Intelligent User Interfaces

Edited by
Joseph W. Sullivan
Sherman W. Tyler
Artificial Intelligence Center
Research and Development Division
Lockheed Missiles and Space Company, Inc.

ACM Press
New York, New York

Addison-Wesley Publishing Company

Reading, Massachusetts • Menlo Park, California • New York
Don Mills, Ontario • Wokingham, England • Amsterdam • Bonn
Sydney • Singapore • Tokyo • Madrid • San Juan • Milan • Paris

ACM Press Frontier Series

The programs and applications presented in this book have been included for their instructional value. They have been tested with care, but are not guaranteed for any particular purpose. The publisher does not offer any warranties or representations nor does it accept any liabilities with respect to the programs or applications.

Library of Congress Cataloging-in-Publication Data

Intelligent user interfaces / [edited by] Joseph W.
Sullivan and Sherman W. Tyler.
p. cm.—(ACM Press frontier series)
Papers from a workshop held Mar. 1988 in Monterey, Calif.
Includes bibliographical references and index.
ISBN 0-201-50305-0
1. Human-computer interaction—Congresses. 2. Artificial
intelligence—Congresses. I. Sullivan, Joseph W. (Joseph William).
1947– . II. Tyler, Sherman W. III. Series.
QA76.9.H85A73 1991
004'.01'9—dc20 90-48188
CIP

1 2 3 4 5 6 7 8 9 10-MA-9594939291

CHAPTER 2

INTELLIGENT MULTI-MEDIA INTERFACE TECHNOLOGY

JEANNETTE G. NEAL *and* **STUART C. SHAPIRO**
Calspan-UB Research Center (CUBRC)

ABSTRACT

The Intelligent Multi-Media Interfaces project is devoted to the application of artificial intelligence methodology to the development of human-computer interface technology that will integrate speech input, speech output, natural-language text, graphics, and pointing gestures for interactive dialogues between human and computer. These dialogues are modeled on the manner in which two people naturally communicate in coordinated multiple modalities when working at a graphics device. The objective is to simplify operator communication with sophisticated computer systems. As part of this project, a knowledge-based interface system, called CUBRICON (the CUBRC Intelligent CONversationalist), is being developed as a proof-of-concept prototype.

This research is supported by the Defense Advanced Research Projects Agency and monitored by the Rome Air Development Center under contract no. F30603-87-C-0136.

CUBRICON possesses the following critical functionality.

1. CUBRICON accepts and understands multi-media input such that references to entities in a natural-language sentence can be accompanied by coordinated simultaneous pointing to the respective entities on a graphics display.
   a. It is able to use a simultaneous pointing reference and natural-language reference to disambiguate one another when appropriate.
   b. It infers the intended referent of a point gesture that is inconsistent with the accompanying natural-language.
2. CUBRICON automatically composes and generates relevant output to the user in coordinated multi-media.
   a. It automatically selects appropriate output media/modalities for expressing information to the user, with the selection based on the nature of the information, discourse context, and the importance of the information to the user's task.
   b. It uses its media/modalities in a highly integrated manner.
   c. It judges the relevance of information with respect to the discourse context and user task and responds in a context-sensitive manner.
   d. It adheres to respected human factor guidelines for human-computer interaction and information presentation, including:
      i. maintaining the context of the user/computer dialogue,
      ii. maintaining consistency throughout a display, and
      iii. maintaining consistency across displays.

2.1 INTRODUCTION

The introduction of improved and advanced processing capabilities into Air Force Command and Control (C2) systems is proceeding at an ever-increasing rate. It is essential that the human-machine interfaces resulting from these developments not be limiting factors that degrade the performance of the system. Many times this issue is either overlooked or handled much like a retrofit after the fact. Much of the general guidance for the introduction of decision support capabilities has been based on the simplified statement that the overall goal is to allocate information processing and decision functions between human and machine in a way that optimizes the use of their respective strengths and compensates for their respective weaknesses. Many aspects of C2 operations are mandated to be human decision processes (e.g., nuclear enablement, fire control, target designation). In many cases, the processes involved and the mechanization of the interface have resulted in performance that is far less than optimal. Considerable R&D has been initiated to support elements of this critical human-machine interface, but attention has only recently been focused on the development of intelligent multi-media interface technology.

This paper discusses an intelligent multi-media interface system called CUBRICON (the CUBRC Intelligent CONversationalist), being developed as part of the Intelligent Multi-Media Interfaces (IMMI) project. The IMMI project is devoted to the application of artificial intelligence methodology to the development of human-computer interface technology that integrates speech input, speech output, natural-language text, graphics, and pointing gestures for interactive dialogues between human and computer. These dialogues are modeled on the manner in which two people naturally communicate in coordinated multiple modalities when working at a graphics device. The objective is to simplify operator communication with sophisticated computer systems. The interface system accepts coordinated graphic/verbal input and composes multi-media output to best convey information to the user.

The critical functionality of CUBRICON is itemized in the next section. Section 2.3 presents related research. Section 2.4 presents the main components of the software system, the I/O devices, and the development environment. Section 2.5 discusses the CUBRICON knowledge sources. Section 2.6 discusses the multi-media language understanding process, and Section 2.7 the multi-media output composition process. An example of a dialogue is presented and discussed in Section 2.8. Section 2.9 presents some future directions for this research, and Section 2.10 summarizes the paper.

2.2 SYSTEM DESIGN

The CUBRICON system design is based on a *unified view of language*. Language is a means of communication, whether verbal, visual, tactile, or gestural. Human beings communicate with each other via written and spoken natural language, frequently supplemented by pictures, diagrams, pointing to objects, and other gestures. The CUBRICON system design provides for the use of a unified multi-media language, defined by an integrated grammar, consisting of textual, graphic, and combined text/graphic symbols. Input and output streams are treated as compound streams, with components corresponding to different media. This approach is intended to imitate, to a certain extent, the ability of humans to simultaneously accept input from different sensory devices (such as eyes and ears), and to simultaneously produce output in different media (such as voice, pointing motions, and drawings). The CUBRICON system includes (1) language parsing and generation to accommodate the compound streams, (2) knowledge representation and inference to provide reasoning ability, (3) knowledge bases and models to provide a basis for its decision-making ability, and (4) automated knowledge-based medium selection and formulation of responses.

The system's critical functionality is as follows. CUBRICON:

- Accepts and understands multi-media input such that references to entities in a natural-language sentence can be accompanied by coordinated simultaneous pointing to the respective entities on a graphics display
  - Is able to use a simultaneous pointing reference and natural-language reference to disambiguate one another when appropriate
  - Infers the intended referent of a point gesture that is inconsistent with the accompanying natural-language
- Automatically composes and generates relevant output to the user in coordinated multi-media
  - Automatically selects appropriate output media/modalities for expressing information to the user, with the selection based on the nature of the information, discourse context, and the importance of the information to the user's task
  - Uses its media/modalities in a highly integrated manner, including simulated parallelism
  - Judges the relevance of information with respect to the discourse context and user task and responds in a context-sensitive manner
  - Adheres to respected human factor guidelines for human-computer interaction and information presentation, including (a) maintaining the context of the user/computer dialogue, (b) maintaining consistency throughout a display, and (c) maintaining consistency across displays

2.3 RELATED RESEARCH

Research and development of artificial intelligence (AI) systems for human-machine interfaces have focused on natural language (NL) text, speech, and graphics primarily in isolation, rather than in integrated interfaces. With the increased functionality and reliability provided by new developments in interface devices such as speech recognition and production systems, high-resolution color and monochrome graphic displays, and pointing devices, as well as the availability of increasingly powerful workstation environments, it is a natural and timely step in the evolution of human-computer interfaces that the media be integrated to meet the information processing needs of the user community.

Computer-based multi-media communication between people has received support via the development of multi-media electronic document systems and mail systems. This research and development includes the electronic document systems at Brown University [Feiner82], an experimental multi-media mail system at ISI [Katz84], a multi-media message system at SRI [Aceves84], and the Diamond message system at BBN [Thomas85]. Hypertext [Conklin87; ACM88; Egan89; Shepard89] provides for multimodal, multidimensional document representation and access.

Intelligent interactive human-computer dialogue via multi-media language (e.g., simultaneous natural-language and graphics) has just begun to be developed. Work has begun on intelligence in interfaces [Neches86; Sullivan91] and, in particular, on the issue of the intelligent use of multiple media and/or modalities for human-computer communication [Arens88; Arens91; Hollan91; Kobsa86; Neal88a; Neal88b; Neal89; Reithinger87; Roth91].

For input, the CUBRICON project focuses on the understanding of natural language accompanied by simultaneous coordinated pointing gestures, particularly the problem of referent identification. Related work includes the development of the TEMPLAR system [Press86] at TRW and XTRA [Allgayer89; Kobsa86] at the University of Saarbrucken. The TEMPLAR system seems to provide only for a pointing gesture to *substitute* for a NL definite reference within a NL sentence, rather than allowing a pointing gesture also to be used *in combination with* a NL reference. In the TEMPLAR system, the NL phrase for the object selected by the point is inserted in the input string to allow the NL parser to complete its processing. Our approach is closer to that of Kobsa and colleagues with the XTRA system, in accepting dual-modality input and applying several knowledge sources for referent identification. Our systems differ in the types of objects that can be targets of pointing gestures, as well as in the knowledge sources that are used.

For output, the CUBRICON project is addressing the problem of having the system select the media/modalities for expressing information to the user, as well as composing the output in the selected media/modalities. Related work includes that of Reithinger [Reithinger87] in generating referring expressions and pointing gestures. The Integrated Interfaces project [Arens88; Arens91] uses a variety of output modalities but does not include speech production or deictic pointing gestures during output. The CUBRICON project and the work of Roth et al. [Roth91] are both concerned with the problems of selecting relevant information to present to the user, composing text and selecting and designing pictures to convey the information, and coordinating the two different modalities. Feiner and McKeown [Feiner89] are also working on the problem of coordinating text and graphics, both displayed on a computer screen, in the generation of explanations to accomplish communicative goals. The CUBRICON project is concerned with the appropriate selection, generation, and coordination of a wider variety of output modalities (e.g., speech, tables, and form output in addition to printed text and graphics).

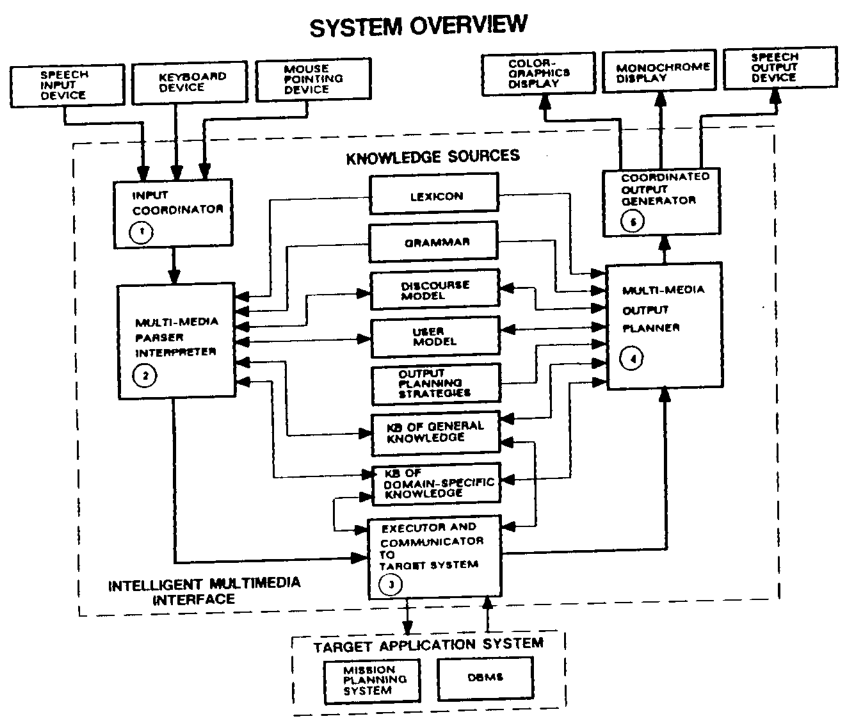
The CUBRICON system includes several knowledge sources (e.g., application-specific knowledge base, discourse model, user model) to generate relevant helpful responses, maintain the discourse context when appropriate, manage its display resources, provide the user and system with the ability to reference the display objects, and use the modalities in coordinated combinations for output generation. Cheikes and Webber [Cheikes88] and Kaplan [Kaplan82] address the problem of generating relevant cooperative responses. The issue of models to support intelligent behavior of interface systems is also addressed by Wahlster [Wahlster91], Mason and Edwards [Mason88], Kass and Finin [Kass91], and Young [Young91].

2.4 SYSTEM OVERVIEW

The CUBRICON team has designed and implemented an integrated user interface system with the functionality described briefly in Section 2.2. Figure 2.1 provides an overview of the software system and hardware I/O devices.

CUBRICON accepts input from three input devices: speech input device, keyboard, and mouse device pointing to objects on a graphics display. CUBRICON produces output for three output devices: high-resolution color graphics display, monochrome display, and speech output device. The primary path that the input data follow is indicated by the numbered modules

FIGURE 2.1
SYSTEM OVERVIEW



in the figure: (1) Input Coordinator, (2) Multi-media Parser Interpreter, (3) Executor/Communicator to Target System, (4) Multi-media Output Planner, and (5) Coordinated Output Generator. The Input Coordinator module accepts input from the three input devices and fuses the input streams into a single compound stream, maintaining the temporal order of tokens in the original streams. The Multi-media Parser/Interpreter is an augmented transition network (ATN) that has been extended to accept the compound stream produced by the Input Coordinator and to produce an interpretation of this compound stream. Appropriate action is then taken by the Executor module. This may be a command to the mission planning system, a database query, or an action that entails participation of the interface system only. An expression of the results of the action is then planned by the Multi-media Output Planner for communication to the user. The Output Planner is a generalized ATN that produces a multi-media output stream representation with components targeted for different devices (e.g., color graphics display, speech device, monochrome display). This output representation is translated into visual/auditory output by the Coordinated Output Generator module. This module is responsible for producing the multi-media output in a coordinated manner in real time (e.g., the Output Planner module can specify that a certain icon on the color graphics display must be highlighted when the entity represented by the icon is mentioned in the simultaneous NL output).

The CUBRICON system includes several knowledge sources to be used during processing, including:

1. A lexicon
2. A grammar defining the language used by the system for multi-media input and output
3. A discourse model
4. A user model
5. A knowledge base of output planning strategies to govern the composition of multi-media responses to the user
6. A knowledge base of information about generally shared world knowledge
7. A knowledge base of information about the specific task domain of tactical air control

These knowledge sources are used for both understanding input to the system and planning/generating output from the system. They are discussed in more detail in the next section.

The CUBRICON system is implemented on a Symbolics Lisp Machine with a mouse pointing device, a color graphics monitor, and a monochrome monitor. Speech recognition is handled by a Dragon Systems VoiceScribe 1000. Speech output is produced by a DECtalk speech production system. CUBRICON software is implemented using the SNePS semantic network processing system [Shapiro79a; Shapiro81; Shapiro87], an ATN parser/generator [Shapiro82a], and Common Lisp. SNePS is a fully intensional propositional semantic network and has been used for a variety of purposes and applications [Maida85; Shapiro82c; Shapiro87; Neal86; Neal87]. SNePS provides:

1. A flexible knowledge representation facility in the semantic network formalism
2. Representation of rules in the network in a declarative form so they can be reasoned about like any other data
3. A bidirectional inference subsystem [Shapiro82b] that focuses attention toward the active processes and cuts down the fanout of pure forward or backward chaining
4. A simulated multiprocessing control structure [McKay80]
5. Special nonstandard connectives [Shapiro79b] to model human reasoning processes
6. Existential, universal, and numerical quantifiers [Shapiro79c]

2.5 KNOWLEDGE SOURCES FOR MULTI-MEDIA LANGUAGE PROCESSING

2.5.1. The Lexicon and Grammar

A lexicon is the collection of all morphemes, tokens, and signals that carry meaning in a given language. The CUBRICON system's lexicon consists of words, graphic figures, and pointing signals. The grammar defines how the morphemes, tokens, and signals of the lexicon can combine to form legal composite language structures. An example of a multimodal language structure that is legal according to the CUBRICON grammar is a noun phrase. A noun phrase consists of the typical linguistic syntax (e.g., determiner, followed by zero or more modifiers, followed by a noun) accompanied by zero or more pointing signals (pointing to objects on the graphics display). The lexicon and grammar together define the multimodal language used by the system.

2.5.2. The Discourse Model

Continuity and relevance are key factors in discourse. Without these factors, people find discourse disconcerting and unnatural. The attentional discourse focus space representation [Grosz78; Grosz85; Grosz86; Sidner86] is a key knowledge structure that supports continuity and relevance in dialogue. CUBRICON tracks the attentional discourse focus space of the dialogue carried out in multi-media language and maintains a representation of the focus space in two structures: a main focus list and a display model.

The main focus list includes those entities and propositions that have been explicitly expressed (by the user or by CUBRICON) via natural language, pointing, highlighting, or blinking. The display model represents all the objects that are "in focus" because they are visible on one of the monitors. CUBRICON is based on the premise that graphics are an integral part of its language, along with natural language and other forms of text and pointing. The CUBRICON system treats objects presented on the graphics displays as having been intentionally "expressed" or "mentioned." All objects on the graphics display are "in focus," and CUBRICON maintains a representation of all these objects in the form of a display model. The display model consists of two levels: (1) a list of windows per monitor and, (2) for each window, a list of all the objects visible in the window. This display model is used in a manner analogous to the use of the main dialogue focus list.

The dialogue attentional focus space representation is used for determining the interpretation of anaphoric references [Sidner86] and definite descriptive references [Grosz81] expressed by the user in natural-language. In the CUBRICON system, the main dialogue focus list is consulted in determining the referent of a pronoun. In the case of a definite reference, if an appropriate referent is not found in the main dialogue focus list, CUBRICON consults the display model. The motivation for this is the fact that when a person expresses a definite reference, such as "the airbase," with just one such object in view (as on a graphics display), and none have been discussed, then the person most likely refers to the one in view, even though he or she may know about several others.

We have not addressed the problem of understanding an input phrase (e.g., "the airbase") when more than one such object is in view and none have been discussed in the human-computer dialogue. The phrase is ambiguous, but a reasonable approach would be to interpret the phrase as referencing the one in view that is most relevant to the user's task.

The discourse model is used during output generation also. When CUBRICON composes a reference for an entity as part of a NL sentence, it consults the discourse model. If the entity is represented in the display model as being visible on one of CUBRICON's windows, then the system uses a deictic dual-media expression to refer to the entity in the output sentence. The deictic expression consists of a phrase, such as "this airbase," and the simultaneous blinking/highlighting of the airbase as the system's means of pointing to the object. If the entity is the most salient of its gender according to the main focus list, CUBRICON uses a pronoun to refer to the entity.

2.5.3. The User Model

Many aspects of a user are highly relevant to interface technology. These aspects include level of expertise in the current task, perspective based on the user's role, his or her value system, degree and nature of impairedness due to fatigue or illness, and preferences concerning mode of communication. Carberry [Carberry87] provides a brief summary of recent research on user modeling. A special issue of *Computational Linguistics* [Kobsa88] provides more in-depth discussion of issues in user modeling. To address all of these different aspects of user modeling is, of course, beyond the scope of this project. The aspects of the user that are most relevant in the CUBRICON system are (1) the importance rating that the user attaches to the different entity types relevant to each given task, which we call the user's *entity rating system;* and (2) the task in which the user is currently engaged.

CUBRICON includes a representation of the user's entity rating system as a function of the task being addressed by the user. For a given task being carried out by the user, the entity rating system representation includes a numerical importance rating (on a scale from 0 to 1) assigned to each of the entity types used in the application task domain. The numerical rating assigned to a given entity type represents the degree of importance of the entity to the user. Associated with the entity rating system is a *critical threshold value;* those entities with ratings above this threshold value are critical to the current task, and those with ratings below the threshold are not. The CUBRICON design provides for the entity rating system representation to change automatically under program control in the following manner: (1) When the user's task changes, the system replaces the current entity rating list with the standard initial rating list for the new task. (2) When the user mentions an entity whose rating is below the critical threshold, its rating is reset to equal the critical threshold to reflect the user's interest in the entity and its seeming relevance to the current task. In the current implementation, CUBRICON performs the second function, but the implementation of the first is not complete.

The user's entity rating system plays an important role in composing responses to the user. The entity rating system representation is used in determining what information is relevant in answering questions or responding to commands from the user. The system is used in selecting ancillary information to enhance or embellish the main concept being expressed and to prevent the user from making false inferences that he or she might otherwise make. The entity rating system is also used in organizing the form in which information is presented.

As an example of the first function, if the user instructs the system to "Display the Fulda Gap region," CUBRICON uses the entity rating system representation to determine what objects within the region should be displayed. If the user is a military mission planner, then displaying all the country cottages in the region, for example, is irrelevant. Thus, the objects that the system selects from its database for display are airbases, missile sites, and targets, for example. Section 2.8 presents examples of the use of this entity rating system representation in interactive dialogue between a user and the CUBRICON system.

CUBRICON includes a simple representation of the current task in which the user is engaged. CUBRICON's mode of response to the user is affected by whether o[illegible] the user's task has just changed. The CUBRICON team is developing a task hierarchy: a decomposition of the user's main tasks into subtasks. This a priori task knowledge can be used by CUBRICON to help track the discourse focus, manage the displays, and anticipate the needs of the user.

2.5.4. Knowledge Bases: General and Domain-Specific

The CUBRICON system includes knowledge bases containing general and domain-specific information. General information includes world knowledge applicable across different task domains; domain-specific information is applicable to the particular task domain of the target information system being used as a "back end" for the interface system. Crucial information included in the knowledge bases is that concerning the presentation or expression of the entities/concepts known to the system. The knowledge base includes information on how to express an entity in its unified verbal/graphic language. This includes the words and symbols used to express an entity, specification of which symbols are appropriate under which conditions, and determination of when particular colors are to be used.

2.6 MULTI-MEDIA LANGUAGE UNDERSTANDING

CUBRICON accepts language consisting of coordinated simultaneous natural language and pointing (via a mouse device) to objects on a graphics display. The use of pointing in combination with natural-language forms a very efficient means of expressing a definite reference. This enables a person to use a demonstrative pronoun as a determiner in a noun phrase and simultaneously point to an entity on the graphics display to form a succinct reference. Thus, a person is able to say "this SAM system" and point to an entity on the display to disambiguate which of several SAM systems is meant. The alternative, using natural language only, would be to say something like "the SAM system at 10.35 degrees longitude and 49.75 degrees latitude" or "the SAM system just outside of Kleinburg." The use of pointing references combined with natural language is efficient since the cognitive process of generating the dual-media reference would be much shorter than the generation of the NL-only reference. The result is a reduction in the cognitive workload for the user.

The CUBRC team has developed a formal grammar defining the syntax of the multi-media language. The grammar is implemented in the form of a generalized ATN. The traditional ATN, which takes a linear textual input stream, has been modified so that it takes a multi-media input stream with components from the different input devices. Input from the devices is accepted and fused into a compound stream, and information is maintained as to which point gesture(s) occurred with (or between) which word(s) of the sentence. Each noun phrase or locative adverbial phrase can consist of zero or more words of text along with zero or more pointing references to objects on the displays (there must be at least one point reference or one word). The pointing input that is a component of a noun phrase or locative adverbial can occur anywhere within the phrase—as the first token(s), between the NL words of the phrase, or as the last token(s).

In the CUBRICON system, four types of objects can be referenced via pointing:

- A geometric point within any window (e.g., a map or graph)
- An entity represented by an icon
- A table entry
- A window on one of the monitors

The following sentences illustrate the type of input accepted by the CUBRICON system. These inputs presuppose that a map is displayed on the color graphics screen, with icons representing various entities. Each "<point>" represents a point to an object or location on one of the graphics displays. The system's responses to such input will be discussed in Section 2.7, where output planning and generation is discussed.

Interrogative

"Where is the 43d Soviet Tank Battalion?"

"What is the mobility of this <point> SAM?"

"Is this <point> the base for these troop battalions <point>$_1$ <point>$_2$ <point>$_3$?"

Imperative

"Display the East-West Germany region."

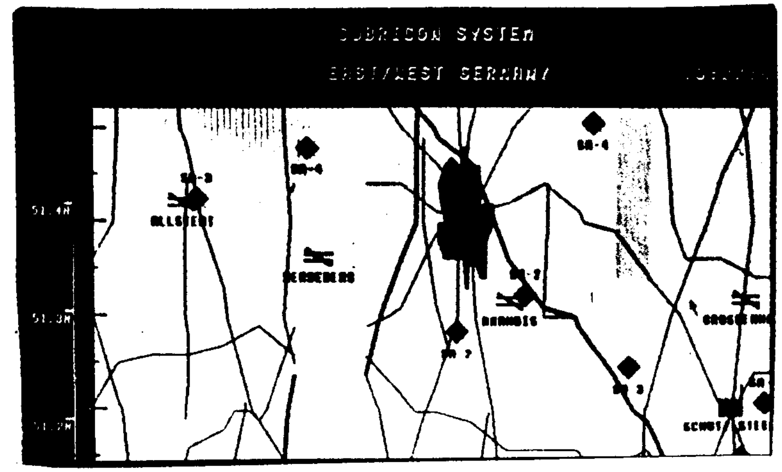
"Display the aimpoints within this <point> airbase."

"Present the OCA1001 mission plan."

Use of such dual-media references entails certain problems, however. A point by the user can be ambiguous if it is to an area where two or more graphical figures or icons overlap. Also, the user may inadvertently miss the object to which he or she intended to point. Figure 2.2 shows overlapping extents for an airbase and a SAM system. If the user points to the overlapping area, the point has two interpretations. However, a NL reference spoken simultaneously with the point can disambiguate the pointing reference. For example, if the phrase "this airbase" is spoken in conjunction with the point to the overlapping extents of Figure 2.2, the dual-media reference is no longer ambiguous. On the other hand, the user may intend to point to the Allstedt airbase but may either inadvertently point to the nearby SAM icon instead or totally miss all the icons with the point gesture. The CUBRICON methodology includes features for handling these problems.

Some systems use default techniques to handle ambiguous pointing. For example:

FIGURE 2.2
RECTANGULAR ICON EXTENTS THAT OVERLAP



1. A point returns the entity represented by the "top" or "foremost" icon, where the system has a data structure it uses to remember the order in which icons are "painted" on the display (i.e., which are further in the background and which are foremost in the foreground).
2. The icons or entities are assigned weights representing importance, and the icon with the largest weight is selected as the interpretation of an ambiguous point.
3. The icon whose "center" is closest to the location pointed to is selected.

Combinations of these techniques can also be used. A serious disadvantage of these point interpretation techniques is that it is difficult, if not impossible, for certain icons to be selected via point references. Such default techniques were deliberately not used in the CUBRICON system. CUBRICON's acceptance of natural language accompanying a point gesture overcomes the limitations of such weak default techniques and provides a more flexible referencing capability.

CUBRICON uses several of the knowledge sources discussed in Section 2.5 in determining the referent of a combined NL and pointing reference. The following examples illustrate the CUBRICON methodology.

In the first example, determination of the referent depends primarily on the use of the task domain knowledge represented in the knowledge base combined with the ancillary visual discourse focus representation called the display model (see Section 2.5.2).

User: "What is the status of this <point> airbase?"

When the phrase "this <point> airbase" is parsed, the system uses the point coordinates to determine which icons are touched by the point. Then the display model is searched to retrieve the semantic network nodes repre- senting the objects graphically displayed by the "touched" icons. Within the knowledge base, the system has a representation of the category to which each object belongs, as well as a representation of the airbase concept. From the hierarchy of the knowledge base, the system determines which of the objects selected by the point gesture are airbases, and it discards the others. If the user has pointed to at least one airbase, the system uses this (these) airbase instance(s) as the referent of the dual-media noun phrase. If the user has pointed at no airbases, the system performs a bounded incremental spatial (geographical) search for an airbase to use as the referent.

The second example entails the use of the syntax and semantics of the sentence processed thus far, along with the knowledge base, to determine the referent of the phrase "this <point>." Here the concept of "mobility" is the critical item of information.

User: "What is the mobility of this <point> ?"

From the display model, the system retrieves the objects represented by the icons that were touched by the point gesture. From the syntax of the noun phrase "the mobility of this <point>" and the semantics of the word "mobility" as represented in the knowledge base, the system deduces that mo- bility is a property (as opposed to a subpart or some other possible relation that could exist between the concepts mentioned) of the object mentioned in the prepositional phrase. The system then determines which of the objects selected by the point gesture have a property called mobility by consulting the knowledge base. The other objects selected by the point gesture are dis- carded. The resulting set is used as the referent of the phrase "this <point>." If the set is empty, the system performs a bounded incremental search in the geographical area of the user's point gesture to find an object with the prop- erty "mobility."

The case frame indicated by the main verb of a sentence is another source of information that CUBRICON can use to determine the referent of a dual- media phrase. The case frame of the main verb can provide constraints on the fillers of the frame slots. A simple example of the use of such constraints is provided in the following example:

User: "Zoom in on this <point>." ,

For CUBRICON, the semantics of the "zoom in" command require ei- ther an object with components that can be presented in enlarged form or a location (that can be represented by a pair of coordinates). If the point gesture touches an icon representing an entity that can be displayed in en- larged form, then the referent of the point is that entity. The system would consult the knowledge base to determine if this is the case. Otherwise, the system assumes that the referent is the location designated by the point, and it displays an enlargement of the area around the point.

Section 2.8 discusses additional examples from a user-computer dia- logue to illustrate the CUBRICON methodology for parsing and inter- preting dual-media input. This topic is discussed more fully in [Neal88b; Neal89].

2.7 PLANNING THE MULTI-MEDIA RESPONSE

The CUBRICON system design provides for output in the form of coordi- nated speech, text displayed on a CRT, and pictorial information on a color graphics display. The system uses the ATN methodology for generation as well as understanding. ATNs have been used to generate text by Simmons & Slocum [Simmons72], in the XCALIBUR system [Carbonell85], and by Shapiro [Shapiro75, Shapiro82a]. In this project, this ATN methodology is being expanded to generate language in the form of text, graphics, or a text/graphics combination from the semantic network knowledge bases.

A multi-media language provides a wider range of choices for the for- mulation of an expression for a given concept than a single-media language, such as natural language used in isolation. There are alternative ways of ex- pressing a given concept in a strictly textual language. For example, George Bush can be referred to as "George Bush," as "the President of the United States," or as "the Commander in Chief of the Armed Forces of the United States." If language is extended to include graphics as well, then Mr. Bush could be referred to by pointing to his picture, by creating/displaying his image as a reference, or by a combination of text and graphics. Thus, a ref- erence can be composed of a proper noun, a noun phrase with modifiers, a graphic form with features of color, size, and location on a display, and so on. Furthermore, the graphic features "modifying" a form or shape can com- municate meaning just as do the adjectives and relative clauses that modify a head noun of a noun phrase.

The Multi-Media Output Planner composes the response that is to be pro- duced to the user by the Output Generator in coordinated multiple modali- ties. The Output Planner determines the media and modalities for expressing the response information to the user, but it then must determine whether the resources are available to do so. If they are not, the Planner must take ap- propriate action to modify the state of the resources, modify the information to be expressed, and/or select different modalities for expressing the infor- mation before composition of the output can be accomplished.

The CUBRICON design includes the following output modalities, in iso- lation or in combination: color graphics/pictorial displays, tables, histograms, written NL prose, spoken natural language, and fill-in-the-blank forms. This list does not exhaust the possibilities, of course, but it provides a good, varied selection with which to "prove our concept" and upon which to build.

Selection of the most appropriate modalities for expressing information in the CUBRICON system is based on the nature and characteristics of the information. Our system design is based on the premise that graphic/pictorial presentation is always desirable. The following is a brief summary of the selection criteria.

1. Color graphics are selected whenever the CUBRICON system knows how to represent the information pictorially.
2. A table is selected when the values of common attribute(s) of several entities must be expressed.
3. A histogram is selected when a quantitative attribute of several entities must be displayed in a comparative form.
4. A predefined form is selected when the task engaged in by the user requires it.
5. Natural-language prose is selected for the expression of a proposition, relation, event, or combination thereof when the knowledge structures being expressed are heterogeneous. Natural language can be presented in either spoken or written form. *Spoken natural language* is selected for
   - Dialogue descriptions to assist the user in comprehending the presented information. These include explanations of graphic displays or display changes and verbal highlighting of objects on the displays (e.g., "The enemy airbases are highlighted in red").
   - Warnings to alert the user of important events that have taken or are about to take place (e.g., new critical information comes into the application system database, and the system notifies user: "The XXX airbase has been damaged by enemy shellfire").
   - Informing the user about the system's activity (e.g., "I'm still working" when the user must wait for output from the system).
   - Short expressions of relatively nontechnical information that can be remembered when presented serially (e.g., a "yes"/"no" answer to a user's question).

   Written natural language is selected for longer technical responses that would strain the user's short-term memory if speech were used (see [Miller56]).

Most frequently, multiple modalities are desirable for expressing a body of information to the user. For example, to inform the user about the movements of a certain tank battalion, a desirable presentation might be a spoken explanation and coordinated drawlng on a graphic map display showing movements of the battalion, as well as a printed textual summary with ancillary information on the monochrome display. The multiple modalities should be selected to complement and enhance one another. Andriole [Andriole86] has used "graphic equivalence" effectively, employing dual displays or split screens to present the same material in different forms to aid user

comprehension and problem solving performance. We are not restricting the system to presenting the *same* material in different forms; instead, our system presents related material or different aspects of a given event or concept in different forms/modalities (as appropriate, based on the nature and characteristics of the information). We are also not restricted to graphic display presentations.

Our top-level output planning process is summarized as follows. This planning process presupposes that the primary relevant information has been obtained for response to the user.

1. For each information item or cluster, determine the ideal modality for its expression. Graphic/pictorial presentation is always desirable. Natural language can always be used as a last resort if no other modality is available.
2. Determine whether the resources are available to express the information as desired. For a color graphics display, find out if the items to be expressed graphically are already present on the display (e.g., if objects of interest in a geographical domain are already displayed on a map). If so, no additions are necessary. If not, is there room to add them in their "natural" position? That is, can the desired objects be inserted in the area already available on the graphics display without changing the area shown, or does the displayed area need to be extended or changed totally? Similar issues arise for a monochrome display. For a speech output device, resources are always available.
3. If the desired resources are not available, modify the state of the resources. The desired resources would be "not available" in at least two cases: (a) The physical device is not functional (e.g., needs repair), or (b) the device (e.g., a display) already contains critical information that cannot be disrupted or covered by a window. For the graphics displays, if all the items to be expressed graphically are not on the display, then the system must compose a new display. Using terminology for a geographical display, the possible cases are:
   - *Zoom out* with intelligent addition of relevant ancillary objects to fill in the new area to maintain consistency throughout the display.
   - *Zoom in* with intelligent addition of relevant objects to create an intelligible display.
   - *Pan* to a different area, maintaining consistency in the types of objects displayed.
   - Employ a combination of these procedures.
   - Display a different disjoint area: (a) Completely replace the display with the new "area," or (b) open a window on the monitor to show new information.
4. If resources are still not available to accommodate the information to be expressed, try modifying the information. Trim the amount of infor-

mation by filtering on the basis of relevance with regard to user model and/or discourse model.

5. If the information still cannot be expressed in the given modality due to insufficient resources, select another modality and go back to step 2.
6. Compose the output, having resolved resource constraints.

The methodology discussed in this section is fully implemented, with the exception of the histogram modality. The processes for the modality selection and for resolution of resource problems are undergoing refinement to maximize generality and extend their functionality. The next section discusses working examples from a dialogue with the system. The examples illustrate some of the functionality of the system.

2.8 SAMPLE DIALOGUES

This section presents short sample dialogues to illustrate the functionality and processing discussed in the previous sections. The dialogues are concerned with mission planning and situation assessment in a tactical air control domain. Consider the following initial user input to the CUBRICON system:

User: "Display the Fulda Gap region."

CUBRICON: (Refer to Figure 2.3)

Speech output:

- Statements to direct the user's attention to the appropriate monitor as information is displayed. Just before the region is displayed on the color graphics monitor:

  "Look at the color graphics screen. The Fulda Gap region is being presented."

  Just before the table is presented on the monochrome monitor:

  "The corresponding table is being presented on the monochrome screen."

Color graphics display:

- Map of the Fulda Gap region with main roads, major cities, waterways, and national boundaries
- Icons representing entities within the Fulda Gap region, that are above the critical threshold according to the entity rating system for the user's task, superimposed on the map

Monochrome graphics display:

- Table of relevant entity attributes for those entities on the map display

The planning and composition of output for the user depends on the nature of the information, the discourse context, and the user model. Since

FIGURE 2.3
THE DISPLAYS COMPOSED BY THE SYSTEM

this sample command is the initial input, the process is simpler than for other cases. The CUBRICON system knows that a region can be represented graphically and therefore chooses graphics as the primary modality for display. Regions are represented in the CUBRICON knowledge base with an associated boundary. The boundary is retrieved by the system, and the main roads, major cities, waterways, and national borders are displayed on the color graphics display. These items are displayed by use of the RADC MAP Display System [Hilton87]. The CUBRICON system then searches its knowledge base for task-specific objects within the region to be displayed. The selection of these objects is based on the user model discussed in Section 2.5, which consists of the user's entity rating system: a task-dependent assignment of importance ratings to the objects in the task domain. The system does not display all entities it knows about in the region, but only those that lie above the criticality threshold for the particular task in which the user is engaged (a subtask of planning an air strike mission). Thus, the system decides to display all airbases, surface-to-air missile (SAM) sites, and critical factories and plants, but not objects such as food production plants or minor industries that are

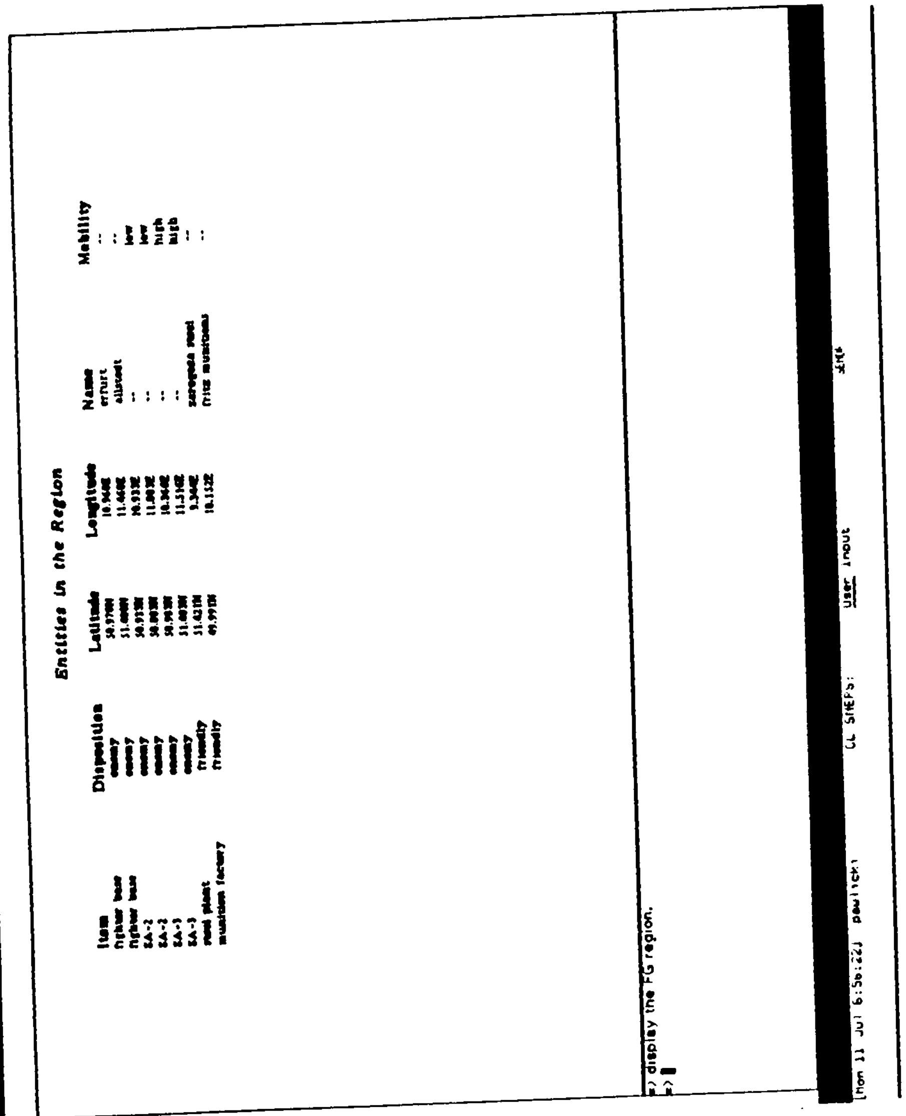


FIGURE 2.3 (Cont.)

not germane to mission planning. The resulting color map display is shown in Figure 2.3*a*.

Based on the information provided by the user model, the system knows the important attributes of each object. These attributes are not displayed or communicated via the map display but have been selected as relevant, so the system must determine a modality for presenting this information. Since this information consists of lists of objects with different values of common attributes, the system determines that a table presentation is appropriate. The system composes a table showing the important attributes of the displayed objects. The number of attributes (columns of the table) is constrained by available space. The resultant table is displayed on the monochrome display and is shown in Figure 2.3*b*.

To further illustrate the CUBRICON system's modality selection and output composition process, we will consider the next user input. The user queries the system about the location of the Dresden airbase in a manner that provides no instruction to the system as to how to present the information (e.g., map, natural language only).

User: "Where is the Dresden airbase?"

CUBRICON: (Refer to Figure 2.4.)

Speech output:

- Statements to direct the user's attention to the appropriate monitor when a major window is presented. As the map is expanded on the color monitor:

  "The map on the color graphics screen is being expanded to include the Dresden airbase."

  As the table is presented on the monochrome monitor:

  "The corresponding table is being presented on the monochrome screen."

- After the map is expanded, a statement to direct the user's attention to the Dresden airbase on the map:

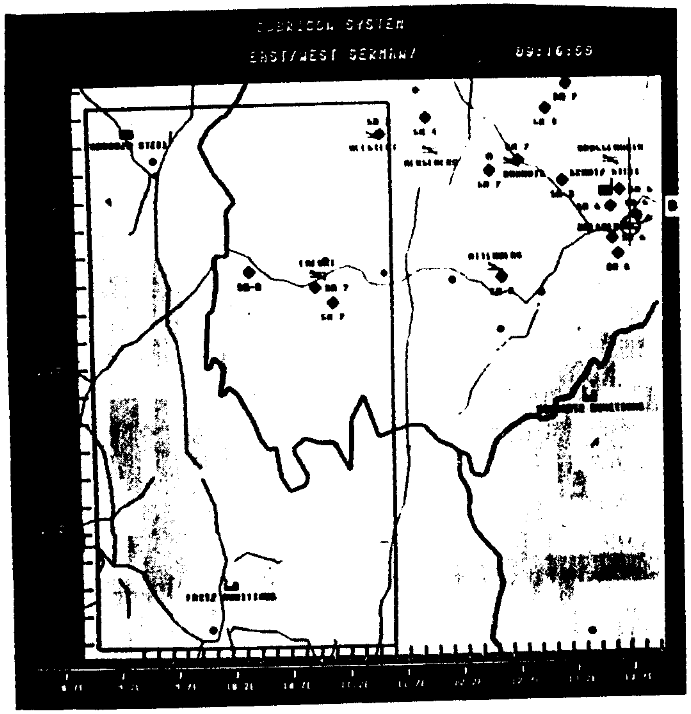
  "The Dresden airbase is located here <point>."

  The word "here" is accompanied by a visual point gesture in the form of a blinking airbase icon and the addition of a pointing text box.

Color graphics display:

- Map of the Fulda Gap region with added area that includes the Dresden airbase
- Main roads, major cities, waterways, and national boundaries (as before, but across the whole map, showing old and new areas)
- Icons representing entities within the new map area displayed that lie above the critical threshold on the entity rating system for the user's task

FIGURE 2.4
MAP AND TABLE MAINTAINING CONTEXT AND CONSISTENCY



- The Dresden airbase icon presented in a blinking and highlighted manner (the blinking is temporary and is coordinated with the speech)

Monochrome graphics display:

- Table of relevant entity attributes (same table as before, but expanded to include the new entities added to the map covering the extended area)

As discussed in Section 2.7, whenever possible the CUBRICON system prefers to present information graphically, with ancillary information presented simultaneously in another modality. Since CUBRICON knows how to display an airbase graphically (it has an icon associated with the class in the knowledge base), and since each particular airbase in the knowledge base has an associated geographical location, the system will display the airbase on the color graphics map, with additional information displayed in another modality. If the Dresden airbase were already displayed on the color map display, the system would choose to blink the particular airbase

FIGURE 2.4 (Cont.)

Entities in the Region

Item	Disposition	Latitude	Longitude	Name	Mobility
fighter base	enemy	51.100N	13.700E	dresden	--
fighter base	enemy	51.350N	11.960E	merseberg	--
fighter base	enemy	51.310N	13.530E	grossenhain	--
fighter base	enemy	50.900N	12.510E	attenberg	--
fighter base	enemy	51.400N	11.460E	allstedt	--
fighter base	enemy	50.970N	10.960E	erfurt	--
fighter base	enemy	51.310N	12.660E	brandis	--
SA-2	enemy	51.203N	12.446E	--	low
SA-2	enemy	51.550N	13.133E	--	low
SA-2	enemy	51.316N	12.716E	--	low
SA-2	enemy	50.803N	11.003E	--	low
SA-2	enemy	50.933N	10.933E	--	low
SA-3	enemy	51.250N	13.100E	--	high
SA-3	enemy	50.950N	12.566E	--	high
SA-3	enemy	51.403N	11.516E	--	high
SA-3	enemy	50.903N	10.366E	--	high
SA-4	enemy	51.473N	12.955E	--	low
SA-4	enemy	51.450N	11.916E	--	low
SA-6	enemy	51.016N	13.386E	--	high
SA-6	enemy	51.066N	13.336E	--	high
SA-6	enemy	51.133N	13.753E	--	high
SA-6	enemy	51.166N	13.320E	--	high
SA-6	enemy	51.216N	13.603E	--	high
steel plant	enemy	51.211N	13.400E	schutz steel	--
steel plant	friendly	51.421N	9.344E	zaregoza steel	--
munition factory	enemy	50.561N	13.320E	schwartz munitions	--
munition factory	friendly	49.991N	10.152E	fritz munitions	--

=> Display the FG region.
=> Where is the Dresden airbase?
=>

[Mon 11 Jul 8:43:02] pawlicki   LL SHEFS: User Input   SEND icon as its way of pointing to the object and would accompany this pointing action with a spoken response. If the Dresden airbase could be added to the current map, CUBRICON would do so and direct the user's attention to the airbase icon. However, the Dresden airbase is outside of the region shown n the map display currently on the color CRT. Therefore, the system must lecide how to show the airbase. What map should be displayed?

In composing a new map on which to display the Dresden airbase, the system has some choices: open a window on the color graphics display showing the area around the Dresden airbase, replace the old map on the CRT with a new map of the area around the Dresden airbase, or compose a new map including both the old map and the region around the Dresden airbase.

An important guideline to which the CUBRICON system tries to adhere is to maintain the context of the user-computer dialogue. With regard to the graphic displays, this means that the system tries to retain the most recently discussed or mentioned objects on the displays in order to maintain continuity in the dialogue. The discourse focus space representations, (see Section 2.5) are the key knowledge sources in this process. The system composes a new map containing the objects on the old map as well as the Dresden airbase. The algorithm that the system uses to determine the boundary for a new map of this type is to find the smallest rectangle that encloses the objects on the current map plus the new objects to be displayed and then add a small border area around all sides. This essentially extends the area shown to include both the old and the new objects.

A key feature of the CUBRICON system is that it can display any rectangular region within a "master map" boundary containing the East-West Germany region. That is, the system can select any degree of longitude for the eastern or western boundary of the map to be displayed and any degree of latitude for the northern or southern boundary (within the confines of the master map). For any such map, as indicated previously, the non-domain-specific items such as roads, major cities, waterways, and national boundaries are displayed by the RADC MAP Display System.

Another important guideline to which the CUBRICON system adheres is to maintain consistency throughout a display to prevent the user from making false inferences about what is or is not located within the region. In the case of our map display, this means that there should be consistency in the types of objects shown across the entire map. If SAMs are displayed in the old region, then they should be displayed in the newly added map area; the same holds for other types of objects. If this is not done, the user would probably infer that there are no SAMs in the new area since none appear on the display. Figure 2.4*a* shows the new map display composed by the CUBRICON system in response to the user's input "Where is the Dresden airbase?" The rectangular outline within the map is used to indicate the previously displayed area. This provides graphic context: the new entities are shown in the context of the previously displayed area. Guided by the consistency principle, the system also modifies the tabular presentation that is on the monochrome display. Both displays are also shown in Figure 2.4.

As discussed in Section 2.6, the CUBRICON system accepts NL input with coordinated pointing to objects on the graphics display. As an example of this multi-media input, the user can enter:

User: "Is this <point> a steel plant?"
where the point touches on or near an icon on the map display.
CUBRICON:
Speech: "No, it is a munitions factory."

CUBRICON decides to respond in natural-language since the information to be expressed is short and relatively nontechnical (refer to the modality selection process discussed in Section 2.7). The system uses a pronoun to refer to the object referenced by the user since it is the most salient object according to the system's discourse focus space list (refer to Section 2.5.2).

The following example illustrates CUBRICON's ability to infer the referent of a pointing gesture that misses the intended icon. It also illustrates CUBRICON's ability to respond in highly integrated speech and graphics.

User: "What is the mobility of this <point>$_1$<point>$_2$<point>$_3$?"
where each point is a mouse point by the user on the map display. The first point touches the wrong icon, and the third touches no icon at all.
CUBRICON:
Speech output with coordinated color graphics:
"The mobility of this <blinking icon>$_1$ sam is low, the mobility of this <blinking icon>$_2$ sam is high, and the mobility of this <blinking icon>$_3$ sam is high."
For each SAM reference, the particular SAM icon blinks as the system's method of pointing to it. Next to each SAM icon, a small text box is added to provide a terse written summary of the requested information.
Monochrome display—small NL window:
"Point 1: The icon you pointed at does not have the property 'mobility,' but a nearby referent has been found."
"Point 3: Your point gesture did not touch any objects on the display; however, a nearby object with the property 'mobility' has been found."

The mouse handler returns representations of the objects touched by the point, if any exist. CUBRICON examines each of the returned objects to determine if it has a mobility property. The first point gesture touched an airbase icon but no SAM icons. CUBRICON deduces that the referent cannot be the airbase, since airbases are not mobile according to its knowledge base. Therefore, the system performs an incremental geographic bounded search around the location of the user's point gesture to find an object with the prop-

CUBRICON correctly interprets the point as referring to a geometric point location, even though the user's point gesture touches some icon on the display. This is due to the NL phrase "this point" accompanying the mouse point gesture. The system interprets the statement as an instruction to magnify an area around the location of the point. The system uses a predefined "radius" to determine the area to magnify. CUBRICON then draws a box around the area to be magnified so that the user can follow its activities. At this point the system has a couple of choices: It can repaint the existing map window with the new area, or it can open a window to show the new area. Since CUBRICON's task model does not indicate that the user's task has changed, it assumes that the original map is still relevant. It therefore opens an additional window to show the new area, as illustrated in Figure 2.5*a*. To present attributes of the entities on the new map that are not communicated via the color graphics map and to maintain consistency, CUBRICON displays a table of the important attributes of the entities visible on the new map. To maintain positioning consistency on the two monitors, CUBRICON places the table window on the monochrome monitor in a position that closely matches the position of the new map window on the color monitor.

In this section we have discussed examples to (1) illustrate some of the key functionality of the CUBRICON system, (2) show how some of the important CUBRICON knowledge sources are used, and (3) show how the important human factor guidelines are applied by the system. The functionality illustrated includes the ability of the system to accept and understand a multi-media noun phrase that consists of a textual noun phrase accompanied by a pointing reference; use natural language to infer the intended referent of an "ill-formed" dual-media expression; select the modalities for presentation of information to the user; select relevant information to present to the user; compose the multi-media presentations, particularly map graphics and tables; and use coordinated speech and pointing gestures for output. The CUBRICON knowledge sources used in the examples of this section include the nature and characteristics of the information to be expressed, the knowledge bases, the discourse model, and the user model. The guidelines discussed in the examples were the maintain-context guideline and the consistency guideline.

2.9 CURRENT STATUS AND FUTURE DIRECTION

As stated in Section 2.2, CUBRICON is implemented on a Symbolics Lisp Machine with a mouse pointing device and both a monochrome graphics monitor and a color graphics monitor. The Dragon Systems VoiceScribe 1000 speech recognition system and the DECtalk speech production system are both fully functional in the CUBRICON hardware suite of equipment.

The CUBRICON software modules discussed are implemented and functioning. These modules include: the lexicon and grammar for parsing and

interpreting input consisting of NL speech or keyboard text with simultaneous coordinated pointing to objects on the graphics displays; the executor that retrieves information from the target information system for presentation to the user; the process of selecting media/modalities for the presentation of information to the user; the processes of intelligently composing map displays and tables; the process of generating simultaneous coordinated speech and pointing gestures during output; and the knowledge sources, including the knowledge bases, discourse model, and user model.

The CUBRICON team is continuing its development of the processes discussed in this paper as well as developing new functionality. Areas in which the team is continuing current work and/or planning future development include:

1. Extending and refining the system's automated process of selecting the appropriate media/modalities for expressing responses to the user
2. Enhancing the user model and developing the task hierarchy
3. Adding additional modalities (e.g., predefined forms for mission planning) to the CUBRICON repertoire
4. Extending and refining the process of determining whether the desired resources are available and, if not, either modifying the state of the resources or the information to be expressed, or selecting different modalities for expressing the information
5. Extending the functionality of the multimodal output generator to produce more sophisticated, highly integrated natural language and graphics at the discourse and sentence level, including the coordinated presentation of time- and space-dependent activities and events (e.g., planned movements of military forces)

2.10 SUMMARY

This paper discussed the architecture of the CUBRICON system, the key functionality of the system, the important knowledge sources used by the system, multi-media language understanding, the system's process for composing multimodal output, examples illustrating some of the system's key functionality and processing methodology, and the current status and future direction of the CUBRICON project.

Key functionality of the system includes the ability to handle simultaneous input in natural language and graphics pointing, automated selection of media/modalities for output presentation, and automated composition of multi-media responses to be conveyed to the user.

The knowledge sources used by the CUBRICON system are the lexicon and grammar, the discourse model, the user model, the knowledge bases of general world and application-specific knowledge, the nature and characteristics of the information to be presented, guidelines for enhancing human comprehension, and the constraints or limitations of the hardware input and output devices.

The system's parsing/interpretation of input consisting of simultaneous coordinated natural-language and graphics pointing gestures was discussed with emphasis on the system's ability (1) to use natural-language to disambiguate pointing to objects on a graphics display and (2) to infer the intended referent of a dual-media expression in which the user misses the object to which he or she intended to point.

The system's output composition process includes selection of appropriate modalities and media, determination of whether resources are available, subsequent modification of resources or information to be expressed (if necessary), modification of selected output modalities/media (if necessary), and composition of the output.

Examples were presented to illustrate some of the key functionality of the CUBRICON system, to show how the CUBRICON knowledge sources are used, and to show how two important guidelines for enhancing human comprehension are applied by the system.

CUBRICON is being developed as part of the Intelligent Multi-Media Interfaces Project. This project is devoted to the development of interface technology that integrates speech, natural-language text, graphics, tables, and pointing gestures for human-computer dialogues. The objective of the project is to develop interface technology that uses the media/modalities intelligently in a flexible, context-sensitive, and highly integrated manner.

Acknowledgments

The other members of the CUBRC team are J. Byoun, Z. Dobes, D. Funke, J. Lammens, M. Summers, and C. Thielman.

The RADC MAP Display System was initially developed by Lt. Mike Hilton (RADC), who graciously arranged for us to use it.

References

[Aceves84] Aceves, J. J. G., Poggio, A., and Eliott, D. 1984. *Research into Multimedia Message System Architecture, Final Report.* Proj. No. 5363. Menlo Park, CA: SRI International.

[ACM88] ACM. 1988. *Communications of the ACM: Special Issue on HyperText*, 31(7).

[Allgayer89] Allgayer, J., Jansern-Winkeln, R., Reddig, C., and Reithinger, N. 1989. Bidirectional Use of Knowledge in the Multi-modal NL Access System XTRA. *Proceedings of IJCAI-89*, Detroit, pp. 1492–1497.

[Andriole86] Andriole, S. J. 1986. Graphic Equivalence, Graphic Explanations, and Embedded Process Modeling for Enhanced Process Modeling for Enhanced User-system Interaction. *IEEE Trans. on Systems, Man, and Cybernetics*, 16 (6), 919–926.

[Arens88] Arens, Y., Miller, L., Shapiro, S., and Sondheimer, N. 1988. Automatic Construction of User-Interface Displays. *Proceedings of AAAI-88*, pp. 808–813.

[Arens91] Arens, Y., Miller, L., and Sondheimer, N. 1991. Presentation Design Using an Integrated Knowledge Base. In present volume.

[Carberry87] Carberry, S. 1987. First International Workshop on User Modeling. *AI Magazine*, 8(3), 71–74.

[Carbonell85] Carbonell, J. G., Boggs, W. M., Mauldin, M. L., and Anick, P. G. 1985. An Integrated Natural Language Interface. In Andriole, S. (ed.), *Applications in Artificial Intelligence.* Petrocelli Books, pp. 227–245.

[Cheikes88] Cheikes, B. A., and Webber, B. L. 1988. The Design of a Cooperative Respondent, *Proceedings of The Workshop on Architectures for Intelligent Interfaces: Elements and Prototypes*, J.W. Sullivan and Tyler, S.W. (eds.), Monterey, CA, March, pp. 3–17.

[Conklin87] Conklin, J. 1987. Hypertext: An Introduction and Survey. *Computer*, 20(9), 17–41.

[Egan89] Egan, D. E., Remde, J. R., Gomez, L. M., Landauer, T. K., Eberhardt, J., and Lochbaum, C. C. 1989. Formative Design-Evaluation of SuperBook. *ACM Transactions on Information Systems*, 7(1), 30–58.

[Feiner82] Feiner, S., Nagy, S., and van Dam, A. 1982. An Experimental System for Creating and Presenting Interactive Graphical Documents. *Trans. Graphics*, 1(1), 59–72.

[Feiner89] Feiner, S., and McKeown, K. R. 1989. Coordinating Text and Graphics in Explanation Generation. *Proceedings of the DARPA Speech and Natural Language Workshop*, Cape Cod, MA, pp. 424–433.

[Grosz78] Grosz, B. J. 1978. Discourse Analysis, in Walker, D., (ed.), *Understanding Spoken Language.* New York: Elsevier North-Holland, pp. 229–345.

[Grosz81] Grosz, B. J. 1981. Focusing and Description in Natural Language Dialogues. In Joshi, A., Webber, B. and Sag, I. (eds.), *Elements of Discourse Understanding.* New York: Cambridge University Press, pp. 84–105.

[Grosz85] Grosz, B. J., and Sidner, C. L. 1985. Discourse Structure and the Proper Treatment of Interruptions. *Proceedings of IJCAI*, pp. 832–839.

[Grosz86] Grosz, B. J. 1986. The Representation and Use of Focus in a System for Understanding Dialogs. In Grosz, B. J., Jones, K. S., and Webber, B. L. (eds.), *Readings in Natural Language Processing.* San Mateo, CA: Morgan Kaufmann, pp. 353–362.

[Hilton87] Hilton, M. L. 1987. Design and Implementation of the MAP Display System. RADC Report.

[Hollan91] Hollan, J., Rich, E., Hill, W., Wroblewski, D., Wilner, W., Wittenburg, K., and Gruden, J. 1991. An Introduction to Hits: Human Interface Tool Suite. In present volume.

[Kaplan82] Kaplan, S. J. 1982. Cooperative Responses from a Portable Natural-Language Database Query System, In Brady, M. (ed.), *Computational Models of Discourse.* Cambridge, MA: MIT Press, pp. 167–208.

[Kass91] Kass, R., and Finin, T. 1991. General User Modeling: A Facility to Support Intelligent Interaction. In present volume.

[Katz84] Katz, A. 1984. An Experimental Internetwork Multimedia Mail System. *Proceedings IFIP 6.5 Working Conf. Computer Message Services.* Nottingham, England.

[Kobsa86] Kobsa, A., Allgayer, J., Reddig, C., Reithinger, N., Schmauks, D., Harbusch, K., and Wahlster, W. 1986. Combining Deictic Gestures and Natural Language for Referent Identification. *Proceedings of the 11th International Conference on Computational Linguistics.* Bonn, FR Germany.

[Kobsa88] Kobsa, A., and Wahlster, W. (eds.), 1988. *Computational Linguistics: Special Issue on User Modeling.* Cambridge, MA: MIT Press.

[Maida85] Maida, A. S., and Shapiro, S. C. 1985. Intensional Concepts in Propositional Semantic Networks. In Brachman, R. J., and Levesque, H. J. (eds.), *Readings in Knowledge Representation.* San Mateo, CA: Morgan Kaufmann, pp. 169–190.

[Mason88] Mason, J. A., and Edwards, J. L. 1988. Explicit Models in Intelligent Interface Design. *Proceedings of the Workshop on Architectures for Intelligent Interfaces: Elements and Prototypes.* Sullivan, J. W. & Tyler, S. W. (eds.) Monterey, CA, March, pp. 151–167.

[McKay80] McKay, D. P., and Shapiro, S. C. 1980. MULTI—A LISP-Based Multiprocessing System. *Conference Record of the 1980 LISP Conference.* Stanford Univ., pp. 29–37.

[Miller56] Miller, G. A. 1956. The Magical Number Seven Plus or Minus Two. *Psychological Review*, 63, 81–97.

[Neal86] Neal, J. G., and Shapiro, S. C. 1986. Knowledge Representation for Reasoning about Language. In Boudreauz, J. C., Hamill, B. W., and Jernigan, R. (eds.), *The Role of Language in Problem Solving.* New York: Springer-Verlag, pp. 27–47.

[Neal87] Neal, J. G., and Shapiro, S. C. 1987. Knowledge Based Parsing. In Bolc, L., (ed.), *Natural Language Parsing Systems.* New York: Springer-Verlag, pp. 49–92.

[Neal88a] Neal, J. G., Bettinger, K. E., Byoun, J. S., Dobes, Z., and Thielman, C. Y. 1988. An Intelligent Multi-media Human-Computer Dialogue System. *Proceedings of the Workshop on Space Operations, Automation, and Robotics (SOAR-88).* Dayton: Wright State University, pp. 245–251.

[Neal88b] Neal, J. G., Dobes, Z., Bettinger, K. E., and Byoun, J. S. 1988. Multi-modal References in Human-Computer Dialogue. *Proceedings of AAAI-88*, pp. 819–823.

[Neal89] Neal, J. G., Thielman, C. Y., Dobes, Z., Haller, S. M., and Shapiro, S. C. 1989. Natural Language with Integrated Deictic and Graphic Gestures. *Proceedings of the DARPA Speech and Natural Language Workshop*, Cape Cod, MA, pp. 410–423.

[Neches86] Neches, R., and Kaczmarek, T. 1986. *AAAI-86 Workshop on Intelligence in Interfaces.* USC/Information Sciences Institute, August 1986.

[Press86] Press, B. 1986. The U.S. Air Force TEMPLAR Project Status and Outlook. *Western Conf. on Knowledge-Based Engineering and Expert Systems*, Anaheim, CA, pp. 42–48.

[Reithinger87] Reithinger, N. 1987. Generating Referring Expressions and Pointing Gestures. In Kempen, G. (ed.), *Natural Language Generation.* Dordrecht: Nijhoff, pp. 71–81.

[Roth91] Roth, S., Mattis, J., and Mesnard, X. 1991. Graphics and Natural Language as Components of Automatic Explanation. In present volume.

[Shapiro75] Shapiro, S. C. 1975. Generation as Parsing from a Network into a Linear String. *AJCL*, microfiche 33, pp. 45–62.

[Shapiro79a] Shapiro, S. C. 1979. The SNePS Semantic Network Processing System. In Findler, (ed.), *Associative Networks: The Representation and Use of Knowledge by Computers.* New York: Academic Press, pp. 179–203.

[Shapiro79b] Shapiro, S. C. 1979. Using Non-standard Connectives and Quantifiers for Representing Deduction Rules in a Semantic Network. Invited paper presented at *Current Aspects of AI Research*, seminar held at the Electrotechnical Laboratory, Tokyo.

[Shapiro79c] Shapiro, S. C. 1979. Numerical Quantifiers and Their Use in Reasoning with Negative Information. *Proceedings of IJCAI*, pp. 791–796.

[Shapiro81] Shapiro, S. C., and SNePS Implementation Group. 1981. *SNePS User's Manual.* Buffalo, NY, SUNY CS Dept.

[Shapiro82a] Shapiro, S. C. 1982. Generalized Augmented Transition Network Grammars for Generation from Semantic Networks. *AJCL*, 8(1), 12–25.

[Shapiro82b] Shapiro, S. C., Martens, J., and McKay, D. 1982. Bi-directional Inference. *Proceedings of the Cognitive Science Society*, pp. 90–93.

[Shapiro82c] Shapiro, S. C., and Neal, J. G. 1982. A Knowledge Engineering Approach to Natural Language Understanding. *Proceedings of ACL*, pp. 136–144.

[Shapiro87] Shapiro, S. C., and Rapaport, W. 1987. SNePS Considered as a Fully Intensional Propositional Semantic Network, *Proc. AAAI-86*, pp. 278–283; in McCalla, G., and Cercone, N. (eds.), *Knowledge Representation*, New York: Springer-Verlag.

[Shepard89] Shepard, S. J. 1989. A New Approach to Hypertext: MINDS. *AI Expert*, 4(9), 69–72.

[Sidner86] Sidner, C. L. 1986. Focusing in the Comprehension of Definite Anaphora. In Grosz, B. J., Jones, K. S., and Webber, B. L. (eds.), *Readings in Natural Language Processing*, San Mateo, CA: Morgan Kaufmann, pp. 353–362.

[Simmons72] Simmons, R., and Slocum, J. 1972. Generating English Discourse from Semantic Networks. *CACM*, 15(10), 891–905.

[Sullivan91] Sullivan, J. W., and Tyler, S. W., (eds.) 1991. *Intelligent User Interfaces*, present volume.

[Thomas85] Thomas, R. H., Forsdick, H. C., Crowley, T. R., Schaaf, R. W., Tomlinson, R. S., Travers, V. M., and Robertson, G. G. 1985. Diamond: a multimedia message system built on a distributed architecture. *IEEE Computer*, December, 65–78.

[Wahlster91] Wahlster, W. 1991. User and Discourse Models for Multimodal Communication. In present volume.

[Young91] Young, R. L. 1991. A Dialogue User Interfaces Architecture. In present volume.