COMET: Generating Coordinated Multimedia Explanations

Steven K. Feiner
Kathleen R. McKeown

Department of Computer Science
Columbia University
New York, New York 10027

INTRODUCTION

COMET (COordinated Multimedia Explanation Testbed) is an experimental system that generates interactive multimedia instructions for equipment maintenance and repair [3, 4]. COMET combines research in natural language generation and in knowledge-based graphics. The form and content of all text and graphics is created interactively. Thus, instructions can be customized on the fly for the individual user and situation.

We are using a radio modeled after the Army AN PRC/119 communications radio as our experimental domain. In our work, we have concentrated on coordinating multiple media. This makes it possible for the text and graphics that COMET generates to work together to communicate information effectively.

INTERFACE

Interaction with COMET is initiated using a menu interface. A user may request an explanation for a specific repair procedure or may specify that troubleshooting help is needed. When help is requested, an underlying diagnostic system is invoked that asks the user to specify failures using the menu. In the course of diagnosing a failure, COMET asks the user to carry out troubleshooting procedures. Each step of a procedure is shown sequentially on the display. The user can request an explanation of any step, or can move forward or backward in the generated explanation, by using the menu interface. Figure 1 shows one step of a multiple-step sequence, generated by COMET, that explains how to install a new battery in the radio. All of the text and graphics in the figure were created by the system in response to the user's query.

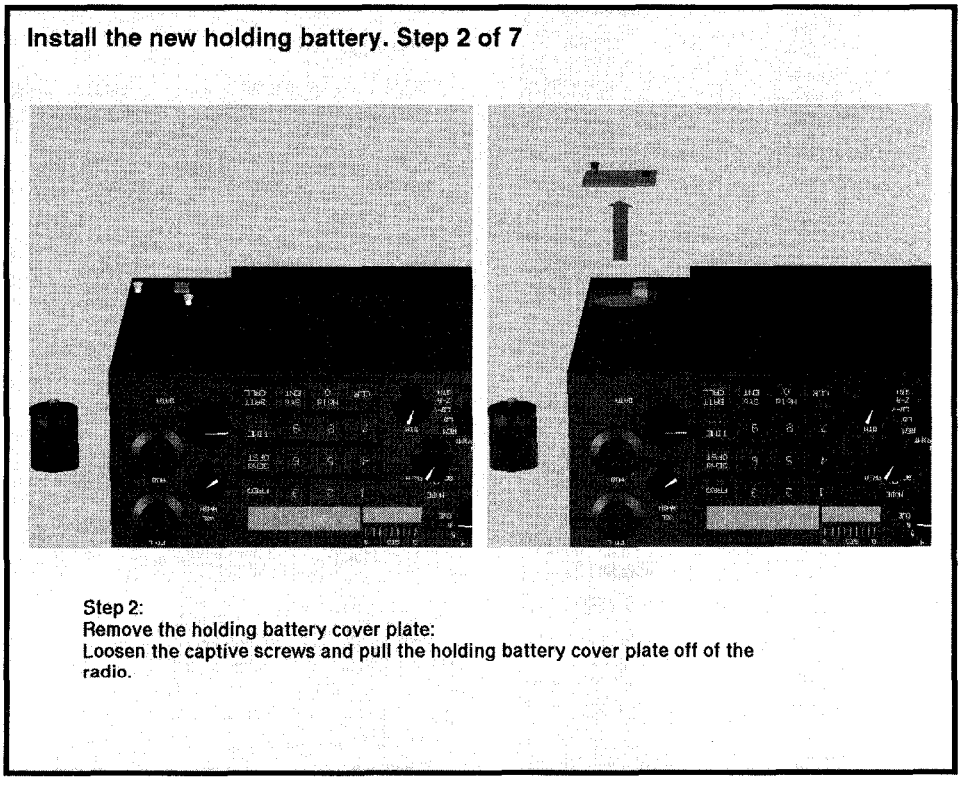

Figure 1: A display generated by COMET.

ARCHITECTURE AND IMPLEMENTATION

COMET's architecture includes an expert-system diagnostic component, several static knowledge bases, a single content planner, a media coordinator, media generators (currently text and graphics), a media layout manager, and interactive typesetting and rendering components. On receiving a request for an explanation, the *content planner* uses text plans, or *schemas* [7], to determine which information from the underlying *knowledge sources* should be included in the explanation. The content planner produces the full content for the explanation, which is represented as a hierarchy of *logical forms* [1]. Each logical form specifies *what* information is to be communicated, but doesn't specify *how* the information is to be communicated. The logical forms are passed to the *media coordinator*. The media coordinator determines which pieces of information are to be communicated in text, and which in graphics. It annotates the logical forms with directives that specify these assignments.

Each annotated logical form is passed to a set of *media generators*. COMET currently has a *text generator* and a *graphics generator*. Each generator creates material in its own medium to express information that the media coordinator has assigned to it. The resulting text and graphics are passed to the *media layout* component, which formats the final presentation, determining where on the display information is to be presented. Real-time rendering and typesetting components then produce the finished display.

Each of COMET's major components runs in parallel in its own process on five networked Sun and HP workstations. COMET has hierarchical models of the radio and its components, and of a number of actions that can be performed in maintaining and repairing the radio. These models are encoded using the LOOM [6] knowledge representation system. Text and menus are displayed using the X Window System, while 3D shaded graphics are displayed using HP's Starbase graphics package. Figure 1 takes 15–20 seconds to generate and display, starting with the initial user request.

RESEARCH FOCUS

Our research has focused on the form of the explanations generated, including the coordination of text and graphics to form effective explanations, as well as the choice of words, sentence structure, picture structure, and graphical style.

Media coordination. By using a single representation of the explanation's content that is provided to all our media generators, we make it possible for each generator to take into account what information is being expressed by the other. COMET's media coordinator performs a fine-grained mapping of information to one or more media, based on the kind of information to be expressed [4]. For example, physical attributes such as size and shape are shown using graphics, whereas abstract properties and relationships such as causality are expressed in text.

Text generation. Our focus in text generation has been on the selection of appropriate vocabulary for the explanation. We have developed a framework for lexical choice using the Functional Unification Formalism (FUF) [5, 2]. As part of this framework, we have identified how previous discourse and the underlying knowledge sources influence lexical choice and have implemented these influences in COMET's lexical chooser. For example, the lexical chooser will use the verb "reinstall" or "return" instead of "install" when instructing the user to install an object that it has previously asked the user to remove.

Graphics generation. Work on graphics generation in COMET has concentrated on the development of an approach for generating technical illustrations of 3D objects, embodied in the rule-based graphics generator IBIS (Intent-Based Illustration System) [8]. Each of IBIS's illustrations is created by an *illustrator*, which designs its illustration to fulfill a set of communicative goals derived from the logical form that is presented to it. The illustrator realizes these goals by creating an *illustration* that includes a set of objects to be depicted and their attributes, a lighting specification that indicates how objects are lit, and a viewing specification that indicates how the 3D objects are to be projected onto the 2D display.

ACKNOWLEDGEMENTS

COMET is supported in part by Defense Advanced Research Projects Agency Contract N00039-84-C-0165, the Hewlett-Packard Company AI University Grants Program, National Science Foundation Grant IRT-84-51438, New York State Center for Advanced Technology Contract NYSSTF-CAT(88)-5, and Office of Naval Research Contract N00014-82-K-0256. COMET is a group effort that has benefited from the research contributions of Michael Elhadad, Dorée Seligmann, Cliff Beshers, Andrea Danyluk, Yumiko Fukumoto, Jong Lim, Christine Lombardi, Jacques Robin, Franck Smadja, Michelle Baker, David Fox, Laura Gabbe, and Tony Weida.

REFERENCES

1. Allen, J. *Natural Language Understanding.* Benjamin Cummings Publishing Company, Inc., Menlo Park, CA, 1987.

2. Elhadad, M. Types in Functional Unification Grammars. *Proc. 28th Meeting of the ACL*, Pittsburgh, PA, June 1990.

3. Feiner, S.K. and McKeown, K.R. Generating Coordinated Multimedia Explanations. *Proc IEEE Conf. on AI Applications*, Santa Barbara, CA, March 5–9, 1990, 290–296.

4. Feiner, S.K. and McKeown, K.R. Coordinating Text and Graphics in Explanation Generation. *Proc. AAAI-90*, Boston, MA, July 29–August 3, 1990, 442–49.

5. Kay, M. Functional Grammar. *Proc. 5th Meeting of the Berkeley Linguistics Soc.*, Berkeley Linguistics Soc., 1979.

6. MacGregor, R. and Brill, D. *LOOM Reference Manual.* Tech. Rep., USC-ISI, Marina del Rey, CA, 1989.

7. McKeown, K.R. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text.* Cambridge University Press, Cambridge, England, 1985.

8. Seligmann, D.D. and Feiner, S.K. Specifying Composite Illustrations with Communicative Goals. *Proc. ACM Symp. on User Interface Software and Technology*, Williamsburg, VA, November 13–15, 1989, 1–9.