INTELLIGENT INTERFACES FOR UNIVERSAL ACCESS: CHALLENGES AND PROMISE

Mark T. Maybury
Information Technology Division
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730, USA
maybury@mitre.org
http://www.mitre.org/resources/centers/it

ABSTRACT

This article outlines some challenges and opportunities in the area of universal access to intelligent interfaces. Technological advances in computing and communication together with increasingly available connectivity provide both new challenges and new opportunities for disadvantaged individuals to become more full participants in society. It is also the case that careful exploitation of these advances can make advantaged workers that much more efficient or effective, for example, when they find themselves in communications and/or computing disadvantaged environments such as while mobile or in remote areas.

1. UNIVERSAL INTERFACES

Growing volumes of information, shrinking product creation cycles, and increasing global competition demand optimal utilization of scare expert human resources. Effective human machine interfaces and information services promise to increase access and productivity for all. Furthermore, governments are moving more actively to promoting universal access. For example, in the United States, Section 508 of the Workforce Reinvestment Act, passed in November 1998, requires that people with disabilities have equity in use of electronic and information technology (Peet 2001).

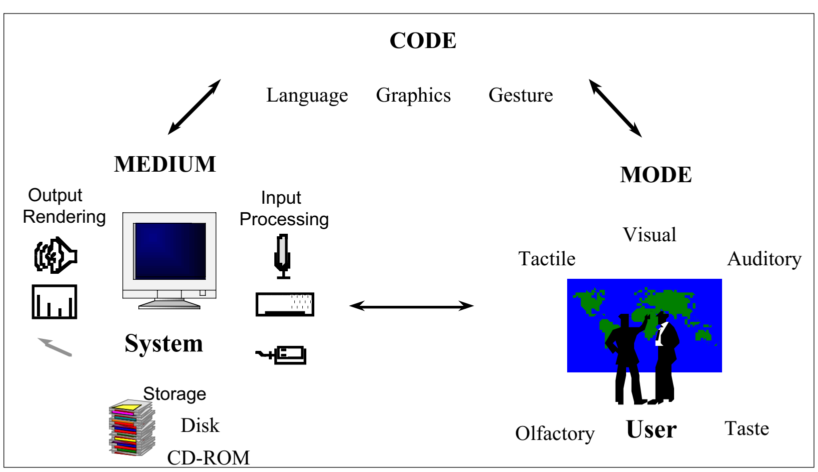

Figure 1. Medium, Mode and Code

To frame our intelligent user interface analysis, we first define some terminology (see Figure 1). Following Maybury and Wahlster (1998), we distinguish the terms medium, mode, and code. By *mode* or *modality* we refer primarily to the human senses employed to process incoming information, i.e., vision, audition, olfaction, touch, and taste. In contrast, we use *medium* to refer to the material object or carrier used for saving or presenting information including computer input/output devices (e.g., microphone, speaker, screen, pointer). As illustrated in Figure 1, medium is material-centered whereas modality is human-centered. We use the term *code* to refer to a system of symbols, syntax, and semantics (e.g., natural language, pictorial language, gestural language) used to encode and transmit information, often via multiple media which may support multiple modes. For example, a natural language code might use typed/written text or speech, which would rely upon visual or auditory modalities and associated media (e.g., keyboard, microphone). In the remainder of this paper we consider challenges and opportunities with multimedia/mode/code interfaces for all.

While today's computer-user interfaces have advanced beyond command line interfaces, they remain principally direct manipulation also called WIMP (windows, icons, menus, and pointing) interfaces. As Figure 2 illustrates, the typical architecture of these systems consists of three primary components. Presentation refers to the widgets – including windows, icons, and menus – that represent the application and are the primary mechanisms of communication and control between the user and system. Dialogue control refers to the data and flow control logic that manages the sensing and action from user interactions with the presentation. Application programming interfaces (APIs) are the routines invoked to perform the underlying system functions (e.g. inserting, deleting, modifying or otherwise manipulating application objects).

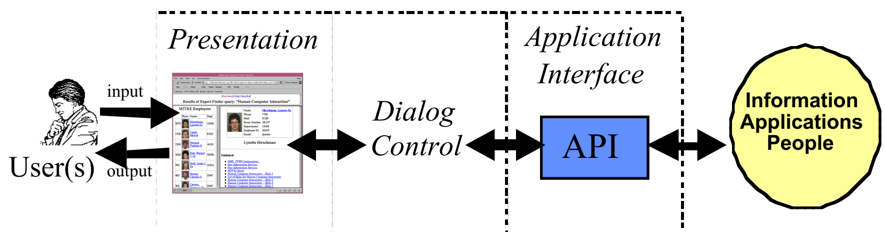

Figure 2. Presentation, Dialog Control and APIs

This is the architecture that is found today's popular graphical user interfaces (e.g., Microsoft Windows, Macintosh, and UNIX windowing systems). Accordingly, today's interface designers have three primary targets for enhancing the universality of human computer interfaces: presentation, dialog control, and APIs. For example, the interface can be customized by providing presentations of dialog control which are tailored to individual user needs (e.g., voice navigation or large, highly contrasted buttons for the visually impaired) or customizing APIs (e.g., limiting available functionality for novice or memory challenged users).

2. INTELLIGENT INTERFACES FOR ALL

Sophisticated interface tailoring requires modeling and reasoning about the domain, task, user, and dialog, as well as interaction media, modes and codes via interactive devices. A grand challenge of universal interfaces is to represent, reason, and exploit these models to more effectively process input, generate output, and manage the dialogue and interaction between human and machine so that we maximize the efficiency, effectiveness, and naturalness, if not joy, of interacting. These key interface functions form the basis of Table 1 that articulates some grand challenges for each of these functions. Table 1 also indicates benefits that accrue for all users if these challenges are addressed, which we briefly discuss in turn.

Interface Function	Grand Challenges	Potential Benefits
Input Analysis	Interpretation of imprecise, ambiguous, and/or partial multimodal input	Flexibility and choice in alternative media input, synergistic input
Output Generation	Automated generation of coordinated speech, natural language, gesture, animation, non-speech audio, generation, possibly delivered via interactive, animated life-like agents.	Mixed media (e.g., text, graphics, video, speech and non-speech audio) and mode (e.g., linguistic, visual, auditory) displays tailored to the user and context.
Dialog Control	Mixed initiative natural interaction that deals robustly with context shift, interruptions, feedback, and shift of locus of control.	Ability to tailor flow and control of interactions and facilitate interactions including error detection and correction tailored to individual physical, perceptual and cognitive differences. Motivational and engaging life-like agents.
Agent/User Modeling	Unobtrusive learning, representation, and use of models of agents (including the user).	Enables tracking of user characteristics, skills and goals in order to enhance interaction.
API	Dealing with increasingly broad and complex application functionality.	Simplification of functionality, possibly limited by user and/or context models. Automated task completion. Task help tailored to situation, context, and user. Mobile and substitutable interfaces for disabled users.

Table 1. Interface Challenges and Benefits

Input Analysis. One of the limitations of current user interfaces is the restricted choice, flexibility and expressiveness of input mechanisms. Today these are typically keyboard and mouse. An example of the limited flexibility is that input from these media must occur using a single input device or sequential input from several devices. An example of limited power is that mouse input is typically constrained to 2D input. Now consider that a user in a hands-eyes busy task (e.g., driving, working with machinery, performing medical diagnosis/procedures) may not have the luxury of using their hands or even eyes to provide input. This limitation can be severe for the motor or visually impaired. Universal interfaces should overcome a user's absolute or situational inability to utilize particular modes (e.g., auditory, visual, tactile) and/or interactive devices (e.g., screen, keyboard, mouse). Advances to broaden the available media to perform similar functions can be important, such as the ability to navigate and/or select options using voice or eye gestures instead of hands and keyboard. This furthermore requires codes and/or artificial languages that capture not only the syntax natural language, graphics, gesture). By providing similar communicative actions across codes, systems can enable users to accomplish the same function across codes (e.g., performing a selection with a typed, spoken, or gestural input). Finally, analysis of the pragmatics (i.e., beliefs, intent, emotions) of inputs can provide even richer models of the user. For example, most speech recognizers ignore intonation associated with signals thus loosing valuable indicators from pitch or intonation patterns of the user's emotional state (e.g., calm, agitated or stressed).

Output Generation. Just as current input devices limit users' expressiveness, so too presentation devices limit the range of expressiveness of computers. Effective presentations require reasoning about communicative intent, selecting content, allocating content to particular codes (e.g., language, graphics), and realizing content on or across particular codes, media and modes. For example, the same intention and propositional content (e.g., a severe weather warning) can be realized in a coordinated fashion in a range of media, modes and codes (e.g., as a non-speech audio siren, as a natural language italicized 12-point Times font email message). Selection can be driven by a range of criteria including but not limited to properties of the message (e.g., importance, urgency, size), the intended recipient (e.g., their physical, perceptual, cognitive, emotional state), the situation (e.g., a small screen public kiosk), and the state of the speaker (e.g., the personality of an animated agent). Limitations in user perceptual modalities (e.g., auditory, visual, tactile) can suggest preferred or required realizations. For example, hearing impaired users often use alarm clocks that vibrate their bed. Microsoft Windows provides user configurable capabilities such as StickyKeys or MouseKeys, ShowSounds, and auditory alerts to support users with tactile, auditory, and visual impairments, respectively. Finally, in perceptual or cognitive overload situations, an effective rendering of information can actually reduce user attention requirements and stress.

Dialog Control & User/Agent Modeling. Even the best input devices and output displays are not useful without control. Whereas current WIMP interfaces primarily support interaction via a standard set of devices including windows, icons, menus and dialogue boxes, these can be tailored to specific user and situational needs. For example, surface display properties of these interactive devices such as their size, or location can enable visual or motor impaired users to perform selections or provide input. In addition, time-outs or displays that change function with time can be sensitive to the reaction times of individual users. For example, elderly users sometimes require more time for query formulation and response generation during human computer dialogues than younger users.

Interfaces with the ability to manage discourse promise more natural interaction that deals robustly with context, interruptions, feedback, and support of mixed-initiative, wherein the locus of control ebbs and flows naturally between system and user. Explicit modeling and management of dialogue would support the ability to tailor the flow and control of interactions and facilitate them using mechanisms such as error detection and correction tailored to individual physical, perceptual and cognitive differences. Doing so requires modeling the user, their (physical, perceptual, and cognitive) characteristics as well as their beliefs, goals, and plans. Representing, reasoning and acting upon these models plays a foundational role in input processing, output generation, and management of dialogue and interaction.

Recent research focusing on life-like animated agents promises more human-like interactions. For example, knowing that the elderly typically have difficulty hearing sounds above 3 kHz can dictate the use of a low-pitched voice in an animated interface agent. While the use of emotion in a life like agent has not yet been shown to increase end user performance, motivational effects have been demonstrated in some applications (André, Rist, and Müller 2000). This could be valuable, for example, for motivation and/or attention disorders. Also, users with time and/or resource limitations may choose to delegate certain functions to an interface agent.

API. Whereas the complexity and breadth of desktop applications can be overwhelming for typical users, they can be disastrous for users with attention deficits. For example, Microsoft Word has over one thousand commands and yet only 20 of these constitute 80% of usage (Linton et al 1999, 2000). Using models of the user and task, it is feasible to limit the range of available functions, to support automated task completion, and to provide help tailored to the user, task and context. In addition, having a standard interface API would support ready substitution for personal interfaces. If used with wireless devices, standard APIs could ease the connectivity of individual devices tailored to the needs of mobility impaired users or those with limited dexterity or control (e.g. Parkinson's patients).

3. CONCLUSION

Truly universal access for all presents grand challenges to the user interface community. It demands a series of scientific advances that will require moving beyond our current interface architectures. Intelligent dialogue control as well as intelligent processing and management of multiple media, modalities and codes is fundamental to this progress and promises increased flexibility in input, processing, and output. Intelligence can be applied to the interface to enable a range of functions, including but not limited to:

1. Understanding imprecise, ambiguous, and/or partial multimodal/media input
2. Generating coordinated, cohesive, and coherent tailored multimodal/media presentations
3. Interaction management that is sensitive to models of the user, domain, task, and context, providing such functions as semi or automated completion of delegated tasks, recovery from miscommunication, tailored interaction styles, and interface adaptation.

Universality will require new modeling advances, such as languages that incorporate partial representation and reasoning, and geospatial and temporal reasoning. Input devices and rendering mechanisms will need to be tailored to the physical, perceptual and cognitive nature of the user, possibly even within individual sessions. They will also need to adapt to the requirements of the environment (e.g., personal computer, public kiosk, mobile digital assistant). Intelligent and adaptive interfaces promise universal access for all across a broad set of domains for work, learning and play.

4. REFERENCES

André, E., Rist, T. and Müller, J. 1999. Employing AI Methods to Control the Behavior of Animated Interface Agents. *Applied Artificial Intelligence*. 13: 415-448.

Boykin, S. and Merlino, M. Feb. 2000. A. Machine Learning of Event Segmentation for News on Demand. *Communications of the ACM*. Vol 43(2): 35-41.

Linton, F., Joy, D., & Schaefer, H-P. 1999. Building User and Expert Models by Long-Term Observation of Application Usage. In J. Kay (Ed.), UM99: User Modeling: Proceedings of the Seventh International Conference (pp. 129-138). New York: Springer Verlag. [Selected data are accessible from an archive on http://zeus.gmd.de/ml4um/]

Linton, F., Joy, D., Schaefer, H-P., & Charron, A. 2000. OWL: A Recommender System for Organization-Wide Learning. Educational Technology and Society. http://ifets.ieee.org/periodical/

Maybury, M., Merlino, A., and Morey, D. 1997. Broadcast News Navigation using Story Segments, ACM International Multimedia Conference, Seattle, WA, November 8-14, 381-391.

Maybury, M. and Wahlster, W. editors. 1998. *Readings in Intelligent User Interfaces*. Morgan Kaufmann Press.

Maybury, M. Feb. 2000. News on demand: Introduction. *Communications of the ACM*. Vol 43(2): 32-34.

Peet, M. 2001. Information Access for the Disabled: the Section 508 Mandate and its Implications for Intelligent Interface Development. In 1st International Conference on Universal Access in HCI. New Orleans, LA .