118

In the Proceedings of the 1989 IEEE AI Systems Conference, George Washington University, Washington, DC, March 1989

Multi-Modal Output Composition for Human-Computer Dialogues

Jeannette G. Neal, Carol Y. Thielman, Douglas J. Funke, Jung S. Byoun

Calspan-UB Research Center (CUBRC)
4455 Genesee Street
Buffalo, NY 14225
(716) 632-7500

Abstract

A knowledge-based human-computer interface system that could minimize the effort required for users to manage and direct the interface, would offer great potential for increasing the user's efficiency in accomplishing his or her task. Several critical features are essential to such a human-computer interface system. These features are the ability to: (1) judge relevance and respond in a context sensitive manner, (2) intelligently select the presentation media and format for best information understanding, (3) use available media and presentation formats in a highly integrated manner, and (4) provide consistency across all system outputs.

The Intelligent Multi-Media Interface Project has been initiated for the purpose of developing knowledge-based approaches to incorporating these features in a multi-media human-computer interface [11]. Specifically, this project is devoted to the development of interface technology that integrates speech input and output, natural language text, graphics, and pointing gestures for human-computer dialogues in a flexible, context-sensitive manner. As part of the Intelligent Multi-Media Interface Project, a knowledge-based interface system called CUBRICON (the CUBRC Intelligent CONversationalist), is being developed as a proof-of-concept prototype. The output planning portion of this prototype system is discussed in this paper.

1. Introduction

The efficiency with which modern, information intensive systems can be used, depends on the human-computer interface (HCI). In fact, the HCI is frequently the most significant factor in limiting the potential of such systems, especially those that must be used under stressful conditions. Modern windowing approaches to interface systems have helped to achieve these goals. However it has been found that a large determinant of the time it takes to solve a problem using windowing systems, is the time spent manipulating the windows themselves [2]. It would seem that a knowledge-based interface system that could minimize the requirement for direct, manual user control of the interface, would offer great potential for increasing the user's efficiency in accomplishing his or her task. Several critical features are essential to such a human-computer interface system. These features are the ability to:

1. Generate output in a context-sensitive manner, based on relevance of the information to the discourse focus and the user task,
2. Select appropriate combinations of modalities and formats for the expression of information, based on the characteristics of the information.
3. Use available multiple media and modalities in a highly integrated manner, and
4. Provide outputs that are consistent across individual and related sequences of system expressions and displays.

Approaches to achieving a human-computer interface that incorporates these critical features are not obvious. Modern systems must convey large amounts of information. This information is often technically complex, dynamically changing, and varied in degree of certainty and completeness. In addition, the humans using the system may be under severe time constraints, and under high stress. Presently, there is no computer system that provides the critical features discussed above. Knowledge-based generation of information by a computer system in multiple media/modalities is recently beginning to be investigated [1,11,12,18]. An interdisciplinary approach which combines modern computer systems and programming approaches, with relevant human factors disciplines is essential.

The Intelligent Multi-Media Interface Project [11] represents a strong step in this direction. Computer scientists are working closely with experts in human-computer interface design, to apply modern artificial intelligence (AI) technology to this problem. This project is devoted to development of interface technology that integrates speech input and output, natural language text, graphics, and pointing gestures in a flexible, highly integrated, context-sensitive manner for human-computer dialogues. The approach is to develop knowledge-based interface technology that presents information in an optimal (or near optimal) fashion, without requiring the human user to engage in activities specifically and directly relating to the control of the computer interface. The determination of the "optimal" presentation is based on human-computer interface design literature (e.g., [8,17]) and, where relevant literature is not available, engineering judgement.

There are obviously major problems in achieving a completely "intelligent" interface: First, we don't completely understand human intelligence. Many theories and models describe how humans organize information and make decisions, but one comprehensive description of the overall process does not exist. Our ability to model such processes is respectively difficult. Second, there is a great deal of

*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.

variance among people and applications. What may be "optimal" for one person, population, or application, is likely to be suboptimal for others. And finally, the experience and research base for designing human-computer interfaces has largely been restricted to systems and applications that do not involve the integration of multi-media or the application of artificial intelligence.

An important goal of the Intelligent Multi-Media Interface Project is to build a prototype intelligent, multi-media human-computer interface system called CUBRICON (the CUBRC Intelligent CONversationalist). It is intended to serve as a proof-of-concept prototype, and a research platform for investigating advanced technology multi-media human-computer interface design. The CUBRICON system 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, text, pointing motions, and drawings). It is planned that through a process of iterative design and research, CUBRICON will evolve into an intelligent human-computer interface that can serve as a front-end to numerous application systems. The current application task domain is that of military tactical air control.

This paper discusses the CUBRICON multi-modal output planning process and the manner in which CUBRICON exhibits the key features listed above. In the next section, we present an overview of the CUBRICON multi-media output capabilities. Section 3 discusses the knowledge sources used for output composition. Section 4 discusses the output planning process, with working example dialogues presented in Section 5 to clarify the methodology and underlying principles. Section 6 briefly presents the status and the future direction for the CUBRICON project and Section 7 summarizes the main ideas of this paper.

2. Overview of CUBRICON's Multi-Media Output Capabilities and Approach

The CUBRICON design includes the following output *media*: speech production device, color-graphics display, and monochrome-graphics display. The *modalities* used by the CUBRICON system include: color-graphic illustrations, tables, histograms, written natural language prose, spoken natural language, and "fill in the blank" forms. This list does not exhaust the possibilities, of course, but provides a good varied selection with which to "test the concept" and upon which to build. These media and modalities are used by CUBRICON in isolation or in coordinated combinations.

In comparison to a single medium language, a multi-modal language provides a wider range of choices for the expression of any given concept. When modalities are combined, a synergism not possible with just one medium is achievable. Even in a strictly verbal language, there are alternatives for expressing a given concept. 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 U.S." If language is extended to include graphics as well, the number of combinations and alternatives increases. For example, to express the locations of several airbases, the choices include: a map of the area showing the individual airbase locations, a table listing the longitude an latitude of all the airbases, or a natural language summary such as "they are all located within 100 miles of Berlin".

The CUBRICON output planning process starts with the selection of the media/modalities for expressing the response information to the user. In making this decision, CUBRICON must consider whether the desired output resources are available from among the potential resources (the two displays and the speech device). If they are not, then the system must take appropriate action to modify the state of the resources, modify the information to be expressed, and/or select different modalities for expressing the information before the composition of the output can be accomplished. In performing this planning process, several knowledge sources are essential. These knowledge sources are discussed in the next section.

3. Knowledge Sources

The knowledge sources within CUBRICON are critical to the process of composing multi-modal output. These are: the content and characteristics of the information to be expressed, discourse model, user/task model, and a knowledge base of information about the application task domain which includes relevant interface information. These knowledge sources are discussed in the following subsections.

3.1 Preferred Presentation Modes Based on Characteristics Of The Information

The characteristics of the information to be expressed are critical to the selection of an appropriate presentation modality. The following list summarizes CUBRICON's criteria for selecting presentation modality based on characteristics of the information:

1. Color-graphics: Selected whenever the information to be presented includes spatial relationships (e.g., maps, schematics) or whenever the information is ammenable to spatial coding (e.g., small hierarchy charts, organization charts). CUBRICON currently considers only spatial relationships. This is accomplished by looking for regional or coordinate references and the existence of iconic symbols associated with information to be presented.
2. Monochrome-graphics: Selected using the same criteria as for color graphics except that when feasible, color graphics is always preferred.
3. Table: Selected when the values of common attributes of several entities must be expressed and a graphic approach is not feasible (e.g., too many attributes to fit on a histogram or the attributes are not quantitative).
4. Histogram: Selected when one or a few quantitative and common attributes of several entities must be displayed in a comparative fashion.
5. Form: A predefined form is selected when the task requires a standardized organization of information in accordance with external needs of the user. For example, when information must be presented in a particular format to facilitate its use within the problem solving realm or process, or to enhance recognition and/or understanding of patterns within the data. An important consideration in the design and requirement for such forms is the existance of population stereotypes among the user population. This aspect of CUBRICON is currently under development.
6. Natural language prose: Selected for the expression of a proposition, relation, event, or combination thereof when the knowledge structures being expressed are heterogeneous, or are peripheral or supplementary to the ongoing dialogue.

Natural language can be presented in either spoken or written form. The following summarizes the selection critieria for spoken versus written language:

Spoken natural language is selected for:

1. Dialogue descriptions to assist the user in comprehending the presented information. These include explanations of graphic displays and display changes as well as verbal highlighting of objects on the displays (e.g., "The enemy airbases are highlighted in red").
2. Warnings to alert the user of important events that have or are about to take place (e.g., new critical information comes into the application system database and the system notifies the user: "close air support is no longer available from airbase XXX").
3. Informing the user about the system's activity (e.g., "I'm still working" when the user must wait for output from the system).
4. Short expressions of relatively non-technical information that can be remembered when presented serially (e.g., a "yes" or "no" answer to user's question).

Written natural language is selected for responses that would strain the user's short term memory if speech were used. This includes longer technical responses and responses to which users may need to refer over a substantial period of time. G.A. Miller [10] provides us with the guideline that short term memory can only hold approximately seven elements.

3.2 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 [4,6,7,16] 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: (1) a main focus list and (2) 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. CUBRICON treats objects presented on the graphics displays as having been intentionally "expressed" or "mentioned." The display model has two levels: (1) a list of windows per monitor and (2) for each window, a list of all the objects that are visible in the given window.

The main focus list primarily serves the process of interpreting anaphoric references [16] and definite descriptive references [5] input via natural language. For example, CUBRICON consults the main focus list when determining the referent of a pronoun. In the case of a definite reference, if an appropriate referent is not found in the main focus list, then 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 knows about several others.

The discourse model is used during output generation also. When CUBRICON composes a reference for an entity as part of a natural language sentence, it consults the discourse model. If the entity is represented in the display model (i.e., it is 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 simultaneous blinking/highlighting of the airbase as its 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.

3.3 The User/Task Model

A user/task model is essential as a basis for judging the relevance and importance of information items to be presented. Carberry [3] provides a brief summary of current research on user modeling. The aspects of the user/task that are most relevant in the CUBRICON system are (1) the importance rating that the user attaches to the different entity types that are relevant to each given task, which we call the user's *entity rating system* and (2) a task hierarchy (if available).

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 zero to one) assigned to each entity type 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 for that task. Associated with the entity rating system is a *critical threshold* value: Those entities with a rating above the critical threshold are considered 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; and (2) when the user mentions an entity whose rating is lower than the critical threshold, then its rating is raised to the critical threshold to reflect the user's interest in the entity and its seeming relevance to the current task from the perspective of the user. It remains at this level until the user's task changes. In the current implementation, CUBRICON performs the second function listed above, 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. (1) The entity rating system representation is used in determining what information is relevant in answering questions or responding to commands from the user. (2) The entity rating 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 might otherwise be made. (3) The entity rating system is also used in organizing the form in which information is presented.

The task hierarchy is a decomposition of the user's main task into subtasks. In a task oriented application system, there is usually some a priori knowledge of the task hierarchy and sequencing. Even though the task hierarchy structure is not absolute in that a user may deviate from the typical roadmap through the tasks, this hierarchy can be used as a valuable knowledge source in tracking the discourse focus and in managing the displays. If the system can anticipate, for example, when a task change will occur and knows the relationship between the tasks, then the system can respond more appropriately, retain relevant information on the displays, and avoid unnecessary disruptive display changes.

The task hierarchy is in development and is not complete. Currently, CUBRICON does include a simple representation of the task in which the user is engaged. This task representation changes when CUBRICON is told that a new task has started. CUBRICON's response composition process is affected by the user's task.

3.4 Knowledge Base: Task Domain And Interface Information

The CUBRICON system includes a knowledge base containing domain-specific and interface information. The domain-specific information is applicable to the mission planning task domain of the target information system using the CUBRICON interface system. The knowledge base contains information about task domain entities such as airbases, missile systems, critical plants and factories, and munitions. The knowledge base also includes essential information concerning how to present or express the various entities via the system's verbal/graphic language. This information includes the words and symbols used to express any given entity, which symbols are appropriate under which conditions, and when particular colors are to be used.

4. Planning The Multi-media Response

The CUBRICON system is being developed so that it embodies the key features discussed at the beginning of this paper. These features are the ability to:

1. Generate output in a context-sensitive manner, based on relevance of the information to the discourse focus and the user task,
2. Select appropriate combinations of modalities and formats for the expression of information, based on the characteristics of the information,
3. Use available multiple media and modalities in a highly integrated manner, and,
4. Provide outputs that are consistent across individual and related sequences of system expressions and displays.

The CUBRICON output planning process is highly dependent on the knowledge sources discussed in Section 3. The input to the CUBRICON output planning module is a list of information items to be expressed. The top level output planning process is summarized below.

1. For each information item or cluster, determine the modality in which it should ideally be expressed. Graphic/pictorial presentation is always desirable. Natural language can always be used as a last resort, if no other modality is available. Section 3.1 discussed the current modalities used by CUBRICON and the criteria for selection of each.
2. Determine whether the resources are available to express the information as desired. Resources: (1) Color graphics display: Are the items to be expressed graphically already on the color display (e.g., objects of interest in a geographical domain may already be displayed on a map)? If so, no additions are necessary. If not, is there room to add them in their "natural" position? (e.g., can the desired objects be inserted in the area already on the display without changing the area shown?) (2) Monochrome display: Similar to the color graphics display. (3) Speech output device: Always available.
3. If the desired resources (i.e., color and/or monochrome display) are not available, modify the state of the resources. The desired resources would be "not available" in at least two cases: (1) the physical device is not functional (e.g., needs repair) or (2) the display device already contains critical information that cannot be disrupted nor covered by a window. If not all the items to be expressed graphically are on the display and it is possible to change the display, then the system must compose a new display. Possibilities:
   - "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.
   - Combination of the above.
   - Create a new display: (i) Completely replace the display with new "area" or (ii) Open a window on the display to show new information.
4. If the display status cannot be adequately modified as per step 3 above, try modifying the information to be expressed: trim the amount of information by filtering on the basis of relevance with regard to the user/task model and/ or the discourse model.
5. If the information can still not be expressed in the given modality due to insufficient resources for the selected modality, then select another modality and go back to step 2.
6. Finish composing the output having resolved resource restraints.

The methodology discussed in this section is fully implemented.

5. EXAMPLES

In this section we present examples of output composed and generated by CUBRICON in the context of a human-computer dialogue. These examples illustrate the output composition process and use of the knowledge sources discussed in the previous sections. The dialogues are concerned with mission planning and situation assessment in a tactical air control domain. These examples are actual working examples of dialogue with the current CUBRICON implementation.

Consider the following user-computer dialogue interchanges:

USER: "Display the Fulda Gap Region."

CUBRICON: (Refer to Figure 1.)

Speech Output:

- Statement 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."

Color Graphics Display:

- Map of Fulda Gap Region with main roads, major cities, waterways, and national boundries.
- Icons representing entities within the Inner Fulda Gap Region, that are above a preset threshold in importance, superimposed on the map.

Speech Output:

- Just before the table is presented on the monochrome monitor: "The corresponding table is being presented on the monochrome screen."

Monochrome Graphics Display:

- Table of relevant entity attributes for those entities that are displayed on the map display.

DISCUSSION:

The planning and composition of output for the user is dependent upon the nature of the information, the discourse context, and user/task model (the knowledge sources discussed in Section 3). Based on its knowledge base and decision making process, CUBRICON 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 the use of the MAP Display System [9]. CUBRICON then searches its knowledge base for task-specific objects within the region that should be displayed. The selection of these objects is based on the entity rating system that is part of the user model discussed in Section 3. CUBRICON does not display all entities that it knows about in the region, but only those that are above the criticality threshold for the user's current task (a sub-task of planning an air-strike mission). Thus the system decides to display all airbases, surface-to-air missile (SAM) sites, critical factories and plants, but not objects such as schools or minor industry that are not germane to mission planning. The resulting color map is shown in Figure 1.

Based on the information provided by the user/task model, CUBRICON knows the important attributes of each object. Since these attributes are not displayed or communicated via the map display but have been selected as relevant, 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 composes a table (as per the modality selection criteria discussed in Section 3.1) showing the important attributes of the displayed objects. The resultant table is displayed on the monochrome display and is shown in Figure 1.

In order to further illustrate CUBRICON modality selection and output composition process, 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, printed natural language, spoken natural language, etc).

USER: "Where is the Dresden airbase?"

CUBRICON: (Refer to Figure 2.)

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."

Color Graphics Display:

- Map of Inner Fulda Gap Region with added area that includes the Dresden airbase.
- Main roads, major cities, waterways, and national boundries (as before but across the whole map, old and new areas).
- Icons representing entities within the map area displayed that are above a preset threshold in importance are superimposed on the map (same as before but across the whole map, old and new areas).
- An airbase icon representing the Dresden Airbase.

Speech Output with coordinated Color Graphics:

- After the map is expanded, statement to direct the user's attention to the Dresden airbase on the map: "The Dresden airbose is located here <point>." The word "here" is accompanied by a visual point gesture in the form of blinking the airbase icon and the addition of a pointing text box.

Speech Output:

- As the table is presented on the monochrome monitor: "The corresponding table is being presented on the monochrome screen."

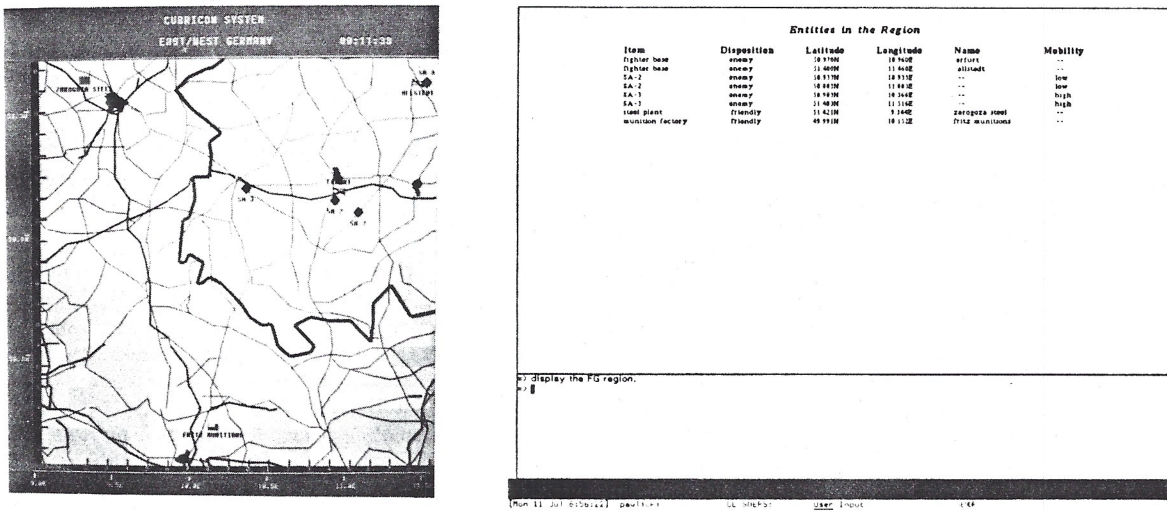

Figure 1. The Displays Composed by the System

Monochrome Graphics Display:

- Table of relevant entity attributes. Same table as before, but expanded to include new entities added to the map.

DISCUSSION:

The information requested is location. The Dresden airbase has an icon representation and a longitude-latitude associated with it in the CUBRICON knowledge base. The preferred modality for presentation is therefore graphical.

In composing a new map on which to display the Dresden airbase, the system has several choices. These include: 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 area around the airbase, or compose a new map including both the old map and the region around the Dresden airbase.

An important principle that CUBRICON tries to follow is to preserve the context of the human-computer dialogue. Since the user task has not changed and there is already a map displayed on the color graphics display, the system expands the displayed area to include the Dresden airbase. CUBRICON selects this option because it provides the requested information while it preserves the display context. When the Dresden airbase icon is added to the map, the system directs the user's attention to the icon using speech output with coordinated graphics, namely blinking the icon and placement of a "pointing text box".

Another important guideline to which the CUBRICON system adheres is to maintain consistency throughout a display so as 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. Similarly for other types of objects. If this is not done, then the user would probably infer incorrectly that there were no SAMs in the new area. The new map display composed by CUBRICON is shown in Figure 2. Guided by the consistency principle, the system also modifies the tabular presentation that is on the monochrome display. This new display is also shown in Figure 2.

The user now asks the system a question phrased exactly like the previous question for purposes of comparison.

USER: "Where is the Stargard airbase?"

CUBRICON:

Color Graphics Display:

- No change.

Monochrome Graphics Display:

- No change.

Natural Language:

- "The Stargard Airbase is outside of the current area of responsibility. It is located in Poland at 16.9°E and 51.3°N."

DISCUSSION:

This example illustrates the flexibility CUBRICON has in selecting from alternative presentation modalities and its ability to measure relevance. Although this question is phrased exactly the same as the previous question, the CUBRICON response is totally different. The Stargard Airbase is well outside of the user's area of responsibility as represented in the CUBRICON knowledge base (i.e., the task model). Therefore, CUBRICON judges that the Stargard airbase is less relevant than the current display and responds verbally without changing the current display.

To demonstrate CUBRICON's ability to generate coordinated NL and graphic pointing gestures, the user queries the system about the mobility of several SAMs.

USER: "What is the mobility of this <point>$_1$ <point>$_2$ <point>$_3$?" where each <point>$_i$ is a mouse point by the user on the map display.

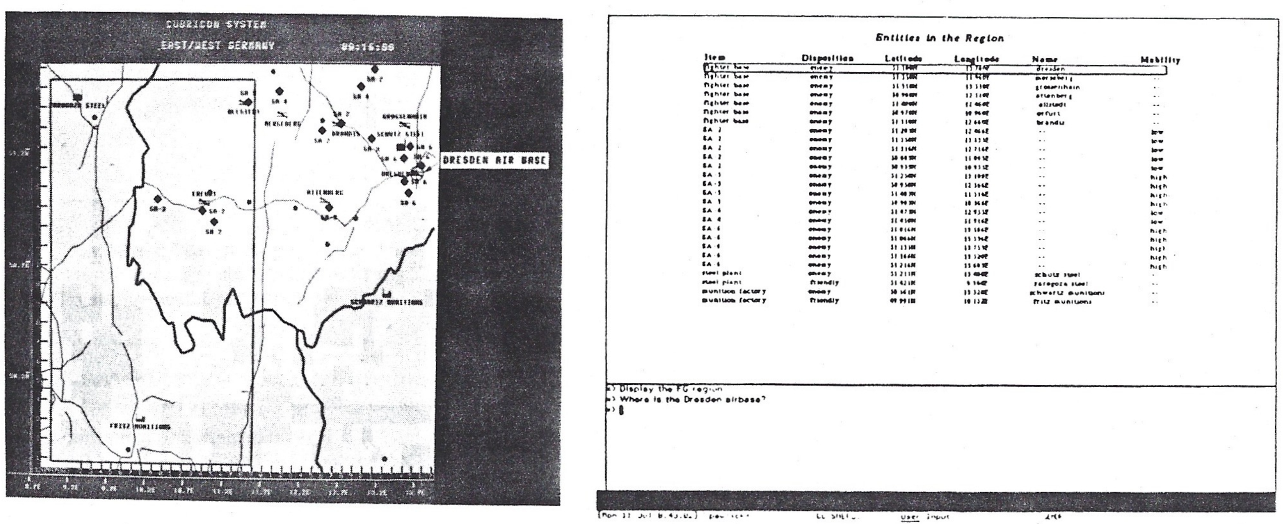

Figure 2. Map and Table Maintaining Context and Consistency

CUBRICON:

Speech Output with coordinated Color Graphics:

- "The mobility of this <blink icon>$_2$ SAM is low and the mobilities of these <blink icon>$_1$ <blink icon>$_3$ SAMs are 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 version of the requested information.

DISCUSSION:

This example demonstrates CUBRICON's ability to generate multi-media output that is highly integrated. When composing a response to the user, the system first selects the modalities in which to respond. As per the criteria discussed in Section 3.1, the system decides to answer in NL since the information is not very voluminous. Since the entities (i.e. the SAMs) are visible on the display map, the system points to each one (by blinking the appropriate icon) as it is referenced in the NL response. Also, since the information is technical and the user may subsequently want to refer back to it, CUBRICON decides to present a terse printed version also. This is done by placing a small text box near each SAM icon with a short statement of its mobility.

6. Current Status And Future Direction

The CUBRICON system is implemented on a Symbolics Lisp Machine with both color and monochrome displays. Speech input is handled by a Dragon Systems VoiceScribe 1000 speech recognition system. Speech output is handled by a DECtalk speech production system. The software is being implemented using the SNePS semantic network processing system [13,15], an ATN parser-generator [14], and Common Lisp.

The CUBRICON system modules and processes discussed in this paper are implemented and functional. These modules include the knowledge sources such as the knowledge base, discourse model, and the user/task model, the executor that retrieves information from the knowledge base, the process of selecting media/modalities for the presentation of information to the user, the output planning process, the processes of intelligently composing map displays and tables, and the process of generating simultaneous coordinated speech and pointing gestures during output. The examples presented in Section 5 are actual working examples that demonstrate some of the CUBRICON capabilities.

The CUBRICON team is continuing its development of the processes discussed in this paper as well as developing new funtionality. Areas in which the team is continuing current work and/or planning future development include: extending and refining the system's automated process of selecting the appropriate media/modalities for expressing responses to the user; enhancing the user/task model; adding additional modalities (e.g., predefined forms for mission planning) to the CUBRICON repertoire; extending the functionality of the multi-modal 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).

7. Summary

Modern information processing and decision-aiding computer systems are complex and require a full range of communications media/modalities to facilitate interaction with the human user. In fact, the power and complexity of many of today's systems have put great pressure on the human-computer interface, often making it the factor that most limits the system's potential. The application of knowledge-based technology to human-computer interfaces may offer great potential for improving the efficiency of the human-computer interface.

This paper discussed the development of a multi-modal human-computer interface system, called CUBRICON. This prototype system will serve as a research platform for investigating how knowledge-based approaches can best be applied to enhance human-computer interfaces. Four critical features of such an interface were introduced along with CUBRICON's initial approach to achieving them. Examples from CUBRICON dialogue have been cited to illustrate these approaches. Table 1 summarizes these critical features along with respective CUBRICON capabilities.

Several knowledge sources have been defined within CUBRICON. These serve as the basis for CUBRICON's multi-media output planning process. These knowledge sources include: characteristics of the information, a model of the ongoing discourse, a user/task model, and a knowledge base of task domain and related interface information. CUBRICON's output planning process includes selection of appropriate media/modalities, selection of ancillary relevant information, determination of whether display resources are available, modification of resources or information to be expressed (if necessary), modification of selected output media/modalities (if necessary), and composition of the output.

CUBRICON will continue to serve as a resource for conducting research on advanced human-computer interface design issues beyond the time-frame of the current program. It is hoped that ultimately CUBRICON will be available as a front-end to operational systems.

8. Acknowledgements

The other members of the CUBRICON team are: Z. Dobes, S. Glanowski, M.S. Summers, and S.C. Shapiro.

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

9. References

1. Arens, Y., Miller, L., Sondheimer, N.K. (1989), "Presentation Planning Using an Integrated Knowledge Base," *Architectures for Intelligent Interfaces: Elements and Prototypes*, J.W. Sullivan and S.W. Tyler (eds.), Addison-Wesley Pub. Co. (forthcoming).
2. Bly, S.A. and Rosenberg, J.K. (1986), "A Comparison of Tiled and Overlapping Windows," *CHI'86 Proceedings*, ACM 0-89791-180-6/86/0400-0101, pp. 101-106.
3. Carberry, S., (1987), "First International Workshop on User Modeling," *AI Magazine*, Vol.8, No.3, pp. 71-74.
4. Grosz, B. J. (1978), "Discourse Analysis," in D. Walker (ed.), *Understanding Spoken Language*, Elsevier North-Holland, New York, pp. 229-345.

5. Grosz, B.J., (1981), "Focusing and Description in Natural Language Dialogues," in *Elements of Discourse Understanding*, A.Joshi, B. Webber, & I.Sag (eds.), Cambridge Univ. Press, pp. 84–105.

6. Grosz, B.J. & Sidner, C.L. (1985), "Discourse Structure and the Proper Treatment of Interruptions," *Proc. of IJCAI*, pp. 832–839.

7. 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 Pulishers, pp. 353–362.

8. Helander, M.G., (1988), *Handbook of Human-Computer Interaction*, Elsevier Science Publishers, B.V., Amsterdam, The Netherlands.

9. Hilton, M.L. (1987), *Design and Implementation of the MAP Display System*, RADC Report.

10. Miller, G.A. (1956), "The Magical Number Seven Plus or Minus Two," *Psychological Review*, 63, pp.81–97.

11. Neal, J.G. & Shapiro, S.C. (1989), "Intelligent Multi-Media Interface Technology," *Architectures for Intelligent Interfaces: Elements and Prototypes*, J.W. Sullivan and S.W. Tyler (eds.), Addison-Wesley Pub. Co. (forthcoming).

12. Neches, R. & Kaczmarek, T. (1986), *AAAI-86 Workshop on Intelligence in Interfaces*, USC/Information Sciences Institute, August, 1986.

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

14. Shapiro, S.C. (1982), "Generalized Augmented Transition Network Grammars for Generation from Semantic Networks," *AJCL*, Vol. 8, No. 1, pp. 12–25.

15. Shapiro, S.C. & Rapaport, W. (1986), "SNePS Considered as a Fully Intensional Propositional Semantic Network," *Proceedings of AAAI-86*, pp. 278–283; in G. McCalla & N. Cercone (eds.) *Knowledge Representation*, Springer-Verlag Pub.

16. Sidner, C.L. (1983), "Focusing in the Comprehension of Definite Anaphora," in M. Brady & R.C. Berwick (eds.) *Computational Models of Discourse*, The MIT Press, pp. 267–330.

17. Smith, S.L. & Mosier, J.N. (1986), *Guidelines for Designing User Interface Software*, MITRE Corporation Technical Report No. 10090, ESD-TR-86-278.

18. Sullivan, J.W. & Tyler, S.W., (eds.) (1989), *Architectures for Intelligent Interfaces: Elements and Prototypes*, Addison-Wesley Pub. Co. (forthcoming).

Table 1

EXAMPLES OF HOW CUBRICON MEETS DESIRED FEATURES OF KNOWLEDGE-BASED USER-COMPUTER INTERFACES

MAJOR FEATURES	EXAMPLE CUBRICON CAPABILITY
1. Judge relevance and respond in a context-sensitive manner.	• Select objects and attributes for display based on discourse and task context. • Preserve context of current display (e.g., map and table displays) when there is no indication of a change in task. • Integrate new information on existing displays in response to user requests. • Respond differently to identical or similar user requests based on relevance to existing context.
2. Intelligently select the presentation media and formats for best information understanding.	• Select media, modalities and formats based on the nature of the information and relevant human factors considerations.
3. Use media and presentation formats in a highly integrated manner.	• Generate pointing gestures in coordination with natural language output. • Present information in more than one modality when different aspects of the information are best expressed using different modalities.
4. Provide consistency across all system outputs.	• Present selected entities across an entire map display, including newly added areas of the map. • Use consistent formats on related displays and displays showing identical types of information.