKE-HOFSTRAND

INTRODUCTION

Video constitutes one of the most direct and contextually rich means for capturing observable behavior in human-computer interaction. However, a major shortfall of this type of visual medium lies in its failure to directly capture those aspects of user behavior that can predictably be associated with human cognition. Goal-directed cognitive user processes such as those that occur in problem-solving, decision-making, and planning activities are by their nature not subject to recording on camera. Traditionally, cognitive researchers have inferred user planning and intent structures by abstracting from relatively long sequences of behavioral cues captured by video protocols.

Luckily, the video medium is not limited to showing the researcher only that which is observable with the human eye. On the contrary, video technology offers us a view into "worlds" which are not directly visible. Supplements and aids to understanding can be superimposed and synchronized with particular "raw" events that are of interest.

SYNCHRONIZED EYE-MOVEMENT TRACE

One such supplement to the basic visual record of human-computer interaction is a synthetic trace of a user's eye-movement and fixation behavior as display features on a computer are scanned. This supplemented video results in a compellingly "readable" record of behavior which shares some of the defining features with good visual presentation screens of so-called interactive training systems, i.e., the use of highlighting techniques with cursor or arrow to point at immediate areas of interest and direct the viewer's attention to that portion of the visual display about which a message is being conveyed.

More importantly, the script for these dynamically highlighted events is authored naturally since it is derived from the eye-movement record of the user test subject. It is this natural series of visual (also auditory) material which presents the rich contextual data that lends itself to rigorous statistical analysis.

PARSING THE ENHANCED VIDEO RECORD

For task-analytical purposes, it is seldom desirable to have the enhanced video representation always conform precisely to its physical analog in terms of the actual passage of time or in terms of the exact chronological sequence of events. Video editing facilities can be used to create highly stylized, discrete summaries of the recorded full events. It is interesting to note that viewers can understand and recognize such stylized event sequences, even if it departs from their prototypical event order, presumably because such orders are embodied as flexible scripts or schemas in the observer's memory (Roske-Hofstrand, 1988).

Generally, the first task for the researcher/editor is to determine where the episode boundaries of the naturally scripted video record lie. This involves recognition of the explicit behavioral markers that signal such transitions. This analytics process differs from the creative authoring or design process used to communicate with an audience in important ways. In the domain of cinema for example, the creative application of transitional markers such as fades, blackouts, and wipes are commonly used to enhance syntactic structure and thus comprehension by the viewer. For the researcher interested in analyzing naturally occuring behavior though, we must attempt a more principled approach.

We may want to borrow techniques first developed in text comprehension. Several of these exist. First, we can collect verbal commentaries from users as they watch their own videotapes immediately after a test session has been recorded. Specific instructions given to the subjects may include (a) to indicate the boundaries between major event "episodes", and (b) to identify and comment on each major action that he or she performed.

It has been our experience that subjects are generally quite fascinated with the visual records of their own performance. As pointed out above, similar techniques have been used in studies of text comprehension where the researcher's goal was to verify underlying schema structures for stories (see Bower, Black, and Turner, 1979). Other techniques (see Graesser, 1981) have included prompting the subject with "why" and "how" questions to determine the goal-subgoal relationships between action elements in common procedural activities.

Another, complementary approach involves scaling the physical distances traversed by the eye-movement trace with respect to real time. Here the actions and episodes are defined based on the results of hierarchical clustering, PF-net, or multi-dimensional scaling analyses.

In any case, these empirically derived actions and episodes can be used as the basic event data for comparing individual users or user groups in human-computer interaction research involving baseline prototype computer systems.

AN APPLICATION IN AIR TRAFFIC CONTROL

One particular application domain for which enhanced video records are currently being analyzed at NASA is air traffic control (see Roske-Hofstrand, Armstrong, & Bergeron, 1988). Of particular interest for our research is the question of how the controller mentally organizes and strategically interacts with various 3-dimensional aircraft traffic patterns that are displayed on a 2-dimensional computer-enhanced radar screen. For example, early analysis of the eye-movement video data obtained from expert controller subjects indicates that controllers tend to organize aircraft not in terms of their actual physical proximity in the airspace, as might be expected since they are tasked with the responsibility to maintain minimum separation. Rather, controllers tend to organize traffic in terms of the aircraft's performance characteristics within particular site-specific, flexible flow corridors. This type of information is invaluable for the design of future interfaces for air traffic controllers. A sample task protocol derived from an eye-movement record is shown in figure 1.

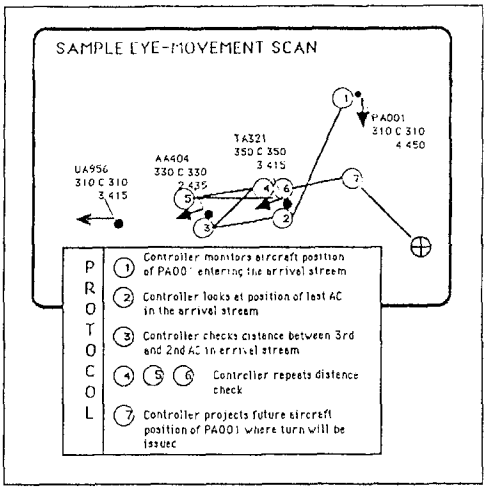

Figure 1.: Sample protocol derived from scan pattern of the enhanced video record. (Arrows indicate aircraft vectors.)

Video records enhanced with eye-movement traces give us access to information not previously accessible since they serve both as stimulus cues for subjects asked to explain their behavior and their work tasks and as complementary, historical databases for naturally occuring cognitive behavior.

EVENT BLOCKS AS A DESIGN TOOL

The judged success of interface design is often subject to the constraints imposed between the user and the computer by the dialogue structure. Dialogue design which includes specification of input modes and interaction features must be matched to the natural sequence of the cognitive task that the user intends to perform in a given application domain. Video records, composed of the empirically parsed event blocks, give the designer

of future systems the unique ability to create alternative dialogue sequences which can be viewed and changed at will without having to consider specific hardware and software implementation concerns. This constitutes real human-centered design! A library of previously recorded event blocks can be used to create task oriented behavioral sequences. This allows for some of the most important design criteria to be made visually explicit and for the more "obvious" problems with particular design decisions to be identified and corrected.

References

Bower, G.H., Black, J.B., & Turner, T.J. (1979). Scripts in memory for text. Cognitive Psychology, 11, 177-220.

Graesser, A.C. (1981) Prose comprehension beyond the word. New York: Springer.

Roske-Hofstrand, R.J. (1988) Discourse understanding of displayed actions: an investigation of the temporal relations of event schemas and their effect on reconstruction. Dissertation: New Mexico State University; Las Cruces, New Mexico; April.

Roske-Hofstrand, R.J., Armstrong, H.B., & Bergeron, H. (1988) Applied cognitive research in air traffic control: a concept paper and research plan. NASA Internal Memorandum, November.