University of Southern California

On the Knowledge Underlying Multimedia Presentations

Yigal Arens, Eduard H. Hovy, and Mirra Vossers
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292

1992
ISI/RR-93- 370

INFORMATION SCIENCES INSTITUTE
310/822-1511
4676 Admiralty Way/Marina del Rey/California 90292-6695

Best
Available
Copy

On the Knowledge Underlying Multimedia Presentations

Yigal Arens, Eduard H. Hovy, and Mirra Vossers
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292

1992
ISI/RR-93- 370

The first author was supported in part by Rome Laboratory of the Air Force Systems Command and the Defense Advanced Research Projects Agency under contract no. F30602-91-C-0081. The second author was supported in part by Rome Laboratory of the Air Force Systems Command under RL contract no. FQ7619-89-03326-0001. The third author, a graduate student at the University of Nijmegen, The Netherlands, spent a research visit of six months at USC/ISI working on this project for her Master's degree. Views and conclusions contained in this report are the author's and not interpreted as representing the official opinion or policy of DARPA, RL, the U.S. Government, or any person or agency connected with them.

REPORT DOCUMENTATION PAGE

FORM APPROVED
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
June 1992

3. REPORT TYPE AND DATES COVERED
Research Report

4. TITLE AND SUBTITLE
On the Knowledge Underlying Multimedia Presentations

5. FUNDING NUMBERS
F30602-91-C-0081 and
FQ7619-89-03326-0001

6. AUTHOR(S)
Yigal Arens, Eduard Hovy, and Mira Vossers

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
USC INFORMATION SCIENCES INSTITUTE
4676 ADMIRALTY WAY
MARINA DEL REY, CA 90292-6695

8. PERFORMING ORGANIZATON REPORT NUMBER
RR-370

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
ARPA
3701 N. Fairfax Drive
Arlington, VA 22203-1714

ROME Laboratories
Rome, New York

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12A. DISTRIBUTION/AVAILABILITY STATEMENT
UNCLASSIFIED/UNLIMITED

12B. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We address one of the problems at the heart of automated multimedia presentation production and interpretation. The media problem can be stated as follows: how does the producer of a presentation determine which information to allocate to which medium, and how does a perceiver recognize the function of each part as displayed in the presentation and integrate them into a coherent whole? What knowledge is used, and what processes? We describe the four major types of knowledge that play a role in the allocation problem as well as interdependencies that hold among them. We discuss two formalisms that can be usedto represent this knowledge and, using examples, describe the kinds of processing required for the media allocation problem.

14. SUBJECT TERMS
multimedia presentations, human-computer interaction, presentation planning

15. NUMBER OF PAGES
28

16. PRICE CODE

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

On the Knowledge Underlying Multimedia Presentations

Yigal Arens, Eduard Hovy, and Mira Vossers[1]

Information Sciences Institute
of the University of Southern California
4676 Admiralty Way
Marina del Rey, CA 90292-6695
Tel: (310) 822-1511
Fax: (310) 823-6714
Email: {ARENS, HOVY}@ISI.EDU

ABSTRACT

We address one of the problems at the heart of automated multimedia presentation production and interpretation. The media allocation problem can be stated as follows: how does the producer of a presentation determine which information to allocate to which medium, and how does a perceiver recognize the function of each part as displayed in the presentation and integrate them into a coherent whole? What knowledge is used, and what processes? We describe the four major types of knowledge that play a role in the allocation problem as well as interdependencies that hold among them. We discuss two formalisms that can be used to represent this knowledge and, using examples, describe the kinds of processing required for the media allocation problem.

[1]This author, a graduate student at the University of Nijmegen, Nijmegen, The Netherlands, spent six months at USC/ISI and has since graduated.

1 The General Problem of Presentations using Multiple Media

When communicating, people almost always employ multiple modalities. Even natural language, which is after all the most powerful representational medium developed by humankind, is usually augmented by pictures, diagrams, etc., when written, or by gestures, hand and eye movements, intonational variations, etc., when spoken. And this preference for multimodality carries over to communication with computational systems, as evidenced by the explosive growth of the field of Human-Computer Interfaces. Since the early dream of Artificial Intelligence — of creating fully autonomous intelligent agents that would interact with people as equals — has proved impossible to achieve in the near term, the thrust of much AI work is on the construction of semi-intelligent machines operating in close symbiosis with humans, forming units. For maximum ease of communication within such units, natural language and other human-oriented media are the prime candidates (after all, computers are easier to program than humans are).

How then can computers construct and analyze such multimedia presentations? A survey of the literature on the design of presentations (book design, graphic illustration, etc.; see [Tufte 90, Bertin 83, Tufte 83]) underscores how this area of communication remains an art and shows how hard it is to describe the rules that govern presentations. But people clearly do follow rules when they use several modalities to construct communications; textbooks, for example, are definitely not illustrated randomly. Psychologists have been studying multimedia issues such as the effects of pictures in text, design principles for multimedia presentation, etc. for many years [Hartley 85, Twyman 85, Dwyer 78, Fleming & Levie 78], although most of their results are too general to be directly applicable in work that is to be computationalized. On the other hand, cognitive science studies of the past few years have provided results which can be incorporated into theories about good multimedia design [Petre & Green 90, Roth & Mattis 90, Mayer 89, Larkin & Simon 86]. They address questions such as whether graphical notations are really superior to text, what makes a picture worth (sometimes) a thousand words, how illustration affects thinking, the characterization of data, etc.

Artificial Intelligence researchers and other computer scientists have been addressing aspects of the problem of automatically constructing multimedia presentations as well. [Mackinlay 86] describes the automatic generation of a variety of tables and charts; the WIP system of [Wahlster et al. 92, André & Rist 92] (and see this volume) plans a text/graphics description of the use of an espresso machine, starting with a database of facts about the machine and appropriate communicative goals, and using text and presentation plans. The COMET system [Feiner 88, Feiner & McKeown 90] plans text/graphic presentations of a military radio using text schemas and pictorial perspective presentation rules. The AIMI system [Maybury 91, Burger & Marshall 91] (and see this volume) plans text/map/tables presentations of database information about military operations and hardware, also using presentation plans. Similarly, the INTEGRATED INTERFACES system [Arens et al. 88] and the CUBRICON system [Neal 90] plan and produce presentations involving maps, text, and menus. Other work is reported in the collections [Sullivan & Tyler 91, Ortony et al. 92].

One lesson that is clear from all this work is the need for a detailed study of the major types of knowledge required for multimedia presentations, encoded in a formalism that supports both their analysis and generation. For the past few years, we have been involved in various studies of one aspect or another of this problem. In particular, we ask: why and how do people apportion the information to be presented to various media? And how do they reassemble the portions into a single message again? This paper contains an overview of some of our results. Section 2 describes our methodology and formalisms. Section 3 provides details about the features and their interdependencies that we have managed to collect, and Section 4 provides some examples of the use of this knowledge.

2 Our Approach and Methodology

2.1 The Problem of Media Allocation

In order to focus our efforts, we have concentrated on the media allocation problem: given arbitrary information and any number of media, how, and on what basis, is a particular medium selected for the display of each portion of the information? This question, a particularization of the question why people use different media and other gestures and movements when they communicate, in our opinion lies at the heart of the general multimedia issue.

Rather than start with a literature study, we here describe the problem from the computational side. In most systems, the media allocation problem is addressed simply by the use of fixed rules that specify exactly what medium is to be used for each particular data type. This is clearly not a satisfactory solution, given the inflexibility and non-portability of such systems. Our approach is a *two-stage generalization* of this straightforward approach. We take an example from a hypothetical data base about ships in a Navy to illustrate. Under the straightforward approach, a typical rule may be:

1. Ships' locations are presented on maps.

Our first generalization is to assign a medium not to each data type, but instead to each feature that characterizes data types. Thus instead of rule 1, we write the rule:

1'. Data duples (of which ships' locations are an example) are presented on maps, graphs, or tables.

Of course, when considering subsets of features, one invariably gets underspecific rules. To provide more specificity we formulate such additional rules as:

2. Data with spatial denotations (such as locations) are presented on media with spatial denotations (such as maps).

However, note that this rule deals not with the medium of maps but instead with a characteristic of this medium. It suggests the second step of the generalization.

The second generalization is to assign characteristics of data not to media, but instead to characteristics of media The two example rules now become:

1. Data duples (of which locations are an example) are presented on planar media (such as graphs, tables, and maps).
2. Data with spatial denotations (such as locations) are presented on media with spatial denotations (such as maps).

In this example, the two rules together suffice to specify maps uniquely as the appropriate medium for location coordinates. Of course, though, one can present the same information using natural language, as in *"the ship is at 15N 79E"*. Thus one is led to rephrase rule 2 to arrive at a more general but very powerful formulation:

2'. Data with specific denotations are presented on media which can convey the same denotations.

Since language, pictures, and maps can carry spatial denotations (while, say, graphs or histograms usually do not), we once again require additional rules in order to specify a unique medium. However, since each of the three mentioned media can be perfectly suitable. In the right context, the rules we formulate might not absolutely prohibit a medium; rather, the rules should be context-dependent in ways which enable the selection of the most appropriate medium. Thus we are led to rules such as:

3. If more than one medium can be used, and there is an existing presentation, prefer the medium/a that is/are already present as exhibits in the presentation.
4. If more than one media can be used, and there is additional information to be presented as well, prefer medium/a that can accommodate the other information too.

Rule 4 has important consequences. If one is to present not only the location of a ship, but also its heading, then both language and a map would do, since both media have facilities for indicating direction (in the case of language, an appositive phrase with the value *"heading SSW"*; and in the case of a map, an icon with an elongation or an arrow). If in addition to this now one adds the requirement to present the nationality of the ship, natural language has such a capability (the adjective *"Swiss"*, say) but due to limitations of the map medium, one of the icon's independent characteristics (say, its color) must be allocated to convey nationality. Of course, this requires the addition of a description of the meaning of the different values the icon's independent characteristic can have (for example, a table of color for nationality). Such additional presentational overhead makes a map a less attractive medium than natural language for presenting a single ship's location/heading/nationality (though possibly not that of several ships together).

We formalize and discuss this point later in more detail. Here it is enough to note that the two-stage generalizations provide collections of rules that relate characteristics of information and characteristics of media in service of good multimedia presentations. In general terms, the medium allocation algorithm required can be described as a constraint satisfaction system, where the constraints arise from rules requiring the features of the information to be presented (i.e., the data) to be matched up optimally with the features of the media at hand.

2.2 The Four Types of Knowledge Required

We illustrated the use of knowledge about media and information type. But what additional factors play a role in multimedia communication?

In our previous work in multimedia human-computer interactions [Arens et al. 92, Vossers 91, Hovy & Arens 91, Hovy & Arens 90, Arens & Hovy 90a, Arens & Hovy 90b], we addressed this question from several angles, trying to build up a library of terms that capture all the factors that play a role in multimedia human-human and human-computer communication. Drawing from an extensive survey of literature from Psychology, Human-Computer Interfaces, Natural Language Processing, Linguistics, Human Factors, and Cognitive Science, (see [Vossers 91]) as well as from several small analyses of pages from newspapers such as the *USA Today* and instruction manuals for appliances such as user manuals for a motor car, a sewing machine, a VCR, and a cookbook, we collected well over a hundred distinct features that play a role in the higher-level aspects of the production and interpretation processes, as well as over fifty rules that express the interdependencies among these features. Where appropriate, we applied the two-step generalization method to come up with features of the right type and at the right level of detail.

These features classify naturally into four major groups:

1. the characteristics of the media used,
2. the nature of the information to be conveyed,
3. the goals and characteristics of the producer, and
4. the characteristics of the perceiver and the communicative situation.

Section 3 provides more details about each type of knowledge resource and the rules interlinking them. Before getting to this section, however, we describe our attempts to find an adequately flexible and powerful representation formalism for the knowledge.

```
  |---------------
X |  X1  X2  X3
Y |  Y1  Y2
Z |  Y2  X2  X3
```

```
                              |-- Y1 ------------
                     /-------Y
                     |        |-- Y2 ----\
     |-- X1 -----{                        |----|
     |               \-- X2 ------------/      |
  ---X                                         Z---
     |                                         |
     |-- X3 -----------------------------------|
```

Figure 1: Equivalent Tabular and Network Representations.

2.3 An Adequate Representation Formalism

Though we did not study all four aspects in equal detail, we needed a representation formalism that could capture the requisite individual distinctions as well as their underlying interdependencies, that was extensible, and that did not hamper our research methodology.

As illustrated in Section 2.1, the two-step generalization process provides features and rules simultaneously. Features and their values we tried to tabulate straightforwardly, until we discovered that the underlying interdependencies between features — for example, the subclassification of some but not all values for a feature into finer classes, or the combination of values from several features to give rise to a new feature — and the interdependencies between rules made the simple tabular format cumbersome. In the spirit of our work on various media, we decided to codify our results in a more visual way, following the paradigm of AND-OR networks of features and values used in Systemic Functional Linguistics to analyze language and write grammars [Halliday 85].

An example table and equivalent network are shown in Figure 1. Processing of the networks is to be understood as similar to discrimination net traversal; one enters the network, makes the appropriate selection(s) at the first choice point(s), records the feature(s) so chosen, and moves along the connecting path(s) to the next choice point(s). In the network, curly brackets mean AND (that is, when entering one, all paths should be followed in parallel) and square brackets EXCLUSIVE OR (that is, at most one path must be selected and followed). Square brackets with slanted serifs are INCLUSIVE OR (that is, zero or more paths may be selected and followed). Whenever a feature is encountered during traversal, it is recorded; the final collection of features uniquely specifies the eventual result.

Using the new notation, our two-stage generalization method could be rephrased as a three-step research methodology: First, we identify the phenomena in some aspect of a presentation (e.g., the fact that the producer usually wants to affect the perceiver's future goals, or the fact that different media utilize different numbers of presentation 'dimensions'); second, we characterize the variability involved in each phenomenon (e.g., a producer may want to affect the perceiver's goals through warnings, suggestions, hints, requests, etc., or language is expressed 'linearly' while diagrams are two-dimensional); and third, we map out the interdependencies among the values of all the phenomena (e.g., the goal to warn selects a feature value 'urgent', and this value is interdependent with values such as 'high noticeability' which are tied to appropriate media such as sound or flashing icons). In the resulting AND-OR networks of interdependencies, each node represents a single phenomenon and each arc a possible value for it together with its interdependencies with other values.

One advantage of the network notation is its independence of process; one can implement the knowledge contained directly in network form, in a traditional rule-based system, or a connectionist one. We maintain the network form because several other presentation-related software at USC/ISI uses the same formalism. The Penman sentence generator [Mann & Matthiessen 83, Penman 88, Hovy 90] and associated text planning system [Hovy et al. 92] contain a grammar of English and various factors influencing text structure all represented as AND-OR networks; sentence generation proceeds by traversing the grammar network from 'more semantic' toward 'more syntactic' nodes, collecting at each node features that instruct the system how to build the eventual sentence (see [Matthiessen 84]). Parsing proceeds by traversing the same network 'backwards', eventually arriving at the 'more semantic' nodes and their associated features, the set of which constitutes the parse and determines the parse tree (see [Kasper & Hovy 90, Kasper 89]). This bidirectionality of processing is an additional advantage of the network formalism.

With respect to multimedia presentation planning and analysis, our overall conceptual organization of the knowledge resources is shown in Figure 2. Each knowledge resource appears as a separate network; the central network houses the interlinkages between the other ones. When producing a communication, the communicative goals and situational features cause appropriate features of the upper three networks to be selected, and information then propagates through the interlinkage network (the system's 'rules') to the appropriate medium networks at the bottom, causing appropriate values to be set, which in turn are used to control the low-level generation modules (the language generator, the diagram constructor, etc.). For multimedia input, a communication is analyzed by identifying its features in the relevant bottom networks for each portion of the communication, and propagating the information upward along the internetwork linkage to select appropriate 'high-level' features that describe the producer's goals, the nature of the information mentioned in that portion, etc. Examples appear in Section 4.

3 The Knowledge Resources

In this section we describe the four major classes of features that influence multimedia presentation planning. In the fifth section we discuss the rules expressing interdependen-

Figure 2: Knowledge Resources that Support Multimedia Communication.

cies among the features in the four classes.

3.1 Characterization of Media

3.1.1 Definition of Terms

The following terms are used to describe presentation-related concepts. We take the point of view of the communicator (indicating where the consumer's subjective experience may differ).

1. Consumer: A person interpreting a communication.

2. Medium: A single mechanism by which to express information. Examples: spoken and written natural language, diagrams, sketches, graphs, tables, pictures.

3. Exhibit: A complex exhibit is a collection, or composition, of several simple exhibits. A simple exhibit is what is produced by one invocation of one medium. Examples of simple exhibits are a paragraph of text, a diagram, a computer beep. Simple exhibits involve the placement of one or more Information Carriers on a background Substrate.

4. Substrate: The background to a simple exhibit. That which establishes, to the consumer, physical or temporal location, and often the semantic context, within which new information is presented to the information consumer. The new information will often derive its meaning, at least in part, from its relation to the substrate. Examples: a piece of paper or screen (on which information may be drawn or presented); a grid (on which a marker might indicate the position of an entity); a page of text (on which certain words may be emphasized in red); a noun phrase (to which a prepositional phrase may be appended). An empty substrate is possible.

5. Information Carrier: That part of the simple exhibit which, to the consumer, communicates the principal piece of information requested or relevant in the current communicative context. Examples: a marker on a map substrate; a prepositional phrase within a sentence predicate substrate. A degenerate carrier is one which cannot be distinguished from its background (in the discussion below the degenerate carrier is a special case, but we do not bother explicitly to except it where necessary. Please assume it excepted).

6. Carried Item: That piece of information represented by the carrier; the 'denotation' of the carrier.

For purposes of rigor, it is important to note that a substrate is simply one or more information carrier(s) superimposed. This is because the substrate carries information as well[2]. In addition, in many cases the substrate provides an internal system of semantics which may be utilized by the carrier to convey information. Thus, despite its name, not all information is transmitted by the carrier itself alone; its positioning (temporal or spatial) in relation to the substrate may encode information as well. This is discussed further below.

7. Channel: An independent dimension of variation of a particular information carrier in a particular substrate. The total number of channels gives the total number of independent pieces of information the carrier can convey. For example, a single mark or icon can convey information by its shape, color, and position and orientation in relation to a background map. The number and nature of the channels depend on the type of the carrier and on the exhibit's substrate.

3.1.2 Internal Semantic Systems

Some information carriers exhibit an internal structure that can be assigned a 'real-world' denotation, enabling them subsequently to be used as substrates against which other carriers can acquire information by virtue of being interpreted within the substrate. For example, a map used to describe a region of the world possesses an internal structure — points on it correspond to points in the region it charts. When used as a background for a ship icon, one may indicate the location of the ship in the world by placing its icon in the corresponding location on the map substrate. Examples of such carriers and their internal semantic systems are shown in Table 1.

Other information carriers exhibit no internal structure. Examples: icon, computer beep, and unordered list.

[2]Note that from the information consumer's point of view, Carrier nd Substrate are subjective terms; two people looking at the same exhibit can interpret its components as carrier and substrate in different ways, depending on what they already know. For example, different people may interpret a graph tracking the daily value of some index differently as follows: someone who is familiar with the history of the index may call only the last point of the graph, that is, its most recent addition, the information carrier, and call all the rest of the graph the substrate. Someone who is unfamiliar with the history of the index may interpret the whole line plotted out as the information carrier, and the graph's axes and title, etc., as substrate. Someone who is completely unfamiliar with the index may interpret the whole graph, including its title and axis titles, as information carrier, and interpret the screen on which it is displayed as substrate.

Carrier	Internal Semantic System
Picture	'real-world' spatial location based on picture denotation
NL sentence	'real-world' sentence denotation
Table	categorization according to row and column
Graph	coordinate values on graph axes
Map	'real-world' spatial location based on map denotation
Ordered list	ordinal sequentiality

Table 1: Internal semantic systems.

An internal semantic system of the type described is always intrinsic to the item carried.

3.1.3 Characteristics of Media

In addition to the internal semantics listed above, media differ in a number of other ways which can be exploited by a presenter to communicate effectively and efficiently. The values of these characteristics for various media are shown in Table 2.

Carrier Dimension: Values: *0D, 1D, 2D.* A measure of the number of dimensions usually required to exhibit the information presented by the medium.

Internal Semantic Dimension: Values: *0D, 1D, 2D, >2D, 3D, #D, ∞D.* The number of dimensions present in the internal semantic system of the carrier or substrate.

Temporal Endurance: Values: *permanent, transient.* An indication whether the created exhibit varies during the lifetime of the presentation.

Granularity: Values: *continuous, discrete.* An indication of whether arbitrarily small variations along any dimension of presentation have meaning in the denotation or not.

Medium Type: Values: *aural, visual.* What type of medium is necessary for presenting the created exhibit.

Default Detectability: Values: *low, medlow, medhigh, high.* A default measure of how intrusive to the consumer the exhibit created by the medium will be.

Baggage: Values: *low, high.* A gross measure of the amount of extra information a consumer must process in order to become familiar enough with the substrate to correctly interpret a carrier on it.

Generic Modality	Carrier Dimension	Int. Semantic Dim.	Temporal Endurance	Granularity	Medium Type	Default Detectability	Baggage
Beep	0D		transient	N/A	aural	high	
Icon	0D		permanent	N/A	visual	low	
Map	2D	>2D	permanent	continuous	visual	low	high
Picture	2D	∞D	permanent	continuous	visual	low	high
Table	2D	2D	permanent	discrete	visual	low	high
Form	2D	>2D	permanent	discrete	visual	low	high
Graph	2D	1D	permanent	continuous	visual	low	high
Ordered list	1D	#D	permanent	discrete	visual	low	low
Unordered list	0D	#D	permanent	N/A	visual	low	low
Written sentence	1D	∞D	permanent	discrete	visual	low	low
Spoken sentence	1D	∞D	transient	discrete	aural	medhigh	low
Animated material	2D	∞D	transient	continuous	visual	high	high
Music	1D	∞D	transient	continuous	aural	med	low

Table 2: Media characteristics.

3.1.4 How Carriers Convey Information

As part of an exhibit, a carrier can convey information along one or more *channels*. For example, with an icon carrier, one may convey information by the icon's shape, color, and possibly through its position in relation to a background map. The number and nature of the channels depends on the type of carrier and the substrate.

The semantics of a channel may be *derived* from the carrier's spatial or temporal relation to a substrate which possesses an internal semantic structure; e.g., placement on a map of a carrier representing an object which exists in the charted area. Otherwise we say the channels is *free*.

Among *free* channels we distinguish between those whose interpretation is *independent* of the carried item (e.g., color, if the carrier does not represent an object for which color is relevant); and those whose interpretation is *dependent* on the carried item (e.g., shape, if the carrier represents an object which has some shape).

Most of the carrier channels can be made to vary their presented value in time. Time variation can be seen as an additional channel which provides yet another degree of freedom of presentation to most of the other channels. The most basic variation is the

Figure 3: Portion of the Media Network: Values for some Text Channels.

alternation between two states, in other words, a flip-flop, because this guarantees the continued (though intermittent) presentation of the original basic channel value.

The fonts and positions of letters and words in a text are also free channels for the words as carriers. Figure 3 contains a fragment of the network describing some possible values for these channels.

3.2 Characterization of Information

In this section we develop a vocabulary of presentation-related characteristics of information.

Broadly speaking, as shown in Table 3, three subcases must be considered when choosing a presentation for an item of information: intrinsic properties of the specific item; properties associated with the class to which the item belongs; and properties of the collection of items that will eventually be presented, and of which the current item is a member. These characteristics are explained in the remainder of this section.

Dimensionality: Some single items of information, such as a data base record, can be decomposed as a vector of simple components; others, such as a photograph, have a complex internal structure which is not decomposable. We define the *dimensionality* of the latter as *complex*, and of the former as the dimension of the vector.

Since all the information must be represented in some fashion, the following rule must hold (where *simple* dimensionality has a value of 0, *single* the value 1, and so on, and *complex* the value ∞):

The Basic Dimensionality Rule of Presentations

□ *Rule:* Dim(Info) $\leq$ Dim(Carrier) + Free Channels(Carrier) + Internal Semantic Dim(Substrate)

Type	Characteristic	Values
Intrinsic Property	Dimensionality	0D, 1D, 2D, >2D, ∞D
	Transience	live, dead
	Urgency	urgent, routine
Class Property	Order	ordered, nominal, quantitative
	Density	dense, discrete, N/A
	Naming	identification, introduction
Set Property	Volume	singular, little, much

Table 3: Information characteristics by type.

In addition, we have found that different rules apply to information of differing dimensions. With respect to dimensionality, we divide information into several classes as follows:

- *Simple:* Simple atomic items of information, such as an indication of the presence or absence of email.
- *Single:* The value of some meter such as the amount of gasoline left.
- *Double:* Pairs of information components, such as coordinates (graphs, map locations), or domain-range pairs in relations (automobile × satisfaction rating, etc.).
- *Multiple:* More complex information structures of higher dimension, such as home addresses. It is assumed that information of this type requires more time to consume.
- *Complex:* Information with internal structure that is not decomposable this way, such as photographs.

Transience: Transience refers to whether the information to be presented expresses some current (and presumably changing) state or not. Presentations may be:

- *Live:* The information presented consists of a single conceptual item of information (that is, one carried item) that varies with time (or in general, along some linear, ordered, dimension), and for which the history of values is not important. Examples are the amount of money owed while pumping gasoline or the load average on a computer. Most appropriate for *live* information is a single exhibit.
- *Dead:* The other case, in which information does not reflect some current state, or in which it does but the history of values is important. An example is the history of some stock on the stock market; though only the current price may be important to a trader, the history of the stock is of import to the buyer.

Urgency: Some information may be designated *urgent*, requiring presentation in such a way that the consumer's attention is drawn. This characteristic takes the values ***urgent*** and ***routine***:

- *Urgent:* This information relates to the user's persistent goals (involving actions which could cause personal injury or property damage, whether an imminent meltdown or a warning to a person crossing the road in front of a car) and must therefore be reinforced by textual devices such as 'boldface', 'capitalization', etc. For more details see [Hovy & Arens 91].
- *Routine:* The normal, non-distinguished case.

Order: Order is a property of a collection of items all displayed together as a group of some kind. Values here are:

- *Quantitative:* This characterizes items belonging to a conceptually and/or syntactically regular but not presentationally ordered set, such as temperature readings for various parts of the country).
- *Ordinal:* This characterizes items of a set ordered according to their semantic denotations (e.g., steps in a recipe).
- *Nominal:* The items are not ordered.

Density: The difference between information that is presented equally well on a graph and a histogram and information that is not well presented on a histogram is a matter of the density of the class to which the information belongs. The former case is ***discrete*** information; an example is the various types of car made in Japan. The latter is ***dense*** information; an example is the prices of cars made in Japan.

- *Dense:* A class in which arbitrary small variations along a dimension of interest carry meaning. Information in such a class is best presented by a modality that supports continuous change.
- *Discrete:* A class in which there exists a lower limit to variations on the dimension of interest.

Naming (function): The role information plays may be defined relative to other information present. A good example is the information that names and introduces, such as that in headings of text sections, titles of diagrams, and labels in pictures. We identify just two of the many types here:

- *Identification:* This information identifies a portion of the presentation, based on an appropriate underlying semantic relation such as between a text label and a picture part; see [Hovy et al. 92].
- *Introduction:* This information identifies and introduces other information by appearing first and standing out positionally.

Figure 4: Fragment of the Information Features Network.

Volume: A batch of information may contain various amounts of information to be presented. If it is a single fact, we call it *singular*; if more than one fact but still little relative to some task- and user-specific threshold, we call it *little*; and if not, we call it *much*. This distinction is useful because not all modalities are suited to present *much* information.

- *Much:* The relatively permanent modalities such as written text or graphics leave a trace to which the consumer can refer if he or she gets lost doing the task or forgets, while transient modalities such as spoken sentences and beeps do not. Thus the former should be preferred in this case.
- *Little:* There is no need to avoid the more transient modalities when the amount of information to present is *little*.
- *Singular:* A single atomic item of information. A transient modality can be used. However, one should not overwhelm the consumer with irrelevant information. For example, to display information about a single ship, one need not draw a map.

The features listed here are only the tip of a large iceberg. They can be subclassified in several ways. One way is by whether the feature is apparent by virtue of the information itself or by its juxtaposition with others, as in Table 3; another way is by its teleological status, as partially shown in Figure 4.

3.3 The Producer's Intentions

Particularly in the field of natural language research, there has been much work identifying and classifying the possible goals of a producer of an utterance — work which can quite easily be applied to multimedia presentations in general.

Automated text generators, when possessing a rich grammar and lexicon, typically require several producer-related aspects to specify their parameters fully. For example, the PAULINE generator [Hovy 88] produced numerous variations from the same underlying representation depending on its input parameters, which included the following presenter-oriented features:

Producer's goals with respect to perceiver: These goals all address some aspect of the perceiver's mental knowledge or state, such as:

- *Affect perceiver's knowledge*: This feature takes such values as teach, inform, and confuse.
- *Affect perceiver's opinions of topic*: Values include switch, reinforce.
- *Involve perceiver in the conversation*: involve, repel.
- *Affect perceiver's emotional state*: Of the hundreds of possibilities we list simply anger, cheer up, calm.
- *Affect perceiver's goals*: Values include activate and deactivate. These goals cover warnings, orders, etc.

Producer's goals with respect to the producer-perceiver relationship: These address both producer and perceiver, for example:

- *Affect perceiver's emotion toward producer*: Values include respect, like, dislike.
- *Affect relative status*: Values here determine formality of address forms in certain languages, etc.: dominant, equal, subordinate.
- *Affect interpersonal distance*: Values such as intimate, close, distant.

For our purposes, we have chosen to borrow and adapt a partial classification of a producer's communicative goals from existing work on Speech Acts. Figure 5 provides a small portion of the network containing aspects of a producer's communicative intentions that may affect the appearance of a presentation (see [Vossers 91] for more details). In this network fragment *warn* is distinguished from *inform* because, unlike inform speech acts, the semantics of warnings involve capturing the attention of the reader in order to affect his/her goals or actions. To achieve this, a warning must be realized using presentation features that distinguish it from the background presentation.

3.4 The Perceiver's Nature and Situation

Our work has only begun to address this issue. Existing research provides considerable material with a bearing on the topic, including especially the work in Cognitive Psychology on issues of human perception which influence the appropriateness of media choices for presentation of certain types of data. A survey and discussion of these results is presented in [Vossers 91]. On the computational side, the abovementioned text generation system [Hovy 88] contains several categories of characteristics of the perceiver, including:

Figure 5: Portion of the Producer Goals Network.

Figure 6: Portion of the Internetwork Linkage.

- *knowledge of the topic*: expert, student, novice.
- *interest in the topic*: high, low.
- *opinions of the topic*: good, neutral, bad.
- *language ability*: high, low.
- *emotional state*: calm, angry, agitated.

3.5 Interdependencies and Rules

The factors that affect multimedia presentations are not independent. Their interdependencies can be thought of as rules which establish associations between the goals of the producer, the content of the information, and surface features of presentations to constrain the options for presenting information (during generation) and disambiguate alternative readings (during interpretation). A small portion of these rules, also represented in network form, appears in Figure 6. Moving from left to right through the network (that is, in the direction of presentation interpretation), one first finds the presentation forms which express the information, then features of the information which are linked to various presentation forms, and finally the producer goals. That formalism is essentially equivalent to standard "Rule" writing, as below. We use one formalism or the other, depending on what we feel is most suitable to the task being addressed.

Below, in traditional form, is a more comprehensive list of rules, organized by characteristics of data being considered for presentation. The terminology is defined in Section 3.2.

Dimensionality

- *Simple:*
  - ☐ *Rule:* As carrier, use a modality with a dimension value of 0D.
  - ☐ *Rule:* No special restrictions on substrate.
- *Single:*
  - ☐ *Rule:* No special restrictions on substrate.
- *Double:*
  - ☐ *Rule:* As substrate, use modalities with internal semantic dimension of 2D.
  - ☐ *Rule:* As substrate, use modalities with discrete granularity (e.g., forms and tables) if information-class of both components is discrete.
  - ☐ *Rule:* As substrate, use modalities with continuous granularity (e.g., graphs and maps) if information-class of either component is dense.
  - ☐ *Rule:* As carrier, use a modality with a dimension value of 0D.
- *Multiple:*
  - ☐ *Rule:* As substrate, use modalities with discrete granularity if information-class of all components is discrete.
  - ☐ *Rule:* As substrate, use modalities with continuous granularity if the information-class of some component is dense.
  - ☐ *Rule:* As carrier, use a modality with a dimension value of at least 1D.
  - ☐ *Rule:* As substrate and carrier, do not use modalities with the temporal endurance value transient.
- *Complex:*
  - ☐ *Rule:* Check for the existence of specialized modalities for this class of information.

Transience

- *Live:*
  - ☐ *Rule:* As carrier, use a modality with the temporal endurance characteristic transient if the update rate is comparable to the lifetime of the carrier signal.
  - ☐ *Rule:* As carrier, use a modality with the temporal endurance characteristic permanent if update rate is much longer.
  - ☐ *Rule:* As substrate, unless the information is already part of an existing exhibit, use the neutral substrate.
- *Dead:*
  - ☐ *Rule:* As carrier, use ones that are marked with the value permanent temporal endurance.

Urgency

- *Urgent:*
  - ☐ *Rule:* If the information is not yet part of a presentation instance, use a modality whose default detectability has the value high (such as an aural modality) either for the substrate or the carrier.
  - ☐ *Rule:* If the information is already displayed as part of a presentation instance, use the present modality but switch one or more of its channels from fixed to the corresponding temporally varying state (such as flashing, pulsating, or hopping).
- *Routine:*
  - ☐ *Rule:* Choose a modality with low default detectability and a channel with no temporal variance.

Density

- *Dense:*
  - ☐ *Rule:* As substrate, use a modality with granularity characteristic continuous (e.g., graphs, maps, animations).
- *Discrete:*
  - ☐ *Rule:* As substrate, use a modality with granularity characteristic discrete (e.g., tables, histograms, lists).

Volume

- *Much:*
  - ☐ *Rule:* As carrier, do not use a modality the temporal endurance value transient.
  - ☐ *Rule:* As substrate, do not use a modality the temporal endurance value transient.
- *Little:*
  - ☐ *Rule:* No need to avoid transient modalities.
- *Singular:*
  - ☐ *Rule:* As substrate, if possible use a modality whose internal semantic system has *low* baggage.

4 Some Examples

In this section we present a few simple examples of how the knowledge and rules outlined earlier can be applied to produce and interpret sample displays. Each example utilizes only a portion of the knowledge resources we have collected.

	Coordinates	Name	Photograph
Information	48N 2E	Paris	Eiffel Tower
Dimensionality	*double*	*single*	*single*
Volume	*little*	*singular*	*singular*
Density	*dense*	*discrete*	*discrete*
Transience	*dead*	*dead*	*dead*
Urgency	*routine*	*routine*	*routine*

Table 4: Example information characteristics.

4.1 Example 1: Identification of Appropriate Modalities

We present three simple tasks in parallel. Given the following:

- **Task:** the task of presenting Paris (as the destination of a flight, say).
- **Available information** (three separate examples): the coordinates of the city, the name *Paris*, and a photograph of the Eiffel Tower.
- **Available modalities:** maps, spoken and written language, pictures, tables, graphs, ordered lists.

The characteristics of the media available appear in Table 2 on page 11, and the characteristics of the information to be presented appear in Table 4.

The allocation algorithm classifies information characteristics with respect to characteristics of modalities, according to the rules outlined in Section 3.2. The modality with the most desired characteristics is then chosen to form the exhibit.

Handling the coordinates: As given by the rules mentioned in Section 3.2, information with a *dimensionality* value of *double* is best presented in a substrate with a *dimension* value of *2D*. This means that candidate substrates for the exhibit are maps, pictures, tables, and graphs. Since the *volume* is *little*, *transient* modalities are not ruled out. The value *dense* for the characteristic *density* rules out tables. The values for *transience* and *urgency* have no further effect. This leaves tables, maps, and graphs as possible modalities. Next, taking into account the rules dealing with the internal semantics of modalities, immediately everything but maps are ruled out (maps' internal semantics denote spatial locations, which matches up with the denotation of the coordinates). If no other information is present, a map modality is selected to display the location of Paris.

Handling the name: The name Paris, being an atomic entity, has the value *single* for the *dimensionality* characteristic. By the appropriate rule (see Section 3.2), the substrate should be the neutral substrate or natural language and the carrier one with *dimension* of *0D*. Since the *volume* is *singular*, a *transient* modality is not ruled out. None

Figure 7: Page from the 1990 Honda manual.

of the other characteristics have any effect, leaving the possibility of communicating the single word **Paris** or of speaking or writing a sentence such as *"The destination is Paris"*.

Handling the photograph: The photograph has a *dimensionality* value *complex*, for which appropriate rules specify modalities with *internal semantic dimension* of ∞D, and with *density* of *dense* (see Section 3.2) — animation or pictures. Since no other characteristic plays a role, the photograph can simply be presented.

This example illustrated how data characteristics can help limit the selection of media appropriate for displaying a particular item. The features we discussed can be used to establish a number of possible display media (or media combinations). Further knowledge can then be applied to make the final media determination.

4.2 Example 2: Rule Simplification and Generalization

This example involves the analysis of a figure taken from the 1990 Honda Accord Owner's Manual page explaining how to adjust the front seat [Honda Manual 90], reproduced in Figure 7.

On first inspection, the section heading **Front Seat** and the label **Pull up** in Figure 7 look very different; indeed, the heading is analyzed as including the features *text-in-text*, *boldface*, *large-font*, *separation*, and *short*, while the label includes the features *text-in-picture* and *short*. But upon following the internetwork linkage rules in Figure 6, both items are seen to serve almost-identical producer goals: *introduce* and *identify*, respectively. Thus they are both instances of the *naming* function (see Figure 4); the features that differ are simply those required to differentiate each item from its background. Thus the operative rule can concisely be expressed as:

□ *Rule:* To indicate the *naming* function, use *short* text which is *distinct* from the background presentation object.

How to achieve distinction is a matter for the individual presentation media, and has nothing to do with the communicative function of *naming* per se. Within a picture, *distinction* is achieved by the mere use of text, while within text, *distinction* must be achieved by varying the features of the surrounding rendering of the language, for example by changing the font type and size or the position of the item in relation to the general text body.

The notion of *distinction*, having crystallized out of the above two presentations, somewhat unexpectedly turns out to be quite generally applicable. Consider the text bullets at the bottom of the figure. Since their function is to *warn* (and not merely to *inform*, which is the purpose of the preceding paragraphs), the text has the feature *bold*. This serves to *distinguish* the warning text from the background, thereby signaling the special force required for a warning. Using the rule stated above, we can now predict that, within the context of a diagram or picture, one can effect a warning simply by placing text within the non-textual substrate.

Thus, though the notion of *distinction* was not explicitly developed for the individual networks influencing presentations, Figure 6 suggested its utility with an appropriate collection of specific features. Its importance was discerned in the course of investigating the internetwork linkage rules and their application to presentations such as this manual page.

The example illustrates the generality of the rules that can be used to generate and parse multimedia presentations, but, when described, it may seem obvious. However, it can only be explained by using such notions as *distinguished/separated* (both the positional/off-text distinctiveness and the realizational/text-vs-graphics distinctiveness) and *communicative function* (one part of the communication serves to name/introduce/identify another part). When one constructs a vocabulary of terms on this level of description, one finds unexpected overlaps in communicative functionality across media.

In the domain of presentations containing text and line drawings, we demonstrated that media selection rules can be written so that the same rule can be used to control the analysis and generation of some aspect of both a diagram and a piece of text. This is extremely significant, in that the resulting parsimony and expressive power of these rules simultaneously motivates the particular representational level we have used and also suggests how the complex task of multimedia communication is achieved with less overhead than at first seemed necessary. The assembly of a vocabulary of media-independent (or at least shared by multiple media) features of the kind we discuss is an important future research task.

5 Conclusion

The enormous numbers of possibilities that appear when one begins to deal with multiple media, as illustrated by the Psychology, Cognitive Science, and automated text generation and formatting work mentioned above, is daunting. We believe that systematic analysis of the factors influencing presentations, such as the types described here, is required before powerful general-purpose multimedia human-computer interfaces can be built. Appropriate formalisms for representing the underlying knowledge may serve to uncover unexpected overlaps of functionality which serve to simplify the rules upon which such interface systems will depend. It appears that the dependency network formalism and feature-based analysis methodology described in this paper hold some promise for untangling the complex issues involved, and, perhaps, may one day help explain why multimedia communication is so pervasive in human interaction.

References

[André & Rist 92] André, E. and Rist, T. 1992. The Design of Illustrated Documents as a Planning Task. German Research Center for AI (DFKI) Research Report.

[Arens et al. 88] Arens, Y., Miller, L., Shapiro, S.C. and Sondheimer, N.K. 1988. Automatic Construction of User-Interface Displays. In *Proceedings of the 7th AAAI Conference*, St. Paul, MN (808–813). Also available as USC/Information Sciences Institute Research Report RR-88-218.

[Arens & Hovy 90a] Arens, Y. and Hovy, E.H. 1990. How to Describe What? Towards a Theory of Modality Utilization. In *Proceedings of the 12th Cognitive Science Conference*, Cambridge, MA.

[Arens & Hovy 90b] Arens, Y. and Hovy, E.H. 1990. Text Layout as a Problem of Modality Selection. In *Proceedings of the 5th Conference on Knowledge-Based Specification* (87–94), RADC Workshop, Syracuse, NY.

[Arens et al. 92] Arens, Y., Hovy, E., and Van Mulken, S. 1992. A Tree-Traversing Prototype that Allocates Presentation Media. Submitted to a workshop.

[Bertin 83] Bertin, J. 1983. *Semiology of Graphics*, (trans. by J. Berg). Madison: University of Wisconsin Press.

[Burger & Marshall 91] Burger, J. and Marshall, R. 1991. AIMI: An Intelligent Multimedia Interface. In *Proceedings of the 9th National Conference on Artificial Intelligence, AAAI-91*, Anaheim, CA (23–28).

[Dwyer 78] Dwyer, F.M. 1978. *Strategies for improving visual learning.* State College, PA: Learning Services.

[Feiner 88] Feiner, S. 1988. An Architecture for Knowledge-Based Graphical Interfaces. ACM/SIGCHI Workshop on *Architectures for Intelligent Interfaces: Elements and Prototypes*, Monterey, CA.

[Feiner & McKeown 90] Feiner, S. and McKeown, K.R. 1990. Coordinating Text and Graphics in Explanation Generation. In *Proceedings of the 8th AAAI Conference* (442–449).

[Fleming & Levie 78] Fleming, M. and Levie, H.W. 1978. *Instructional Message Design: Principles from the Behavioral Sciences.* New Jersey: Educational Technology Publications.

[Halliday 85] Halliday, M.A.K. 1985. *An Introduction to Functional Grammar.* Baltimore: Edward Arnold Press.

[Hartley 85] Hartley, J. 1985. *Designing Instructional Text.* (2nd edition). Great Britain: Kogan Page Ltd.

[Honda Manual 90] Honda Accord: 1990 Owner's Manual. Japan: Honda Motor Co. Ltd.

[Hovy 88] Hovy, E.H. 1988. *Generating Natural Language under Pragmatic Constraints.* Hillsdale, NJ: Lawrence Erlbaum Associates.

[Hovy 90] Hovy, E.H. 1990. Natural Language Processing at ISI. *The Finite String* 16(4) (37–42).

[Hovy & Arens 90] Hovy, E.H. and Arens, Y. 1990. When is a Picture Worth a Thousand Words? — Allocation of Modalities in Multimedia Communication. Presented at the *AAAI Symposium on Human-Computer Interaction*, Stanford University.

[Hovy & Arens 91] Hovy, E.H. and Arens, Y. 1991. Automatic Generation of Formatted Text. In *Proceedings of the 10th AAAI Conference*, Anaheim, CA.

[Hovy et al. 92] Hovy, E.H., Lavid, J., Maier, E., Mittal, V., and Paris, C.L. 1992. Employing Knowledge Resources in a New Text Planning Architecture. In *Proceedings of the 6th International Workshop on Language Generation*, Trento, Italy (to appear).

[Kasper 89] Kasper, R.T. 1989. Unification and Classification: An Experiment in Information-Based Parsing. In *Proceedings of the International Workshop on Parsing Technologies*, Pittsburgh, PA.

[Kasper & Hovy 90] Kasper, R.T. and Hovy, E.H. 1990. Integrated Semantic and Syntactic Parsing using Classification. In *Proceedings of the DARPA Speech and Natural Language Workshop*, Pittsburgh, PA.

[Larkin & Simon 86] Larkin, J.H., and Simon, H.A. 1986. Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science* 11(1) (65–99).

[Mackinlay 86] Mackinlay, J. 1986. *Automatic Design of Graphical Presentations*. Ph.D. dissertation, Stanford University.

[Mann & Matthiessen 83] Mann, W.C. and Matthiessen, C.M.I.M. 1983. Nigel: A Systemic Grammar for Text Generation. Research Report RR-83-105, USC/ISI.

[Matthiessen 84] Matthiessen, C.M.I.M. 1984. Systemic Grammar in Computation: The Nigel Case. In *Proceedings of 1st Conference of the European Association for Computational Linguistics*, Pisa, Italy. Also available as USC/ISI Research Report RR-84-121.

[Maybury 91] Maybury, M.T. 1991. Planning Multimedia Explanations using Communicative Acts. In *Proceedings of the 9th National Conference on Artificial Intelligence, AAAI-91*, Anaheim, CA (61–66).

[Mayer 89] Mayer, R.E. 1989. Systematic Thinking Fostered by Illustrations in Scientific Text. *Journal of Educational Psychology* 81 (240–246).

[Neal 90] Neal, J.G. 1990. Intelligent Multi-Media Integrated Interface Project. SUNY Buffalo. RADC Technical Report TR-90-128.

[Ortony et al. 92] Ortony, A., Slack, J., and Stock, O. (eds). 1992. *Computational Theories of Communication and their Applications*. Berlin: Springer Verlag.

[Penman 88] The Penman project. 1988. *The Penman Primer, User Guide, and Reference Manual*. Unpublished USC/ISI documentation.

[Petre & Green 90] Petre, M. and Green, T.R.G. 1990. Is Graphical Notation Really Superior to Text, or Just Different? Some Claims by Logic Designers about Graphics in Notation. *Proceedings of the ECCE-5*, Urbino, Italy.

[Roth & Mattis 90] Roth, S.F. and Mattis, J. 1990. Data Characterization for Intelligent Graphics Presentation. *CHI'90 Proceedings* (193–200).

[Sullivan & Tyler 91] Sullivan and Tyler (eds). 1991. *Intelligent User Interfaces*. New York: ACM Press.

[Tufte 83] Tufte, Edward R. 1983. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

[Tufte 90] Tufte, Edward R. 1990. *Envisioning Information*. Cheshire, CT: Graphics Press.

[Twyman 85] Twyman, M. 1985. Using Pictorial Language: A Discussion of the Dimensions of the Problem. In Duffy, T.M. & Waller, R. (Eds.), *Designing Usable Texts*. Florida: Academic Press (245–312).

[Vossers 91] Vossers, M. 1991. Automatic Generation of Formatted text and Line Drawings. Master's thesis, University of Nijmegen, The Netherlands.

[Wahlster et al. 92] Wahlster, W., André, E., Bandyopadhyay, S., Graf, W., Rist, T. 1992. WIP: The Coordinated Generation of Multimodal Presentations from a Common Representation. In *Computational Theories of Communication and their Applications*, A. Ortony, J. Slack, and O. Stock (eds), Berlin: Springer Verlag.