In the Proceedings of the GIS/LIS '89 Conference, Orlando, Florida, November, 1989.

CUBRICON: A MULTI-MODAL USER INTERFACE [1]

J.G. Neal, C.Y. Thielman, Z. Dobes
S.M. Haller, S. Glanowski, S.C. Shapiro
Calspan-UB Research Center (CUBRC)
P.O. Box 400, 4455 Genesee Street
Buffalo, NY 14225

ABSTRACT

The CUBRICON project is devoted to the development of knowledge-based interface technology that integrates speech input, speech output, natural language text, geographic maps, graphics, and pointing gestures for interactive dialogues between human and computer. The objective is to provide both the user and system with modes of expression that are combined and used in a natural and efficient manner, particularly when presenting or requesting information about objects that are visible, or can be presented visibly, on a graphics display. This paper focuses on those aspects of the CUBRICON human-computer interface (HCI) that pertain to the use of geographic displays. This paper discusses unique interface capabilities that CUBRICON provides including the ability to: (1) accept and understand multi-media input such that references to entities in (spoken or typed) natural language sentences can be accompanied by coordinated simultaneous pointing to the respective entities on a graphics display; (2) dynamically compose geographic map displays including the determination of the boundary of the region to be displayed, selection of the relevant entities to display in the region, and selection of the location and size of the window in which to place the map display; (3) dynamically compose and generate multi-modal language that combines NL with deictic gestures and graphic expressions; synchronously present the spoken natural language and coordinated pointing gestures and graphic expressions.

1 INTRODUCTION

The CUBRICON project [Neal88a, Neal88b, Neal89] is devoted to the development of knowledge-based interface technology that integrates speech input, speech output, natural language text, geographic maps, graphics, and pointing gestures for interactive dialogues between human and computer. The objective is to provide both the user and system with modes of expression that are combined and used in a natural and efficient manner, particularly when presenting or requesting information about objects that are visible, or can be presented visibly, on a graphics display. The goal of the project is to develop interface technology that uses its media/modalities intelligently in a flexible, highly integrated manner modelled after the manner in which humans converse in simultaneous coordinated multiple modalities.

The interface technology developed as part of this project has been implemented in the form of a prototype system, called CUBRICON (the CUBRC Intelligent CONversationalist). Although the application domain used to drive the research for the CUBRICON project is that of tactical Air Force mission planning, the interface technology incorporated in CUBRICON is applicable to domains with similar communication characteristics and requirements.

[1]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.

This paper focuses on those aspects of the CUBRICON human-computer interface (HCI) that pertain to the use of geographic displays. This paper discusses unique interface capabilities that CUBRICON provides including the ability to:

- accept and understand multi-media input such that references to entities in (spoken or typed) natural language sentences can be accompanied by coordinated simultaneous pointing to the respective entities on a graphics display; use simultaneous pointing and natural language references to disambiguate one another when appropriate; infer the intended referent of a point gesture which is inconsistent with the accompanying natural language;
- dynamically compose geographic map displays including the determination of the boundary of the region to be displayed, selection of the relevant entities to display in the region; and selection of the location and size of the window in which to place the map display;
- dynamically select modalities and present information as appropriate for the nature and importance of the information and the discourse context; in particular, for objects that are visible on a map display, be able to use spoken natural language output accompanied by real time coordinated pointing gestures and graphic expressions.

The next section of this paper presents an overview of the CUBRICON design and discusses related research. Section 3 discusses the CUBRICON multi-modal input understanding process. Section 4 discusses the CUBRICON map composition and transformation process. Section 5 discusses the CUBRICON multi-modal output presentation processes. Section 6 discusses future directions for this research and Section 7 presents a summary of the paper.

2 SYSTEM OVERVIEW

The CUBRICON design provides for the use of a unified multi-media language, by both the user and system, for communication in a dialogue setting. 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).

An overview of the CUBRICON software system and hardware I/O devices is presented in Figure 1. CUBRICON accepts input from three input devices: speech input device, keyboard, and mouse device. CUBRICON produces output for three output devices: high-resolution color-graphics display, high-resolution monochrome display, and speech production device. The primary path that the input data follows is indicated by the modules that are numbered in the figure: (1) Input Coordinator, (2) Multi-Media Parser-Interpreter, (3) Executor-Communicator to Target System, (4) Multi-Media Output Planner, and (5) the 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 a generalized augmented

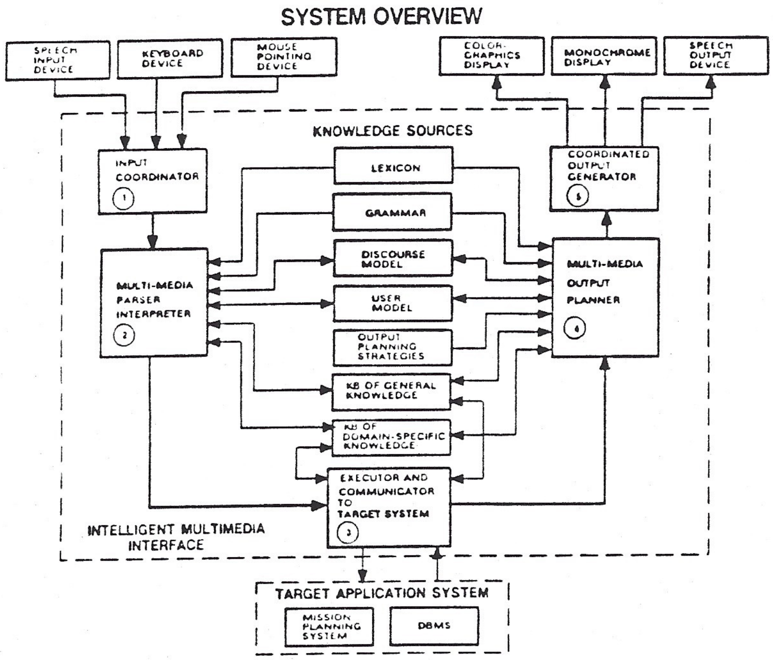

Figure 1: System Overview

transition network (GATN) that has been extended to accept the compound stream produced by the Input Coordinator and produce an interpretation of this compound stream. Appropriate action is then taken by the Executor module. This action 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 uses a GATN that produces a multi-media output stream representation with components targeted for the different output devices. This output representation is translated into visual/auditory output by the Output Generator module. This module is responsible for producing the multi-media output in a coordinated manner in real time (e.g., the 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 natural language output).

The CUBRICON system includes several knowledge sources that are used for both understanding input and composing output. The knowledge sources include: a lexicon, a grammar defining the multi-modal language used by the system for input and output, a discourse model, a user model, and a knowledge base of task domain and interface information. The latter knowledge sources are discussed briefly in the following paragraphs.

The *knowledge base* consists of information about the task domain of tactical Air Force mission planning. This knowledge base includes information about concepts such as SAMs, air bases, radars, and missions as well as related HCI concepts such as verbal/graphical expressions for the domain concepts.

The *discourse model* is a representation of the attentional focus space [Grosz86] of the dialogue carried out in multi-modal language. It consists of (1) a main focus list that includes those entities and propositions that have been explicitly expressed (by the user or by CUBRICON) via natural language and/or graphic/pointing gestures and (2) a display model that includes a representation of all the objects (windows and their contents) that are "in focus" because they are visible on one of the two CRT screens.

The *user model* [Kobsa88] consists of an *entity rating module* that includes a task-dependent representation of the relative importance of all the entity types known to the system and an algorithm for modifying these ratings depending on task and dialogue activity.

Key features of the CUBRICON design, discussed in this paper, include (1) the integration of NL and graphics in a unified language that is defined by a multi-modal grammar, (2) the generation of synchronized speech and graphics in real time, and (3) the relevance-based map composition/transformation process. The integration of NL and graphics in a unified language distinguishes this research from other approaches to multi-modal interface technology [Sullivan88, Arens89]. The Integrated Interface system [Arens88] and the XTRA system [Kobsa86, Allgayer89] are two of the most relevant. The Integrated Interface system is a multi-modal system in that it uses both maps and NL for the presentation of information to the user. The system provides information about the status and movements of naval platforms and groups in the Pacific Ocean. The system displays NL in text boxes positioned on a map display near the relevant objects. The system does not use a multi-modal language, however. The language generated is purely NL with no integrated graphics. The XTRA system is a multi-modal interface system which accepts and generates NL with accompanying point gestures for input and output, respectively. In contrast to the XTRA system, however, CUBRICON supports a greater number of different types of pointing gestures, greater variety in the types of objects that can be targets of pointing gestures, and greater flexibility in the frequency and coordination of pointing gestures relative to the simultaneous corresponding NL. CUBRICON also includes graphic gestures (i.e., certain types of simple drawing) as part of its multi-modal language, in addition to pointing gestures. Furthermore, CUBRICON addresses the problem of coordinating NL (speech) and graphic gestures during both input and output.

CUBRICON software is implemented on a Symbolics Lisp Machine using the SNePS semantic network processing system [Shapiro79, Shapiro86], a GATN parser-generator, and Common Lisp. Speech recognition is handled by a Dragon Systems VoiceScribe 1000. Speech output is produced by a DECtalk speech production system.

The following sections discuss CUBRICON's input understanding and output composition processes and their use of the knowledge sources discussed above.

3 MULTI-MODAL LANGUAGE UNDERSTANDING

People commonly and naturally use coordinated simultaneous natural language and graphic gestures when working at geographic displays. These modes of communication combine synergistically to form an efficient language for expressing definite references and locative adverbials. One of the benefits of this multi-modal language is that it eliminates the need for the lengthy definite descriptions that would be necessary for unnamed objects if only natural language were used. Instead, a terse reference such as "this SAM" (surface-to-air missile system) accompanied by a point to an entity on the display can be used. CUBRICON accepts such NL accompanied by simultaneous coordinated pointing gestures. The NL can be input via the keyboard, the speech recognition system, or a mixture of both. CUBRICON has the flexibility to accept zero or more pointing gestures per noun phrase or locative adverbial and such a phrase may contain no words, just the pointing gestures.

Just as natural language used alone has shortcomings, so also does the use of pointing gestures alone. Pointing used alone has the following problems: (1) a point gesture can be ambiguous if the point touches the area where two or more graphical figures or icons overlap or (2) the user may inadvertently miss the object at which he intended to point. To handle these pointing problems, some systems use default techniques such as having a point handler return the entity represented by (a) 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) or (b) the icon whose "center" is closest to the location on the screen/window touched by the point. A serious disadvantage of such default point-interpretation techniques is that it is difficult, if not impossible, for certain icons to be selected via a point reference.

CUBRICON's acceptance of dual-media input (NL accompanied by coordinated pointing gestures) overcomes the limitations of the above weak default techniques and provides an efficient expressive referencing capability. The CUBRICON methodology for handling dual-media input is a decision-making process that depends on a variety of factors such as the *types* of candidate objects being referenced, their *properties*, the *sentential context*, and the *constraints on the participants* or fillers of the semantic case frame for the verb of any given sentence. CUBRICON's decision-making process draws upon it's knowledge sources discussed briefly in Section 3.

We present a few brief examples to illustrate CUBRICON's referent determination process. This process handles the problems listed above: ambiguous point gestures and point gestures that are inconsistent with the accompanying NL. First we discuss ambiguous point gestures. In each of the following examples, assume that the <point> represents a point gesture with a device such as a mouse and each point gesture can be ambiguous (i.e., it can touch more than one icon).

Example 1: USER: "What is the status of this <point> airbase?"

From the icons touched by the point, the display model is searched for the semantic representation of the objects which were graphically represented by the touched icons. From the hierarchy of the knowledge base, the system determines which of the objects selected by the point gesture are of the type mentioned in the accompanying verbal phrase ("airbase" in the example sentence) and discards the others.

Example 2: USER: "What is the mobility of these <point>$_1$ <point>$_2$ <point>$_3$?"

Example 2 illustrates that CUBRICON enables the user to use more than one point gesture per phrase. In contrast to Example 1, no object type is mentioned in the noun phrase corresponding to the point gestures. In this case, CUBRICON can use a mentioned property (e.g., mobility) to select from amoung the candidate referents of the point gesture. CUBRICON accesses the display model to retrieve the semantic representations of the objects touched by each of the user's point gestures, and then determines which of these objects have property "mobility" using the knowledge base of application information.

Example 3: USER: "Enter this <point-map-icon> here <point-form-slot>."

Example 3 illustrates that CUBRICON enables the user to use point gestures in conjunction with more than just one phrase of a sentence and that the point gestures may access different types of windows, even on different CRTs. In Example 3, the user's first point gesture touches an object on a map display on the color-graphics CRT and the second selects a slot of the mission planning form on the monochrome CRT. Two of CUBRICON's features are critical to its ability to process the sentence of Example 3: First, the display model contains semantic representations of all the objects displayed visually in each of the windows of each CRT, and second, all objects and concepts in the CUBRICON system are represented in a single knowledge representation language, namely the formalism of the SNePS knowledge base. This knowledge base is shared by all the modules of the CUBRICON system. Suppose that the <point-map-icon> selects the Nuernberg airbase on the map and the <point-form-slot> touches the "origin airbase" slot on the mission planning form. CUBRICON's response to the input of Example 3 would be to build the knowledge base structure which represents the assertion that Nuernberg is the airbase from which the particular mission will be flown.

As mentioned previously in this section, in addition to being ambiguous, another problem that can arise with point gestures is that the user may inadvertently miss the object at which he intended to point. In this case, the point gesture will be inconsistent with the accompanying natural language phrase, meaning that the natural language part of the expression and the accompanying point cannot be interpreted as referring to the same object(s) (e.g., the user says "this airbase" and points to a factory or points at nothing, missing all the icons). CUBRICON includes methodology to infer the intended referent in this case. CUBRICON uses the information from the sentence, parsed and interpreted thus far, as filtering criteria for candidate objects. The system performs a bounded incremental search around the location of the user's point to find the closest object(s) that satisfy the filtering criteria. If one is found, then the system responds to the user's input (e.g., command or request) and also issues an advisory statement concerning the inconsistency. In the event that no qualified object is found in the vicinity of the user's point, then a response is made to the user to this effect.

4 KNOWLEDGE-BASED MAP COMPOSITION/TRANSFORMATION

CUBRICON uses the following modalities for presenting information and responses to the user: color geographic maps, tables, "fill in the blank" forms, spoken natural language, printed natural language, text-boxes, and pointing gestures. When information is to be presented (expressed) to the user as output, CUBRICON selects the appropriate combination of modalities for presenting the information and then composes and generates the output to the user. The decision-making process for selecting the appropriate modalities for the given information is discussed in more detail in [Neal88a] and [Neal88b].

The design of CUBRICON has been based on the premise that users should concentrate on their task domain activities. They should *not* need to waste their mental or temporal resources on control and management of the interface, particularly windows and displays. Such control and management should be handled by the interface system in an intelligent manner. The interface system should present relevant information appropriately so as to eliminate the need for the user to manipulate or control the interface. With regard to geographic map displays, an important aspect of CUBRICON's processing capability is its decision-making logic for deciding when and how to create and transform map displays.

Relevancy is a critical factor for an HCI to consider when determining what information to present and how to present it to the user. In CUBRICON's map manipulation process, relevancy plays an important role (1) in determining what type of map transformation to perform and (2) in determining what objects to display on a map.

One aspect of relevancy in the CUBRICON system pertains to the user's task. CUBRICON keeps track of the task that the user is working on and registers the user's transition from one task to another. CUBRICON's process for deciding on an appropriate map transformation takes into consideration whether or not the user's task has just changed. If the user's task has not changed, then CUBRICON assumes that the current main map is still relevant and tries to keep it in the same window, subject to some expansion or contraction. If the user's task has just changed, however, then CUBRICON will move the map from the main window to a secondary window and "repaint" the main window with a new appropriate map area.

Relevancy is also important in selecting the objects to display in a map region. Frequently, sophisticated application systems include one or more massive databases and, indeed, the databases may be shared by more than one application system. When a system such as CUBRICON selects objects from the database for display on a map, it should be discriminating in its selection. Not all the available objects should be selected from the database for display, since this could result in an unnecessarily cluttered and confusing map. Instead, only the relevant objects should be displayed. The technique used in the CUBRICON system to accomplish this relies on the use of the *entity rating system* of the user model (discussed in Section 3). When composing maps, CUBRICON displays only those objects above the critical importance threshold for the user's current task. Thus, for an Offensive Counter Air (OCA) planning task, CUBRICON would display all airbases, SAM sites, critical factories and plants, but not objects such as schools or minor industry.

The operations on geographic maps are listed below. In general, CUBRICON's decision-making process has been designed with the goal of maintaining context for the user and helping the user understand the transition from one map to another. For each of the map transformations, CUBRICON presents the new map in the context of the previously displayed map. In communicating the map transitions, CUBRICON uses a "region boundary box" to outline or highlight a region that is a sub-region of another. Objects to be displayed on each new map are selected according to their importance to the user's task, as discussed above.

- Create: a new map is created and displayed in a window on the color-graphics screen.

- Zoom out: the area shown in a map window is extended to include appropriate additional area of interest to the user. A "region boundary box" is superimposed on the new map to show the boundary of the map that was previously displayed. This helps the user understand the transformation from previous map to new map display.

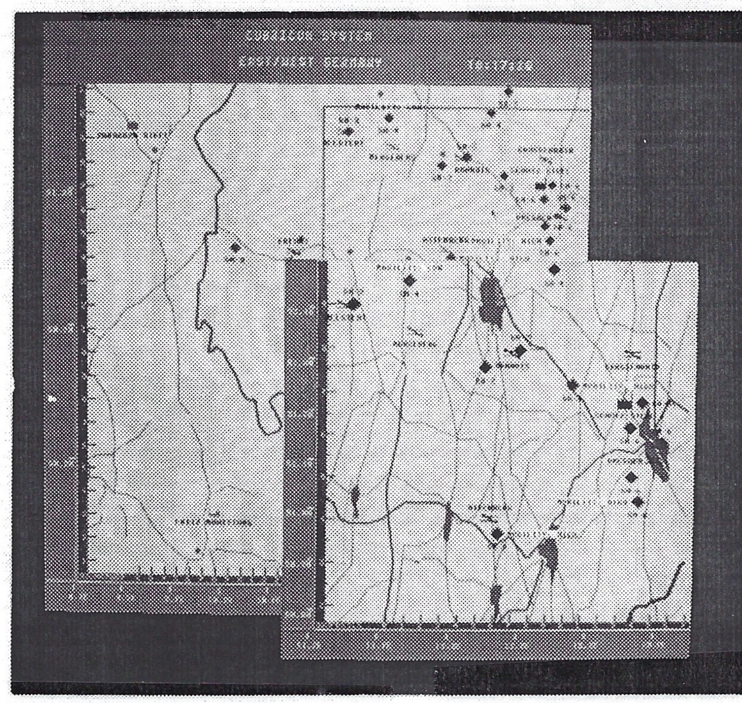

Figure 2: Color-Graphics Screen After Minor Zoom In Operation

- Major zoom in: a sub-region (specified by either the user or the system) of the current main map is enlarged. The map transition is performed as follows: CUBRICON first moves the map currently displayed in the main window to a secondary window and adds a "region boundary box" to this secondary window showing the sub-region that is to be enlarged. The enlarged version of the sub-region is then displayed in the main window.

- Minor zoom in: a sub-region (specified by either the user or the system) of the current main map is enlarged. The map transition is performed as follows: CUBRICON superimposes a "region boundary box" on the main map showing the outline of the region to be enlarged. An enlarged version of the designated region is then displayed in a secondary window of appropriate size. Figure 2 shows the CUBRICON color-graphics screen after a minor zoom-in operation has been performed.

- Pan to a new region. The map transition is performed as follows: in one of the secondary windows, CUBRICON displays a map region whose boundary is the smallest rectangle enclosing the old map in the main window and the new region to be displayed; CUBRICON then shows region boundary boxes designating (1) the old region that was in the main window and (2) the new region to be displayed. The new region is displayed in the main map window.

Not all the information about objects on a map can be displayed graphically. Color and other graphical attributes of visual objects can be used to convey information (e.g., red for enemy and blue for friendly objects in a military application), but the amount of information that can be represented at any one time using graphical attributes is very limited. In order to accomodate additional relevant information, CUBRICON will compose and display a table of such information along with the corresponding map display. Selection of relevant attributes is based on the user model (discussed in Section 3) which includes a representation of the relevant attributes per object type.

For output presentation, CUBRICON is not limited to the use of maps and tables as discussed above. In addition, CUBRICON can present output using spoken natural language, printed natural language, text boxes, graphic drawing, pointing gestures, and combinations thereof. The next section focuses on CUBRICON's multi-modal language generation.

5 MULTI-MODAL LANGUAGE GENERATION

Just as CUBRICON accepts NL accompanied by deictic and graphic gestures during input, CUBRICON can generate multi-modal language output that combines NL with *deictic gestures* and *graphic expressions.* An important feature of the CUBRICON design is that NL and graphics are incorporated in a single language generator providing a unified multi-modal language with speech and graphics synchronized in real time. CUBRICON's use of deictic gestures and graphic expressions are discussed in the following paragraphs.

Deictic gestures are combined with appropriate NL during output to guide the user's visual focus of attention. CUBRICON uses deictic gestures with speech only, since a pointing gesture needs to be temporally synchronized with the corresponding verbal phrase. The coordination between the point gesture and co-referring verbal phrase is lost if printed text is used instead of speech. CUBRICON also displays a printed version of the response in a window on the CRT, but the deictic expressions used in the speech/graphics version are replaced with definite descriptions without point gestures.

During language generation, in order to compose a reference for an object,

1. if the object is represented by an icon on the display, then CUBRICON generates a NL expression for the object and a simultaneous coordinated graphic gesture that points to its icon.

   If the object has an individual name or identifier, then CUBRICON uses its name or identifier (e.g., "the Merseberg airbase") as the NL expression

   else CUBRICON generates an expression consisting of a demonstrative pronoun followed by the name of an appropriate class to which the object belongs (e.g., "this SAM", "these SAMs") as the NL expression.

2. if the object (call it X) is not represented by an icon on the display, but is a component of such a visible object (call it Y), then CUBRICON generates a phrase that expresses object X as a component of object Y and uses a combined deictic-verbal expression for object Y as described in the above case. For

example, if CUBRICON is generating a reference for the runway of an airbase called Merseberg and an icon for the airbase is visible on the map (the airbase as a whole is represented visibly, but not its parts), then CUBRICON generates the phrase "the runway of the Merseberg Airbase" with a simultaneous point gesture that is directed at the Merseberg airbase icon on the map.

CUBRICON combines *graphic expressions* with NL output when the information to be expressed is, at least partially, amenable to graphic presentation. In the current CUBRICON implementation, the type of information that falls in this category includes (1) locative information and (2) path traversal information. Due to the space limitations of this paper, we briefly discuss only the locative case.

When generating locative information about some object (call it the figure object [Herskovits85]), CUBRICON selects an appropriate landmark as the ground object [Herskovits85], determines a spatial relationship between the figure and ground object, and generates a multi-modal expression for the locative information including the spatial relationship. When selecting the ground object, CUBRICON selects a landmark such as a city, border, or region, that is within the current map display (i.e., does not require a map transformation). If possible, CUBRICON uses a landmark that is in focus by virtue of its having been already used recently as a ground object. If a new landmark must be used as a ground object, then CUBRICON selects the landmark that is nearest the figure object. CUBRICON derives a spatial relation between the ground object and figure object that it represents in its knowledge base. This relation includes (1) the direction from the ground object to the figure object and (2) the distance if the distance is greater than 0.04 of the window width. If the distance is less than 0.04 of the window width, then the figure object appears to be right next to the ground object. This criterion for deciding whether to include distance as part of the relation reflects the tendency for people to omit a distance measure when the distance is small relative to the geographic area under discussion and to say something like "just northeast of" instead of stating a distance explicitly.

As an illustrative example, the user may ask about the location of a particular object, such as the Fritz Steel plant. The system then uses the steel plant as the figure object, selects a ground object, and derives a spatial relation between ground object and figure object as discussed above. The multi-modal response is given below.

USER: "Where is the Fritz Steel plant?"

CUBRICON: "The Fritz Steel plant is located here <point>, 45 miles southwest of Dresden <graphic-expression>."

The <point> consists of a gesture that points out the Fritz Steel plant icon to the user via a gesture that uses a combination of blinking, highlighting, circling the icon and the attachment of a pointing label-box that identifies the icon. The <graphic-expression> is a visual presentation of the spatial relation between the figure object (Fritz steel plant) and the ground object (Dresden city). The graphic expression consists of an arrow drawn from the Dresden city icon to the steel plant icon, a label stating the distance, and a label identifying the city (the steel plant should already be labeled).

CUBRICON's multi-modal language generation is also discussed in [Neal89].

6 FUTURE DIRECTIONS

There are numerous worthwhile areas and ideas to be investigated and developed to advance this research. We briefly discuss two of these areas:

CUBRICON is currently being extended so that it accepts a larger vocabulary of graphic drawing gestures as part of the user's multi-modal input. An integrated language consisting of both verbal and graphic "tokens" can be used for both referencing objects that the system already knows about as well as explaining and defining new concepts to the system. Such a multi-modal input language should be especially useful for the definition and explanation of geographical and spatial concepts to a system that would then use the concepts for geographical applications. We are currently focusing on adding polylines to the set of graphic gestures that CUBRICON accepts. Polylines can be used to approximate free-hand drawing and thereby give the user great expressive power.

We are also planning to conduct a research program to investigate the problem of user gestures that are not synchronized with their corresponding NL phrases. We are interested in the characteristics of the phenomenon: to what degree are gestures of different types out of sync, how frequently does the phenomenon occur, is there a correlation between characteristics of the phenomenon and characteristics of the corresponding natural language? We also plan to investigate methods that would enable the system to decide which phrase of the accompanying natural language input is the co-referring phrase for any pointing gesture that is not synchronized with its co-referring phrase.

7 SUMMARY

The CUBRICON project is devoted to the development of knowledge-based interface technology that integrates speech input, speech output, natural language text, geographic maps, graphics, and pointing gestures for interactive dialogues between human and computer. The objective is to provide both the user and system with modes of expression that are combined and used in a natural and efficient manner, particularly when presenting or requesting information about objects that are visible, or can be presented visibly, on a graphics display.

The CUBRICON design provides for the use of a unified multi-media language, by both the user and system, for communication in a dialogue setting. CUBRICON's main I/O processing modules have access to several knowledge sources or data structures, including one modeling each of (1) the application domain, (2) the discourse, and (3) the user.

CUBRICON provides unique interface capabilities including the ability to: (1) accept and understand multi-media input such that references to entities in (spoken or typed) natural language sentences can be accompanied by coordinated simultaneous pointing to the respective entities on a graphics display; (2) dynamically compose geographic map displays including the determination of the boundary of the region to be displayed, selection of the relevant entities to display in the region, and selection of the location and size of the window in which to place the map display; (3) dynamically compose and generate multi-modal language that combines NL with deictic gestures and graphic expressions; synchronously present the spoken natural language and coordinated pointing gestures and graphic expressions.

8 REFERENCES

[**Allgayer89**] Allgayer, J., Jansen-Winkeln, R., Reddig, C., & Reithinger, N. 1989. Bidirectional Use of Knowledge in the Multi-Modal NL Access System XTRA. *Proc. of IJCAI-89*, Detroit, MI, pp. 1492-1497.

[**Arens88**] Arens, Y., Miller, L., & Sondheimer, N.K. 1988. Presentation Planning Using an Integrated Knowledge Base, in *Architectures for Intelligent Interfaces: Elements and Prototypes*, J.W. Sullivan & S.W. Tyler (eds.), Addison-Wesley, pp. 93-108.

[**Arens89**] Arens, Y., Feiner, S., Hollan, J., & Neches, R. (eds.) 1989. *A New Generation of Intelligent Interfaces, IJCAI-89 Workshop*, Detroit, MI.

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

[**Haller89**] Haller, S.M. 1989. Technical Report: Spatial Relations and Locative Phrase Generation in a Map Context. Computer Science Department, State University of New York at Buffalo.

[**Herskovits85**] Herskovits, A. 1985. Semantics and Pragmatics of Locative Expressions. *Cognitive Science*, 9:341-378.

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

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

[**Neal88a**] Neal, J.G. & Shapiro, S.C. 1988. Intelligent Multi-Media Interface Technology, in *Architectures for Intelligent Interfaces: Elements and Prototypes*, J.W. Sullivan & S.W. Tyler (eds.), Addison-Wesley, pp. 69-91.

[**Neal88b**] Neal, J.G., Dobes, Z., Bettinger, K.E., & Byoun, J.S. 1988. Multi-Modal References in Human-Computer Dialogue, *Proc. AAAI-88*, St. Paul, MN, pp. 819-823.

[**Neal89**] Neal, J.G., Thielman, C.Y., Funke, D.J., & Byoun, J.S. 1989. Multi-Modal Output Composition for Human-Computer Dialogues. *Proc. of the AI Systems in Government Conference*, George Washington Univ., Wash. D.C., pp. 250-257.

[**Shapiro79**] Shapiro, S.C. 1979. The SNePS Semantic Network Processing System, in *Associative Networks - The Representation and Use of Knowledge by Computers*, N. Findler (ed.), Academic Press, pp. 179-203.

[**Shapiro86**] Shapiro, S.C. & Rapaport, W.J. 1986. SNePS Considered as a Fully Intensional Propositional Semantic Network, in *The Knowledge Frontier, Essays in the Representation of Knowledge*, N. Cercone & G. McCalla (eds.), Springer-Verlag, pp. 263-315.

[**Sullivan88**] Sullivan, J.W. & Sherman, W.T. (eds.) 1988. *Architectures for Intelligent Interfaces: Elements and Prototypes*, Addison-Wesley Pub. Co.