Mapping Instructions Onto Actions: A Comprehension-Based Model of Display-Based Human–Computer Interaction

Muneo Kitajima[a] and Peter G. Polson[b]

[a]National Institute of Bioscience and Human–Technology, 1-1 Higashi Tsukuba Ibaraki 305, Japan

[b]Institute of Cognitive Science, University of Colorado, Boulder, CO 80309-0345, USA

1. COMPREHENSION-BASED MODELS

This paper describes a cognitive model, LICAI, **LI**nked model of **C**omprehension-based **A**ction planing and **I**nstruction taking, that simulates the cognitive processes involved in comprehending and following hints, and successfully performing steps by exploration [5][6]. The cognitive processes specified in LICAI are implemented using the *construction–integration (CI) architecture* developed by Kintsch [1]. The CI architecture is symbolic-connectionist and has been applied successfully to model cognitive processes such as text comprehension, word problem solving, and action planning. In the *construction* phase, a CI model generates a connectionist network that includes alternative meanings of the current text or alternative actions that can be performed on the interface display. In the *integration* phase, a CI model spreads activation among the constructed networks and selects a contextually appropriate alternative. Links in the network are established by common symbols; when two nodes share symbols, they are connected.

The CI architecture has evolved from a detail model of skilled, text comprehension — a highly automated collection of cognitive processes that make use of massive amounts of knowledge stored in long-term memory [1][2]. This is a very different foundation from the other cognitive architectures. For example, one of the primary foundations of Soar [7] was the General Problem Solver, a model of deliberate cognition (i.e., problem solving and action planning) in situations where the problem solver has limited background knowledge.

2. THE LICAI MODEL

LICAI, shown in Figure 1, simulates the processes involved in reading instructions and performing a task on the interface by exploration with three different processes. These processes are expressed by various CI cycles.

a) E-mail: kitajima@nibh.go.jp; related papers available from WWW: http://www.aist.go.jp/NIBH/~b0544/

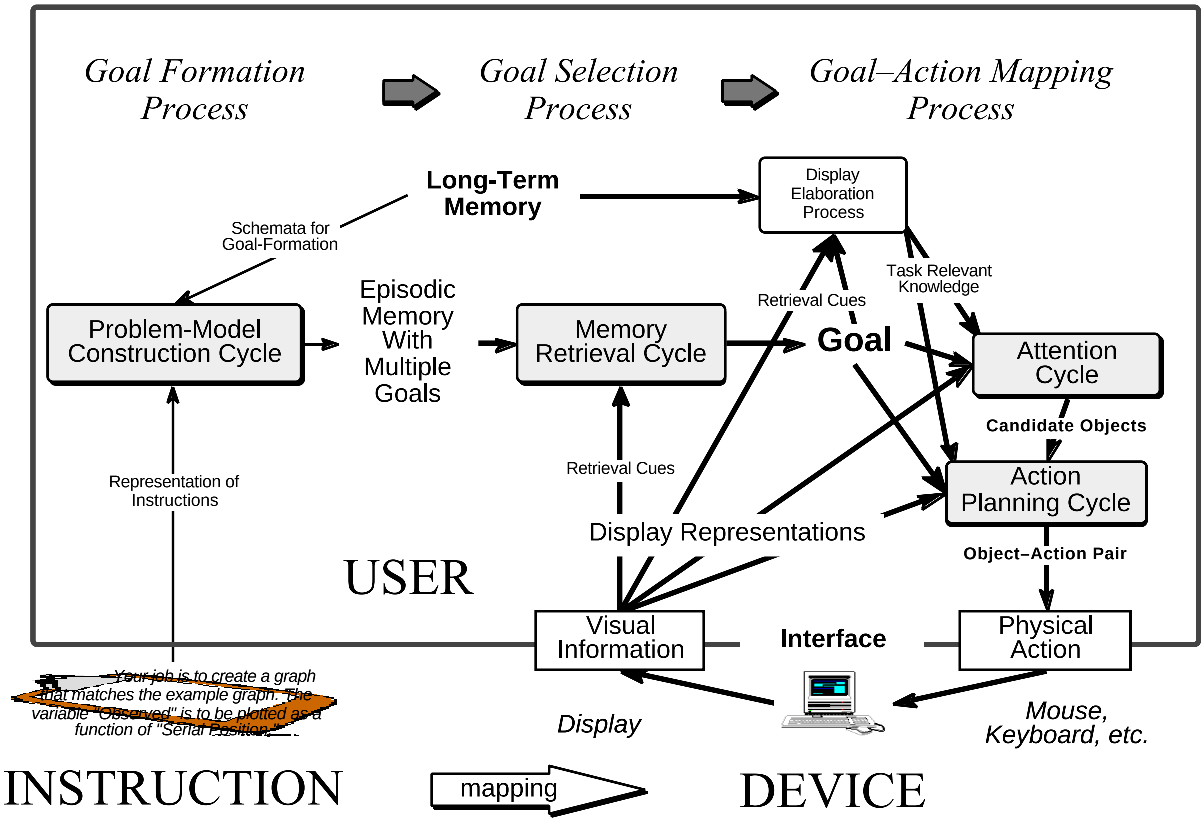

Figure 1. An illustration of LICAI [6], adapted from [3].

2.1. Goal Formation Process

When reading instructions, a user attempts to extract goals that should be accomplished on an interface. LICAI assumes that this *goal formation process* is analogous to solving word problems, and models it as a problem-model construction CI cycle, which is a strategic form of the basic text-comprehension process that generates representations specialized for interacting with devices; that is, the goals that control solution of a task described in instructions. Instructions are processed by executing a single CI cycle for each sentence. In the construction phase, LICAI generates a network that includes semantic representations of a sentence as well as elaborations that translate the semantic representation into goals. In the integration phase, LICAI selects a single meaning for the sentence and links this representation with the memory representation of earlier parts of the text. Thus, after reading the entire text, the memory contents represent the result of instruction comprehension. If the text contains descriptions of multiple goals, LICAI stores them in episodic memory.

2.2. Goal Selection Process

After reading instructions, the user tries to select a goal from the episodic memory. LICAI assumes that this *goal selection process* is done by a single CI cycle using the current application display as retrieval cues. In the construction phase, the episodic memory and the current interface display constitute a network. In the integration phase, a goal consistent with the current interface display is selected; that is, the most highly activated goal is selected. Since the sources of activation are the nodes representing the display and the pattern of links in the network is largely determined by the argument overlap mechanism, a goal that overlaps the currently visible screen objects is likely to be selected. For example, if the representation of the goal includes matching labels on any screen objects, it will be selected.

2.3. Goal–Action Mapping Process

After selecting a goal, the user tries to generate a sequence of one or more actions that will accomplish the selected goal. This *goal–action mapping process* involves two CI cycles, *the attention cycle* and *the action-planning cycle*, which is a generalization of the model of skilled, display-based, action planning developed by Kitajima and Polson [4].

2.3.1. Display Elaboration Process

The initial display representations contain only limited information about the identity of each screen object and its appearance, including visual attributes (e.g., color, highlighting). The poor display representations are augmented by retrieving relevant knowledge from long-term memory. This display elaboration process is simulated by a random memory sampling process: the retrieval cues are the current goal and the propositions representing the current display. The elaboration process is stochastic. The probability that each cue retrieves particular information in a single memory retrieval process is proportional to the strength of the link between them. LICAI carries out multiple memory retrieval in a single elaboration process. A parameter controls the number of times each argument in the display and goal representations is used as retrieval cues. The predictions and implications that follow from this stochastic elaboration process is discussed in detail by Kitajima and Polson [4].

2.3.2. Attention Cycle

The elaborated display representation is the model's evaluation of the current display in the context defined by the current goal. In the goal–action mapping process, the model first limits its attention to three screen objects out of ~100 objects displayed on the screen by applying an attention CI cycle. These screen objects are candidates for the next action to be operated upon. During the construction phase, a network is generated that consists of nodes representing the goals, the screen objects and their elaborations, and candidate object nodes of the form *Screen-Object-X is-attended*. Any screen objects are potential candidates. During the integration phase, the conflict is to be resolved. The sources of activation are the goals and the screen objects. The targets are the candidate object nodes. When the spreading activation process stabilizes, the model selects the three most highly activated candidate object nodes. These nodes represent screen objects to be attended to during the next action-planning cycle.

2.3.3. Action-Planning Cycle

The second CI cycle is an action-planning cycle. As preparation for constructing a network, the candidate objects carried over from the preceding cycle are combined with any possible actions to form object–action pairs of alternatives. The model considers all possible actions on each candidate object. Examples would include *Single-click Object23*, *Double-click* Object23, *Move Object23*, *Release Object23*, and the like, where *Object23* represents one of the candidates.

During the construction phase, the model generates a network that includes the goals, the screen objects and their elaborations, and representations of all possible actions on each candidate object. During the integration phase, the sources of activation are the goals and the screen objects, and the targets are the nodes representing the combinations of object–actions. At the end of the integration phase, the model selects the most highly activated object–action pair whose preconditions are satisfied as the next action to be executed. The action representations include conditions to be satisfied for their execution. The conditions are matched against the elaborated display representations. Some conditions are satisfied by the current screen, others by information that was retrieved from long-term memory in the display elaboration process. For example, the model cannot select an action *double-click a document icon* unless the icon is currently pointed at by the mouse cursor, and the information "the icon can be double clicked" is available. Observe that if information about a necessary condition is missing from an elaborated display representation, the model *cannot* perform that action on the *incorrectly* described object.

3. NATURE OF INSTRUCTION–ACTION MAPPINGS

LICAI describes the underlying mechanism that controls users' instruction mapping onto interface actions. A series of CI cycles depicts the various component processes executed. Mapping from instructions to an action is successful if the goal formation processes generate the correct goal, if the goal selection processes select that goal, and if the goal–action mapping processes generate the correct action sequence. LICAI predicts that the instruction–action mapping processes will become more difficult with longer instructions, more screen objects, and/or an increase in possible actions. LICAI suggests that interacting with a novel application by following instructions will become difficult unless the instructions are worded carefully and the interface is designed to facilitate exploration (see [3] for more detailed discussion).

REFERENCES

1. Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review*, **95**, 163-182.
2. Kintsch, W. (in press). *Comprehension: A paradigm for cognition*. Cambridge University Press.
3. Kitajima, M. (in press). Successful technology must enable people to utilize existing cognitive skills. In *Cognitive Technology: Methods and Practice*. Elsevier.
4. Kitajima, M., & Polson, P.G. (1995). A comprehension-based model of correct performance and errors in skilled, display-based human–computer interaction. *International Journal of Human–Computer Systems*, **43**, 65-99.
5. Kitajima, M., & Polson, P.G. (1996). A comprehension-based model of exploration. In *Proceedings of CHI'96 Conference on Human Factors in Computing Systems,* 324-331.
6. Kitajima, M., & Polson, P.G. (in press). A comprehension-based model of exploration. *Human–Computer Interaction.*
7. Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.