HUMAN COMPUTER INTERACTION:

STATE OF THE ART AND FURTHER DEVELOPMENT IN THE INTERNATIONAL CONTEXT- NORTH AMERICA

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

1. INTRODUCTION

Intelligent human computer interaction[1] promises to support more sophisticated and natural input and output, to enable users to perform potentially complex tasks more quickly, with greater accuracy, and to improve user satisfaction. Intelligent interfaces are becoming increasingly important as users face increasing system complexity and information overload, as expert staff levels decrease, and with corporation's increasing requirements for systems that are adaptive to global commerce and heterogeneous user populations. These systems are typically characterized by one or more of the following properties (Maybury 1993, 1999; Maybury and Wahlster 1999):

1. *Multimodal input* – they process potentially ambiguous, impartial, or imprecise combinations of mixed input such as written text, spoken language, gestures (e.g., mouse, pen, dataglove) and gaze.
2. *Multimodal output* – they design coordinated presentations of, e.g., text, speech, graphics, and gestures, which may be presented via conventional displays or animated, life-like agents.
3. *Interaction management* – mixed initiative interactions that are context-dependent based on system models of the discourse, user, and task

This new class of interfaces promises *knowledge or agent-based dialogue*, in which the interface gracefully handles errors and interruptions, and dynamically adapts to the current context and situation. The overarching aim of intelligent interfaces is to both increase the interaction bandwidth between human and machine (e.g., by increasing interactive media and modalities) and at the same time increase interaction effectiveness by improving the quality of interaction. For example, by explicitly monitoring user attention, intention, and task progress, an interface can explain why an action failed, predict a user's next action, warn a user of undesirable consequences of actions, or suggest possible alternative actions. This article outlines the state of the art, remaining challenges and potential benefits of this critical technology area.

[1] See http://www.mitre.org/resources/centers/it/ maybury/iui99/index.htm for an on line tutorial.

Human Computer Interaction. International Status Conference, Saarbruecken, 26-27 October 2001.
Sponsored by Federal Ministry of Education and Research, Federal Ministry of Economics and Technology

2. DRIVERS

Before we consider the state of the art, why all the fuss about interfaces? A number of factors have increased their role and importance. These include:

- *Speed, efficiency, and joy*. Organizations are pressured to enhance worker productivity and use of scarce resources (especially human capital). Effective interfaces can have dramatic influences on (especially knowledge) worker productivity. Moreover, enjoyable interfaces have been shown to reduce stress and decrease errors.

- *Ubiquity and mobility* – Computers are now not only everywhere but in everything – at work, at home, at play, in cars … even in our sporting equipment. New wireless platforms such as cell phones, PDAs, and wearable computers have enabled mobile computing, affording new interaction opportunities and challenges.

- *Personalization* – Mass customization has become a global phenomena. For example, today increasing numbers of purchased automobiles (7% in the US; 19% in England; 60% in Germany) are custom built to individual specifications. [2] Analogously, models of user motor, perceptual and cognitive abilities allow us to specialize interaction, either by allowing users to adapt the interface or have the system automatically adapt it based on interaction context (e.g., localizing language to a specific country, customizing a display for a handheld device).

- *Graying of society*. One factor in the recent thrust toward so-called "Interfaces for All" has been dealing with aging populations that require increases in affordances for users who may require enhanced visual, auditory or physical interaction (e.g., slower response times in dialogues) (Peet 2001). In the United States, this has led to "Universal Access" legislation (Section 508) which requires that "federal agencies provide employees and members of the public who have disabilities access to electronic and information technology that is comparable to the access available to employees and members of the public who are not disabled". Section 508 builds upon the World Wide Web Consortium (W3C)' s Accessibility Initiative (WAI) which has developed Accessibility Guidelines (www.w3c.org/WAI)

3. STATE OF THE ART

As Table 1 details, intelligent human computer interaction includes analysis of input, generation of output, management of interaction, modeling and adaptation to the user, and support for interaction with the underlying system. The table outlines the evolution of synergistic multimodal input, coordinated multimodal output, tailored dialog control, and detailed tracking and reacting to

[2] Mass Customization. *The Economist*. July 14-20, p. 63. Report of the International Car Distribution Program (ICDP).

models of the user and discourse. In addition to supporting a much richer range of interaction styles, interface advances enable the user to do things they perhaps could not otherwise.

Interface Function	**State of the Art**	**Grand Challenges**	**Benefits of Advances**
Input Analysis	Sequential keyboard and two dimensional mouse or touch screen input. Limited spoken language input.	Interpretation of imprecise, ambiguous, and/or partial multimodal input	Flexibility and choice in alternative media input, synergistic input, robust interpretation.
Output Generation	Canned presentations utilizing primarily graphics and text. Single document, mono-lingual summarization.	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. Life like animated characters.
Dialog Control	Pre-scripted interactions with standard dialogue presentations (e.g., windows, menus, buttons)	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	Limited models of user interests (e.g., via explicitly solicited user models). Recommender technology.	Unobtrusive learning, representation, and use of models of user/agents, including models of perception, cognition, and emotion.	Enables tracking of user characteristics, skills and goals in order to adapt and enhance interaction.
API	Variable specification of underlying application functionality. Move toward component based architectures.	Addressing increasingly broad, interdependent, 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

Several areas a worth highlighting as key interface trends to watch. These include the growth of agent communication languages, the introduction of affect into the interface, and the growing focus on awareness and knowledge management, each of which we briefly describe.

- *Agent communication languages*. Advances in tools and techniques for control of knowledge-rich components is being advanced by specific architectures such as the Open Agent Architecture (OAA, http://www.ai.sri.com/~oaa) but also by government initiatives such as the Distributed Agent Markup Language (www.daml.org) and the semantic web (www.w3.org/2001/sw, www.semanticweb.org).

- *Affective interfaces*. Recognizing and expressing mood and emotion via the interface has received increased interest. This could come, for example, in the form of detecting delight or stress via language, speech, and gesture or expressing emotional displays via an interactive life-like agents. It could also be as practical as detecting and effecting drowsiness in a car-driver interface.

- *Awareness*. The explosion of Instant Messaging (IM) and associated presence information has increased user desire for information regarding user identity, physical and virtual location, activity (e.g., idle, working), availability, and communication capability (e.g., platform, interactive devices, network connectivity). In addition to awareness of individual characteristics, there also is a need for awareness of the emergence and tracking of group activity and roles participants play (e.g., who is the leader, facilitator, key contributor).

- *Knowledge Management.* Strongly related to awareness are areas necessary to support knowledge access, including:
  - *Expert Discovery*: Modeling, cataloguing and tracking of distributed organizations and communities of experts.
  - *Knowledge Discovery*: Identification and classification of knowledge from unstructured multimedia data.
  - *Knowledge Sharing*: Awareness of and access to enterprise expertise and know-how.

Finally, human computer interaction often occurs in the context of information seeking tasks, so we briefly visit the state of the art in related technologies. In 2001, with the importance of information seeking on the web, the best (automated) information retrieval systems that have been formally evaluated by the National Institute for Standards and Technology[3] return documents relevant to a particular subject with around 80% precision but low recall. Manual query specification performs better than automated query expansion, however, automated relevance feedback is nearly indistinguishable from manual methods. Translingual document and passage retrieval is an active area of research and system building. Systems can automatically (across several languages) identify entities (e.g., person, organization, location names) at over 90% accuracy and relations among entities (e.g., a person lives-at some location) at approximately 70-80% accuracy (measured in terms of precision and recall)[4]. Automated document summarization[5] can reduce documents to 20% of their source size without information loss, saving users 50% of task time. This has been evaluated across multiple systems in both English

---

[3] The Text Retrieval Conference (trec.nist.gov), run by NIST, has been held annually since 1992.

[4] See tutorial at www.ai.sri.com/~appelt/ie-tutorial.

[5] See www.mitre.org/resources/centers/it/maybury/summarization/summarization.htm for tutorial.

and Japanese. Finally, a new class of systems can respond to a simple factual (English) questions by returning answers from relevant documents at 75% accuracy.

4. US INITIATIVES

A number of public initiatives are focused on advancing the state of human computer interaction. For example, advances in information retrieval are reported in the annual Text REtrieval Conference (trec.nist.gov). Results from the previous two TRECs can be found at trec.nist.gov/pubs/trec8/index.track.html (TREC8) and trec.nist.gov/pubs/trec9/index.track.html (TREC9). Detecting the emergence of new topics of concern and tracking them is the focus of the Topic Detection and Tracking (TDT) Evaluation Project (www.nist.gov/speech/tests/tdt/index.htm) and also a track within the TREC Conference. Other advances are reported in international journals and conferences, such as those sponsored by ACM special interest group in information retrieval (SIGIR, www.acm.org/sigir).

Several initiatives focus on enhancing information extraction (e.g., see NIST Information Extraction web page at www.itl.nist.gov/iaui/894.02/related_projects/muc/index.html). Government organizations have joined together in the Automated Content Extraction (ACE) program (www.nist.gov/speech/tests/ace). The Translingual Information Detection Extraction and Summarization (TIDES) program (www.darpa.mil/ito/research/tides, tides.nist.gov) is aimed at advancing precisely its name.

A new focus on answering questions from sources is the focus of the AQUAINT program (www.ic-arda.org/solicitations/AQUAINT). Results of the first Q&A track in TREC8 can be found at citeseer.nj.nec.com/346894.html and at trec.nist.gov. Finally, DARPA Communicator is focused on the creation of shared infrastructure for spoken dialogue systems (http://fofoca.mitre.org). The EU AMITIES http://www.dcs.shef.ac.uk/nlp/amities) project focuses on call center automation.

5. CONCLUSION

Truly universal access requires fundamental advances across the interface functionalities described in Table 1. This includes needs in the areas of interpretation, generation, and interaction management. At the same time, valuable enhancements even to today's interfaces are possible with recent advances. To a significant extent, interface research, has not benefited from the same kind of corpus-based, community evaluation which has resulted in steady performance improvements in speech, information retrieval, and information extraction. In these situations, common grand challenge problems, shared data sets, resources and tools as well as comparative system evaluations have increased community learning. Technology advancement has attracted additional investment and spin-off commercialization as the technology matures. We should seek similar data set formulation and task based evaluation for core interface functions.

6. 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.

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. T. (ed.) 1993. *Intelligent Multimedia Interfaces*. AAAI/MIT Press. 405 pp. ISBN 0-262-63150-4 (www.aaai.org:80/Press/Books/Maybury1, mitpress.mit.edu/book-home.tcl?isbn=0262631504)

Maybury, M. T. 1999. Keynote. Intelligent Multimedia for the New Millennium. Proceedings of Eurospeech '99. Budapest, September 6-9, 1999. vol 1. p. KN1-15.

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

Mani, I. and Maybury, M. (eds.) 1999. *Advances in Automatic Text Summarization*. MIT Press, Cambridge

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.