In: RUIU, San Francisco: Morgan Kaufmann, 1998, pp. 1- 13.

Intelligent User Interfaces: An Introduction

This introduction describes the need for intelligent user interfaces (IUIs), specifies the intended purpose and use of this collection, outlines the collection's scope, and defines basic terminology used in the field. After outlining the theoretical foundations of intelligent user interfaces, this introductory section describes the current state of the art and summarizes the structure and contents of this collection, which addresses some remaining fundamental problems in the field.

1. MOTIVATION

The explosion of available materials on corporate, national, and global information networks is driving the need for more effective, efficient, and natural interfaces to support access to information, applications, and people. This is exacerbated by the increasing complexity of systems, the shrinking of task time lines, and the need to reduce the cost of application and interface development. Fortunately, the basic infrastructure for advanced multimedia user interfaces is rapidly appearing or already available. In addition to traditional public telephone networks, cable, fiber-optic, wireless and satellite communications are rapidly evolving with the aim of serving many simultaneous users through a great variety of multimedia communications (e.g., video, audio, text, data). Rapidly advancing microprocessor and storage capabilities, coupled with multimedia input and output devices integrated into workstations and portable machines, provide a dizzying array of potential for personal and personalized multimedia interaction.

Interface technology has advanced from initial command line interfaces to the established use of direct manipulation or WIMP (windows, icons, menus, and pointing) interfaces in nearly all applications. Even some of the first computing systems incorporated graphical displays and light pens as pointing devices (Everett et al. 1957). The next generation of interfaces, often called "intelligent," will provide a number of additional benefits to users, including adaptivity, context sensitivity, and task assistance. As with traditional interfaces, principled intelligent interfaces should be learnable, usable, and transparent. In contrast, however, intelligent user interfaces promise to provide additional benefits to users that can enhance interaction, such as:

- Comprehension of possibly imprecise, ambiguous, and/or partial multimodal input
- Generation of coordinated, cohesive, and coherent multimodal presentations
- Semi- or fully automated completion of dele gated tasks
- Management of the interaction (e.g., task completion, tailoring interaction styles, adapting the interface) by representing, reasoning, and exploiting models of the user, domain, task, and context

In addition to these end-user benefits, new model-based interface tools promise to help user interface designers and developers decrease the time, expense, and level of expertise necessary to construct successful user interfaces.

In search of these benefits, governments, industry, and academia have emphasized the importance of the human-machine interface in the global information economy. For example, the United States Digital Library, European Telematics, and the Japanese Human Interface Programs are all well funded in the long term, but their continued success will require researchers and managers to rapidly acquire the standard literature and train in the latest interface tools and techniques. As an example, the $500 million, 10-year Real World Computing (RWC) Program initiated in 1992 (Tsukuba, Japan) focuses on pattern/symbol processing and includes an emphasis on multimodal interfaces integrating gesture, speech, and body language. Academic and commercial advances in human-computer interface technology (Baecker, Grudin, Buxton, and Greenberg 1995) has dramatically improved interaction with computers; however, these efforts are necessary but not sufficient to address the preceding challenges. Instead, a new class of interfaces is required that goes beyond the current tripartite interface model of application, dialogue, and presentation. This collection of papers points the way toward interfaces that model the situation, task, user, discourse, and media and that enable model-based specification and generation of interfaces, agent-based interaction, and integrated multimodal input and output. Unlike traditional human-computer interfaces, intelligent interfaces are those that represent and reason about the user, domain, task, media, and situation. A number of applications are emerging, ranging from mail filters (Maes, Section VII of this volume) to office assistants (Horvitz 1997) to speaking and listening interface agents (Nagao and Takeuchi, Section VII of this volume).

2. PURPOSE AND USE

The purpose of this collection is multifold. First, it is intended to motivate and define the field of intelligent user interfaces. Second, it is intended to capture and place into context key developments in this field. Third, it is intended to serve as a stimulus for continued research into the many interesting and challenging problems that remain. Finally, a principal goal of this collection is to bridge the gap between scientists and engineers working in the distinct but interdependent fields of human-computer interaction and intelligent user interfaces/artificial intelligence. We hope the collection will also serve to foster scientific interchange among individuals working in both theory and applications, and, as such, the collection reflects a mix of these activities.

This collection can be used as: a key reference source for students, researchers, and practitioners of IUI or as a text in user interface classes or advanced graduate seminars. To satisfy these purposes, the book is organized around the key areas of IUI: input analysis, output generation, user- and discourse-adapted interaction, agent-based interaction, model-based interface design, and intelligent interface evaluation. In addition to a traditional author and keyword index, we also provide a two-dimensional content index in Section 5.7, to facilitate tailored access to relevant content for a range of purposes: research, analysis, or teaching.

Articles were chosen from a broad range of sources, including journals, conference proceedings, workshop notes, and previous book collections. Each article was nominated by a member of the editorial board and evaluated by multiple reviewers considering selection criteria of quality, significance, originality, clarity, and relevance as well as special considerations such as historical and sustained influence and difficulty of acquisition. Each article is followed by a brief reflection written by the original authors, indicating important developments that have influenced, have resulted from, and/or have followed the publication of their work, including key follow-up publications by the authors and others. After providing the scope and definitions of the field, we overview its brief history and then provide summaries of the key sections of the readings.

3. SCOPE AND DEFINITIONS

Intelligent user interfaces (IUIs) are human-machine interfaces that aim to improve the efficiency, effectiveness, and naturalness of human-machine interaction by representing, reasoning, and acting on models of the user, domain, task, discourse, and media (e.g., graphics, natural language, gesture). As a consequence, this interdisciplinary area draws upon research in and lies at the intersection of human-computer interaction, ergonomics, cognitive science, and artificial intelligence and its subareas (e.g., vision, speech and language processing, knowledge representation and reasoning, machine learning/knowledge discovery, planning and agent-modeling, user and discourse modeling). Whereas previous collections have focused on related enabling technologies such as text processing (Grosz, Sparck Jones, and Webber 1986; MUC-6 1995), spoken language processing (Waibel and Lee 1990), human-computer interaction (Baecker, Grudin, Buxton, and Greenberg 1995), user modeling (Kobsa and Wahlster 1989), artificial intelligence (Webber and Nilsson 1985), knowledge representation (Brachman and Levesque 1985), and planning (Allen and Hendler 1990), intelligent human-computer interaction requires a synergistic integration of these areas. This collection complements previous works focused on human-computer interaction, multimedia or intelligent interfaces (Blattner and Dannenberg 1992, Sullivan and Tyler 1991), and intelligent multimedia interfaces (Maybury 1993).

Figure 1 illustrates a high-level architecture of intelligent user interfaces and, as such, defines many of the subareas in the field. These include analyzing and interpreting input, designing and rendering output, managing the interaction, and representing and reasoning about models that support intelligent interaction. An example of a model is a user model, more generally, an agent model (e.g., that could represent the user, system, intermediary, addressee, etc.). The "intelligence" in IUIs that distinguishes them from traditional interfaces is indicated in bold in Figure 1—it includes mechanisms that perform automated media analysis, design, and interaction management. Thus, the collection does not address input and output devices and drivers for input processing and output rendering.

As the dotted regions in Figure 2 illustrate, traditional user interfaces distinguish only three models: presentation, dialog, and application. Refinements beyond these three models that are found in IUIs include explicit models of the user, discourse and domain, input analysis and output generation, and mechanisms to manage the interaction, such as fusing and interpreting imprecise, ambiguous, and/or inaccurate input, controlling the dialog progression, or tailoring presentation output to the current situation.

Research so far has shown that it is possible to adapt many of the fundamental concepts developed to date in computational linguistics and discourse theory in such a way that they become useful for multimedia user interfaces as well. In particular, semantic and pragmatic concepts like communicative acts, coherence, focus, reference, discourse model, user model, implicature, anaphora, rhetorical relations, and scope ambiguity take on an extended meaning in the context of multimodal communication. As Figure 3 illustrates, artificial intelligence has much to contribute to user interfaces, including the use of knowledge representations for model-based interface development tools, the

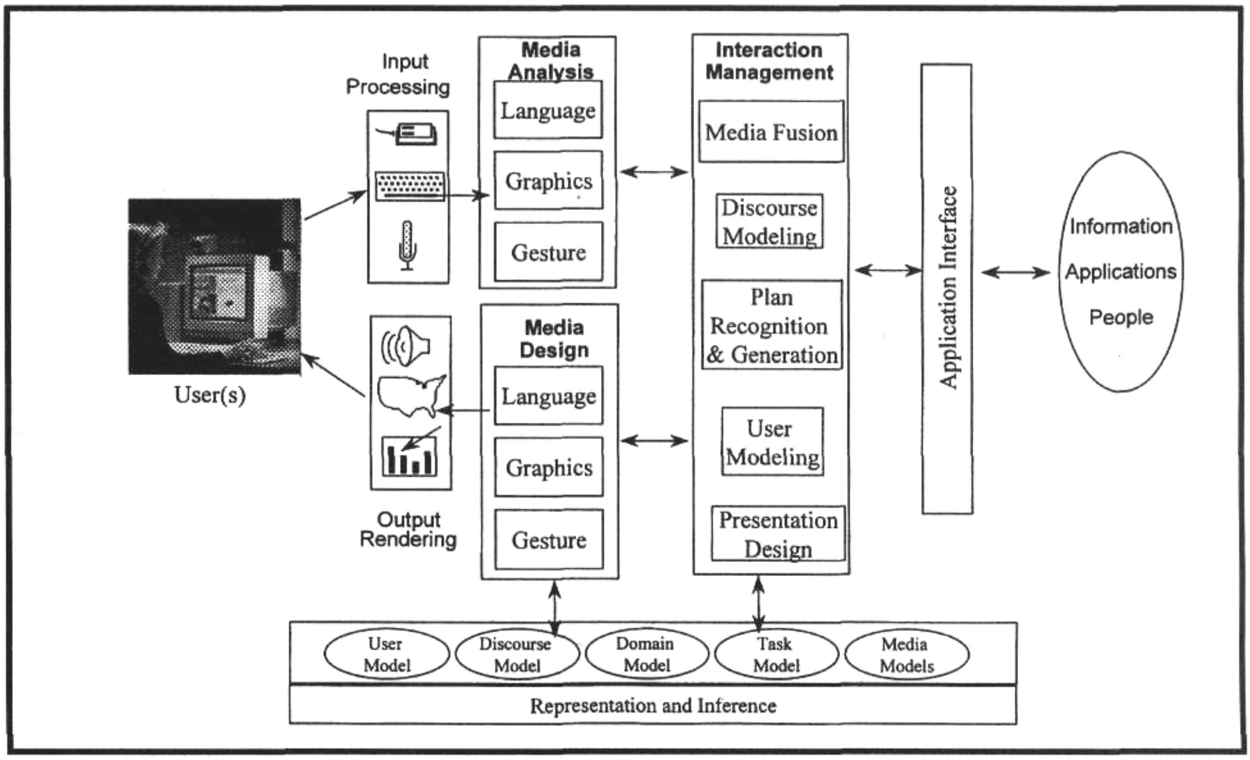

Figure 1. Architecture of Intelligent User Interfaces

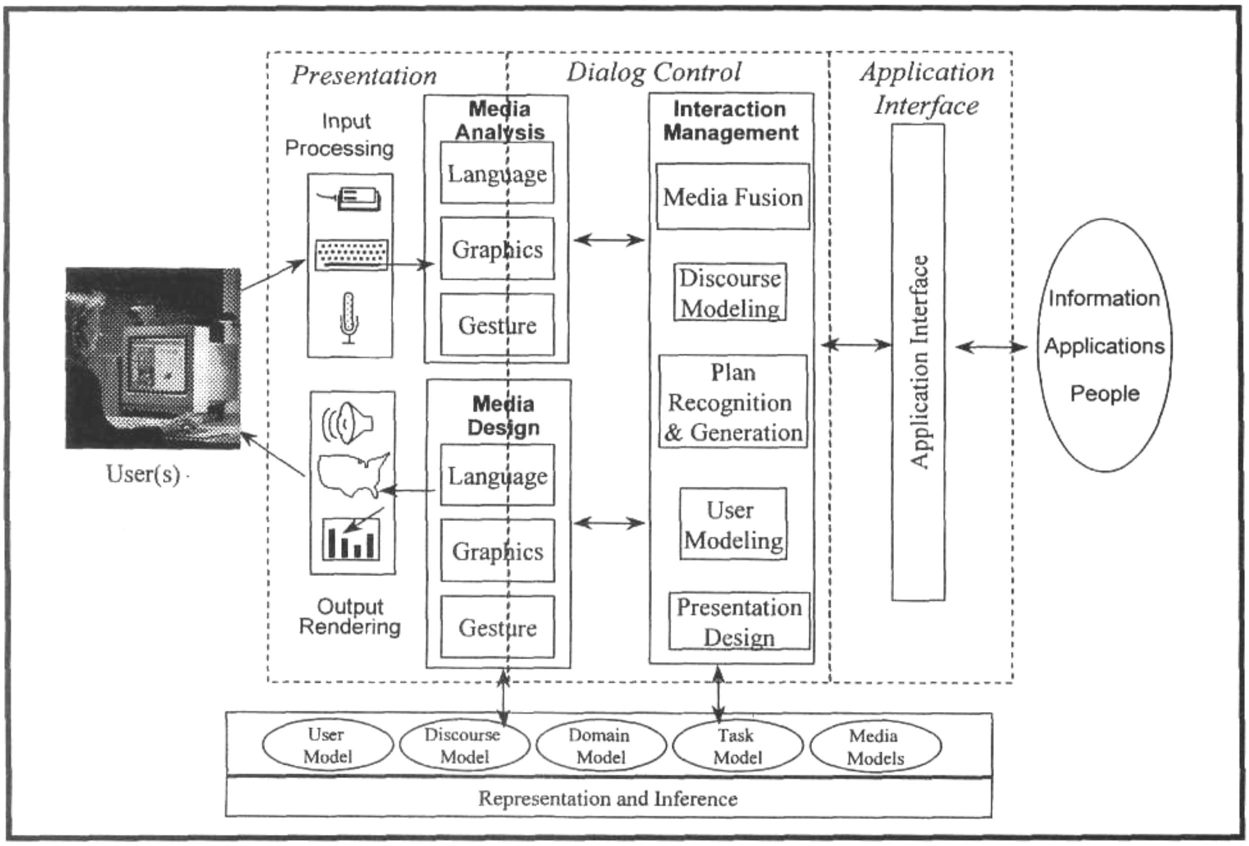

FIGURE 2. Current Interface Practice and Its Relation to IUI

application of plan generation and recognition in dialog management, the application of temporal and spatial reasoning to media coordination, the use of user models to tailor interaction, and so on. We will detail these theoretical and technical foundations in the following.

As shown in Figure 3, more effective, efficient, and natural human-computer or computer-mediated, human-human interaction will require both automated understanding and generation of multimedia. Fluent conversational interaction demands explicit models of the user, discourse, task, and context. It also requires a richer understanding of media in its use both in the interface to support interaction with the user and in access to content by the user during a session.

Because of widespread terminology confusion, we begin with a clarification of the terms *medium* and *mode.* By *mode* or *modality,* we refer primarily to the human senses employed to process incoming information: vision, audition, olfaction, touch, and taste. In contrast, *medium* refers to the material object (e.g., the physical carrier of information such as paper or CD-ROM) used for presenting or saving information and, particularly in the context of human-computer interaction, to computer input/output devices (e.g., microphone, speaker, screen, pointer). We use the term *code* to refer to a system of symbols (e.g., natural language, pictorial language, gestural language). For example, a natural language code might use typed or written text or speech, which in turn would rely upon visual or auditory modalities and associated media (e.g., keyboard, microphone). It is important to note, however, that especially the terms *media* and *mode* are frequently used ambiguously in the literature. Indeed, in this collection we will use them interchangeably when their distinction is not important.

Medium, mode, and code are related nontrivially (see Figure 4). First, a single medium may carry several modalities and, in turn, codes. For example, a piece of paper may support both language and graphics codes just as a visual display may support text, images, and video. Likewise, a single code may be supported by many media and modalities. For instance, language can be supported visually (i.e., written language) and aurally (i.e., spoken language)—in fact, spoken language can have a visual component (e.g., lip reading). Analogously, a user of a multimedia CD-ROM is interacting with a physical me-

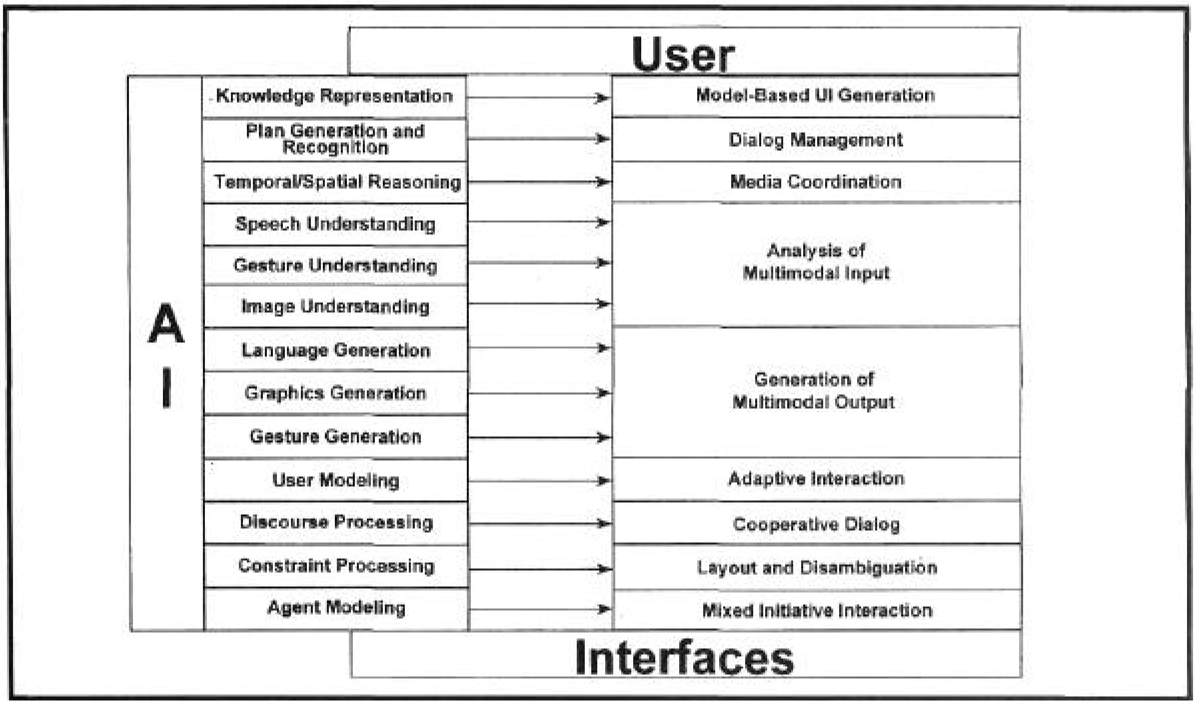

FIGURE 3. AI Meets User Interfaces

dium used to store information captured in a variety of codes (e.g., language, graphics) using multiple modalities (e.g., auditory, visual) and using various input/output media (e.g., mouse, display, speaker). A multimedia document on the CD-ROM might include text, graphics, speech, and video that affect several modalities, for example, visual and auditory perception of natural language, visual perception of images (still and moving), and auditory perception of sounds. Finally, this multimedia and multimodal interaction occurs over time. Therefore, it is necessary to account for the processing of discourse, context shifts, and changes in agent states over time.

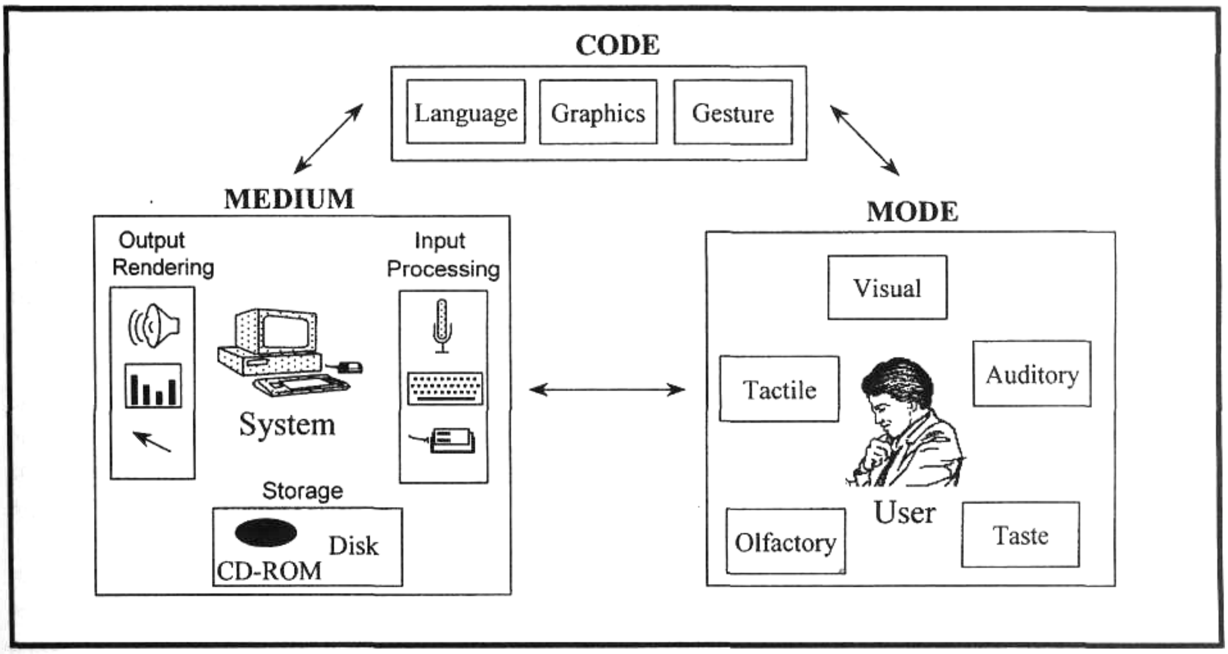

FIGURE 4. Medium, Mode, and Code

The new generation of intelligent multimodal systems (Maybury 1993, 1995) goes beyond the standard canned text, predesigned graphics, and prerecorded images and sounds typically found in commercial multimedia systems of today. A basic principle underlying these so-called intellimedia systems' is that the various constituents of a multimodal communication should be generated on the fly from a common representation of what is to be conveyed without using any preplanned text or images; that is, the principle is "no generation without representation." It is an important goal of such systems not simply to merge the verbalization and visualization results of a text generator and a graphics generator but to carefully coordinate them in such a way that they generate a synergistic improvement in communication. Such multimodal presentation systems are highly adaptive since all presentation decisions are postponed until runtime. The quest for adaptation is based on the fact that it is impossible to anticipate the needs and requirements of each potential user in an infinite number of presentation situations.

Figure 5 indicates the key processes and exemplifies some systems that have addressed multimodal information processing, including media generation and media conversion. The large arrows indicate where processing typically begins, that is, from formal representations such as a data or knowledge base or from the media themselves (e.g., text, graphics, or images). Key processes include *verbalization* (moving from formal representations or graphics or images to text) and *visualization* (from representations or text to graphics or images). Several systems have focused on multimodal presentation generation, designing, and realizing coordinated text, speech, graphical, and cartographic presentations. As the dial in the middle indicates, these systems raise the opportunity to select between a scale from an entirely linguistic to a completely visual presentation.

Multimedia dialog prototypes have been developed in several application domains, including CUBRICON to support mission planning (Neal et al., Section I of this volume), XTRA for tax form preparation (Wahlster, Section V of this volume; Kobsa et al. 1986), MMI2 for network management (Binot et al. 1990), AIMI for air mission planning (Burger and Marshall, Section V of this volume), and ALFresco to enable art history information exploration (Stock, Section V of this volume). Typically, these systems parse mixed and asynchronous multimedia input and generate coordinated multimedia output. They also attempt to maintain coherency, cohesion, and consistency across both multimedia input and output. For example, these systems often support integrated language and deixis for both input and output. They

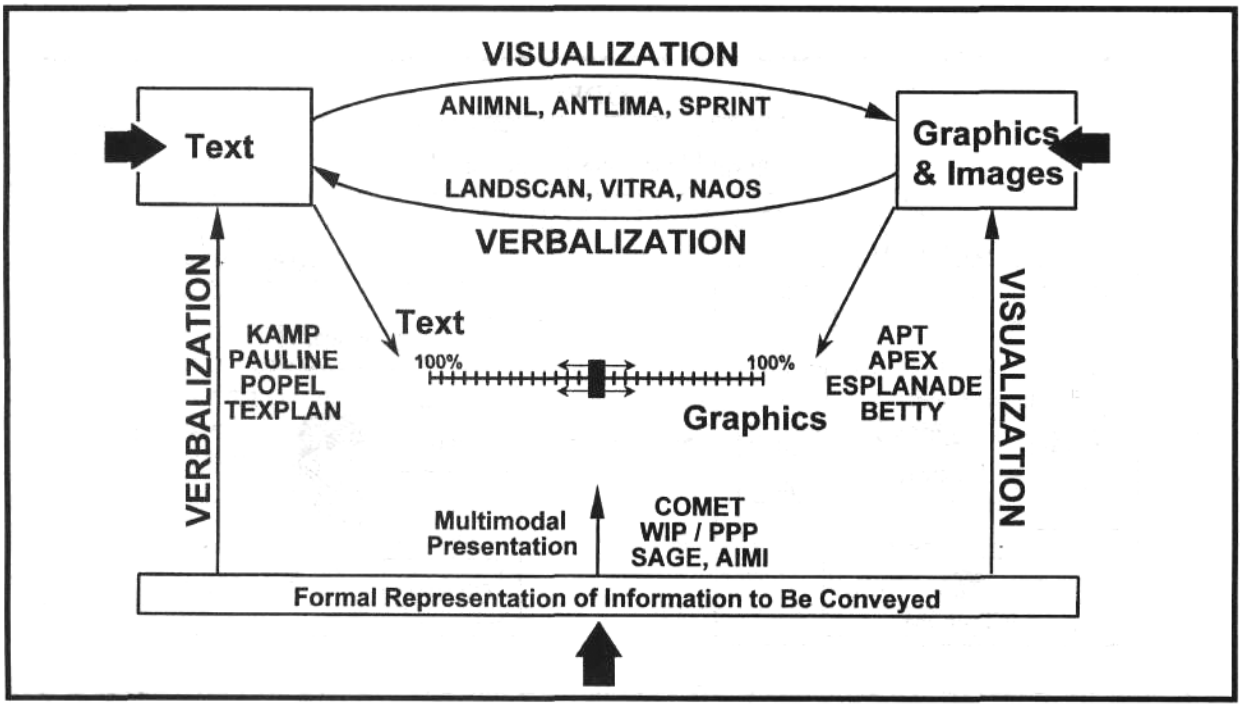

FIGURE 5. Key Processes in Multimedia Processing

extend research in discourse and user modeling (Kobsa and Wahlster 1989) by incorporating representations of media to enable media reference, cross-reference, and reuse over the course of a session with a user. These enhanced representations support the exploitation of user perceptual abilities and media preferences as well as the resolution of multimedia references (e.g., "Send this plane there" articulated with synchronous gestures on a map).

The details of discourse models in these systems, however, differ significantly. For example, CUBRICON represents a global focus space ordered by recency whereas AMI represents a focus space segmented by the intentional structure of the discourse (i.e., a model of the domain tasks to be completed). Although intelligent multimedia interfaces promise natural and personalized interaction, they remain complicated and require specialized expertise to build. One practical approach to achieving some of the benefits of these more sophisticated systems without the expense of developing full multimedia interpretation and generation components was achieved in ALFresco (Stock, Section V of this volume), a multimedia information kiosk for Italian art exploration. By adding natural language processing to a traditional hypermedia system, ALFresco achieved the benefits of hypermedia (e.g., organization of heterogeneous and unstructured information via hyperlinks, direct manipulation to facilitate exploration) together with the benefits of natural language parsing (e.g., direct query of nodes, links, and subnetworks, which provides rapid navigation). Providing a user with natural language query within a hypertext system helps overcome the indirectness of the hypermedia web as well as disorientation and cognitive overhead caused by large amounts of semantically heterogeneous links (e.g., part-of, class-of, instance-of, or elaboration-of). In addition, as in other systems previously mentioned (e.g., CUBRICON, XTRA), ambiguous gesture and language can yield a unique referent through mutual constraint. Finally, ALFresco incorporates simple natural language generation that can be combined with more complex canned text (e.g., art critiques) and images. Reiter, Mellish, and Levine (Section II of this volume) analogously integrate traditional language generation with hypertext to produce hypertext technical manuals.

Whereas practical systems are possible today, the multimedia interface of the future may have facilities that are much more sophisticated. These interfaces may include humanlike agents that converse naturally with users, monitoring their interaction with the interface (e.g., keystrokes, gestures, facial expressions) and the properties of those interactions (e.g., conversational syntax and semantics, dialog structure) over time and for different tasks and contexts. Equally, future interfaces will likely incorporate more sophisticated presentation mechanisms. For example, Pelachaud, Badler, and Steedman (1996) characterize spoken language intonation and associated emotions (anger, disgust, fear, happiness, sadness, and surprise) and from these use rules to compute facial expressions, including lip shapes, head movements, eye and eyebrow movements, and blinks. Finally, future multimedia interfaces should support richer interactions, including user and session adaptation (Schneider-Hufschmidt et al. 1993), dialog interruptions, follow-up questions, and management of the focus of attention.

In summary, as Figures 1 through 5 illustrate, principal areas of intelligent interface research include

- *Analysis* of input (e.g., spoken, typed, and hand written language; gestures, including hand, eye, and body states and motion)
- *Generation* (planning or realization) of coordinated output
- *Modeling* of the user, discourse, task, and situation and *interaction management,* including possible tailoring of interaction to the user, task, and/or situation

As such, we distinguish these functions in the organization of this collection, described later in this introductory section.

4. THE ROOTS OF INTELLIGENT USER INTERFACES

Enabling conversational interaction with computers has been a vision since the creation of the first computers. In part stimulated by attempts to pass the Turing test, a number of initial efforts attempted to literally simulate conversation with computers using pattern matching to select possible responses from a conversational database (e.g., McCarthy's ADVICE, Weizenbaum's ELIZA, Colby's PARRY). Other efforts focused on specific aspects of conversation, most notably the focus on natural language interfaces. We refer the reader to *Readings in Natural Language Processing* (Grosz, Sparck Jones, and Webber 1986), which outlines the history of natural language processing research, including theoretical and computational investigations into tasks, discourse, attention, beliefs, and plans in support of both analysis and generation of natural language. Intelligent user interfaces have benefited from a rich interaction between theoretical developments, such as Grice's work on implicatures and Austin and Searle's work on speech acts, as well as practical application areas, such as interfaces to data bases, intelligent tutoring, and automated interface design. Figure 6 captures some of the important events in the emergence of the discipline of IUIs, including the appearance of the first international workshops and conferences, specialized collections, and the emergence of commercial products and standards. Depicted is the creation of natural language interfaces in the seventies and eighties (including natural language processing toolkits), agents in commercial products in the nineties, and a standard reference model (SRM) for intelligent multimodal presentation systems (DVIMPS) in 1998 (Bordegoni et al. 1998). The nineties are also characterized by increasing scientific advances (many captured in this collection), tool developments (e.g., Kobsa et al.'s BGP-MS user modeling shell (Section V of this volume)), and commercial applications such as e-mail filters (Maes, Section VII of this volume) and Bayesian-based user models for Microsoft's Office Assistant (Horvitz 1997).

5. STATE OF THE ART: AN OVERVIEW OF THE READINGS

This readings collection is organized around solutions to the key elements introduced in the architecture outlined at the beginning of this introductory section: input analysis, output generation, user and discourse modeling, model-based interfaces, agent interaction, and evaluation. We briefly describe each of these in turn, referring the reader to the section introductions for summary overviews of the included papers.

5.1. Analysis of Input

The chapters in Section I of the collection focus on supporting intelligent input processing. Motivated by the observation that human-human communication is multimedia, multimodal, and multicodal (including spoken and written language, gesture, gaze), these papers investigate enriching human-machine interactions using such capabilities. Collectively, the authors illustrate how supporting integrated input from multiple sources can simultaneously enhance communication efficiency, effectiveness (e.g., speed and accuracy), and naturalness. They provide technical solutions that support interpretation of parallel, imprecise, and ambiguous multimedia input. The papers also illustrate the context-dependent, multifunctional nature of multimodal input, providing rich ground for future research. Taken together, these papers advance our ability to enable humans to utilize the full extent of their linguistic, gestural, and gaze input, with the attendant benefits.

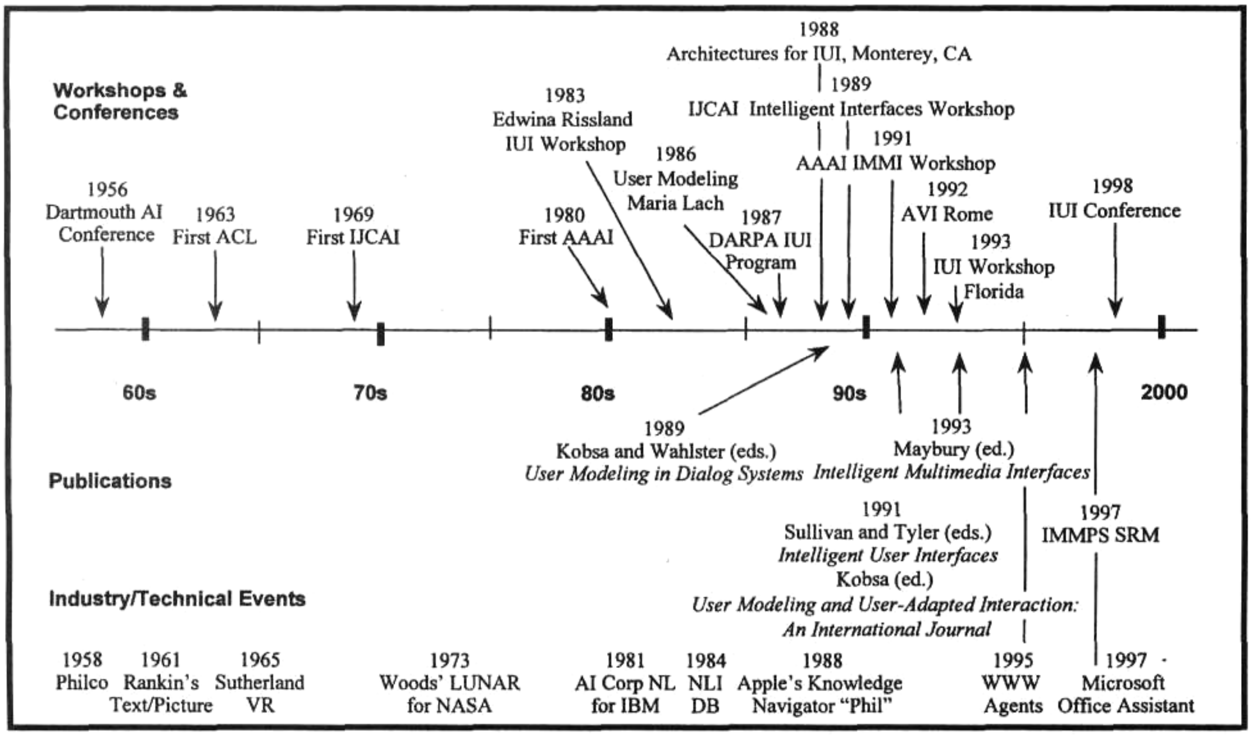

FIGURE 6. Historical Emergence of Intelligent User Interfaces

5.2. Generation of Output

The next three sections of the book address semi- or fully automatic generation of coordinated multimedia output. Designing and realizing coherent and cohesive multimedia presentations can be subdivided into several co-constraining processes, which include the determination of communicative intent, the selection of content to achieve this, its grouping/structuring and ordering, its allocation to a particular code (e.g., text versus graphics), its realization in a coordinated fashion across media, and finally its layout. Ideally, the generation process is tailored to the context, task, and user. Investigators have explored these processes in several domains using a range of algorithms. This area is accordingly divided into three sections: "Multimedia Presentation Design," "Automated Graphics Design," and "Automated Layout."

5.2.1. Multimedia Presentation Design

The second section of the book addresses multimedia presentation design, which encompasses the tasks of content selection, media allocation, media realization, and layout. The papers suggest these processes are interdependent and must be closely coordinated during generation to ensure cohesive and coherent output. Together, the papers also argue for the importance of key knowledge sources in these processes, including models of the information and media, the user, the discourse context, the producer, and the interface itself. They further demonstrate generalizations of text-linguistic notions, such as coherence, speech acts, anaphora, and rhetorical relations for multimedia presentation design. Each paper contributes concrete but widely ranging technical approaches and solutions to these tasks (e.g., employing templates, rules, plans, constraints).

5.2.2. Automated Graphics Design

Whereas several articles in Section II address automated generation of natural language text, this third section looks at the automatic design and realization of two- and three-dimensional graphics from structured data (see Maybury 1994 for pointers to eight surveys and six specialized collections on natural language generation and Grosz, Sparck Jones, and Webber 1986).

These papers collectively aim to elucidate not only how but why graphics should be designed. Moving beyond descriptive works of graphical design practice **that** consider examples of successful graphics and make observations on how to avoid ambiguous, confusing, or imprecise and misleading graphics, these efforts aim at formal and prescriptive theories of graphics design.

A number of factors have motivated researchers to seek automated graphic designers. Currently, application designers are forced to anticipate and predesign every possible data and presentation situation. Moreover, in order to create effective graphics, developers need to be design experts, which is often not the case.

Several features characterize the papers. First, they all focus on a move toward explicit representation of graphical presentation knowledge. Second, they support explicit choices among graphical encoding mechanisms that reason about the expressiveness and effectiveness of underlying representations and resulting presentations. Finally, the authors increasingly focus on representation of knowledge of the user, task, and context and its exploitation to generating more effective, tailored presentations while decreasing required user expertise. Whereas some papers focus principally on information characteristics and data graphics, others consider how differences in users' goals impact the effectiveness of designed graphics. Still others exploit perceptual operations that yield more rapid results than cognitive operations (e.g., arithmetic, comparison); for example, grouping and ordering information as well as encoding it using color, shading, and layout to support "preattentive" and sometimes parallel visual search to enable both more accurate and more efficient task performance.

5.2.3. Automated Layout

The fourth section of the book addresses the layout of media objects, which has a strong influence on the attentional structure of multimodal communication. A change in the layout of a multimedia document does not necessarily change the meaning of the document but certainly changes the focus of attention of the reader. Multimedia presentations are too dynamic and come too fast to have the layout of every visual presentation designed manually so that automated layout becomes a necessity. In addition, automated layout may help to adapt an interface to the screen or window size of a user as well as to the user's perceptual abilities and preferences. The articles in this section survey the most important techniques for automated multimedia layout, including approaches based on rules, constraints, or simulated annealing.

5.3. User and Discourse Models

Section V of the collection addresses the adaptation of interfaces to the user and context of the interaction, specifically addressing the acquisition, tracking, and utilization of models of the user and discourse. The articles address user modeling issues, including stereotypes, plan- and goal-based user models, system initiative, and user modeling shells. The articles also address the modeling and use of models of discourse for such tasks as planning explanations, answering follow-up questions in the context of prior discourse, and supporting interruption. The articles cover a wide range of application areas, including interactive consultation (e.g., recommending books or guiding software use), user- and context-adaptive hypertext (e.g., art exploration), and multimedia interfaces to decision support. A number of innovations in these systems include the extension of user and discourse models to multimedia interfaces (e.g., to process multimodal deixis), the use of incremental explanation planning, and interleaved design and realization.

5.4. Model-Based Interfaces

Section VI of the collection describes efforts to create tools that decrease both the time and the expertise required to create interfaces through automation or design assistance. These efforts go beyond user interface toolkits by separating dialog control from application code and teasing out presentation and style decisions from the toolkit code libraries. They are also distinct from user interface management systems in that they make finer-grained distinctions and provide more powerful design tools to interface developers. Much research effort has been focused on more crisply defining the functional areas of the interface in order to support declarative expression and modularization of interface functionality. These papers make a few key contributions. First, they move toward declarative specifications of interfaces, refining the distinctions among models and processes associated with the domain, the application, the user-machine dialog control, and the presentation. Second, they promise increased portability and ease of evolution as maintenance and extension is done within a more formal framework. Finally, they enable new forms of designer support, such as automated design critique, refinement, and implementation.

5.5. Agent-Based Interaction

The papers in Section VII consider the use of agents in the interface. Important questions explored include, What can and should an agent do? How they should do it? and How, when, and why should they interact with the user when doing it? Agents promise to decrease human workloads and make the overall experience of interaction less stressful and more productive. Agents may assist by decreasing task complexity, bringing expertise to the user (in the form of expert critiquing, task completion, coordination), or simply providing a more natural environment with which to interact. The papers in this section report examples of each of these and also describe open architectures for building agent-based multimodal interfaces, the use of agents to express system and discourse status via facial displays, and the multimodal communication between animated computer agents.

5.6. Empirical Evaluation

The final section focuses on IUI evaluation. Whereas community-based evaluation using standard corpora and tasks has been applied in several areas related to intelligent interfaces (most notably, DARPA evaluations in speech, starting with Hirschman 1989; information extraction, e.g., MUC-6 1995; and information retrieval, e.g., TREC-1, Harmon 1993), relatively little evaluation has been systematically performed on IUIs. This section attempts to collect the best examples to foster more rigorous development of and widespread use of evaluation in the future. Important dimensions of the problem include considering human-human versus human-computer communication, spoken versus written communication, unimodal versus multimodal communication, and direct versus mediated communication.

As the papers in this section illustrate, the evaluator and analyst have at their disposal a range of tools, such as Wizard-of-Oz experiments, simulations, and instrumentation of live environments to evaluate a range of metrics using a variety of quantitive and qualitative measures and evaluation methodologies (e.g., corpus based, task based).

5.7. Content Index

Because many of the papers address issues that cut across section distinctions, Table 1 provides cross-references to facilitate access to chapters according to the following categories:

1. Media input and output data types investigated (e.g., text, speech, graphics, gesture), which, as a result, indicate if the investigations examine cross-stream or multiple media processing
2. The underlying models (e.g., of the user, discourse, task, situation) that are created, maintained, and exploited
3. Representational devices utilized, such as numerical-, rule-, plan-, model-, or agent-based processing
4. Application areas addressed, (e.g., decision/design support, information access or creation, training)

SECTION	Input Analysis					Presentation Design						Graphics Design					Layout					User & Discourse Models								Model-Based Design						Agents					Evaluation			
CHAPTER	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	2.6	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	4.5	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	6.1	6.2	6.3	6.4	6.5	6.6	7.1	7.2	7.3	7.4	7.5	8.1	8.2	8.3	8.4
Text	✖	✖	✖	✖		✖	✖	✖	✖	✖	✖	✖					✖	✖	✖			✖	✖	✖	✖	✖	✖	✖	✖		✖	✖	✖	✖	✖		✖	✖	✖	✖		✖	✖	✖
Speech	✖	✖	✖	✖																																		✖	✖	✖		✖	✖	✖
Audio/Music																													✖															
Graphics	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖		✖	✖	✖			✖	✖		✖	✖	✖	✖	✖	✖	✖	✖	✖		✖	✖	✖		✖	✖	✖	✖
Maps																													✖									✖				✖	✖	✖
Gesture	✖	✖	✖	✖		✖	✖	✖																✖				✖	✖									✖	✖	✖	✖		✖	
Gaze				✖	✖																																		✖	✖				
Discourse		✖	✖	✖	✖	✖	✖	✖					✖											✖		✖	✖	✖	✖		✖	✖	✖	✖	✖				✖	✖				✖
User			✖	✖	✖		✖	✖		✖				✖	✖	✖						✖	✖	✖	✖	✖	✖	✖	✖										✖	✖				
Task			✖	✖		✖	✖	✖		✖				✖	✖	✖						✖	✖	✖	✖	✖	✖		✖	✖	✖	✖	✖	✖	✖	✖	✖		✖	✖	✖			✖
Media						✖	✖	✖	✖	✖	✖								✖		✖			✖					✖										✖		✖		✖	
Statistical					✖				✖																✖																			
Rule-Based	✖					✖			✖	✖	✖	✖	✖	✖	✖	✖	✖		✖	✖	✖	✖			✖			✖	✖								✖			✖				
Plan-Based		✖	✖	✖			✖	✖										✖					✖			✖	✖																	
Model-Based																								✖	✖					✖	✖	✖	✖	✖					✖	✖	✖			✖
Agent-Based																																				✖		✖	✖	✖			✖	
Design Assist												✖	✖	✖		✖	✖	✖	✖	✖	✖									✖	✖	✖	✖	✖	✖		✖							
Decision Support	✖	✖	✖	✖	✖																	✖		✖		✖	✖		✖									✖			✖	✖	✖	✖
Help						✖	✖		✖	✖	✖												✖		✖											✖			✖					
Info Access					✖	✖	✖	✖	✖	✖	✖													✖	✖			✖								✖		✖	✖	✖	✖	✖	✖	✖
Training						✖			✖		✖		✖																	✖								✖						

TABLE 1. Content-Based Index

6. CONCLUSION

Intelligent interfaces promise to improve the quality of interaction for all who interact with computers—at work and at play. They promise

- more *efficient* interaction—enabling more rapid task completion with less work.
- more *effective* interaction—doing the right thing at the right time, tailoring the content and form of the interaction to the context of the user, task, dialog.
- more *natural* interaction—supporting spoken, written, and gestural interaction, ideally as if interacting with a human interlocutor.

When these interfaces are created in a model-based fashion, modifying their behavior will require model changes, not reprogramming. This will reduce the time, cost, and expertise required to develop interfaces and, at the same time, will facilitate the creation of more principled interfaces. Intelligent interface technology will be essential to effective information interaction in the future. For example, better interaction via the web (Brusilovsky 1996) has been identified as a challenging problem, and intelligent web sites in the future promise to discover user and group skills and interests, tailor presentations, and automatically improve web site interfaces (Perkowitz and Etzioni 1997). In short, this area has the potential to improve the quality and effectiveness of interaction for everyone who communicates with a machine in the future. To achieve these benefits, however, we must overcome the remaining fundamental problems outlined in the chapters herein.

7. RESOURCES

A number of resources for teachers, students, and researchers contain additional information about this subject area. Several relevant collections of papers are cited in the references at the end of this introductory section, notably, *Intelligent User Interfaces* (Sullivan and Tyler 1991), *Intelligent Multimedia Interfaces* (Maybury 1993), and *Readings in Human-Computer Interaction* (Baecker et al. 1995). Maybury (1995) provides a summary of research in multimedia parsing and generation. Key journals (and associated World Wide Web sites) in which new results in intelligent user interfaces appear include the international journals of *Human-Computer Interaction {www.parc.xemx. com/istl/projects/HCI),* including a special issue on multimedia interfaces (Oviatt and Wahlster 1997), *User Modeling and User-Adapted Interaction (umuai.informatik.uni-essen.de), Artificial Intelligence (www.elsevier.com/locate/artint), Cognitive Science,* and *ACM Transactions on Graphics* as well as more general forums such as *Communications of the ACM* and *IEEE Computer.*

In addition to journals and books, a series of conferences and workshops can provide additional sources, such as the annual ACM-sponsored International Conference on Intelligent User Interfaces *(sigart.acm.org/iui99/),* the International Conference on User Modeling, the International Workshop on Advanced Visual Interfaces (AVI), and the User Interface Systems Technology (UIST Conference). Proceedings from a number of annual or semiannual conferences typically contain sessions on intelligent user interfaces, such as the conference of the Association for Computing Machinery Special Interest Group on Computer Human Interaction *(www.acm.org /sigchi),* the American Association of Artificial Intelligence (AAAI) National Conference on Artificial Intelligence, the European Conference on Artificial Intelligence (ECAI), and the International Joint Conference on Artificial Intelligence (IJCAI). There are also many related conferences and specialized workshops in subdisciplines, including speech and language processing, user and discourse modeling, multimedia, and intelligent training systems.

Finally, materials are increasingly available online, such as an on-line tutorial on intelligent multimedia interfaces *(www.mitre.org/resources/centers /advanced_info/mark.html);* an on-line survey, "State of the Art in Human Language Technology" *(www.cse.ogi.edu/CSLU/HLTsurvey);* a study by **the** National Research Council on Every Citizen Interfaces *(www.nap.edu/readingroom/books/screen);* a **human-**computer interaction index *(is.twi.tudelft.nl/hci);* and the Electronic Transactions on Artificial Intelligence *(www.ida.liu.se/ext/etai/indexframe.html),* **which** includes a special area on intelligent user interfaces *(www.dfki.de/~andre/etai/colloqb.html).* **Additional** pointers are available from the ACL SigMedia special interest group in multimedia language processing *(www.dfki.de/sigmedia).* Related government initiatives include the European Intelligent Information Interfaces program *(www.i3net.org)* and DARPA's Intelligent Collaboration and Visualization program *(snad.ncsl.nist.gov/~icv-ewg/).*

8. REFERENCES

Allen, J., and Hendler, J. (eds.). 1990. *Readings in Planning.* San Francisco: Morgan Kaufmann.

Arens, Y.; Feiner, S.; Foley, J.; Hovy, E.; John, B.; Neches, R.; Pausch, R.; Schorr, H.; and Swartout, W. 1991.

Intelligent User Interfaces. (ISI/RR-91-288). University of Southern California Information Sciences Institute Research Report, Marina del Rey, CA.

Baecker, R.; Grudin, J.; Buxton, W.; and Greenberg, S. 1995. *Readings in Human-Computer Interaction: Toward the Year 2000* (second edition). San Francisco: Morgan Kaufmann.

Binot, J-L.; Falzon, P.; Perez, R.; Peroche, B.; Sheehy, N.; Rouault, J.; and Wilson, M. D. 1990. Architecture of a Multimodal Dialogue Interface for Knowledge-Based Systems. In *Proceedings of Esprit '90 Conference,* 412-433. Dordrecht, Netherlands: Kluwer Academic Publishers.

Blattner, M., and Dannenberg, R. (eds.). 1992. *Multimedia Interface Design.* New York: ACM Press.

Bordegoni, M.; Faconti, G.; Feiner, S.; Maybury, M.; Rist, T.; Ruggieri, S.; Trahanias, P.; and Wilson, M. 1997. A Standard Reference Model for Intelligent Multimedia Presentation Systems. *The International Journal on the Development and Application of Standards for Computers, Data Communications and Interfaces, Special Issue on Intelligent Multimedia Presentation Systems.* Vol. 18, Nos. 6 and 7. Amsterdam: Elsevier Science.

Brachman, R. J., and Levesque, H. J. (eds.) 1985. *Readings in Knowledge Representation.* San Francisco: Morgan Kaufmann.

Brusilovsky, P. 1996. Methods and Techniques of Adaptive Hypermedia. In Brusilovsky, P., and Vassileva, J., (eds.), *Special Issue on Adaptive Hypertext and Hypermedia, User Modeling and User-Adapted Interaction* 6 (2-3): 87-129.

Everett, R.; Zraket, C. A.; and Bennington, H. 1957. SAGE: A Data Processing System for Air Defense, In *Proceedings of the Eastern Joint Computer Conference,* Washington, DC, December. IRE (now IEEE), 148-155. Reprinted in Everett, R.; Zraket, C. A.; and Bennington, H. 1983. SAGE: A Data Processing System for Air Defense. *Annals of the History of Computing,* 5(4): 330-339. October.

Grosz, B. J., Sparck Jones, K., and Webber, B. 1986. *Readings in Natural Language Processing.* San Francisco: Morgan Kaufmann.

Hirschman, L. (chair), *Proceedings of the Speech and Natural Language Workshop,* February 1989. San Francisco: Morgan Kaufmann.

Horvitz, E. 1997. Agents with Beliefs: Reflections on Bayesian Methods for User Modeling. Invited Talk, Sixth International Conference on User Modeling, Chia Laguna, Sardinia, June 2-5. *(http://www.research.microsoft.com /research/dtg/horvitz/lum. htm)*

Kobsa, A.; Allgayer, J.; Reddig, C; Reithinger, N.; Schmaucks, D.; Harbusch, K.; and Wahlster, W. 1986. Combining Deictic Gestures and Natural Language for Referent Identification. In *Proceedings of 11th International Conference on Computational Linguistics,* Bonn, Germany, 356-361.

Kobsa, A., and Wahlster, W. (eds.). 1989. *User Models in Dialog Systems.* Berlin: Springer-Verlag.

Maybury, M. T. 1993. *Intelligent Multimedia Interfaces.* Menlo Park, CA/Cambridge, MA: AAAI/MIT Press.

Maybury, M. T. 1994. Automated Explanation and Natural Language Generation. In Sabourin, C, (ed.), *Computational Text Generation,* Bibliography, Montreal: Infolingua, 1-88.

Maybury, M. T. 1995. Research in Multimedia Parsing and Generation. In McKevitt, P. (ed.), *Artificial Intelligence Review: Special Issue on the Integration of Natural Language and Vision Processing,* 9(2-3): 103-127.

Maybury, M. T. 1997. *Intelligent Multimedia Information Retrieval.* Menlo Park, CA/Cambridge, MA: AAAI/MIT Press.

MUC-6. 1995. *Proceedings of the Sixth Message Understanding Conference.* Advanced Research Projects Agency Information Technology Office, Columbia, MD, November 6-8.

Oviatt, S., and Wahlster, W. (eds.). 1997. *Special Issue on Multimodal Interfaces. Human-Computer Interaction: A Journal of Theoretical, Empirical, and Methodological Issues, of User Science and of System Design,* 12(1, 2). Mahwah, NJ: Lawrence Erlbaum Associates.

Pattabhiraman, T., and Cercone, N. (eds.). 1991. *Computational Intelligence: Special Issue on Natural Language Generation,* 7(4) November. National Research Council of Canada.

Pelachaud, C; Badler, N. I.; and Steedman, M. 1996. Generating Facial Expressions for Speech. *Cognitive Science,* 20(1): 1-46.

Perkowitz, M., and Etzioni, O. 1997. Adaptive Web Sites: An AI Challenge. In *Proceedings of Fifteenth International Joint Conference on Artificial Intelligence (IJCAI '97),* Nagoya, Japan, August 23-29, 16-21.

Schneider-Hufschmidt, M.; Klihme, T.; and Malinowski, U. (eds.). 1993. *Adaptive User Interfaces: Principles and Practice.* Amsterdam: Elsevier.

Sullivan, J., and Tyler, S. (eds.). 1991. *Intelligent User Interfaces.* New York: ACM Press.

TREC-1. 1993. First Text REtrieval Conference (TREC-1), NIST, Harmon, Donna K. (ed.). NIST Special Publication 500-207, March.

Waibel, A., and Lee, K. 1990. *Readings in Speech Recognition.* San Francisco, CA: Morgan Kaufmann.

Webber, B. L., and Nilsson, N. J. (eds.). 1985. *Readings in Artificial Intelligence: A Collection of Articles.* San Francisco: Morgan Kaufmann.