RL-TR-96-120
Final Technical Report
July 1996

VIRTUAL ENVIRONMENTS IN COMMAND AND CONTROL FOR ULTRA-HIGH DEFINITION DISPLAYS

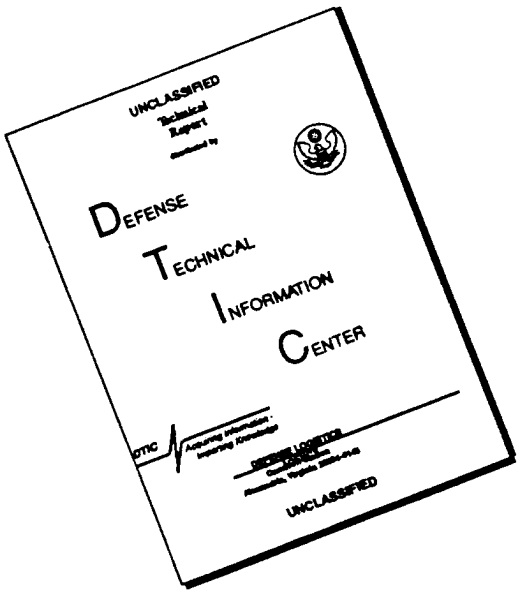
Massachusetts Institute of Technology

Sponsored by
Advanced Research Projects Agency

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Materiel Command
Rome, New York

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be releasable to the general public, including foreign nations.

RL-TR- 96-120 has been reviewed and is approved for publication.

APPROVED: Richard T. Slavinski

RICHARD T. SLAVINSKI
Project Engineer

FOR THE COMMANDER: John A. Graniero

JOHN A. GRANIERO
Chief Scientist
Command, Control & Communications Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify Rome Laboratory/ (C3AB), Rome NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

VIRTUAL ENVIRONMENTS IN COMMAND AND CONTROL FOR ULTRA-HIGH DEFINITION DISPLAYS

Nicholas P. Negroponte
Richard A. Bolt

Contractor: Massachusetts Institute of Technology
Contract Number: F30602-92-C-0141
Effective Date of Contract: 01 July 1993
Contract Expiration Date: 30 June 1995
Short Title of Work: Virtual Environments in Command and Control for Ultra-High Definition Displays
Period of Work Covered: Jul 93 - Jun 95

Principal Investigators: Nicholas P. Negroponte
Phone: (617) 253-5960
Richard A. Bolt
(617) 253-5897

RL Project Engineer: Richard T. Slavinski
Phone: (315) 330-2805

This research was supported by the Advanced Research Projects Agency of the Department of Defense, and by the Joint National Intelligence Development Staff (JNIDS), and was monitored by Richard T. Slavinski, RL/C3AB, 525 Brooks Rd, Rome NY 13441-4505.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1996	Final Jul 93 - Jun 95

4. TITLE AND SUBTITLE

VIRTUAL ENVIRONMENTS IN COMMAND AND CONTROL FOR ULTRA-HIGH DEFINITION DISPLAYS

5. FUNDING NUMBERS

C - F30602-92-C-0141
PE - 62708E
PR - H931
TA - 00
WU - 01

6. AUTHOR(S)

Nicholas P. Negroponte and Richard A. Bolt

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Massachusetts Institute of Technology
20 Ames Street
Cambridge MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Advanced Research Projects Agency*
3701 North Fairfax Drive
Arlington VA 22203-1714

Rome Laboratory/C3AB
525 Brooks Rd
Rome NY 13441-4505

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

RL-TR-96-120

11. SUPPLEMENTARY NOTES

Rome Laboratory Project Engineer: Richard T. Slavinski/C3AB/(315) 330-2805
*Additional sponsor - Joint National Intelligence Development Staff (JNIDS)

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The MIT Media Laboratory pursued three coordinated streams of research: the development of graphically intelligent tools and principles to support the interactive creation of symbolic information landscapes; the integration of such landscapes with pictorially convincing virtual environments; and the enabling of multi-modal natural language communication with the virtual environment display and its contents via combinations of speech, manual gesture, and gaze. It was intended that these streams converge in the final year of the overall project in the form of an ultra-high definition, seamlessly tiled wall-sized display (datawall).

14. SUBJECT TERMS

Symbolic information landscapes, Multi-modal interaction, Virtual environment, Large format displays, Intelligent agents

15. NUMBER OF PAGES

98

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Table of Contents

EXECUTIVE SUMMARY

The work under this contract reflects the second year activities under three coordinated streams of investigation:

- the development of graphically intelligent tools and principles to support the interactive creation of *symbolic information landscapes*;
- the *integration* of such landscapes with pictorially convincing virtual environments;
- the enabling of *multi-modal natural language communication* with the virtual environment display and its contents via combinations of speech, manual gesture, and gaze.

These three streams are projected in a 3rd research year to converge in the context of an ultra-high definition, seamlessly tiled wall-sized display (DataWall).

* * * * * *

Summary of 2nd year activities

Creation of symbolic information

This second year saw progress in the following areas:

Large Scale High-Resolution Display

- Developed a software library that allows GL (SGI Graphics Language) programs to be split into multiple displays.
- To augment visual display with sound, developed a sound server and client library to provide sound services on multiple UNIX platforms.

Selective Information Filtering

- Developed techniques for organizing, structuring, and visualizing large numbers of independently authored information objects ("Galaxy of News").

- Developed a compact representation of information relationships, known as an "Associative Relation Network" or ARN, that can be used to organize and visually present information to users.

- Developed an algorithm to construct an information hierarchy based on an Associative Relation Network (ARN).

Adaptive Graphics

- Developed techniques for automatic construction of "immersive information spaces" that allow for people to fluidly interact with information while browsing and searching according to their individual preferences.

- Designed methods for genetically evolving personal information seeking agents that explore distributed information bases to find information that is personally relevant to an individual or community, and visualize the results of that search (note that this is design only at this stage).

- Typographic Space: We have continued to develop software which supports investigating the use of typography in three-dimensional space. Several applications examples have been developed in order to pursue this investigation.

Understanding Large Complex Information-Bases

- Investigated new ways of aiding people to interact with large complex information bases through: 1) automatically analyzing information objects and deriving structures that convey relationships between information elements; projecting the multi-

dimensional structural model into a three-dimensional representation; defining possible user interactions and their effect upon the display of objects and contexts - e. g., an information object will display more detail when close up than from far away, and will foreground and background different information from different points of view.

- Explored ways to let a knowledge seeker to move through virtual time and space to explore connections between artifacts of philosophy, painting, music, literature, science, and political events of a pivotal time in world history, here the years from 1906 to 1918. This virtual space continually constructs and reconstructs itself based on the knowledge seeker's movements through and within it, much like the process of moving through the conceptual spaces of our minds as we construct meaning.

- Explored way to actively engage people in exploring complex, multimedia data bases using dynamic storytelling techniques.

Intelligent Design Tools

- Abstraction for multimedia temporal expression: an abstraction for multimedia temporal expression has been developed as a conceptual tool which the communication designer can "think with."

- Decentralized model of design: Theories developed in distributed problem solving and multi-agent systems have been applied to represent complex and dynamic behavior of communication design. We have developed a theoretical framework for a decentralized model of design, and we are starting to implement an experimental computer system based on the framework.

- Framework for the development of intelligent design systems: We have created a framework which guides and evaluates our development of intelligent design systems. A new role of designers as well as new requirements for design systems are involved.

- Knowledge acquisition system for graphic design applications: Performed studies to clarify the roles of designers and design systems in context of computer-based media and graphic design. Proposed a framework to guide and evaluate our in-house development of intelligent design systems.

Design of intelligent dynamic information display.

- Developed a theoretical model of dynamic design which includes a multiagent model of design thinking, plus an abstraction of temporal visual form which provides a language to describe the graphical behavior of design agents in terms of their dynamic activities, rather than the traditional method which uses fixed attributes.

- Implemented software which automatically adjusts color differences caused by simultaneous color contrast, and examined the effectiveness of adjustment in terms of visual communication.

MEDIAte: An Intelligent Authoring Environment for Information Tools

Spatial Parsing and Generation Using Relational Grammar

- Automatic Presentation: Relational Grammars have been developed which can support the personalization of information display based on display environment, user's task, and user's personalized style of presentation.

- Interactive Support For Design: We have adapted Relational Grammars to support interactive design suggesting improvements and enhancements to designs as they progress.

- Rule Editing By Demonstration: A rule editor was explored that will allow the creation, modification, and enabling/disabling of grammar rules by demonstration.

- Substrate Advancements: The system incorporates linking of graphics to dynamic simulations and underlying applications. The graphics then change to reflect the values and actions within the simulation. An interactive constraint system has also been incorporated into the system.

Browsing Very Large Display Spaces

- We have developed a technique, dubbed the *macroscope*, for browsing very large display spaces at multiple scales and resolutions. This technique, an alternative to the traditional solution of zoom and pan, is based on zooming and panning in multiple translucent layers. We have experimented with this work in the domain of geographic maps, display of hierarchical file structures, and other application areas.

- Developed an agent, dubbed Letizia, which operates in tandem with a conventional Web browser such as Mosaic or Netscape. The agent tries to anticipate what items may be of interest to the user by using simple heuristics to model what the user's browsing behavior might be.

Symbolic Information Landscapes

- Information Visualization: Researched and developed new models for displaying and interacting with complex information. Explored the potential of 3D spatial representations of information-bases for more sophisticated (i.e. non-hierarchical, non- linear) comprehension of multi-dimensional information. Testbed data bases included: financial data on seven mutual funds; consumer information about automobiles; demographic information.

- Researched and developed new design methods and working prototypes for information spaces that address the need for "mass customization." Also examined the problem of designing systems to generate forms for information spaces which can dynamically adapt both to information content and to the interests of the user.

- Developed sets of tools and skills, including: linear programming and "data mining" techniques; 3-D fonts; constraint- and rule-based programming; object-oreinted and graphics programming skills.

- Developed an Abstract 3D News Browser which provides a visualization of internet news in a 3D space organized by news group, relative importance, age, and length of article.

- Developed GeoSpace: An Intelligent Multilayered Mapping Environment for Visualizing Complex Spatial Data. The cartographic display is modeled using an Spreading Activation Network which enables the system to respond dynamically to high level user-queries. An integrated learning mechanism enables the system to reconfigure its domain knowledge according to individual user preferences.

- We have explored the potential of techniques rooted in *stage magic* for information presentation, including such aspects as: atmosphere, attention, surprise, continuity, and user preferences.

Emboding virtual space to enable the understanding of information

- We attempted to introduce into the 3D information environment a sense of scale and point-of-view, specifically through *metaphor theory* as it relates to the body and physical space. We explored as well findings from a study of the use of metaphor and abstraction in architecture. Emphasis is upon the creative use of metaphor as a tool for visual communication to move from the concrete to the abstract, as in going from the

notion of the "desktop metaphor" with its folders and documents to a generalized sense of information organization and presentation. We also explored the notion of the body (of the user) as a mode of experiencing virtual worlds, both as bodily experience influencing the way we encounter abstract ideas and "disembodiment" as a mode of travel in virtual space and time.

Integration of symbolic information with virtual environment

Virtual Actors

- Continued development of testbed—dubbed "WavesWorld"—for building virtual actors and designing and debugging their behaviors. Built our testbed on top of sophisticated commercial systems, leveraging industry standards and speeding our own development. Particular emphasis was placed on problem of multiple designers constructing virtual actors; the infrastructure we implemented promotes reuse of both geometry and behaviors of our virtual actors.

- Defined and implemented a common dynamic language (dubbed "eve") to describe the shape, shading, state and behavior of objects and actors in a virtual environment (VE). This language allows parts of actors to be described in a uniform way, promoting a "black box" approach to constructing increasingly more sophisticated virtual actors.

- Integrated industry standard digital time-based output (QuickTime) into our system, permitting documentation of work more completely and build up catalogs of virtual actors' competencies and behavior.

- Strategy is to fully integrate the language into the VE development environment.to allow us to quickly build and iterate over a variety of virtual actors competent in a

variety of domains and tasks, as well as share actor parts and behaviors among character designers.

Intelligent Virtual Camera

- Developed a framework for exploring intelligent camera controls in a 3D virtual environment. Developed and evaluated a methodology for designing the underlying virtual camera controls based on an analysis of what tasks are to be required in a specific environment. Once an underlying camera framework is built, a variety of interfaces can be connected to the framework. Prototypical virtual environments covering several application domains have been used to exercise and evaluate these ideas, including a virtual museum, a sporting event, and a conversational dialog between two virtual actors. In each of these applications, we have identified some of the visual tasks that need to be performed, presented a paradigm for encapsulating those tasks into virtual camera modules, and described in detail the underlying mechanisms that make up the camera module for navigating within these environments.

Multimodal natural language

In the second year, we made significant progress in the following areas:

Body model

- Developed a "body model" to incorporate system knowledge of the user's bodily position and actions. This body model is device-independent, and offers 2 encodings of the user's body:
    - positional information for upper body parts
    - joint angle information

- The body model is modular on the network, so that other working groups about the Lab can also tap into information about the user's movements. It works in a *client server* mode, to make information about user upper-body position and actions available to other modules upon request. These modules can stipulate what sub-sets of data will be delivered from the server, and at what data rates.

Gesture

- Developed a new scheme for the low-level processing of gesture input . The raw data from the glove sensors is processed to intermediate levels, that is, recorded as positions and actions, but short of interpretation as to meaning.

- The gesture analyzer produces gestural frames, like simplified "cartoons" of the structure of the users body actions, including eye movements. We plan to use the joint angles from the body model to do the first level segmentation of gestural input into movements into gestural units or building blocks.

History

- In the spatio-temporal representation, we added the *temporal* dimension allowing the system to record actions and events as they occur. This encoding can then be used to resolve references backwards in time: the user can describe an event, and the system can go back in the x,y,z,t stream and determined which past event or action the user is referring to. The schema should be adaptable for forward-looking temporal references, such as when we stipulate some system action contingent upon some criteria to be met at a future time. This process involves the setting up of a model of the anticipated event, and executing it when the conditions are met (e.g., when the red vehicle crosses the road, do such-and -such).

Part base interface to graphics

- We continued work on a part-base approach to object representation. This approach has 2 important features:
    - it is object-oriented, in that it allows the user to talk about objects in the scene. The schematization of objects and relations in the scene allows for the abstraction of spatial information out of the graphics, which allows for their integration with descriptions from speech and gestures.
    - it supports an hierarchical structure of the scene structure, and it allows to describe changes to the scene, new states, and new events by imposing new scene strictures on the database.
- This part-base approach allows a very-high level description of user actions or changes to the graphics domain, and the interface language used during the resolution of reference is compatible to the spatio-temporal system.
- We are now in a position to implement the part-base interface on top of the SGI "Inventor" system. A big advantage to doing this is that the SGI platform seems to be the graphics platform of choice for the project (e.g., the VLW is using SGI graphics engines).

Projected work in year three

The third year will primarily be devoted to integrating the three streams of work:

- Creation of symbolic information landscapes
- Integration of symbolic information with virtual environment
- Multi-modal natural language communication with the virtual environment display

developed by the participating groups. This will include the implementation of the DataWall (see following special note), and the development of command and control scenarios that demonstrate the seamless integration of 3D models, symbolic graphics and multi-modal natural language communication.

Special note on *DataWall*:

The scale of DataWall we had originally hoped for, that is, 8x10K, is at present unattainable. We had hoped to couple an array of Texas Instruments Digital Micromirror Device (DMD) projection displays with Silicon Graphics Inc. (SGI) Reality Engines to produce, or at least to approximate, the DataWall we would have liked to achieved. With the rate of development and availability of such technology at best uncertain, we developed as an alternative strategy the following approach based upon the Hughes Liquid Crystal Light Valve Projector.

The conventional Hughes Liquid Crystal Light Valve Projector uses a CRT imaged through a relay lens onto the photosensor of a photo-activated Image Light Amplifier. Our proposed approach, developed by Mr. Ronald MacNeil, would be to replace the CRT with multiple tiled Kopin Inc.'s c-Si AMLCD displays, which allow a more compact and less costly package. The inherently low pass filtering feature of the ILA removes any artifacts due to the seam between images without sacrificing image sharpness.

Our goal would be to develop a real time tiled display for digital data at 2048 by 3840 pixel resolution at 30 fps, with high brightness, high sharpness, large scale and perfect seamlessness. It would be suitable for a range of projection situations, from theater-like situations using front projection to more confined rear projection situation display rooms using folded optics. In tandem, we would develop the interface hardware to allow both full motion digital cinema resolution and a full motion synthetic composite of such sources as computer graphics, and live video teleconferencing over digital cinema image frames.

Such a display would be suitable for such DOD purposes as:

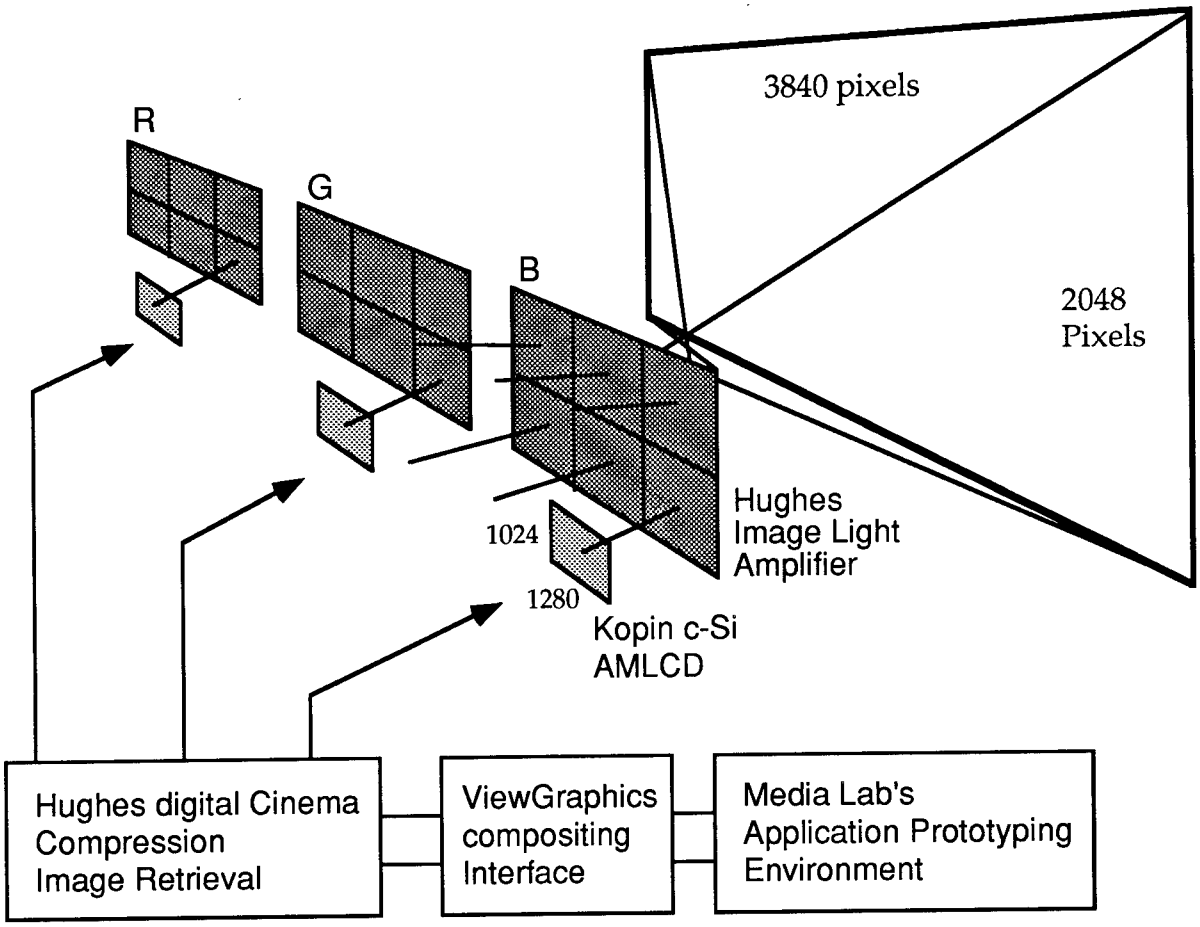
- Automated mapping and Location visualization
- Interactive Simulation and Training
- Command and Control, sensor fusion, aircraft situation assessment displays, where the data could include multiple live video feeds, real time computer graphics, all composited over preproduced dynamic backgrounds.

As to development strategy: Hughes Inc. would develop the 6 tile prototype projector, and the digital image data compression and retrieval hardware. Kopin Inc., of Taunton, MA, currently produces a VGA resolution version of the AMLCD arrays, and is about to bring to market the 1280 by 1024 AMLCD image array, which will be used for this prototype. ViewGraphics Inc., of Palo Alto CA., specializes in very high performance digital frame stores for use in digital cinema prototyping. They would develop the interface to Hughes' image retrieval hardware.

We note a significant event with respect to the prospect to successful ***tiling*** with such a display concept: A successful test of seamlessly tiling two Kopin VGA size AMLCD images was accomplished in November '94 by Don Mead at Hughes Data Compression Studio. Observers could not find the seam between the two tiles and image quality was not reduced across the seam.

Below is a schematic of this proposed system:

SYSTEM DIAGRAM



INTRODUCTION

This report attempts to summarize detailed research progress documented in various publications, reports, articles, and/or academic theses which are identified herein.

Purpose of the project:

The purpose of this project is to research and develop in prototype a command and control setting wherein the commander/decision-maker communicates in *concurrent speech, gesture, and gaze* with a rich *symbolic information landscape* integrated with the *virtual environment*, and rendered on an ultra-high definition, seamlessly tiled wall-sized display. The project will also research and prototype a unified, graphically intelligent multiple-media information authoring environment.

Project personnel:

Personnel at the MIT Media Laboratory directly involved in the conduct of this research are:

Dr. Richard A. Bolt, Senior Research Scientist and member of the Media Lab's Perceptual Computing Group.

Prof. Muriel R. Cooper, Professor of Visual Studies and Director of the Lab's Visual Language Workshop. (Prof. Cooper died suddenly and unexpectedly on May 26, 1994.)

Mr. Ronald L. MacNeil, Principal Research Associate at the Media Laboratory and co-founder with Prof. Muriel Cooper of the Visible Language Workshop.

Dr. David L. Zeltzer, Principal Research Scientist at the Research Laboratory of Electronics, MIT.

CREATION OF SYMBOLIC INFORMATION

Relevant Personnel:

Work under five sub-topics of managing visual complexity was completed under the supervision of the late **Prof. Muriel R. Cooper**, Professor of Visual Studies and Director of the Lab's Visual Language Workshop, and **Mr. Ronald L. MacNeil**, a Principal Research Associate at the Media Laboratory and co-founder with Prof. Cooper of the Visible Language Workshop.

* * * * * * *

Large Scale High-Resolution Display

Splitting Graphical Library (GL) programs into multiple displays

To demonstrate the feasibility of driving a multidisplay data wall from a single application, we developed a technique for distributing Graphical Library GL rendering operations among multiple SGI (Silicon Graphics, Inc.) machines. The library works by intercepting each GL call and distributing it to a set of SGI machines that have been arranged in a two dimensional array. The key aspects of this operation are determining the viewing transformations for each of the displays in the 2D grid, and synchronizing the updates to the displays.

Sound server and client library

A sound server and client library interface was developed to provide distributed sound services on multiple UNIX platforms. The client interface sound library allows an application to connect to a sound server, create sound objects, and play them. The library also allows for presentation control of sound information, such as play, pause, stop, fade in, fade out, volume, balance, and so forth. The client library allows the application to

connect to multiple servers simultaneously, which can be used to create spatial audio effects. The sound server can accept connections from multiple clients, and can manage their resources separately.

Selective Information Filtering

Galaxy of News

We developed a system, called Galaxy of News, for organizing, structuring and visualizing large numbers of independently authored information objects. The Galaxy of News system embodies a scalable approach to visualizing and navigating through large quantities of independently authored pieces of information, in this case news stories. It combines the effective aspects of both searching and browsing, and the ability to switch between these modes of operation seamlessly within a single interface. The system automatically organizes disconnected articles into dynamically formed groups, based on the content of the articles, that allow quick access to related information and the ability to quickly understand the relationships between articles.

The Galaxy of News, in effect, creates a new medium, an abstract information space, providing interactive navigation and intuitive access to correlated information. The Galaxy of News information spaces form structured, multidimensional, interactive environments where the information objects contained within the space determine the underlying structure. A powerful *relationship construction engine* utilizes an associative relation network to automatically build implicit links between related articles. To enhance the understanding of the space and its contents, the multidimensional information spaces constructed can change scale, orientation, perspective, representation and presentation as the user navigates through the space. Users interact with these information spaces through visual dialogs where actions have implicit meaning, e.g., moving forward in the space

indicates more specific detail is desired and moving backward indicates a desire for less detail, and more abstraction.

The Galaxy of News project investigates several information access and visualization principles, including:

1) pyramidal visualization of news objects to provide progressive refinement of news information;

2) visual clustering of news elements based on the content of news articles to provide structured information access;

3) semantic zooming and panning, where zooming is synonymous with searching or filtering, and panning is synonymous with browsing;

4) fluidity of interaction to understand and maintain the context of the information being presented;

5) animation and motion to illustrate relationships between news elements;

6) dynamic visual cues to aid in the navigation through an abstract news information space;

7) dynamic visual presentation of information to present the proper quantity of information at each instance of interaction and to eliminate distracting clutter.

These principles define an outline for building a structured hierarchical representation of news, whereby the upper portions of the pyramid consist of general descriptions or abstractions of the lower levels which contain more detail. Pyramidal representation offers news readers the ability to go through a process of glancing, to investigating, to reading details in a fluid and selective manner, while maintaining context of where they are in the process.

The Associative Relation Network (ARN)

At the heart of the Galaxy of News system is a mechanism for learning the relationships between news articles. This mechanism is called an Associative Relation Network, or ARN. This mechanism uses reinforcement techniques to capture the relationships between symbols extracted from documents. The relationship between symbols contained in an ARN define the relationship between documents.

An ARN works by maintaining weighted relationships between symbols contained in the network. An ARN is described as follows: For a given set of documents D , there exists a set of symbols S. The frequency of occurrence for symbol Si, is defined as

$$ci = 7\ Dx : \{Si, ...\}$$

where, Dx : {Si, ...} denotes a document containing Si

The weighted relationship between Si and Sj in a symmetric network is defined as

$$wi,j = wj,i = 7\ Dx : \{Si, Sj, ...\}$$

where, Dx : {Si, Sj, ...} denotes a document containing both Si and Sj

With an ARN, the documents reinforce the associative weights between symbols that represent the relationships between documents.

An ARN can be used to construct an information space to allow people to explore the relationships between documents. Further, this representation can be used to construct an abstracted information hierarchy as described below.

Constructing an information hierarchy based on an ARN

The Galaxy of News System just described learns the relationships between articles contained within a news base. This is done by parsing articles contained in a news base,

extracting symbols that describe the contents of the documents, and inserting them into an ARN, as described above. In effect, the documents are used to reinforce the relationships between symbols. The resulting ARN is then processed to extract an information hierarchy that forms the basis of an information pyramid used to visualize the news base.

The information hierarchy is extracted from the ARN by using the following recursive process:

1. Search through the ARN and find all the statistically independent symbols

2. For each independent symbol, find all the symbols statistically dependent on the independent symbol

3. For each set of dependent symbols, find the independent symbols

4. Repeat steps 2 and 3 until all the dependent symbols are independent of one another.

The information hierarchy resulting from this process is used to progressively refine the presentation of information to the user.

Adaptive Graphics

Learning and using user preferences

The Galaxy of News system described above also learns user's preferences as he or she navigates through the information space and reads the news. The user preferences are also represented by a ARN that is separate from the ARN that represents the relationships between the articles. When the user zooms up to and reads a news article, the system passively inserts the associations contained within the articles read into the user preference ARN. Additionally, the user can actively reinforce his or her preference for a

particular article. Conversely, if the user does not like a particular article, the symbol associations are decreased in the network with negative reinforcement. This may result in negative associations.

The user can control the user preference learning process through the use of two parameters: a learning rate; an aging or forgetting rate. The learning rate is used to control the level of weighting when associations are added into the user preference ARN. The aging rate is used to periodically dissipate the relationships between associations. As the user preference ARN ages, some associations and/or symbol counts may drop below thresholds, and are removed from the network. Hence, associations that are not continually reinforced will eventually be removed from the network. Thus, the preference ARN tracks the current preferences of the user.

Genetically evolving personal information-seeking agents

We designed an approach to exploring the World Wide Web with genetically evolved info-seekers. These info-seekers would evolve to find information specific to individual users or groups of users. The design of the system included definition of a proposed genotype and evolution process to evolve classes of info-seekers with different objectives that cooperate to meet a common goal.

The basic structure of the proposed approach is to create a set of agents that explores and searches through the World Wide Web, deriving the structure of information and presenting the results of what they find to the user. As the results are presented to the user, the system would learn the users interests. These learned user interests would in turn guide the agents' exploration and search.

In this approach, we would use a genetic algorithm (GA) to evolve a set of info-seeker agents. The genotype of the agents controls how an info-seeker searches through a distributed information base, in this case the World Wide Web (WWW). The WWW is a

special information base because the documents it contains provide the interconnection of the information. The interconnection is provided by the hyperlinks. These hyperlinks can essentially provide a search path to explore the distributed information base. In the proposed algorithm the links are not used to derive the structure of the information; rather, they are simply used to move between documents in the information collection or search process.

The role of each info-seeker agent (controlled by a genotype) is to traverse links, parse the contents at a node, and build up an Associative Relation Network (ARN) that describes the relationships and hence the structure of the information contained within the distributed infobase. This ARN and the associated documents are reported back to a server local to a user. A front-end application presents the reported structure of the information to the user. This application allows the user to navigate through a constructed information space that has been created from the structure derived from the content of the articles. Note that the information space constructed is not the same as the original links contained in the documents.

When the user navigates through the information space and reads documents, the system learns the relational structure of the information that the user prefers. This learned relational structure can then be used as a fitness function for the genetic algorithm that generates phenotypes, and hence new agent sets.

Typographic Space

We have continued to investigate an interactive and dynamic universe of worlds with landscapes of typographic, spatial, and symbolic information. The user moves through this galaxy, browsing the generalities or exploring the detail of complex ideas and relationships. The infinite universe of three-dimensional information is a truly dynamic

interface—one that can allow the user unprecedented freedom and control if it is well conceived and designed.

We have explored new metaphors and models—not simply for the acquisition of data, but also to help people to think better—in order to make complex and dynamic information accessible. Various experimental works lead us to propose a solution— "Information Landscape," a space that allows the user to peruse information by navigating in and out, and then to investigate specific portions in both two and three dimensions. By using elements with a wide range of sizes, we enable both macro- and micro-readings of the information.

Understanding Large Complex Information-Bases

The Mind's Eye: Understanding Large Complex Information-Bases through Visual Discourse

We have been investigating new ways of interacting with the computer that help people gain an understanding of a large complex information-base as they progress through structured navigation and access of information that forms a structured visual discourse. Conceptually, this involves developing a process where the computer will explore an information-base, build a structural representation of the information corpus, and then allow a person to explore the information through visual representations constructed during an interactive visual dialogue. From this perspective, the user's exploration of the information base is through the mind of the computer, hence the concept name "The Mind's Eye." The focus of this research is 1) the process of constructing meaningful visualizations of information objects in a 3D space and the relationships between information objects within a given context, and 2) the types of interactions the user will have with the information objects, and how these interactions form a visual discourse.

More specifically, we are developing a grammar that formally defines the rules for the dialog.

We have been developing a computational process that enables dynamic exploration of information based a conceptual framework being developed on the Millennium Project (described below), and on the approach explored with the Galaxy of News system [Rennison, 1994]. There are two primary objectives of the computational process: 1) to automatically compute conceptual structures that describe information, and 2) to dynamically present the conceptual structure to the users to aid them in understanding the information.

We have been developing a system that automatically analyze a corpus of information to derive conceptual structures that aid understanding the relationships between information objects that represent concepts that span both space and time. These conceptual structures include categorical structures, hierarchical structures, relational structures, radial structures, linear quantity scales, and foreground-background structures. Each of these structures help us understand the relationships between information elements. We have also been developing a framework for projecting conceptual structures into virtual information spaces. This approach also defines relationships between conceptual structures and the five information organization structures proposed by Richard Wurman [Wurman, 1989]. These include

- Location
- Alphabet position
- Time
- Category
- Hierarchy

In addition, we have defined the correspondence between conceptual structures and metaphorical image schemas, and between conceptual structures and metaphorical

mappings. We have defined computational structures to represent the conceptual structures.

We have also developed methods for dynamically presenting the information to the user through a process that interactively unfolds over time. We call this process "Visual Discourse." There are two important aspects of visual discourse: 1) how the conceptual structures are mapped to virtual space such that they convey meaning, and 2) how the computer interprets user interaction.

The computational process consists of the following steps:

1. Analyzing the information-base to construct a representation of the relationships between the information objects, namely analyzing the underlying structure of the information-base

2. Presenting the information relationships in a 3D virtual space that provides a particular contextual view on the information, and

3. Interpreting user movements and actions in the 3D virtual space to dynamically query for additional information and dynamically reconstruct the virtual space to show the relationships between the objects returned from the query.

Each of these steps is discussed in detail in the following sections.

Automatic Structure Analysis

We have defined a process for analyzing information objects and deriving structures that convey relationships between the information elements. Information elements in this definition include the original information objects as well as features that are extracted from the information objects. We use the extracted features to analyze the structure of the information objects (as described below). The ultimate objective of the structure analysis

process is to construct structures that correlate to cognitive structures such as categories, hierarchical structures, relational structures, and radial structures [Lakoff, 87]. These structures will in turn be used in the process of mapping the structural relations into a visual space that is presented to the user. These conceptual structures aid the user in navigating through the virtual information spaces, as well as aid in understanding the relationships between information elements. It is through these conceptual structures that we are able to understand the relationships between events and artifacts that span place and time. The way we understand these relationships is through their projection into a virtual space.

The conceptual structures are derived through the following process:

1. Filtering the original set of information objects to a reduced subset

2. Extracting key features from the reduced set of information objects

3. Constructing a representation that captures the structural relationships between extracted features and the underlying information objects

4. Processing the structural relationship representation to extract computational structures that correspond to conceptual structures.

We describe each of these steps in the following subsections.

Information Object Filtering

The first step of the structure analysis process is to filter the original set of objects to a reduced set. This essentially establishes the initial or global context for a discourse. This filtering process is based on an initial condition specified by the user. Queries for information objects are either made explicitly, via a dialog box, or more powerfully through implicit interaction within an information space.

Extracting Key Features

The second stage of the analysis process is to extract key features from the information objects. These features include such information as the dates/duration that an event occurred, location of an event that occurred, and sets of symbols that describe the information object. The symbols in this case refer to elements such as nouns, noun phrases, verbs, verb phrases, and so forth. They may also include constructs such as Universal Record Locators (URLs) and names.

We allow for three levels of feature definition: 1) features extracted from the content or body of the information object, 2) features defined by an object annotator, and 3) features associated with the object by the end user, or knowledge seeker. Each of these features are treated separately and the user has control over how the system applies them in constructing the information spaces.

The features fall into two categories: general properties and structural relations. General properties include information such as size, date/time, location, and so forth. The general properties of the information objects vary according to the type of object. For example, information objects that pertain to artifacts may contain a size of the artifact, date produced, location produced, who created it, and so forth. Information objects that pertain to events would not include a size (unless some conceptual size can be specified), the date may be specified as a duration, the location may be specified as a region that may change over time, etc. Structural information consists of sets of symbols that indirectly bind an information object to other related objects.

We extract key symbols and symbol sets from the contents of textual information via one of three techniques. First, we provide a mark-up language that allows authors or annotators to explicitly embed specifications of AssociationSets in the body of an information object description file. These AssociationSets can have a hierarchical

structure. This hierarchical structure is similar to the sentence-paragraph-section-chapter-book type structures that bind words together, but operates on the principles of association as opposed to grammatical structures. Second, we can use automatic text indexing techniques based on symbol frequencies to extract keywords from a text document. And, third, we can use a part-of-speech tagger [Brill, 1992] to identify the nouns, noun phrases, verbs, and so forth.

Constructing Relationship Representation

Once we have extracted important features from the documents, we use these features to construct a representation that captures the emergent relationships between the information objects. A key element of our research is to find emergent structural properties that are not globally defined, but rather emerge out of the amalgamated properties of the individual objects. Hence, we do not impose a global structure on the information spaces, they are derived automatically from the contents of the information-bases through this bottom-up structuring process.

We are specifically using associative relations that define co-occurrences of symbols as the basis for our structural representation [Rennison, 1994]. In addition, we also use temporal-causal relationships, and geographical and absolute temporal parameters (as specified by the authors of the information objects) to build a representation of the underlying structure. As described above, each information object can contain a set of dates, a set of locations, and associated sets of symbols (AssociationSets). When these sets of symbols, dates, and locations are inserted into the core representation they strengthen weights between the symbols, dates and locations.

The main element of our representation is an Associative Relation Network (ARN) [Rennison, 1994]. An ARN captures the relationships between symbols contained within information objects. The relationships between symbols contained in an ARN define the

relationships between information objects. An ARN maintains weighted relationships between symbols contained in the network, and the relationship between symbols and the information objects to which they relate.

An ARN defines an N-dimensional space that contains N^2-N terms. The basis vectors of the space are defined by the symbols. Associated with each basis vector (i.e. symbol) is a vector that defines the relationship between itself and all the other basis vectors (i.e. symbols). With an ARN, the information objects reinforce the associative weights between symbols that represent the relationships between information objects. A symbol also forms a link between objects. However, the link that the ARN forms between documents is not a simple index between information objects. It contains structural information that determines the strength of the relationship between the objects.

The ARN described above is also used to capture relative temporal relationships between information objects, and implicitly the cause and effect relationships between information objects. Our current information object mark-up language allows authors and annotators to specify sets of symbols that the subject of the information object was influenced by, and a set of symbols that the subject of information object influenced. Each of the influenced by symbols are associated with each of the symbols that describe the information object, and these relationships are maintained in a separate ARN that also maintains the temporal distance between associated symbols. We call this extended ARN a Temporal ARN, or TARN. Likewise, each of the influenced symbols are associated with each of the symbols that describe the information object. This relationship is maintained in a separate TARN.

In addition, each symbol in the representation has a reference to all the locations and times that the symbol occurred as defined by an information object. Likewise, each location and time has a reference to associated symbols, and back to the information objects that contain the location or time. The locations are also stored in a geographic

database that facilitates quick filtering and searching of either symbols or information objects. Times are stored in a temporal database that facilitates quick filtering and searching for related symbols and information objects.

The primary utility of this representation is the ability to compute probability, similarity, and distance measures between symbols and information objects. These measures are used in computing categorical classifications, fuzzy clustering, hierarchical structures and sorted lists. These complex representations are dynamically processed to extract these structural relationships which are implicitly maintained by the representation just described. This process is discussed in the next section.

Computing Conceptual Structures

The most important step of the structuring process is deriving computational structures that correspond to conceptual structures that implicitly define structural relationships between information elements. We specifically compute the following computational structures:

- graph--where each node in the graph corresponds to a category and linked nodes correspond to related symbolic categories
- acyclic directed graphs--where each node in the graph corresponds to a symbolic category and linked nodes correspond to symbolic sub-categories
- fuzzy cluster graphs--where each node in the graph corresponds to a symbolic category and linked nodes correspond to related symbolic categories such that the node is the central theme (as in a conceptual radial structure)
- sorted lists--where each node represents a place in some linearly ordered sequence or scale.

A brief description on how these are computed is provided below.

A graph is generated from an ARN. Essentially an ARN represents a graph structure; however, since this structure has a very high dimension, it can be pruned by applying a similarity threshold. This process simply removes nodes from the ARN that fall below the similarity threshold.

We use several techniques to compute acyclic directed graphs [Rennison, 1995a]. These techniques fall into two categories: clustering and probabilistic sorting. Within these two categories we use two primary techniques: top-down and bottom-up. The clustering algorithms use similarity and distance measures calculated from an ARN. The probabilistic sorting techniques use probabilities measures computed from an ARN [Rennison, 1994]. The following recursive process describes one of the techniques we use to compute acyclic directed graphs:

1. Search through the ARN and find all the statistically independent symbols

2. For each independent symbol, find all the symbols statistically dependent on the independent symbol

3. For each set of dependent symbols, find the independent symbols

4. Repeat steps 2 and 3 until all the dependent symbols are independent of one another.

The information hierarchy resulting from this process is used to aid the user in navigating through information structures. This process essentially defines a technique for abstracting and generalizing. As the philosopher William James noted "we acquire knowledge through a process of differentiating characteristics. This process of differentiation is based on finding dissociations between elements" [Arnheim, 1969]. This process captures the essence of this objective.

Currently, we compute a fuzzy cluster graph by first computing an acyclic directed graph using a top-down probabilistic approach. Then, we apply a clustering algorithm using each node in the graph as a centroid and searching for all symbols that fall within the range of the symbol, where the range is defined as the farthest distance from the node symbol to a child symbol.

The result of the computational processes described above is a set of computational structures that map to conceptual structures. What follows is a description of how these computational structures are used to construct spaces that reflect the underlying conceptual meaning.

Space Building

The presentation aspect of the Visual Discourse process consists of projecting the multi-dimensional structural model into a three-dimensional visualization. Because of the high dimensionality of the underlying space (a direct correlation to the number of features extracted from the information objects), it is not possible, or at least not meaningfully intelligible, to project the entire underlying space into a 3D representation directly. The construction of the projection, therefore, must be carefully considered. The projection should be a direct representation of the cognitive structures derived from the information objects. Our objective is to generate dynamic virtual spaces that correspond to the mental spaces we continually construct during natural language exchanges.

We have defined a model and process for projecting the structural information into a 3D space. The process is dependent upon the type of view, or the conceptual viewpoint, on the information for a given space. Currently, we have parameterized the types of spaces that can be generated according to:

- location
- alphabetical position (though the use of this constraint is limited)
- time (absolute, e.g. at time X, and relative, e.g. before, after)
- category
- hierarchy

or, as Wurman [Wurman, 1989] terms them collectively by first letter, LATCH. These parameters may be specified individually, or by combinations. For example, a space can be generated to illustrate the temporal relationships between information elements (which may include combinations of the original information objects, and features extracted from the information objects). Or, a temporal relationship may be combined with a geographical relationship. Specification of these parameters essentially define the context in which the information elements are positioned in space. Some particularly meaningful contexts include the following:

- Categorical
- Categorical-Temporal (absolute)
- Categorical-Temporal (relative)
- Categorical-Geographical
- Categorical-Geographical-Temporal

Constructing 3D Information Spaces

The multidimensional structural representation of our information-base allows our system to dynamically generate meaningful sets of information objects that adapt to our

continuous queries, as expressed by our continuous movements in the information space. In order to dynamically explore and interact with these information sets, we have to display them in such a way that invites investigation and allows for intuitive interaction. To this end, a 3D space builder automatically constructs information contexts from a list of information objects and a list of extracted features (such as keywords) which are also displayed as graphical objects. An information context is displayed as an enclosure that contains the set of information and feature objects.

All information objects contained in a 3D information context are assigned a context-specific XYZ location, XYZ axial rotation, scale, color, and transparency based on a mapping of each one of these display attributes to an appropriate information content attribute. In addition, each information object displays different representations of itself relative to our position and orientation in space.

Another task of our 3D space builder is the generation of transitional spaces. Transitional spaces are connectors from one context container to another. A transition between a context that is contained inside another (i.e. the information object list of the new context is a subset of the old information object list) is experienced like a power-of-ten shift or an infinite zoom.

Interaction Interpretation

An important aspect of meaning communication, and hence understanding information-bases, is the dynamic process of shifting point-of-view and shifting context. As Fauconnier clearly delineates, a central theme in meaning construction is access through conceptual connections that define mappings between source and target domains. In our computational environment we define a context to be a set of information objects and the relationships between them. A context is represented and presented to the knowledge seeker as a container and a set of contained objects, where the container defines the

relationship between the objects. We define a context shift to be either global or local. Establishing a global context implies filtering or refiltering the original information objects into a working subset of information objects. For example, we may wish to establish a global context to be all objects "in the geographic area of France during the period of 1911 to 1912." Local context shifts imply a change in conceptual viewpoint on the subset of information objects and illustrate a new set of relationships between the subset of objects. For example, we can shift between a categorical view, to a categorical-temporal view, to a geographical-temporal view. Key questions are: How does the user indicate these context shifts? How are these context shifts executed? These are some of the main issues of our research.

We have outlined an approach to these questions. It defines the possible user interactions and their effect on the display of objects and contexts as well as the underlying information representation. An information object will display more detailed information up close than it will from far away, for example, and will foreground and background different information from different points of view [Strausfeld, 1995]. We have also defined the possible user interactions which consist of movement of self and manipulation of objects and defined how we interpret user actions based on the cognitive models based on our conceptual framework. We use these actions to change the current context based on the this interpretation of the user's actions.

Other Projects Exploring Visual Discourse

The Millennium Project

On the Millennium Project, we are exploring ways to provide a knowledge seeker the ability to move through virtual time and space to explore and discover the connections between artifacts of philosophy, painting, music, literature, science, and political events of a pivotal time in world history: the years from 1906 to 1918. This virtual space

continually constructs and reconstructs itself based on the knowledge seeker's movements through and within it, much like the process of moving through the conceptual spaces of our minds as we construct meaning.

The conceptual framework for this research is based on linguistics and cognitive science. We are addressing how our concepts of embodied cognitive models and visual discourse assist us in designing and building a computational environment that enables people to understand large bodies of information.

To give an idea of what the Millennium project is about, consider 1912. In 1912 the S. S. "Titanic" sank on it's maiden voyage, Woodrow Wilson won the U. S. presidential election, Sun Yat-sen founded Kuomintang (Chinese National Party), C. G. Jung published "The Theory of Psychoanalysis," Edwin Bradenburger invented a process for manufacturing cellophane, and Marcel Duchamp painted "Nude descending a Staircase." How, if at all, do these events relate to one another? Where, when and what were the confluences of ideas and people that influenced the outcome of these events? How do we acquire the knowledge to understand the complex associations between people and ideas, across time and place, based on the artifacts and events they created?

We are searching for ways to simulating an expert visual interlocutor to address these issues. In order to create this expert we need to do three things:

1. Build up a representation of how each information object in our database relates to all other pieces of information

2. Project the representation onto a virtual space that resembles a mental space

3. Enable interaction with us, as users, by reading our movements in the space and responding by dynamically restructuring the space.

The Millennium Database

Our database consists of a set of files that contain information objects that describe events, artifacts, people and ideas pertaining to the years 1906-1918. They are displayed as 3D text objects that sometimes include images, video clips, or sounds. Each of these information objects contain annotations that describe the properties of the information objects. These basic properties include:

1. date
2. location
3. associations (term, object)
4. cause-effect relationships
5. size measurements

This work has been documented in Rennison & Strausfeld, 1995.

Elastic Boston

In the Elastic Boston project, conducted with Prof. Glorianna Davenport and several other members of the Media Lab, the object is to build a database about the history of the Boston Artery. This includes historical events surrounding the construction of the current artery, and plans and controversies concerning the construction of the harbor tunnel and the destruction of the current elevated artery. Given this collection of information, we are exploring ways to actively engage people in exploring this information using dynamic storytelling techniques.

Intelligent Design Tools

Abstraction for Multimedia Temporal Expression

An abstraction for multimedia temporal expression has been developed as a conceptual tool which the communication designer can "think with." This model is also intended to

be a basis for the development of software tools that support a designer in programming, or representing, expressive behavior of multi-media objects. The proposed model borrows concept from the performing arts, such as dance and music, in order to extend the static nature of design into dynamic and continuous design.

The model is still a theoretical one, and further work is necessary in order to evaluate its practical value. Currently, a software design support tool based on the model is being developed. This software will be used as an experimental apparatus to evaluate the proposed model. This model also will be used as a basis for the representation language that is used to describe behaviors of design elements in the decentralized design system described below.

Decentralized model of design

Theories developed in distributed problem solving and multi-agent systems have been applied to represent complex and behavior of design solutions for electronic communication, where both information and interaction are dynamic. We have developed a theoretical framework for a decentralized model of design, preliminary to implementing an experimental computer system based on the framework.

The model considers a design solution to be a continuous system consisting of a collection of smaller design systems, or design elements, which are called "design agents." Design specifications, or design strategies, are specified for individual design elements, which are considered autonomous and cooperative. The hypothesis behind this approach is that the decentralized model is more natural for a designer to "think with," particularly when a design solution involves many temporally expressive design elements, since design specification can be encoded locally. This model uses the abstraction of temporal expression described above as a basis for representing each design agent's expressive behavior.

A theory of multi-agent systems proposed by M. Singh (Singh, 1991) has been adopted, together with a software engine, implemented in LISP, which allows us to develop a multi-agent system. In addition, various low level tools have been implemented in order to support high quality graphics. The next step will be to integrate various modules implemented in the past year to develop experimental design systems with several application domains.

Framework for the development of intelligent design systems

In the visual design of computer-based media, designers often find it impossible to solve individual design problems by themselves, and instead must encode a method of designing in the form of a computer program. We argue that research in the development of generative design systems has not emphasized the role of designers and has failed to evaluate design systems within the broader picture of visual communication.

This study clarifies the roles of designers and design systems in the context of computer-based media and graphic design, and identifies research issues involved in their development. The proposed framework is intended to help us guide and evaluate our development of intelligent design systems.

Knowledge acquisition system for graphic design applications

Graphic designers and other visual problem solving experts now routinely use computer-based image-editing tools in their work. Recently, attempts have been made to apply learning and inference techniques from artificial intelligence techniques to graphical editors in order to provide intelligent assistance to design professionals. The success of these attempts will depend on whether the programs can successfully capture the design knowledge of their users. But what is the nature of this knowledge? Because AI techniques have usually been applied in such areas as medicine or engineering rather than visual design, little is known about how design knowledge might differ from knowledge

in other fields. This work reports the results of a knowledge engineering study to try to understand how knowledge is communicated between humans in graphic design {Lieberman, 1995a]. Nowhere is the process of design communication more critical than in teaching beginning designers, since the effectiveness of the communication is crucial to the success of the student. This study examined books intended to teach graphic design to novices, and tried to analyze the nature of the communication with a view toward applying the results to a knowledge acquisition system for graphic design applications.

Design of intelligent dynamic information display

Typographic Performance: Design solutions as emergent behaviors of active design agents

Development of theoretical model: Multiagent model

We developed a theoretical model of dynamic design that provides a model, along with a language, with which the visual designer can think during the course of designing. The theoretical model consists of two parts. The first is a multiagent model of design in which a design solution is considered an emergent behavior of a collection of active design agents-performers-each of which is responsible for presenting a particular aspect of information. The second is an abstraction of temporal visual form which provides a language to describe the graphical behavior of design agents in terms of their dynamic activities, rather than the traditional method which uses fixed attributes.

We expect that this new model of design will advance the field of graphic design in the realm of electronic communication, by providing a means of dialogue between a designer and dynamic artifacts, as well as a communication tool between designers.

Multiagent Design System: "perForm"

In order to enable the computational generation of dynamic design solutions, we have implemented a software tool, called "perForm," which simulates parallel activities of multiple design agents. "perForm" provides a simple lisp-based programming language which allows designers to describe behaviors of individual design agents,, based on the proposed model.

"perForm" consists of three parts: a multiagent simulation engine, a graphics library, and an agent description language. The multiagent simulation engine was built based on the multiagent theory proposed by Singh (1991). In this system, the agent is described in terms actions and strategies. It allows designers, or programmers, to specify individual agent's behavior and it simulates the parallel nature of multiagent interaction. The system is independent of realizations, and can be used with any graphical, or auditory environment. Since it is designed as a general system, it is independent of particular types of multiagent systems.

The graphic library supports a visual realization of agents. The aim of this library is to use high quality typography in order to examine the theory with design solutions that reflect the reality. The agent description language is a set of macros that are used define agents' ability and behavior, which include actions, strategies, and sensors. The agent description language is a subclass of the multiagent system, and is designed to describe dynamic design agents.

Experimental Design Solutions

"perForm" is being used to develop a series of experimental design solutions for concrete design problems in order to (1) illustrate the user of the theory in context, (2) help evaluate and examine the theory, (3) debug "perForm" software. We created a new

version of Dynamic News Display using "perForm," and simultaneously finding and fixing perForm software.

Adjusting simultaneous contrast for dynamic information display

We also worked on a small project that concerns the problem caused by simultaneous color contrast effects for the design of dynamic information displays [Ishizaki, 1995, 1994]. Simultaneous contrast effects are a known phenomenon; humans perceive the same physical color differently depending on its surrounding color, or background. These color differences caused by simultaneous contrast are particularly problematic in information graphics where colors convey meanings. On a computer-based dynamic display, such as weather and air traffic, since background color or position of graphical element is difficult to predict at run-time, the simultaneous contrast effect becomes a problem for reliable communication.

We implemented software which automatically adjusts color differences caused by simultaneous color contrast, and examined the effectiveness of adjustment in terms of visual communication. The results of the research showed that the automatic adjustment improves the visual design on information display, as well as the flexibility of the color choice.

MEDIAte: An Intelligent Authoring Environment for Information Tools (JNIDS)

Spatial Parsing and Generation Using Relational Grammar

Automatic presentation

Relational Grammars can support the personalization of information display based on display environment, user's task and personalized style of presentation.

Grammars to encode Style: WIRED vs Scientific American

The automatic layout of magazine's table of contents now has data from two issues of *WIRED* and three issues of *Scientific American*. A grammar encodes the basic style of each magazine, and one can view the WIRED data in the standard format or switch to the alternative *Scientific American* style. Similarly, one can view *Scientific American* in its normal style or the altered (*WIRED*) state. By using the grammars to encapsulate the style of the different magazines' table of contents, grammars illustrate how they might support dynamic presentation styles.

Temporal vs spatial resolution in automatic layout of dynamic documents

This grammar suggests that the display of information can be sensitive to the environment and context of output. The trade-off between high-resolution color displays and small personal digital assistants is suggested with the display of information from *Popular Mechanics* home repair procedure and data from an on-line training manual. Two separate presentation techniques are used, and involve hyperlinks from displayed buttons to the information or an interactive slider which control's the automatic presentation of the information.

Grammars sensitive to users task

Minor filtering of information to display the elements relevant to the task at hand has also been demonstrated. Characterization of tasks and user modeling with preference information can help guide the parser to present more or less detailed information that becomes more personalized to the user's specific problem.

Relational Grammars for the automatic presentation of information

We created a final demonstration using Relational Grammars for the automatic presentation of information [Weitzman & Wittenburg, 1994]. When presented on a high

resolution color display, a three page presentation is constructed. An alternate presentation is automatically constructed when viewing the same information on different hardware, such as a personal digital assistant (e.g., a Newton). This same three page presentation that included QuickTime movies was transformed into a six page presentation without movies (i.e., the Newton could not play QuickTime movies). In the final presentations, the user has the control to view particular pages of interest. This is accomplished with a dynamic slider that controls which pages the system presents.

Architecture of Information: Interpretation and presentation of information in dynamic environments

Design of information presentation is undergoing significant changes. Documents are information interfaces that must dynamically reconfigure themselves based on their content, the medium in which they are displayed, and the intended use of the information they present. Increases in computational power and the increased bandwidth of interconnected networks provide greater access to information. These factors, combined with the realization that not all of this information can now be pre-designed, necessitate new tools and techniques to ensure the effective presentation of computer-based information.

We have exploited the structure of information to support the design of dynamic documents. From this structure, visual languages are created which support the process of building an Architecture of Information [Weitzman, 1995]. Relational Grammars, an extension to traditional string languages, is the formalism in which these visual languages are constructed. This formal approach affords a number of different interaction techniques, three of which we examined in this research. First, information is automatically presented from predefined languages. This dynamic layout reconfigures the same information accounting for the constraints of different delivery environments. Second, the authoring of information is supported by incremental improvements during

the design process. These improvements help the user explore the design space with incremental design decisions. Third, these visual languages are constructed by demonstration. An authoring tool to modify these languages without coding is presented in Weitzman [1995].

Interactive support for design

Relational Grammars support interactive design, suggesting improvements and enhancements to designs as they progress.

Interactive Improver-Based Scenario

In this scenario, the designer works with a generic grammar that will support "good" design augmented with a domain dependent grammar. This generic grammar has rules for alignment and sizing of elements. Then, a second set of domain specific rules fire on top of the groupings that are created.

The domain specific grammar will build composites and understand the output requirements of the domain (i.e., HTML files for Mosaic, high quality layout for on-line viewing, PDA layout for handheld devices, etc.). These multiple grammars will then support the application in a much more modular fashion. These interpretations carry with them the necessary domain "semantics". Multiple, ambiguous interpretations of the resulting improvements are presented continuously to the designer.

Within the current scenario, an article for the *WIRED* table of contents is constructed with the end delivery application being Mosaic. The designer places an image and some text around it. These are grouped when sizing and alignment rules fire. Then, table of contents entries are formed when the designer selects from the alternatives presented. This second step includes the semantics of the realization in Mosaic.

Rule editing by demonstration

A rule editor is being explored that will allow the creation, modification, and enabling/disabling of grammar rules by demonstration.

Rule Editor

This research is also focused on how users will be able to create and modify rules by demonstration. Currently, two windows visualize a rule: a "before" window, which displays the lexical and composite categories used in rule formation along with their relationships; an "after" window, which shows the "articulation" of these categories for graphic presentation. Constraints are used to lay out both windows. The after window uses the constraints that will be enforced in the final presentation. The ability to modify rules and then to immediately incorporate them in the grammar definition is possible for simple rules. An "improver" grammar helps to "clean up" the rule output and regroup the elements into a new arrangement for presentation.

Rule enabler and disabler

The user can also enable and disable individual rules and rule sets. These rule sets may be related to the output media or delivery device (such as Mosaic), and include the necessary semantics for generating a presentation. This permits a user to indicate to the system what is important during the parsing process.

Substrate advancements

The system incorporates linking of graphics to dynamic simulations and underlying applications. The graphics then change to reflect the values and actions within the simulation. An interactive constraint system has also been incorporated into the system.

Linking display graphics to underlying simulation or application.

The use of dynamic icons has been incorporated into the low level substrate of this research. This feature takes advantage of simulation-based icons in the interface, and permits any display attribute (color, position, etc.) to be modified based on underlying simulation or application values. In addition, the user can modify application values through interaction with the interface components.

For instance, this feature will enable the grammars to produce elements in a presentation that can monitor a "clock" variable and display themselves according to a sequence of events and user interaction. These dynamic icons were used in the presentation of a home repair procedure. This procedure had three major steps and a number of substeps. The substeps also included images of the substep repair. By linking all of the elements to a simulation variable that indicated the current time of presentation, elements dynamically change their visibility. A slider icon is presented to the user which modifies the time of presentation. Elements appear and disappear dynamically according to this presentation variable.

Interactive Constraint Mechanism

DeltaBlue, an incremental constraint solving program, has been incorporated as the underlying constraint satisfaction algorithm. This substrate support includes both the spatial and temporal constraints for the presentation of multimedia documents.

Browsing Very Large Display Spaces

The Macroscope

The traditional solution for interactively browsing a very large display space is to *zoom* and *pan*. Invariably, though, the sense of context is lost upon zooming-in. Also, sequential applications of zoom-in and zoom-out operations may become tedious. We

have researched an alternative technique, dubbed the *macroscope*, based upon zooming and panning in multiple translucent layers. It should comfortably permit browsing continuously on a single image, or set of images in multiple resolution, on a scale of at least 1 to 10,000 (Lieberman, 1994).

The book and film *Powers of Ten* (Morrison and Eames, 1994; 1982; 1978) try to instill an appreciation of the scale of the physical world by a succession of images each on a scale differing from the next by factors of ten, from atomic to galactic perspectives. Each image shows the scale of the next smaller image by a rectangular viewfinder placed at its center. *Powers of Ten* brings together in a single work phenomena that occur over a wide range of spatial scales, and forces the viewer to think about the relationships between them.

The *macroscope* takes the visual device of *Powers of Ten* and compounds it, using the computer's ability to combine, change, and display images. Consider a typical zooming-in operation. The user can choose a smaller subset of the screen, which we will call the viewfinder. The zooming operation blows up the viewfinder to fill the entire image. A zoom-out operation does the inverse transformation.

Ordinarily, after the zoom operation, the viewer loses the context of where the blown-up image came from. The *macroscope* approach is to make the zoomed-in and zoomed-out views *share the same physical screen space* by displaying them in multiple translucent layers. Recent experiments by Colby and Scholl (1991) have shown that it is feasible to combine multiple layers of information on a single display, using translucency, focus and other image processing techniques to visually combine layers while retaining the integrity of the individual components.

For example, starting with a map of the south central US, and selecting a rectangle, we construct a two-layer macroscope that focuses in on the Oklahoma-North Texas area.

Superimposed on the original map is an enlarged image of the viewfinder area. The value is in interactively controlling the multiple layers: change of position of the rectangle corresponds to a pan operation; change of size of the rectangle corresponds to a zoom operation.

The visual effect of zooming and panning is that of imagining the visible screen as a window onto a much larger translucent virtual sheet displayed "in front of" the background. Zooming corresponds to "stretching" or "shrinking" the sheet and panning corresponds to "sliding" the sheet. While one layer is moving relative to another, the layers are much more easily distinguished.

Dynamically adjusting the translucency levels between layers, as in Colby and Scholl (1991), is a key technique for selectively emphasizing or de-emphasizing information under user control. Emphasizing the background layers aids in orienting yourself in a large space; emphasizing the foreground layers gives precise control over the close-up view. The *macroscope* interface is designed to give the user dynamic control over the relative emphasis of layers simultaneously with the zooming and panning operations. The technique is not restricted to just 2 layers; in our work, we have gone on to displaying a third layer with satisfactory results.

Letizia: a web-browsing agent

The recent explosive growth of the World Wide Web and other on-line information sources has made critical the need for some sort of intelligent assistance to a user who is browsing for interesting information. Past solutions have included automated searching programs such as WAIS or Web crawlers that respond to explicit user queries. Among the problems of such solutions are that the user must explicitly decide to invoke them, interrupting the normal browsing process, and the user must remain idle waiting for the search results. This work introduces an agent, Letizia, which operates in tandem with a

conventional Web browser such as Mosaic or Netscape [Lieberman, 1995b]. The agent tracks the user's browsing behavior -- following links, initiating searches, requests for help -- and tries to anticipate what items may be of interest to the user. It uses a simple set of heuristics to model what the user's browsing behavior might be. Upon request, it can display a page containing its current recommendations, which the user can choose either to follow or to return to the conventional browsing activity.

Symbolic Information Landscapes

Information Visualization

We have researched and developed new models for displaying and interacting with complex information. Also, we have explored the potential of 3D spatial representations of information-bases for more sophisticated (i.e. non-hierarchical, non- linear) comprehension of multi-dimensional information.

To this end, a prototype was developed for displaying financial data. In this project, financial data on seven Mutual Funds is displayed on a 3D data grid that can be sliced or sectioned by the user with dynamic intersecting planes. Charts and graphs can also be generated by the user to analyze financial performance of the Funds on the basis of numerous features such as risk and annual rate of return.

A second prototype was developed for displaying a database of consumer information, such as automobile information in a 3D space. This application uses the 3D location of each data item (automobile, in this case) to express its value relative to other data items based on 3 (x, y, z) parameters such as cost, mpg, and consumer rating. The user can change the parameters dynamically to construct many different spatial representations of the database.

A third prototype was developed for displaying demographic information geographically. Scrolling planar maps allow the user to filter data (by density of subscribers to a hypothetical on-line service) interactively.

Design of Adaptive Information Spaces

We researched and developed new design methods and working prototypes for information spaces that address the need for "mass customization." We also examined the problem of designing systems to generate forms for information spaces which can dynamically adapt both to information content and to the interests of the user.

To this end, a method that uses Artificial Evolution to evolve 3D information environments was conceptualized. An experimental system was developed that dynamically evolves 3D spaces consisting of simple forms. Future research will explore design systems for information spaces that can learn fitness criteria through user interaction.

An interface agent that tracks a user's interest in financial data was designed and implemented as part of the Financial Space project. The agent tracks how long the user analyzes one particular Mutual Fund and asks the user if they would like more information on that Fund. If the user is interested, the agent presents some significant piece of information about the Fund (e.g., it has the highest net assets of all the funds) and restructures the information space to best communicate this information.

An autonomous agent was conceptualized and enacted to autonomously read patterns and trends in data and to visualize them effectively for the user.

Development of Tools and Skills

Sets of tools and skills were developed, including:

- Linear programming and "data mining" techniques were studied to expand skills for extracting knowledge from databases. Interest was generated in the use of visualization techniques that, in themselves, extract a kind of knowledge or structure from the information.

- A 3D stroked font for the SGIs was written. This font runs on all the lower end machines because it does not require texture memory.

- Experience with constraint-based and rule-based programming was gained. Experiments were done using Delta-Blue and Clips.

- Object-oriented and graphics programming skills were improved with the use of C++ and Iris GL.

Abstract 3D News Browser

The goal of this project was to design and implement an abstract three-dimensional news space that would enable easy browsing of a set of filtered articles. The articles are filtered by an agent according to users preferences. The news articles are represented on a plane, with color coded three-dimensional bars corresponding to different news stories. Articles are positioned on the plane based on their age and relative importance. The height of the bar gives a measure of the length of each article. The user can browse around the abstract space to gain an overview of the news articles presented.

GeoSpace: A Behavior based approach for Exploring Large Information Spaces

Visualizing complex spatial information as in a map, where users can easily access, view, and discover interesting relationships in the data is a hard problem. It usually requires a continuous shift of visual attention and an excessive burden on the cognitive capacities of a person. Our goal was to create a system, GeoSpace, that was capable of the following:

- Enable users to interact with a dynamic complex visual environment to identify quickly a graphic object such as a street, county label, highway, etc.;
- Provide a mechanism whereby the display reacts to a series of user requests while maintaining overall context, so that users don't get lost in the process;
- Integrate a learning mechanism that would enable the graphic display to be molded according to different users preferences.

Our focus was to have more reactive information spaces that would behave and adapt to users requests dynamically. We designed and implemented a prototype system in the mapping domain which enables users to interact by specifying queries to which the system responds. This research effort can benefit many other domains such as news, education, etc. where the information space is complex, such that users get lost in the process of data exploration.

The current system consists of multiple layers of data, all or some of which will be visible at a given moment in time. The data is visually represented according to domain knowledge constructed by the designer of the system. An activation spreading network is used to maintain the current state of the display, and is also responsible for enhancing certain regions of the map by changing transparency, color and typographic size of graphic elements.

We also started to develop a new version of the system, called GeoSpace-II, using U. S. census TIGER database, in order to examine the system using more complex and realistic set of information. Significant effort was put into the development of software which translates the TIGER database into a meaningful form so that domain knowledge-base can be automatically created. Since the new database was very large, we also optimized the activation spreading mechanism.

Research in this domain has been reported in Ishizaki and Lokuge (1995) and in Lokuge and Ishizaki (1995).

Techniques from magic in adaptive 3-D information spaces

We speculatively explored the idea of adaptive three-dimensional complex information spaces based on the principles and techniques of magic [Lokuge & Ishizaki, 1995]. Stage magicians regularly create fascinating information spaces which enrich users knowledge as well as entertain them during the process of interaction. Those familiar to the realm of magic appreciate the distinction between a "trick" and a "illusion." An illusion is the "big picture," the magical effect which embodies a plethora of techniques germane to making the impossible happen. A trick on the other hand is just one such technique, and is nothing but the secret behind the illusion.

After exploring various magic-based techniques, we narrowed the principles to include:

- Create Atmosphere by framing context
- Focus of Attention
- The element of surprise
- Continuity
- Adapt to users preferences

Based on these principles, we developed a preliminary model of creating interactive presentations, which showed the relationships between the constituent parts. The essential idea was to create framing contexts, prior to presenting information, so that users would be more receptive when they perceive complex information space. In addition, the idea of using focus of attention was further explored within the context of GeoSpace. Activation Spreading techniques have proved to be very useful for achieving smooth continuous transitions so that users attention is drawn to the most relevant piece of information. This

idea was explored in detail by building GeoSpace II. (Cf. previous section on GeoSpace, and GeoSpace II.)

We also explored a theme dubbed MediaMagic, a conceptual space of different activities for the Boston area. Instead of organizing the activities according to the geographic space as in GeoSpace, they are organized by conceptual closeness to each other. The idea was to have the geographic map transform to the MediaMagic space as the user zooms into an activity of interest. The image on the map becomes a cube, with each face representing different views of the same information. The user can traverse through this information space to gain access to other related activities. This prototype enabled us to explore and test ideas related to guiding a persons attention in a dense display, timing the visual presentation of information, violating expectations, and the notion of continuous transformations.

Embodying Virtual Space to Enable Understanding of Information

With current advancements in real-time 3D computer graphics and animation hardware it is now possible to create virtual information spaces through which users can move and interact. The move into virtual space, besides providing greater possibilities for interaction with information, also introduces a number of interesting and challenging problems for designers:

> How does the space and interaction with it enable understanding of the information?
>
> How does the user get a sense of the magnitude or the boundaries of the space?
>
> How does the user not get lost in the space?

One way to address these problems is by introducing into the 3D information environment a sense of scale and point-of-view. This can be done by drawing on bodily intuition and the idea, evidenced in language, that understanding is structured by bodily

experience. Such an approach involves an iterative 3-phase process of background research, development of design and evaluation criteria, and experimentation that will result in design principles, methods, and prototype software for interactive 3D information design. Specifically, the approach applies concepts from metaphor theory as it relates to the body and physical space as well as findings from a study of the use of metaphor and abstraction in architecture [Strausfeld, 1995b].

Approach

This research, reported in detail in the thesis by Strausfeld [Strausfeld, 1995b], has centered around the role of the body in under- standing complex and abstract information. The primary domains for this work are linguistic metaphor theory, and architecture. This research supports the hypothesis that by embodying virtual space we can better enable understanding of complex information.

Metaphor

How can designers correlate parameters particular to information to parameters particular to virtual space? Metaphor provides a solution [Ortony, 1993]. Designers often use metaphor as a tool for visual communication. The purpose of metaphor in the context of electronic media is to orient users in an initially new and characteristically abstract domain. A metaphor can help import the concrete into the abstract because we are generally more comfortable with what we can see and touch. Applying the concrete metaphor of the desktop, for example, to the relatively abstract operating system of the Macintosh generates expectations and understandings from the user about tangible ideas such as folders and documents. Metaphors enable interface designers to make decisions about how to represent information in a consistent and clear way.

The problem of representing the abstract with the concrete is the problem of language. We encounter this problem every time we attempt to express ideas outside the realm of

the physical world. In *Metaphors We Live By*, Lakoff and Johnson expose the way language allows us to implicitly (and often sub-consciously) reference our physical and cultural experience in the world to express or understand abstract concepts or ideas. Lakoff and Johnson show that language is based on a conceptual system that is metaphorical in nature. They write: "The essence of metaphor is understanding and experiencing one kind of thing in terms of another." [Lakoff & Johnson, 1980]

Metaphor can allow us, then, to bridge the gap between information and concrete virtual space. Most of our fundamental metaphorical concepts, in fact, are organized in terms of spatialization metaphors. We say, "things are looking up" or "I'm feeling down" or "we need to work through this." We understand these statements because our physical and cultural experiences provide the basis for the underlying metaphors. If most of our conceptual metaphors are spatialized, then we tend to structure and understand things abstract, like information, the way we structure and understand physical space.

Although there may be many spatial metaphors that apply to information, we here focus on scale and point-of-view. To clarify, scale and point-of-view are not metaphors themselves. Point-of-view is part of the metaphorical concept "understanding is seeing". Scale is an image schema that Johnson views as basic to the quantitative and qualitative aspects of our cultural and physical experience. [Johnson, 1987]

In an interview for *International Design,* Prof. Muriel Cooper of the MIT Media Laboratory emphasized the "power of abstraction" as an alternative to metaphor [Abrams, 1994] Abstraction can also be used as an effective tool for visual communication or expression because it provides the viewer a freedom of interpretation that metaphor does not. Abstraction is the opposite of metaphor in that it is about defying reference to anything concrete. Abstract artists since Kandinsky have attempted to achieve complete non-referentiality through abstraction. Although many would argue that total non-referentiality is impossible, many artists were successful in achieving such a high degree

of abstraction that their work has multiple references or possible interpretations. In general, the more abstract the organizing concept, the more potentially adaptable the concept is to a viewer or user. As designers, we need to strike the right balance between metaphor and abstraction.

Body

Because it is through the body that we experience the physical world, the body is implicit in spatial metaphors. Mark Johnson's book *The Body in the Mind* examines the way in which our bodily experience directly influences the way we structure and understand abstract ideas.[Johnson, 1987]

Johnson's theory that our embodiment is the key to dealing adequately with questions of meaning and reason raises some interesting questions about the role of the body in virtual space. One of the great advantages of virtual space appears to be disembodiment: the ability to move through space without the constraints of size and weight, to see through objects and to fly through time.

References/Publications/Theses

Abrams, Janet. Muriel Cooper's Visible Wisdom, *I. D. The International Design Magazine,* September October 1994, pp. 48-55, 96-97.

Arnheim, Rudolf. *Visual Thinking*. 1969, University of California Press, London.

Brill, Eric. A Simple Rule-Based Part of Speech Tagger. In: *Proceedings of Third Conference on Applied Natural Language Processing*. Trento, Italy: ACL, 1992.

Colby, G. and Laura Scholl. Transparency and Blur as Selective Cues for Complex Visual Information. In: *Proceedings of Image Handling and Reproduction Systems Integration Conference,* San Jose, CA, February 24-March 1, 1991

Fauconnier, Gilles. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge, England: Cambridge University Press, 1994.

Ishizaki, Suguru. [1994] Adjusting Simultaneous Color Contrast Effect for Dynamic Information Display. In: *Proceedings of IS&T and SID's 2nd Color Imaging Conference,* November, 1994.

Ishizaki, Suguru. Color adaptive graphics: What you see in your color palette isn't what you get. In: *ACM SIGCHI '95 Companion,* May 1995.

Ishizaki, Suguru. and Lokuge, Ishanta. Intelligent interactive dynamic maps. In: *Proceedings of AutoCarto 12*, February, 1995.

Johnson, Mark. *The Body in the Mind: the bodily basis of meaning, imagination, and reason*. Chicago: The University of Chicago Press, 1987.

Lakoff, G. *Woman, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago, Illinois: University of Chicago Press, 1987.

Lakoff, G., and M. Johnson. *Metaphors We Live By*. Chicago, Illinois: University of Chicago Press, 1980.

Lieberman, Henry. The Visual Language of Experts in Graphic Design. In: *Proceedings of IEEE Symposium on Visual Languages,* Darmstadt, Germany, September 1995a.

Lieberman, Henry. Letizia: An Agent That Assists Web Browsing. In: *Proceedings of the International Joint Conference on Artificial Intelligence,* Montreal, August 1995b.

Lieberman, Henry. Powers of ten thousand: navigating in large information spaces. In: *Proceedings of ACM Conference on User Interface Software Technology*, Marina del Rey, CA, November, 1994.

Lokuge, Ishanta. and Ishizaki, S. GeoSpace: An interactive visualization system for exploring complex information space. In: *ACM SIGCHI '95 Proceedings*, May 1995.

Morrison, Philip & Phyllis, and the Office of Charles and Ray Eames. *Powers of Ten: About the Relative Size of Things in the Universe*. New York: Scientific American Library, 1994.

Morrison, Philip and Phyllis, and C. and R. Eames. *Powers of Ten*. Scientific American Press, 1982.

Morrison, Philip and Phyllis, and C. and R. Eames. *Powers of Ten*. Pyramid Films, Santa Monica, CA, 1978.

Ortony, Andrew (Ed.). *Metaphor and Thought*. Cambridge, England: Cambridge University Press, 1993.

Rennison, E. Galaxy of News: An Approach to Visualizing and Understanding Expansive News Landscapes. In: *Proceedings of UIST 1994*. Marina Del Ray, California.

Rennison, E. *The Mind's Eye: An Approach to Understanding Large Complex Information-Bases through Visual Discourse*. MS Thesis. Massachusetts Institute of Technology. Cambridge, MA., 1995a.

Rennison, E. and L. Strausfeld, The Millennium Project: Constructing a Dynamic 3+D Virtual Environment for Exploring Geographically, Temporally and Categorically Organized Historical Information. In: *Proceedings of the Conference on Spatial Information Theory*. Vienna, Austria. September, 1995b.

Singh, Munindar P. Group ability and structure. In Y. Demazeau and J.-P. Muller (eds.) *Decentralized A. I.* vol. 2. Elsevier Science Publishers B. V./ North-Holland, Amsterdam, Holland, 1991.

Strausfeld, L. *Embodying Virtual Space to Enable Understanding of Information.* MS Thesis. Massachusetts Institute of Technology. Cambridge, MA. November, 1995.

Strausfeld, Lisa. Financial Viewpoints: Using Point- of-View to Enable Understanding of Information. In: *CHI '95 Conference Companion,* Boulder, CO, May 1995.

Weitzman, Louis Murray. *The architecture of information : interpretation and presentation of information in dynamic environments.* MIT Doctoral dissertation, MIT, 1995.

Weitzman, Louis and Wittenburg, Kent. Automatic Presentation of Multimedia Documents Using Relational Grammars, In: *ACM Multimedia '94,* San Francisco, CA. October 15-20, 1994.

Wurman, Richard Saul. *Information Anxiety.* New York: Bantam Books, 1989.

INTEGRATION OF SYMBOLIC INFORMATION WITH VIRTUAL ENVIRONMENT

Relevant Personnel:

Work in this area was completed under the direction of **Dr. David L. Zeltzer**, Principal Research Scientist at the Research Laboratory of Electronics, MIT.

* * * * * * *

We have continued development of a testbed for building virtual actors and designing and debugging their behaviors. We have built our testbed on top of sophisticated commercial systems, leveraging industry standards and speeding our own development. We have placed particular emphasis on the problem of multiple designers constructing virtual actors, and the current infrastructure we have implemented promotes reuse of both geometry and behaviors of our virtual actors.

Much of this year has been spent on defining and implementing a common dynamic language (which we call "eve") to describe the shape, shading, state, and behavior of objects and actors in a virtual environment (VE). This language allows parts of actors to be described in a uniform way, promoting a "black box" approach to constructing increasingly more sophisticated virtual actors.

We have recently integrated industry standard digital time-based output (QuickTime) into the system, allowing us to document the work more completely, and to build up catalogs of virtual actors' competencies and behavior. We expect this to play an ever-increasing role in the user interface of the development environment.

We have been working on fully integrating the language into the VE development environment. This will allow us to quickly build and iterate over a variety of virtual

actors, competent in a variety of domains and tasks, as well as share actor parts and behaviors among character designers.

In addition to the above work, we have developed a framework for exploring intelligent camera controls in a 3D virtual environment. As part of this research, we have developed and evaluated a methodology for designing the underlying virtual camera controls based on an analysis of what tasks are to be required in a specific environment. Once an underlying camera framework is built, a variety of interfaces can be connected to the framework. Prototypical virtual environments covering several application domains have been used to exercise and evaluate these ideas, including a virtual museum, a sporting event, and a conversational dialog between two virtual actors. In each of these applications, we have identified some of the visual tasks that need to be performed; we have presented a paradigm for encapsulating those tasks into virtual camera modules; and, we have described in detail the underlying mechanisms that make up the camera module for navigating within these environments.

WavesWorld: A Parallel, Distributed Testbed for Developing and Debugging Autonomous Behaviors

Virtual environments (VEs), whether for education, training, or entertainment, are still almost exclusively constructed by hand. Ideally, we would like interactive simulations to construct themselves from a set of high level commands:

- "What happens to the material when the temperature reaches 400 degrees?"
- "Show me how to fix the engine."
- "What happened to Brenda last week?"

In attempting to build such a task-level VE system, we believe that too little attention has been focused on the issues surrounding the process whereby the virtual actors themselves are designed, built, and debugged. The process of debugging our autonomous, virtual

actors is vital to building them (cf. Zeltzer, 1991). If we cannot change what we have built to meet whatever specifications we have agreed on, the process of constructing specific virtual actors for specific scenarios is doomed. Also, any tools for building virtual actors must be embedded in a supportive, multi-modal development environment which supports rapid prototyping and a tight equivalent to the "design-implement-debug" cycle of traditional software engineering.

Other important issues are: how can we build virtual actors whose behavior, shape, and shading are plug-and-play? How will disparate craftspeople and artisans work together to create autonomous virtual actors? What sort of protocols between structural and behavioral components will allow parts of one character to be reused successfully in another? Finally, how will all this work in the high-bandwidth, distributed, computational milieu of the near future in which computation, communication and content converge?

For the past five years we have been developing a software testbed for exploring these ideas. This system—WavesWorld—is a collection of software for designing, building, and debugging distributed simulations. The intent of this work is to allow a user to quickly build a graphically-simulated character that can act autonomously in a networked virtual environment.

WavesWorld uses a planning algorithm (based on independent work by Profs. Pattie Maes and David Zeltzer) which has been significantly extended for dealing with asynchronous and parallel execution. A comprehensive sensing structure has also been added to the planner, complementing the high-level controls the reactive planner provides by allowing time varying sampling rates to be integrated into the perception mechanisms. In addition, this system integrates computational economics at the lowest level, such that all processes in WavesWorld are "bid" at run-time. Finally, a networked, multi-modal, interactive development environment that supports debugging as well as design is essential to building complex VE systems. To this end, WavesWorld has been seamlessly

integrated into the NeXTSTEP development environment, which acts as a front-end to a large set of heterogeneous computing resources.

The system uses a modeling language based on RenderMan as its intermediate scene description, which allows it access to high quality, photorealistic rendering, as well as utilizing various high end graphics boxes to accelerate real-time preview using QuickRenderMan and OpenGL.

Current research is centered around building "malleable media," which are objects containing parametric shape, shading, and behavior, along with all the conventional user interface elements to allow a world builder to intelligently manipulate it. One of the difficulties in building complex virtual environments with complex virtual actors is the sheer bulk of modeling (shape, shading and behavior) that must be done initially. Conventional "clip objects" are useful sometimes, but too often they don't quite fit the bill. What is needed is some sort of "clip art with knobs" or, as we say, "malleable media."

For example, you want to situate some activity in a room. Does it have windows? Which walls are they on? Does it have doors? What kind? What sort of material is covering the walls? Where are the lights? Is it a wood floor or a linoleum floor? All of these are questions that a scenario developer would like to answer, but with current software tools, such attributes are bound very early in the VE development cycle. We have designed a modeling language to encapsulate shape and shading information about the VE which is tightly integrated into both sophisticated and portable user interface development environments such as NeXTSTEP and TK. This allows a developer to build a complex parametric model of some object in the VE, and then construct an appropriate set of user interface widgets to set the parameters of the model. This provides a uniform interface and toolset for constructing virtual actors out of clip objects that have geometry and

behavior, and which are presented to the designer in a uniform, multi-modal user interface.

We have also built conversion tools to allow models from most commercial modeling packages to be automatically translated into our modeling language. This allows us the best of both worlds —an in-house, programming solution with access to models developed on commercial software packages.

We have used a prototypical world to exercise our structural and behavioral modeling and visualization tools. This world—SanderWorld—is based on an exemplar task for evaluating autonomous robot systems, described by Charniak and McDermott (Charniak and McDermott, 1985, pp. 487-489) and later used by Maes (Maes, 1989). The scenario concerns a simple robot confronted with a classic planning dilemma.

Intelligent Camera Control in a Virtual Environment

Current interest in so-called immersive interfaces and large-scale virtual worlds serves to highlight the difficulties of orientation and navigation in synthetic environments, including abstract "data spaces" and "hypermedia" as well as more familiar modeled exterior and interior spaces. This is equivalent to manipulating a viewpoint—a synthetic camera—in and through the environment. Nearly all prior work in the field, however, has focused on techniques for directly manipulating the camera. In our view, this is the source of much of the difficulty. Direct control of the six degrees of freedom (DOFs) of the camera (or more, if field of view is included) is often problematic and forces the human VE participant to attend to the interface and its "control knobs" in addition to— or instead of—the goals and constraints of the task at hand. If the intention of the human VE participant is, e.g., to observe some object X, then allowing him or her to simply tell the system, "Show me object X" is a more direct and productive interface.

This is an instance of TASK LEVEL interaction. In earlier work, we characterized the levels of abstraction at which one can interact with virtual objects and processes, and we described the varying "access panels" one obtains. Here, we will describe a system for specifying behaviors for virtual cameras in terms of task level goals and constraints. As in our earlier work on camera control, we make task level control available as well as enabling various direct manipulation metaphors.

We share the view of many in the user interface community that one of the first steps in interface design should be a task analysis of the application. While this may be a difficult exercise in and of itself, it allows us to identify with reasonable confidence the objects and operations we should provide at the interface, and to specify the necessary software abstractions. While it is impossible to completely describe human behavior at the visual interface for all applications, our analysis suggests that the generic visual operations we need to support involve:

- orientation -- i.e., visual comparison of ego- centric and exocentric coordinate frames;
- navigation from point to point;
- exploration of unknown areas; and
- presentation to external observers.

Here we will describe one of the applications we have chosen in which to implement these ideas, one which we feel is a visually rich domain—that of an art museum. The museum contains both two- and three-dimensional objects spatially arranged in many different rooms. We chose the museum application because it is a kind of spatial information space within which we can formulate a task level description fairly easily. Based on the chosen task domain, we interviewed several architects, museum designers, and interactive exhibit designers to find out for what basic tasks they might want

assistance. This formed the basis for the task analysis that underlies the framework for the virtual museum system. Further details can be found in Drucker (1994) and in Drucker and Zeltzer (1994).

System design

The overall structure of the Virtual Museum system is based on a framework for specifying and controlling the placement and movement of virtual cameras. This framework is proposed as a formal specification for many different types of camera control.

The central notion of this framework is that camera placement and movement is usually done for particular reasons, and that those reasons can be expressed formally as a number of constraints on the camera parameters. We identity these constraints based on analysis of the tasks required in the museum. The entire framework involves a network of camera modules which encapsulate user control, constraints, and branching conditions between modules. The work presented here does not cover the entire framework, but concentrates on the components of individual camera modules, some of the types of constraints for the camera, and different interfaces that can be built to the system. A more complete description of the entire framework is available in Drucker (1994).

Our concept of a camera module is similar to the concept of a shot in cinematography. A shot represents the portion of time between the starting and stopping of filming a particular scene. Therefore, a shot represents continuity of all the camera parameters over that period of time. The unit of a single camera module requires an additional level of continuity, that of continuity of control of the camera. This requirement is added because of the ability in computer graphics to identically match the camera parameters on either side of a cut, blurring the distinction of what makes up two separate shots. Imagine that the camera is initially pointing at character A and following him as he moves around the

environment. The camera then pans to character B and follows her for a period of time. Finally, the camera pans back to character A. In cinematic terms, this would be a single shot since there was continuity in the camera parameters over the entire period. In our terms, this would be broken down into four separate modules. The first module's task is to follow character A. The second module's task would be to pan from A to B in a specified amount of time. The third module's task would be to follow B. And, finally, the last module's task would be to pan back from B to A.

The notion of breaking this cinematic shot into 4 modules does not specify implementation, but rather a *formal description* of the goals or constraints on the camera for each period of time. Most of the modules that are present in the virtual museum are fairly straightforward, and could be implemented in many different fashions. It is only the most complicated modules, e.g., those that handle moving along a computer generated path, that show the utility of the framework, since they combine complex movements with multiple other constraints.

The Virtual Museum project capitalizes upon the W3D system, an extension to the 3D virtual environment software testbed developed at MIT. The Virtual Museum system is structured to emphasize the division between the virtual environment database, the camera framework, and the interface that provides access to both. The system contains the following elements.

- A general interpreter that can run pre-specified scripts or manage user input. The interpreter is an important part in developing the entire runtime system. Currently, the interpreter used is TCL, with the interface widgets created with TK. Many commands have been embedded in the system, including the ability to do dynamic simulation, visibility calculations, finite element simulation, matrix computations, and various database inquiries. By using an embedded interpreter, we can do rapid prototyping of a virtual environment without sacrificing too much performance since a great deal of

the system can still be written in a low level language like C. The addition of TK provides convenient creation of interface widgets and interprocess communication. This is especially important because some processes might need to perform computationally intensive parts of the algorithms; they can be off loaded onto separate machines.

- A built-in renderer. This subsystem can use either the hardware of a graphics workstation (currently Silicon Graphics (SGI) and Hewlett-Packard (HP) workstations are supported), or software to create a high quality anti-aliased image.

- An object database for a particular environment. In this case, the database is the virtual museum which has pre-calculated colors based on radiosity computations which the W3D system supports. The database also contains information about the placement and descriptions of all artwork within the museum.

- Camera modules. Essentially, the camera modules encapsulate the behavior of the camera for different styles of interaction. They are pre-specified by the user and associated with various interface widgets. Several widgets can be connected to several camera modules. The currently active camera module handles all user inputs and attempts to satisfy all the constraints contained within the module, in order to compute camera parameters which will be passed to the renderer when creating the final image. Currently, only one camera module is active at any one time, though if there were multiple viewports, each of them could be assigned a unique camera.

There are 7 different types of interface widgets that can be used to control the camera within the museum. These different widgets illustrate different styles of interaction based on the task level goals of the user.

Camera modules

The generic camera module will contain the following components:

- the local state vector. This must always contain the camera position, camera view normal, camera "up" vector, and field of view. State can also contain values for the camera parameter derivatives, a value for time, or other local information specific to the operation of that module. While the module is active, the state's camera parameters are output to the renderer.

- initializer. This is a routine that is run upon activation of a module. Typical initial conditions are to set up the camera state based on a previous module's state.

- controller. This component translates user inputs either directly into the camera state or into constraints, there can be at most one controller per module.

- constraints to be satisfied during the time period that the module is active. Some examples of constraints are:

    - maintain the camera's up vector to align with world up.

    - maintain height relative to the floor.

    - maintain the camera's gaze (i.e., view normal) toward a specified object.

    - maintain the camera's position on a collision free path through world.

In this system, a constraint can be viewed simply as a black box that produces values for some degrees-of-freedom (DOFs) of the camera. The constraint solver combines these constraints to come up with the final camera parameters for a particular module. Some constraints are desired values for a degree of freedom, for example, specifying the up vector for the camera or the height of the camera. Some involve calculations that might

produce multiple DOFs, such as adjusting the view normal of the camera to look at a particular object. Some, like the path planning constraint, are quite complicated, and construct a path through the environment based on an initial and final position. This allows the user to see objects within the museum based on some spatial context or sequence. At any one time step, the path planning constraint still produces only 2 DOFs for the camera: the x & y position in world space. In the virtual museum system, modules are activated by selecting the corresponding interface widget. The selected widget also passes information to the controller of the module.

Path planning

The most complicated constraint in the current framework is used to achieve automatic navigation. The path planning process is decomposed into several sub-algorithms, many of which can be precomputed in order to speed calculation as much as possible. First, a general description of the overall process is given, then more detailed descriptions of each sub-algorithm follow. The problem of traveling from one point in the museum to another point is first decomposed into finding which doors to travel through. A node to node connectivity graph is pre-computed based on the accessibility between adjacent rooms in the environment. Accessibility can either be indicated by hand, or by an automatic process which uses a rendered image of the building floor, clipped at door level, and a simple visibility test between points on either side of a door. This visibility graph can be updated based on special accessibility requirements (such as handicapped access between rooms).

Traversing the graph is done by a well known graph searching technique called A* (Hart *et al,* 1968). The A* process produces a list of straight-line node-node paths. Paths then need to be computed between each node to avoid obstacles within each room.

This algorithm is optimized for finding paths that originate or terminate at a doorway, so another algorithm must be used to navigate from one point to another point within a room. This second algorithm can also deal with a partially dynamic environment as opposed to the strictly static environment discussed in the first algorithm. Finally, a method for generating a guided tour through the environment has also been developed.

Summary

We have presented an overall framework for exploring camera controls in a 3D virtual environment. Special constraints based on an analysis of task requirements can be designed and combined with a host of other constraints for camera placement. Interfaces can be connected to the system to explore human factors issues while maintaining a consistent underlying structure. We feel that it is important to separate the underlying framework which can incorporate task level requirements from the user interface.

Future work can be in several different directions. More efficient path planning algorithms can be substituted into the camera module framework as they are implemented. In particular, algorithms to deal with totally dynamic environments would be useful. One common task in many virtual environments is the presentation of the information to a third party observer. While the path planning constraint goes toward convenient automatic presentation, a number of other considerations must be made, including the difficult problem of editing a single move into several, smaller cuts. We are incorporating a variety of constraints from cinematography into the camera framework, and work is progressing on techniques that combine those constraints in a meaningful fashion.

References/Publications/Theses

Charniak, Eugene and Drew McDermott. *Introduction to Artificial Intelligence*. Reading, MA: Addison-Wesley, 1985.

Drucker, S. M. *Intelligent camera control for graphical environments*. Ph. D. Thesis, 1994, Massachusetts Institute of Technology, Cambridge, MA.

Drucker, S. M. and D. Zeltzer. "Intelligent Camera Control in a Virtual Environment." In: *Proceedings of Graphics Interface '94*, Banff, Alberta, Canada, May 16-20, 1994.

Hart, P. E., N. J. Nilsson and B. Raphael. "A Formal Basis for the Heuristic Determination of Minimum Cost Paths." *IEEE Transactions on Systems Science and Cybernetics*, 4(2), 100-107 (1968).

Maes, P. "How to Do the Right Thing." *Connection Science*, Vol. 1, No. 3, 1989.

Zeltzer, D. (1991). "Task Level Graphical Simulation: Abstraction, Representation and Control." In: *Making Them Move: Mechanics, Control and Animation of Articulated Figures*, N. Badler, B. Barsky and D. Zeltzer, eds., San Mateo CA, Morgan Kaufmann, pp. 3-33.

Zeltzer, D. "Motor Problem Solving for Three Dimensional Computer Animation", Proceedings of Proc. L'Imaginaire Numerique, (May 14-16, 1987).

MULTIMODAL NATURAL DIALOGUE

Relevant Personnel:

Work in this area was completed under the direction of **Dr. Richard A. Bolt,** Senior Research Scientist and member of the Media Lab's Perceptual Computing Group.

* * * * * * *

In the second year, we made significant progress in the following areas:

Body model

There can be many ways to track the position and dynamics of a person's body, from devices which are attached directly to the body, such as an arrays of light-emitting diodes (LED's) which are tracked by stereo-cameras, magnetic space-sensors and gesture tracking gloves (such as we use), to image processing cameras which are situated away from the wearer, and may use techniques such as neural net image processing. Regardless of manner in which dynamic and positional information about the body is captured at the "sensor" level, the main interest for our main multimodal interpreter module is to have a dynamic picture of the body's position and motion.

Our initial body-tracking system was simply a left-right pair of DataGloves™, which relied upon optic fiber technology to measure the curvature of the fingers; now, we are using a left-right pair of CyberGloves™, by Ascension Technologies, which relies upon small strain gauges which embedded in the surface of the glove material. Later, we may be using cameras at a distance, outfitted with image-processing software, to capture the body image. Yet more efficient and less obtrusive technologies may come along in the future. Across all of these technologies, and indifferent to them, the need is for a data level in the system for the registration of body dynamic and position which is

independent of the particular sensing technology used, and thus insulates the balance of the system upstream of the body sensors from any particular form of sensor.

This last year saw the development of a Body Model to incorporate system knowledge of the user's bodily position and actions. The limb position readings are taken from a small magnetic space sensing cube attached to the upper cuff of either CyberGlove™, that is, on the distal end of the forearm. A third space-sensing cube is attached to inside of a lightweight jacket worn by the user, and positioned at the nape of the neck. (We have been experimenting with the placement of this third cube, trying it out at different positionings along the axis of the wearer's spine.) A fourth and final cube is worn on the user's head, outrigged from our head borne eyetracker apparatus; this cube is used as well, along with the position (attitude) of the user's eye within the head, to determine the wearer's point-of-regard in the surround.

This body model is, as noted, device-independent, and offers 2 encodings of the user's body:

- positional information for upper body parts
- joint angle information

The body model is modular on the network, so that other working groups about the Lab can also tap into information about the user's movements. The body model works in a *client server* mode, in that it makes available information about user position and actions upon request from system modules external to it. The client systems can stipulate what sub-sets of data are will be delivered from the server, and at what data rates.

Gesture

We developed a new scheme for the low-level processing of gesture input . The underlying approach we have adopted, in general, to the processing of gesture data is to

not attempt to match the input to "templates;" rather, the raw data from the glove sensors is processed to intermediate levels, that is, recorded as positions and actions, but short of interpretation as to meaning.

The reason for this is not bind the user to any particular form of gesture to express any specific intention, particularly in the realm of "iconic" gesture, wherein the hand stands for some thing or action. For example, in our videotape "ICONIC," presented at the 1994 CHI Conference in Boston, MA, we showed the user creating an item—a teapot— and placing it in position on a tabletop. The user then says, "Turn the teapot." The system's interpreter module first takes in the spoken words via its speech recognizer unit (The HARK speech recognition system, by BBN Systems, Inc., of Cambridge, MA), and analyzes it as to speaker intent. The command is coherent, yet incomplete: which way to turn the teapot? The strategy of the interpreter is to actively seek the balance of the meaning by searching the space of gesture: what the user's hands were doing when they uttered the command. (All input to our system—speech, gestures, and eye fixations— are time-stamped so that they may be temporally matched.)

In our video, the user, concurrent with saying "Turn the teapot," makes a two-handed, counter-clockwise gesture—somewhat like holding a large bowl with the two hands apart, and moving them along circular paths opposite one another, as if turning a steering wheel. However, a "stirring" gesture with one hand, or a jar-cap twist with one hand had ought to be sufficient, as well as (ideally) any other gestural pattern which favors one direction about the teapot's central axis over the other.

Similarly, should the user say, for example, "...close that door..." and there are two (graphical) doors on the display which happen to be open, it is not a matter of insisting, when examining gestural input to complete the meaning of the sentence, that the user be pointing to the vicinity of either door in the canonical hand-pointing gesture: hand elevated and extended, index finger extended, the other fingers and thumb furled tightly

into a fist. The residual uncertainty after the interpreter examines the spoken input is one bit: is it the door to the left, or the door to the right? Thus, *any simple action on the part of the user would be sufficient:* a twist of one hand towards either side; a twitch of either hand; a twist or shrug of the shoulder; a jerk of the head—all or any of these, or even the canonical index-finger pointing gesture of either hand. The value added in this non-template approach is that the user need not memorize or practice or any single gesture to indicate this or that. They simply "behave" in the way they naturally or spontaneously would; the system is deliberately designed to be "proactive" in seeking in the realm of gesture, and of glance as well, any movement or position of the hand(s) or the eyes that could resolve the uncertainty of the reference initiated in words.

Graphical context also plays an active part in interpreting the user's gesture. The graphical imagery on the display screen and the items and objects thereon portrayed affect the manner and style of gestures that the user is apt to employ when attempting to refer to those items and objects. For example, in our ICONIC videotape, the user sets up the teapot in the first place by saying "On the table (two-handed gesture in near-body space defining the side edges of the table)...place a teapot (closed hand placed in spot of the table's 'surface')...". The two-handed gesture defining the left/right edges of the tables contains an interesting detail: either hand is momentarily positioned such that palms are facing side-by-side, the implied width of the table apart, but the fingers, while held together, are also folded at a 90-degree angle at the joint where they meet the palm. In other words, the shape of either hand "cups" either outer corner of the table. This is, indeed, of the essence of "iconic" gesture," wherein the hands stand for the item referenced (McNeill, 1992, p.12+).

As for symbolic gestures, like the "thumb's up" sign, that have a societal-wide conventional, perhaps the best approach there is template matching; a particular, standard hand configuration is the hallmark of such gestural signals. In a confined and limited

context, particularly where rapid system response is mandatory, this may be the better approach. But, for the balance of coverbal gestures which are spontaneous and non-uniform, a template approach is apt to confine the user to pre-formed gestural repertoires, which they must perforce learn prior to using the system. That the system lets them do whatever, and then works hard to make sense of whatever actions were performed seems the better approach for a non-restrictive system that lets the user "be themselves."

We note, in any case, that our gesture analyzer produces gestural frames, like simplified "cartoons" of the structure of the users body actions, including eye movements. We use the joint angles from the body model to do the first level segmentation of gestural input into movements into gestural units or building blocks.

History

In the spatio-temporal representation, we added the *temporal* dimension allowing the system to record actions and events as they occur. This encoding can then be used to resolve references backwards in time: the user can describe an event, and the system can go back in the x,y,z,t stream and determined which past event or action the user is referring to.

Examples of this kind of referencing might be:

"Go back to when...(looking a/o pointing to locations on a map display)...the tank column crossed that road...".

Go back to when the sortie came in like this (swooping hand gesture from top-right to lower left of terrain map display)..."

The schema should be adaptable for forward-looking temporal references, such as when we stipulate some system action contingent upon some criteria to be met at a future time. This process involves the setting up of a model of the anticipated event, and executing it

when the conditions are met (e.g., when the red vehicle crosses the road, do such-and -such).

Unresolved issues and problems associated with the addition of the time dimension includes: how much to keep of the original interaction data. That is, do we record the spatial coordinates of the 3-dimensional path in user body space of the hand as each gesture is made—so as to allow matching to a performed *action*—or, do we instead simply record the trail over the terrain as that trail is mapped from the gestural indication. Put another way, do we enable reference to a past gesture, or to some thing or action referenced by a past gesture? Perhaps the key to this is to attempt to assess whether the user is apt to refer *their own past act* —in effect, "Go back to when I did this (they re-enact some past speech-gesture-glance act)...", or whether the user is apt to refer to the consequences of some past act of their, rather than to the act per se. To be optimally effective, the system ideally ought let the user refer to some past happening, or to some anticipated one, through either style of expression; like avoidance of "templates" at the level of gestural interpretation, this "be ready for anything" stance would maximize the expression options of the user, and not burden them with having to recall specifics before they may express themselves to the system.

Part base interface to graphics

We continued to work on a part-base approach to object representation. This approach has 2 important features:

> 1) it is object-oriented, in that it allows the user to talk about objects in the scene. The schematization of objects and relations in the scene allows for the abstraction of spatial information out of the graphics, which allows for their integration with descriptions from speech and gestures.

2) it supports an hierarchical structure of the scene structure, and it allows to describe changes to the scene, new states, and new events by imposing new scene strictures on the database.

This part-base approach allows a very-high level description of user actions or changes to the graphics domain, and the interface language used during the resolution of reference is compatible the spatio-temporal system.

We are now in a position to implement the part-base interface on top of the SGI "Inventor" system. A big advantage to doing this is that the SGI platform seems to be the graphics platform of choice of choice for the project (e.g., the VLW is using SGI graphics engines).

References/Publications/Theses

Koons, David B., Carlton J. Sparrell. ICONIC: Speech and depictive gestures at the Human-Machine Interface. In: Catherine Plaisant (ed.), *CHI '94 Human Factors in Computing Systems: Conference Companion*. CHI '94, Boston, MA, April 24-28, 1994, page 453-454. Also presented as a **videotape** in the CHI '94 Formal Video Program.

McNeill, David. *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press, 1992.

*U.S. GOVERNMENT PRINTING OFFICE: 1996-710-126-47017

DISTRIBUTION LIST

addresses	number of copies
RICHARD T. SLAVINSKI ROME LABORATORY/C3AB 525 BROOKS RD. ROME NY 13441-4505	25
MIT MEDIA LAB ATTN: DR. R. BOLT 20 AMES ST. CAMBRIDGE MA 02139	5
ROME LABORATORY/SUL TECHNICAL LIBRARY 26 ELECTRONIC PKY ROME NY 13441-4514	1
ATTENTION: DTIC-OCC DEFENSE TECHNICAL INFO CENTER 8725 JOHN J. KINGMAN ROAD, STE 0944 FT. BELVOIR, VA 22060-6218	2
ADVANCED RESEARCH PROJECTS AGENCY 3701 NORTH FAIRFAX DRIVE ARLINGTON VA 22203-1714	1
ROME LABORATORY/C3AB 525 BROOKS RD ROME NY 13441-4505	1
NAVAL WARFARE ASSESSMENT CENTER GIDEP (QA50) ATTN: RAYMOND TADROS PO BOX 8000 CORONA CA 91718-8000	1
WRIGHT LABORATORY/AAAI-2, BLDG 635 2185 AVIONICS CIRCLE WRIGHT-PATTERSON AFB OH 45433-7501	1

AFIT ACADEMIC LIBRARY/LDEE 1
2950 P STREET
AREA B, BLDG 642
WRIGHT-PATTERSON AFB OH 45433-7765

WRIGHT LABORATORY/MLPO 1
ATTN: R. L. DENISON
BLDG 651
3005 P STREET, STE 6
WRIGHT-PATTERSON AFB OH 45433-7707

AL/CFHI, BLDG 248 1
ATTN: GILBERT G. KUPERMAN
2255 H STREET
WRIGHT-PATTERSON AFB OH 45433-7022

AIR FORCE HUMAN RESOURCES LAB 1
TECHNICAL DOCUMENTS CENTER
DL AL HSC/HRG, BLDG 190
WRIGHT-PATTERSON AFB OH 45433-7604

AUL/LSE 1
BLDG 1405
600 CHENNAULT CIRCLE
MAXWELL AFB AL 361126+24

US ARMY SPACE & STRATEGIC 1
DEFENSE COMMAND
CSSD-IM-PA
PO BOX 1500
HUNTSVILLE AL 35807-3801

NAVAL AIR WARFARE CENTER 1
6000 E. 21ST STREET
INDIANAPOLIS IN 46219-2189

COMMANDING OFFICER 1
NCCOSC RDTE DIVISION
ATTN: TECHNICAL LIBRARY,CODE 0274
53560 HULL STREET
SAN DIEGO CA 92152-5001

COMMANDER, TECHNICAL LIBRARY 1
4747000/C0223
NAVAIRWARCENWPNDIV
1 ADMINISTRATION CIRCLE
CHINA LAKE CA 93555-6001

SPACE & NAVAL WARFARE SYSTEMS
COMMAND, EXECUTIVE DIRECTOR (PD30A)
ATTN: MR. CARL ANDRIANI
2451 CRYSTAL DRIVE
ARLINGTON VA 22245-5200 1

COMMANDER, SPACE & NAVAL WARFARE
SYSTEMS COMMAND (CODE 32)
2451 CRYSTAL DRIVE
ARLINGTON VA 22245-5200 1

US ARMY MISSILE COMMAND
AMSMI-RD-CS-R/DOCUMENTS
RSIC BLDG 4484
REDSTONE ARSENAL AL 35898-5241 2

LOS ALAMOS NATIONAL LABORATORY
PO BOX 1663
REPORT LIBRARY, P364
LOS ALAMOS NM 87545 1

AEDC LIBRARY
TECHNICAL REPORTS FILE
100 KINDEL DRIVE, SUITE C211
ARNOLD AFB TN 37389-3211 1

COMMANDER
USAISC
ASHC-IMD-L, BLDG 61801
FT HUACHUCA AZ 85613-5000 1

AFIWC/MSO
102 HALL BLVD, STE 315
SAN ANTONIO TX 78243-7016 1

DIRNSA
R509
9800 SAVAGE ROAD
FT MEADE MD 20755-6000 1

NSA/CSS
K1
FT MEADE MD 20755-6000 1

ESC/IC 1
50 GRIFFISS STREET
HANSCOM AFB MA 01731-1619

PHILLIPS LABORATORY 1
PL/TL (LIBRARY)
5 WRIGHT STREET
HANSCOM AFB MA 01731-3004

THE MITRE CORPORATION 1
ATTN: E. LABUPE
D460
202 BURLINGTON RD
BEDFORD MA 01732

OUSD(P)/DTSA/DUTD 2
ATTN: PATRICK G. SULLIVAN, JR.
400 ARMY NAVY DRIVE
SUITE 300
ARLINGTON VA 22202

SDIO/S-PL-RM 1
ATTN: CAPT JOHNSON
THE PENTAGON
WASH DC 20301-7000

SDI TECHNICAL INFORMATION CENTER 1
1735 JEFFERSON DAVIS HIGHWAY #703
ARLINGTON VA 22202

ESD/ATN 1
ATTN: COL LEIB
HANSCOM AFB MA 01731-5000

NAVAL AIR DEVELOPMENT CTR 1
ATTN: DR. MORT METERSKY
CODE 300
WARMINSTER PA 189974

ESD/XTS 1
ATTN: LT COL JOSEPH TOOLE
HANSCOM AFB MA 01731

USA-SDC CSSD-H-SBE 1
ATTN: MR. DOYLE THOMAS
HUNTSVILLE AL 35807

HQ AFSPACECOM/DOXP 1
ATTN: CAPT MARK TERRACE
STOP 7
PETERSON AFB CO 80914

USSPACECOM/J5B 1
ATTN: LT COL HAROLD STANLEY
PETERSON AFB CO 80914

NTB JPO 1
ATTN: MR. NAT SOJOUNER
FALCON AFB CO 80912

NASA LANGLEY RESEARCH CENTER 1
ATTN: WAYNE BRYANT
MS578
HAMPTON VA 23665-5225

AL/LRG 1
ATTN: MR. M. YOUNG
WRIGHT-PATTERSON AFB OH 45433-6503

AL/CFHV 1
ATTN: DR. B. TSOU
WRIGHT-PATTERSON AFB OH 45433-6503

AL/HEO 1
ATTN: MAJ M. R. MCFARREN
WRIGHT-PATTERSON AFB OH 45433-6503

WL/KTO 1
ATTN: DR. D. HOPPER
WRIGHT-PATTERSON AFB OH 45433-6503

AFIT/ENG 1
ATTN: LT COL M. STYTZ
WRIGHT-PATTERSON AFB OH 45433-6503

USA ETL 1
ATTN: MR. R. JOY
CECOM-C4-D
FT BELVOIR VA 22060

SPAWAR 1
ATTN: MR. J. PUCCI
CODE 32410
2511 JEFFERSON DAVIS HIGHWAY
WASH DC 20363-5100

ATZL-CDC-S 1
ATTN: CAPT CHARLES ALLEN III
FT LEAVENWORTH KS 66027-5300

NUSC 1
ATTN: MR. L. CABRAL
NEWPORT RI 02841-5047

NOSC (CODE 414) 1
ATTN: MR. P. SULTAN
271 CATALINA BLVD
SAN DIEGO CA 92151-5000

NTSC (CODE 251) 1
ATTN: MR. D. BREGLIA
12350 RESEARCH PARKWAY
ORLANDO FL 32826-3224

NASA 1
ATTN: DR. J. ROBERTSON
LANGLEY RESEARCH CENTER
MAIL CODE 152E
HAMPTON VA 23665-5225

NAVAL OCEAN SYSTEMS CENTER 1
ATTN: GLEN OSGA, PHD
CODE 444
SAN DIEGO CA 92152

ARI FORT LEAVENWORTH 1
ATTN: MAJ ROB REYENGA
P. O. BOX 3407
FT LEAVENWORTH KS 66027-0347

ARI FIELD UNIT 1
ATTN: JON J. FALLESEN
P. O. BOX 3407
FORT LEAVENWORTH KS 66027-0347

RL/ESOP 1
ATTN: MR. JOSEPH HORNER
HANSCOM AFB MA 01731

JOINT NAT'L INTEL DEVELOP STAFF 1
ATTN: CAPT JOHN HILBING
4600 SILVER HILL ROAD
WASH DC 20389