



CHI 2006 Workshop Proceedings
Sunday, April 23rd 2006

What is the Next Generation of Human-Computer Interaction?

Robert J.K. Jacob

Department of Computer Science
Tufts University
Medford, MA, 02155 USA

jacob@cs.tufts.edu

<http://www.cs.tufts.edu/~jacob/workshop/>



Workshop Abstract

What is the Next Generation of Human-Computer Interaction?

Robert J.K. Jacob
Department of Computer Science
Tufts University
Medford, MA 02155 USA
jacob@cs.tufts.edu

Keywords

Virtual reality, ubiquitous computing, tangible interfaces, reality-based interfaces, next-generation interfaces, non-WIMP interfaces.

ACM Classification Keywords

H.1.2 [Models and Principles]: User/Machine Systems--human factors; H.5.2 [Information Interfaces and Presentation]: User Interfaces

MOTIVATION

Is there an emerging next generation of human-computer interaction or rather simply “a thousand points of light” of disparate and unrelated innovative new developments? This workshop will bring together researchers in a range of emerging new areas of HCI to look for common ground and a common understanding of a next generation of user interfaces. If we consider command-line interfaces as the first generation, then direct manipulation and the graphical user interface defined a distinct new generation of user interfaces[5] that is still the state of practice today. Unlike the early days of graphical user interfaces, research in HCI today is developing on many fronts, making the next generation more difficult to connect and define. Yet, much current research appears to be moving away from the screen based GUI, in a related general direction.

The key components of this next generation are found in a variety of loosely-related current research areas in HCI—or, more broadly, interaction design or human-information design:

- virtual and augmented reality
- ubiquitous, pervasive, and handheld interaction
- tangible user interfaces
- lightweight, tacit, passive, or non-command interaction
- perceptual interfaces

- affective computing
- context-aware interfaces
- ambient interfaces
- embodied interfaces
- sensing interfaces
- eye-movement based interaction
- speech and multi-modal interfaces.

This workshop seeks ways to tie some of these and others together intellectually and to try to define a next generation from them. We will look for unifying ideas, frameworks, and theories that provide common ground for discussing, analyzing, connecting, inventing, comparing, and making predictions about emerging new interaction styles and interface designs. We also hope to use such a framework provide some explanatory power for understanding what makes particular new interfaces better or worse. In addition, it could help identify gaps or opportunities to develop a future research agenda suggested by holes or “sweet spots” in a new taxonomy.

STARTING POINT: REALITY-BASED INTERACTION

As a starting point for discussion, we will introduce the notion of natural or realistic or reality-based interfaces as a thread to connect new developments in HCI. This notion focuses on the ways in which the new interfaces leverage users' built-in abilities. These interfaces increasingly draw their strength from exploiting the user's pre-existing skills and expectations from the real world rather than trained computer skills. For example, navigating through a conventional computer graphics system requires a set of learned commands, such as keywords to be typed in or function keys to be pressed. By contrast, navigating through virtual reality exploits the user's existing real-world “navigational commands:” positioning the head and eyes, turning the body, and walking toward something of interest.

Perhaps basing the interaction on the real world reduces the mental effort required to operate the system because the user is already skilled in those aspects of the system. For casual use, this reduction can speed learning; for use in situations involving information overload, time pressure, or stress (e.g., surgery, disaster recovery), this reduction of overhead effort could improve performance.

A unifying characteristic for much next generation HCI may thus be that it increasingly taps into the users' abilities and pre-existing knowledge. Direct manipulation moved user interfaces toward more realistic interaction with the computer; next generation, reality-based interfaces push further in this direction, increasing the realism of the interface objects and allowing the user to interact even more directly with them.

We could also take this approximate notion of “realistic” or “natural” and make it more precise—perhaps by focusing on the pieces of knowledge or skills that a system requires its user to know. This could lead to a notional checklist of the knowledge the user needs, which may help in discussing and connecting different new user interfaces. However, there are many kinds of things that the user already knows. Moving the head to change point of view is one. The user may already know more arcane facts such as pressing the Alt-F4 keys will close a window. It seems intuitively better to exploit the more “basic,” more built-in knowledge that the user learned in infancy (or perhaps was born with) than to exploit more recently learned, less innate knowledge, like the Alt-F4 keys. We could explore how to measure reality-based vs. non-reality-based knowledge on a more continuous scale. This requires a way to rate the degree of reality or innate-ness for a piece of knowledge. One way is by when the user learned it; and we conjecture that younger is better. Information that is deeply ingrained in the user seems somehow more robust, perhaps more highly practiced, and should take less effort to use than information learned recently. These questions may be explored in the workshop.

Another side of this issue is that reality-based is typically not sufficient. A useful interface will rarely entirely mimic the real world, but will necessarily include some “unrealistic” or artificial features and commands. In fact much of the power of using computers comes from this “multiplier” effect, the ability to abstract from or go beyond a precise imitation of the real world.

WORKSHOP GOALS

Our goal is to find common elements for understanding and discussing a next generation of HCI and to build a community of researchers to consider this topic explicitly, in contrast to many recent developments in new interaction styles, which have thus far tended to proceed independently on unconnected and unrelated fronts.

We will use the notion of reality-based interaction to provide a concrete starting point for the workshop. We will

begin by considering whether we can use that to tie together developments in next generation interaction styles into the beginning of a useful conceptual framework. From there, depending on the participants' views and contributions, we will extend or expand this approach as well as introducing alternative opposing or complementary approaches to the problem. Participants will be invited to extend, expand, discredit, or replace this initial approach, but it will provide a concrete starting point with which to agree or disagree. Current research at Tufts is fleshing out the reality-based approach with analyses of reality-based knowledge and skills needed for different interaction styles, and we will provide our latest work as input to start the discussion.

We will thus invite participants to present:

- their current new interface designs or research that they see as part of next-generation interaction
- alternative frameworks or theories to the reality-based approach
- refinements and elaborations of it
- ideas for how to test frameworks and concepts we develop
- psychological evidence or theories
- ideas for new designs inspired by gaps or opportunities uncovered by this thinking.

We hope a key contribution will be that ideas emerging from the workshop will serve as a lens or common language for viewing, discussing, comparing, and advancing proposed innovative new interface developments and technologies—to provide some coordinate axes on which to put them into perspective and organize them.

A second goal is to produce a research agenda for new work both in gaps suggested by our frameworks and in ways to evaluate or validate our frameworks. The initial conclusions we draw from these for simple examples may turn out to be true, but not surprising. We will focus on ways to make and test theory-based predictions that are less obvious, as a better way to test our theories.

Our final goal is to create a community of HCI researchers who are thinking specifically about connecting their research to other developments in next generation interaction. We hope the results will generate ideas that give the HCI community a new and more explicit way of thinking about and connecting next generation interaction and will suggest a research agenda for future work in this area.

PARTICIPANTS & EXPECTED COMMUNITY INTEREST

The workshop will welcome researchers working in areas such as those listed above (virtual and augmented reality, ubiquitous, pervasive, and handheld interaction, tangible user interfaces, etc.) and in particular: participants researching and developing things they view as part of next generation interaction; participants with ideas or approaches

for describing or defining next generation interaction; and participants with ideas for how to elaborate, formalize, test, evaluate, or expand on the reality-based starting point and to replace it with new approaches.

To date, few researchers have addressed this issue explicitly, but several have discussed sub-areas and made contributions toward it. People who have attempted to explain or organize these new styles of user interfaces have tended to concentrate more on individual classes or groups of new interfaces than on concepts that unify the classes. The time is ripe to start a discussion that connects such work. For example, Ullmer and Ishii provide a framework for tangible interfaces[6]; Fishkin, Moran, and Harrison propose the concept of embodied interfaces[3]; Bellotti, Back, Edwards, Grinter, Henderson, and Lopes define sensing interfaces, and raise a set of key problems[2]; and Beaudouin-Lafon's Instrumental Interaction model sheds new light on post-WIMP interfaces[1].

We are also pursuing this area at Tufts, under an NSF grant, and teaching a course in Fall 2005 on "Reality-based Interaction: Understanding the Next Generation of User Interfaces," which will provide new work as input to the workshop. The NSF project will also serve as a nexus for continuing and collecting work in this topic after the workshop.

Finally, work that helped define the GUI generation is a model for us. Shneiderman took a set of disparate new user interfaces and unified them through their common characteristics[5]. Hutchins, Hollan, and Norman then explained their power and success of these interfaces with a theoretical framework[4]. Our hope is to take a first step in that direction for the emerging generation.

ACKNOWLEDGEMENTS

We thank our collaborators Andrew Afram, Eric Bahna, Georgios Christou, Audrey Girouard, Michael Horn, Leanne Miller, Michael Poor, Orit Shaer, Erin Treacy, Larissa Winey, and Jamie Zigelbaum of the Computer Science Department at Tufts as well as Caroline Cao and Holly Taylor of Tufts, Leonidas Deligiannidis of University

of Georgia, Frank Ritter of Penn State University, and Hiroshi Ishii and Sile O'Modhrain of the MIT Media Lab.

And we thank the National Science Foundation for support for our research project on this topic (NSF grant IIS-0414389).

REFERENCES

1. M. Beaudouin-Lafon, "Instrumental Interaction: An Interaction Model for Designing Post-WIMP User Interfaces," Proc. ACM CHI 2000 Human Factors in Computing Systems Conference, pp. 446-453, Addison-Wesley/ACM Press (2000).
2. V. Bellotti, M. Back, W.K. Edwards, R.E. Grinter, A. Henderson, and C. Lopes, "Making Sense of Sensing Systems: Five Questions for Designers and Researchers," Proc. ACM CHI 2002 Human Factors in Computing Systems Conference, pp. 415-422, ACM Press (2002).
3. K.P. Fishkin, T.P. Moran, and B.L. Harrison, "Embodied User Interfaces: Toward Invisible User Interfaces," Proc. of EHCI'98 European Human Computer Interaction Conference, Heraklion, Crete (1998).
4. E.L. Hutchins, J.D. Hollan, and D.A. Norman, "Direct Manipulation Interfaces," in User Centered System Design: New Perspectives on Human-computer Interaction, ed. by D.A. Norman and S.W. Draper, pp. 87-124, Lawrence Erlbaum, Hillsdale, N.J. (1986).
5. B. Shneiderman, "Direct Manipulation: A Step Beyond Programming Languages," IEEE Computer, vol. 16, no. 8, pp. 57-69 (1983).
6. B. Ullmer and H. Ishii, "Emerging Frameworks for Tangible User Interfaces," in Human-Computer Interaction in the New Millenium, ed. by J.M. Carroll, Addison-Wesley/ACM Press, Reading, Mass. (2001).

Papers

Special Guest

A Second Path to HCI Innovation:
Generative Theories Tied to User
Needs

Ben Shneiderman University of Maryland

A. Frameworks and Surveys

A.1 On Beyond GUI

Steven Feiner Columbia University

A.2 Tangible User Interfaces

Hiroshi Ishii MIT Media Laboratory

A.3 An Encompassing View on
Tangible Interaction: A Framework

Eva Hornecker University of Sussex, UK and
University of Canterbury, New Zealand

A.4 From "Human-Computer
Interaction" to "The Use and Design of
Digital Artefacts"

Lars Erik Holmquist Viktoria Institute,
Sweden

A.5 What is the Next Generation of
Human-Computer Interaction?

George Fitzmaurice Autodesk, Toronto

A.6 Interaction Design for Active
Bodies: Two Themes

Scott R. Klemmer Stanford University
Bjorn Hartmann Stanford University
Leila Takayama Stanford University

B. Broader Perspectives, Psychological Considerations

B.1 Opportunities and Challenges for
HCI Design and Research

Jim Hollan University of California, San Diego

B.2 Matching Human Abilities and
Limitations: The Basis for Better
Future User Interfaces

David Kieras University of Michigan

B.3 Coping with Information Overload in the New Interface Era

Steven M. Drucker Microsoft Research, Redmond, WA

B.4 Some Musings On Building and Showing Future UI Prototypes

Andrew D. Wilson Microsoft Research, Redmond, WA

C. New Interaction Styles

C.1 Moving Towards Naturalistic Interactions With Invisible Interfaces: Can "Digital Simplicity" Be Achieved?

Beverly L. Harrison Intel Research Seattle

C.2 Egocentric Interaction

Thomas Pederson Umea University, Sweden

C.3 Core Tangibles and Tangible Visualizations: prospects for tangible convergence and divergence

Brygg Ullmer Louisiana State University

C.4 Computational Corporeality: Computers, Materials, and the Body

Amanda Parkes MIT Media Laboratory

C.5 The Future of Interface Design, Through a Child's Eyes

Hayes Raffle MIT Media Laboratory

C.6 Frames of Reference for Portable Device Interaction

Sile O'Modhrain Queen's University, Belfast

C.7 Embodied Interfaces: The Next Generation of HCI?

Chris Creed University of Birmingham, UK
Russell Beale University of Birmingham, UK

D. New Interface Designs and Systems

D.1 Here Or There Is Where? Haptic Egocentric Interaction With Topographic Torch

Mike Bennett UCD Dublin
Sile O'Modhrain Queen's University, Belfast
Aaron Quigley UCD Dublin

D.2 Personalized Interactions with Reality Based Interfaces

Angela Chang MIT Media Laboratory
Hiroshi Ishii MIT Media Laboratory

D.3 Reality-Based Object Movement Techniques for 3D

Darius Dadgari York University, Toronto
Wolfgang Stuerzlinger York University, Toronto
Ji-Young Oh University of Arizona

D.4 Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction

Desney S. Tan Microsoft Research, Redmond, WA

D.5 The Future of Mobile Device Research in HCI

Jacob O. Wobbrock Carnegie Mellon University

D.6 Federated Devices: Augmenting Mobility with Ambient Resources

Frankie James SAP Research, Palo Alto, CA

D.7 Reflecting on the Potential of VR for the Future of HCI

Timo Ropinski University of Munster, Germany
Frank Steinicke University of Munster, Germany
Klaus Hinrichs University of Munster, Germany

D.8 Pervasive Interaction - Using movement and speech to interact with computers

Thomas Riisgaard Hansen University of Aarhus, Denmark
Jakob E. Bardram University of Aarhus, Denmark
Mads Sogaard University of Aarhus, Denmark

D.9 Managing Extrinsic Costs via Multimodal Natural Interaction Systems

Rebecca Lunsford Natural Interaction Systems, Portland, OR
Ed Kaiser Natural Interaction Systems, Portland, OR
Paulo Barthelmess Natural Interaction Systems, Portland, OR
Xiao Huang Natural Interaction Systems, Portland, OR
David McGee Natural Interaction Systems

E. Tools and Development Techniques

E.1 An Approach for Developing Ubiquitous Augmented Reality Systems

Bernd Bruegge Technical University of Munich
Asa MacWilliams Technical University of Munich
Asim Smailagic Carnegie Mellon University

E.2 Reflecting on Production Tools and Methods for Mixed Reality Design

Steven Dow Georgia Institute of Technology
Blair MacIntyre Georgia Institute of Technology

E.3 Exploring Activity-Based Computing: Interaction Styles, Models and Tool Support

Yang Li University of Washington
James A. Landay University of Washington and Intel Research Seattle

F. Other Issues and Innovations

F.1 World of WarCraft, the role-playing game and the deconstruction of the classical model. The construction of environments and multiform narrations.

Mar Marcos Molano Universidad Complutense de Madrid
Maria Luisa Garcia Guardia Universidad Complutense de Madrid
Michael Santorum Gonzalez Universidad Complutense de Madrid

F.2 An Argument for Using Techniques Found in Audio Games as the Next Generation of User Interfaces for Blind People

Chris Hofstader

F.3 Considering the Next Generation(s) of Human Computer Interface: A conceptual framework or theory

Fred Welter ITT

Participants

Robert J.K. Jacob
(Organizer)

Tufts University
jacob@cs.tufts.edu

Ben Shneiderman
(Special guest)

University of Maryland
ben@cs.umd.edu

Mike Bennett

UCD Dublin
mike.bennett@ucd.ie

Bernd Bruegge

Technical University of Munich
bruegge@cs.cmu.edu

Angela Chang

MIT Media Laboratory
anjchang@media.mit.edu

Chris Creed

University of Birmingham, UK
C.P.Creed@cs.bham.ac.uk

Darius Dadgari

York University, Toronto
dariusd@cs.yorku.ca

Steven Dow

Georgia Institute of Technology
steven@cc.gatech.edu

Steven M. Drucker

Microsoft Research, Redmond, WA
sdrucker@microsoft.com

Steven Feiner

Columbia University
feiner@cs.columbia.edu

George Fitzmaurice

Autodesk, Toronto
fitzmaurice@acm.org

Thomas Riisgaard Hansen

University of Aarhus, Denmark
thomasr@daimi.au.dk

Beverly L. Harrison

Intel Research Seattle
beverly.harrison@intel.com

Jim Hollan

University of California, San Diego
hollan@cogsci.ucsd.edu

| | |
|----------------------------|--|
| Lars Erik Holmquist | Viktoria Institute, Sweden leh@viktoria.se |
| Eva Hornecker | University of Sussex, UK and University of Canterbury, New Zealand eva@ehornecker.de |
| Hiroshi Ishii | MIT Media Laboratory ishii@media.mit.edu |
| Frankie James | SAP Research, Palo Alto, CA frankie.james@sap.com |
| David Kieras | University of Michigan kieras@eecs.umich.edu |
| Scott R. Klemmer | Stanford University srk@cs.stanford.edu |
| Yang Li | University of Washington yangli@cs.washington.edu |
| Rebecca Lunsford | Natural Interaction Systems, Portland, OR rebecca.lunsford@naturalinteraction.com |
| David McGee | Natural Interaction Systems davidrmcgee@hotmail.com |
| Sile O'Modhrain | Queen's University, Belfast sile@qub.ac.uk |
| Amanda Parkes | MIT Media Laboratory amanda@media.mit.edu |
| Thomas Pederson | Umea University, Sweden top@cs.umu.se |
| Aaron Quigley | UCD Dublin aaron.quigley@ucd.ie |
| Hayes Raffle | MIT Media Laboratory hayes@media.mit.edu |
| Timo Ropinski | University of Munster, Germany ropinski@math.uni-muenster.de |

Mads Sogaard

University of Aarhus, Denmark
madss@daimi.au.dk

Desney S. Tan

Microsoft Research, Redmond, WA
desney@microsoft.com

Brygg Ullmer

Louisiana State University
ullmer@cct.lsu.edu

Andrew D. Wilson

Microsoft Research, Redmond, WA
awilson@microsoft.com

Jacob O. Wobbrock

Carnegie Mellon University
jrock@cs.cmu.edu

**Student volunteers and
participants**

Audrey Girouard
agirou01@cs.tufts.edu

Michael Horn
mhorn01@cs.tufts.edu

Leanne Miller
Leanne.Miller@tufts.edu

Orit Shaer
orit_shaer@yahoo.com

Erin Treacy
Erin.Treacy@tufts.edu

Jamie Zigelbaum
jamie.zigelbaum@tufts.edu

A Second Path to HCI Innovation: Generative Theories Tied to User Needs

Ben Shneiderman
Department of Computer Science
University of Maryland
College Park, MD 20742 USA
(301) 405-2680
Email ben@cs.umd.edu
homepage <http://www.cs.umd.edu/~ben>

CHI 2006 Workshop Position Paper:
What is the Next Generation of Human-Computer Interaction?

The most frequent source of innovation in many fields, including human-computer interaction (HCI), is by proposing a novel technology approach, such as tangible, embedded, embodied, ubiquitous, or pervasive user interfaces. Other popular proposals in recent years include context-aware, perceptual, mobile, and handheld. The resulting systems are sometimes described as invisible or disappearing, and may include attributes such as affective, ambient, and implantable. Media features can also lead to innovations such as visualization, animation, sonification, haptics, gestural, and multi-modal. These sources of innovation are effective in suggesting new products and services, but technology-centered approaches need to be supplemented by user-centered approaches to provide a more complete guide to the future.

The second path to HCI innovation starts from user-centered approaches and leads to novel products and services based on a deep understanding of human needs. I explored these approaches earlier in *Leonardo's Laptop: Human Needs and the New Computing Technologies* (MIT Press, 2002) by proposing a generative theory of HCI tied to an Activities and Relationship Table. The activities included collect, relate, create, and donate, while the relationships were described as self, family & friends, colleagues & neighbors, and citizens & markets. The resulting 4 x 4 table can be helpful in suggesting new products and applications, tied to new market segments.

This Activities and Relationship Table is not a yellow brick road or golden path to produce new products and services, but can be helpful in structuring innovation by guiding designers to opportunities to serve different markets with novel services. For example, this generative theory could be applied to personal medical devices that support health monitoring to enable users to collect information, such as blood pressure, temperature, or insulin levels. The generative nature of theory guides innovators to consider convenient sharing of information with healthcare professionals or family members, enabling users to analyze their own data or and compare their readings with others with similar age, gender, and health conditions.

These generative theories are a compass to guide designers to think about user needs. These needs may be tied to personal use domains, as they are in *Leonardo's Laptop*, such as education,

healthcare, e-commerce, and government services. Other fresh discussions of user needs are emerging in many circles as evidenced by the appearance of conferences and workshops on topics such as persuasive computing, that is meant to alter user habits about healthcare, nutrition, or personality, and human sexuality, that covers educational, social, and therapeutic themes.

Other second path topics that show growth are enhancing trust and privacy in e-commerce website design, and supporting empathy while controlling anger in online communities. These affective issues deal with emotional relationships between two people, and enable more fruitful outcomes for patient support groups with nurturance for caregivers, as well as more effective mediation and conflict resolution.

Online security researchers are increasingly aware of the HCI issues tied to responsibility, accountability and anonymity in discussions of spam, viruses, and hacker attacks. While anonymity is important for some forms of freedom of expression, it becomes a shield to hide behind for those bent on destructive and violent use of the internet. Enhancing responsibility and accountability could accelerate acceptance of e-commerce applications, as ebay has done so successfully with its trust management and mediation methods.

The negative emotional side of human nature is surfacing in HCI design discussions, as issues such as anger, credibility, risk, and fear become more common. Discussion groups can quickly be disrupted by a single hostile act and chat rooms become dangerous for children who become victims for child abusers.

The reality of user frustration is also gaining attention in studies that reveal how much time is wasted by most computer users [1, 2] and how often they fail to accomplish their goals in e-learning, e-commerce, and e-healthcare or other web sites. Long delays, dropped phone lines, and dead links are widespread, as are application crashes and operating system problems. Improvements in reliability would do much to improve the user experience.

Other generative theories come from discussions of universal usability [3] that propose innovation in three areas:

Technology variety: Support broad range of hardware, software, and network access
{e.g. fast/slow computers, fast/slow networks, small/large screens }

User diversity: Accommodating users with different skills, knowledge, age, gender, disabilities, literacy, culture, income, etc
{e.g. design for screen readers, user control over font size, contrast, etc. }

Gaps in user knowledge: Bridge the gap between what users know and what they need to know
{e.g. novices/experts, newcomers, cross cultural visitors }

Universal usability concepts guide innovators by reminding them of the diverse needs of novices vs experts, so that they consider multi-layer designs of interfaces. They also remind users to consider the differences between users of consumer electronics (cellphone, music players, digital cameras, etc.) and professional workstations (air traffic control, radiology, programming environments), as well as the differences between self-confident highly literate users of multiple

devices in developed nations and frightened users with low literacy of wholly new technologies in developing nations.

In summary, technology-oriented thinking is one path to innovation for advanced user interfaces, but a second path is defined by generative theories that guide thinking in structured ways. These orderly approaches, based on taxonomies of human values, needs, roles, and tasks may also be helpful in shaping the future.

- [1] Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., and Shneiderman, B., Determining causes and severity of end-user frustration, *International Journal of Human-Computer Interaction* 17, 3 (2004), 333-356.
- [2] Lazar, J., Jones, A., and Shneiderman, B., Workplace user frustration with computers: An exploratory investigation of the causes and severity, *Behaviour & Information Technology* (to appear, 2006).
- [3] Shneiderman, B., Universal Usability: Pushing human-computer interaction research to empower every citizen, *Communications of the ACM* 43, 5 (May 2000), 84-91.
- [4] Shneiderman, B., Promoting universal usability with multi-layer interface design, *ACM Conference on Universal Usability*, ACM Press, New York (2003), 1-8.

A. Frameworks and Surveys

A.1 On Beyond GUI

Steven Feiner Columbia University

A.2 Tangible User Interfaces

Hiroshi Ishii MIT Media Laboratory

A.3 An Encompassing View on
Tangible Interaction: A Framework

Eva Hornecker University of Sussex, UK and
University of Canterbury, New Zealand

A.4 From "Human-Computer
Interaction" to "The Use and Design of
Digital Artefacts"

Lars Erik Holmquist Viktoria Institute,
Sweden

A.5 What is the Next Generation of
Human-Computer Interaction?

George Fitzmaurice Autodesk, Toronto

A.6 Interaction Design for Active
Bodies: Two Themes

Scott R. Klemmer Stanford University
Bjorn Hartmann Stanford University
Leila Takayama Stanford University

On Beyond GUI

A Position Paper for the CHI 2006 Workshop “What is the Next Generation of Human-Computer Interaction?”

Steven Feiner
Dept. of Computer Science
Columbia University
New York, NY 10027
feiner@cs.columbia.edu

Introduction

In HCI, as in any other technical field, trying to predict the future has always been fashionable, even though it rarely proves accurate. In this position paper, I will not try to predict explicitly where I think things will go next, but will instead review some of the themes underlying the research in Columbia’s Computer Graphics and User Interfaces Lab, insofar as we believe that these themes characterize what we think future user interfaces should, and ultimately will, do better than they do now.

Wearability

What better theme to start with than *wearability*? Especially since it is missing from the list of research areas that appear in the workshop call for participation. Over the past decade, many researchers have delineated and explored some of the key ways in which wearable user interfaces differ from ones that are merely mobile or hand-held [Rhodes 1997; Mann 1997], including “always on” performance, minimization of conventional manual input, sensitivity to environmental context, and the perception by the user and others that the device is intimately associated with (and, in some cases, considered to be “part of”) the user.

These distinctions are, in part, tied up with the idea that wearable user interfaces are active during our normal interactions with the world, and therefore must avoid needlessly distracting us from the world. One way to address this is to make interaction, when possible, *eyes-free*—avoiding the need for the user to visually monitor a display, as we do with conventional GUIs, instead of attending to the world around them. Even if the computer’s display is overlaid visually on the user’s view of the world through augmented reality [Feiner 2002], the need to watch virtual material in conjunction with or instead of the real world can be burdensome. This is especially true of interaction techniques that control a cursor to select items or manipulate widgets such as buttons and sliders—these are classic examples of replacing dedicated devices that do not always require visual interaction (e.g., physical buttons and sliders in familiar configurations) with the lowest-common denominator of mouse or stylus-driven virtual devices whose users must visually track a cursor.

An alternative approach employs *cursorless* interaction techniques [Blaskó & Feiner 2004a]. Some of these techniques directly emulate devices such as buttons or sliders through interactions on dedicated areas of a touch-sensitive surface. *Tactile landmarks*, such as the raised bezel of a watch computer [Blaskó & Feiner 2004b], as shown in Figure 1, or the edge of a touchpad, guide the user’s fingers to the desired locations, without the need for visual feedback for the interaction technique itself.

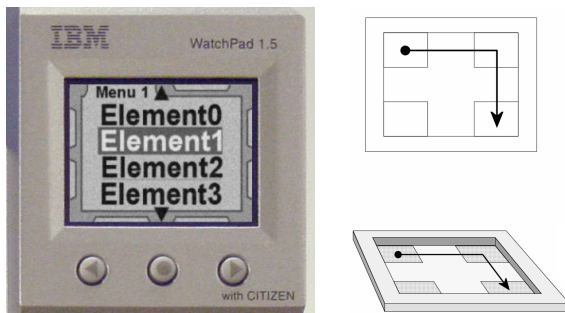


Figure 1. Cursorless menu design for research prototype watch computer. The user’s finger can easily follow the tactile landmarks established by the raised bezel (shown schematically on the right) to traverse the menu without visual interaction, once the user is sufficiently familiar with the menu design. (From [Blaskó & Feiner 2004b].)

The move to mobile and wearable computing has a number of other implications [Feiner 1999]. For example, the potential for same-place/same-time collaboration is greatly increased when users are mobile, rather than physically isolated at separate, stationary computers. Furthermore, since mobile users can freely and quickly move in and out of each other’s presence, this encourages transient, evanescent interactions, some of which may transpire in less time than it takes to turn on and off a “conventional” hand-held device.

Heterogeneity

As displays proliferate, it is becoming common to use them in concert. Desktop GUI users often spread their desktops across multiple displays, and it is difficult to find a clamshell mobile phone that doesn’t have two displays (or even three, counting the display on the Bluetooth earpiece). But, there are much more interesting possibilities than tiling several of the same or similar displays to emulate a single seamless desktop, or placing displays on different surfaces of a phone. We have used the term *hybrid user interface* [Feiner & Shamash 1991] to refer to the synergistic combination of heterogeneous user interface technologies, whether displays or other interaction devices, to take advantage of the best features of each.

In our earliest hybrid user interface prototype [Feiner & Shamash 1991], the user wore a tracked, see-through, head-worn display through which they viewed the stationary flat panel display of a desktop computer. The flat panel display served as the conventional desktop of an X11 window manager. However, when users moved windows to the edge of that display, they could continue onto the surface of a virtual hemispherical surround that was presented on the see-through head-worn display, effectively extending the desktop beyond the bounds of the flat panel. That virtual space of this secondary display was centered about the user’s head and tangent to the flat panel primary display. While the secondary display was of lower quality than the primary display, it was much larger, since the user could look around to view different portions of it. Thus, the flat panel acted as a high-resolution informational “sweet spot” inset within the larger and coarser context of the head-worn display. Other approaches to combining heterogeneous displays have included embedding a high-resolution flat panel within a lower-resolution projection screen [Baudisch et al. 2001], viewing part of a 3D model through an opaque stereo head-worn display while seeing a lower-resolution projection of the entire model through peripheral vision [Ilie et al. 2004], and moving a tracked hand-held Tablet PC on the surface of a larger projected tabletop to selectively view additional detail [Benko et al. 2004]. In all these examples, the displayed material does not overlap because display of the context is suppressed in the overlap area, either through blanking it in software or physical obstruction.

In contrast, a hybrid user interface can also use heterogeneous displays and interaction devices to display and interact with complementary overlapping content. For example, EMMIE [Butz et al. 99] uses tracked, see-through head-worn displays to embed a variety of other displays within a virtual visual “ether.” This allows a 2D search dialog, displayed and manipulated on a tracked, hand-held tablet, to be overlaid with 3D leader lines terminating on objects within the environment that fulfill the search criteria. Similarly, Sandor and colleagues [Sandor et al. 2005] connect tracked physical interaction devices to the graphical objects that they control on desktop displays by overlaying labeled 3D leader lines viewed on a tracked, see-through head-worn display.

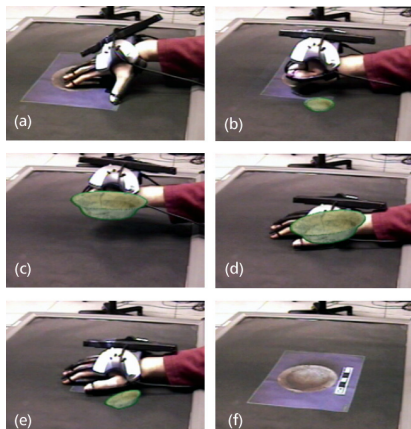


Figure 2. Frames from a sequence of cross-dimensional pull and push gestures that transition virtual objects between 2D and 3D display and interaction technologies. (a) Pull gesture begins by selecting the 2D object and beginning to form a grabbing gesture. (b) The 2D object disappears, while the 3D object appears from the table. (c) Holding the 3D object. (d) Push gesture begins by pressing on the table through the 3D object. (e) The 3D object disappears and the 2D object appears. (f) The 2D object, projected on the table. (Table surface is covered with black paper to provide a darker background for imaging the projected display through the live tracked video see-through display used to create these images.) (From [Benko et al. 2005].)

Benko and colleagues [Benko et al. 2005] make it possible for objects to be viewed and manipulated using 2D or 3D displays and interaction techniques and smoothly transferred between these different dimensionalities. A 2D touch-sensitive projected tabletop is viewed through a 3D volume that is displayed on a tracked, stereo, see-through head-worn display and manipulated with a tracked glove. As shown in Figure 2, virtual objects can be “pushed” or “pulled” between the displays using *cross-dimensional gestures* that are recognized through simultaneous tracking of the user’s hands using the 2D touch-sensitive surface and the 3D tracked glove.

Adaptivity

Long-lived and short-lived interactions, played out in a changing environment, with a dynamic mix of users, displays, devices, call out for the ability to sense and adapt to changing context. For example, even the relative motion of a user within a static 3D environment can dramatically alter what is visible at any given time. We use the term *view management* [Bell et al. 2001] to refer to the automated dynamic layout of information that is presented to the user (e.g., virtual representations of physical objects and textual annotations) to enforce desired spatial relationships (e.g., to prevent more important material from being obscured by less important material), as demonstrated in Figure 3. This can be made possible by the use of efficient algorithms that analyze the projection of the scene from the user’s viewpoint to determine an appropriate position and size for each item controlled by the system. While the example of adaptivity provided here addresses only layout, many other aspects of user interfaces, ranging from the content presented, to the displays and devices used, could be adaptively modified in response to contextual changes, to create more effective user interfaces. A key challenge is to make dynamic adaptation comfortable, understandable, and, in general, desirable to users, rather than confusing and unwelcome.

Conclusions

In the course of writing this position paper, I looked back at one that I wrote for the *SIGGRAPH 1990 Workshop on Software Architectures and Metaphors for Non-WIMP User Interfaces*. While I was unsurprised that several of the research problems that I had discussed then remain far from solved today, I was struck by one way in which our field has measurably matured. In 1990, yearly conferences relevant to HCI were few and far between. Outside of *CHI* and the barely two-year-old *UIST*, work on specific kinds of user interfaces could be presented at well established, but far broader, conferences, such as *SIGGRAPH*, *Eurographics*, *Graphics Interface*, and *AAAI*, or at a few relatively



Figure 3. View management applied to the labeling of a virtual campus model and the layout of a meeting agenda and annotated building model, in a collaborative augmented reality system. The layout also automatically ensures that the line of sight is kept clear between the user at the center of the figure and the user from whose view the image is shown. (Photographed through a tracked, see-through, head-worn display.) (From [Bell et al. 2001].)

new specialist venues, such as *CSCW* and *I3D*. Now, some fifteen years later, many of the research areas listed in the present workshop call for papers and in this position paper have their own yearly conferences that are, in many cases, as long-lived (or longer) now as the eight-year-old *CHI* was then: *VR*, *VRST*, *ICAT*, *ISMAR*, *UbiComp*, *MobileHCI*, *PerComm*, *MUM*, *ICMI*, *AVI*, *ISWC*, and *IUI*, to name just some of the better known ones. While we still have a long way to go, the number of researchers working on these problems has increased tremendously, and there are substantial, individually cohesive (although, oftentimes, collectively disparate), research communities.

Acknowledgments

Many current and former members of the Columbia University Computer Graphics and User Interfaces Lab contributed to the ideas expressed here, including Blaine Bell, Hrvoje Benko, Gábor Blaskó, Andreas Butz, Tobias Höllerer, Edward Ishak, Blair MacIntyre, Alex Olwal, and Christian Sandor. This work was funded in part by NSF Grants IIS-01-21239 and IIS-00-82961, Office of Naval Research Contracts N00014-04-1-0005, N00014-99-1-0394, N00014-99-1-0683, and N00014-99-1-0249, National Library of Medicine Contract R01 LM06593-01, Air Force Research Lab Contract FA8650-05-2-6647, and gifts from IBM, Mitsubishi Electric Research Laboratories, and Microsoft Research.

References

- Baudisch, P., Good, N., and Stewart, P. Focus plus context screens: Combining display technology with visualization techniques. *Proc. UIST 2001 (ACM Symp. on User Interface Software and Technology)*, Orlando, FL, November 11–14, 2001 (*CHI Letters*, vol. 3, no. 2), 31–40.
- Bell, B., Feiner, S., and Höllerer, T. View management for virtual and augmented reality. *Proc. UIST 2001 (ACM Symp. on User Interface Software and Technology)*, Orlando, FL, November 11–14, 2001 (*CHI Letters*, vol. 3, no. 2), 101–110.
- Benko, H., Ishak, E., and Feiner, S. Collaborative mixed reality visualization of an archaeological excavation. *Proc. ISMAR 2004 (IEEE and ACM Int. Symp. on Mixed and Augmented Reality)*, Arlington, VA, November 2–5, 2004, 132–140.
- Benko, H., Ishak, E., and Feiner, S. Cross-dimensional gestural interaction techniques for hybrid immersive environments. *Proc. IEEE Virtual Reality 2005*, Bonn, Germany, March 12–16, 2005, 209–216.
- Blaskó, G. and Feiner, S. Single-handed interaction techniques for multiple pressure-sensitive strips. *ACM CHI 2004 Extended Abstracts*, Vienna, Austria, April 24–29, 2004, 1461–1464.
- Blaskó, G. and Feiner, S. An interaction system for watch computers using tactile guidance and bidirectional segmented strokes. *Proc. ISWC 2004 (IEEE Int. Symp. on Wearable Computers)*, Arlington, VA, October 31–November 3, 2004, 120–123.
- Butz, A., Höllerer, T., Feiner, S., MacIntyre, B., and Beshers, C. Enveloping users and computers in a collaborative 3D augmented reality. *Proc. IWAR '99 (IEEE and ACM Int. Workshop on Augmented Reality)*, San Francisco, CA, October 20–21, 1999, 35–44.
- Feiner, S. The importance of being mobile: Some social consequences of wearable augmented reality systems. *Proc. IWAR '99 (IEEE and ACM Int. Workshop on Augmented Reality)*, San Francisco, CA, October 20–21, 1999, 145–148.
- Feiner, S. Augmented reality: A new way of seeing. *Scientific American*, 286(4), April 2002, 34–41.
- Ilie, A., Low, K.-L., Welch, G., Lastra, A., Fuchs, H., and Cairns, B. Combining head-mounted and projector-based displays for surgical training. *Presence*, 13(2), April 2004, 128–145.
- Mann, S. Smart clothing: The wearable computer and wearcam. *Personal Technologies*, 1(1), March 1997, 21–27.
- Rhodes, B. The wearable remembrance agent: A system for augmented memory. *Personal Technologies*, 1(4), December 1997, 218–224.
- Sandor, C., Olwal, A., Bell, B., and Feiner, S. Immersive mixed-reality configuration of hybrid user interfaces. *Proc. ISMAR 2005 (IEEE and ACM Int. Symp. On Mixed and Augmented Reality)*, Vienna, Austria, October 5–8, 2005, 110–113.

Tangible User Interfaces

Hiroshi ISHII
Tangible Media Group
MIT Media Laboratory

Introduction

Where the sea meets the land, life has blossomed into a myriad of unique forms in the turbulence of water, sand, and wind. At another seashore between the land of atoms and the sea of bits, we are now facing the challenge of reconciling our dual citizenships in the physical and digital worlds. Our visual and auditory sense organs are steeped in the sea of digital information, but our bodies remain imprisoned in the physical world. Windows to the digital world are confined to flat square screens and pixels, or "painted bits." Unfortunately, one can not feel and confirm the virtual existence of this digital information through one's hands and body.

Imagine an iceberg, a floating mass of ice in the ocean. That is the metaphor of Tangible User Interfaces. A Tangible User Interface gives physical form to digital information and computation, salvaging the bits from the bottom of the water, setting them afloat, and making them directly manipulatable with human hands.

From GUI to TUI

People have developed sophisticated skills for sensing and manipulating their physical environments. However, most of these skills are not employed in interaction with the digital world today. A Tangible User Interface (TUI) is built upon those skills and situates the physically-embodied digital information in a physical space. Its design challenge is a seamless extension of the physical affordance of the objects into digital domain (Ishii and Ullmer, 1997; Ullmer and Ishii, 2000).

Interactions with digital information are now largely confined to Graphical User Interfaces (GUIs). We are surrounded by a variety of ubiquitous GUI devices such as personal computers, handheld computers, and cellular phones. The Graphical User Interface (GUI) has been in existence since the 70's and the first appeared commercially in the Xerox 8010 Star System in 1981 (Smith, 1982). With the commercial success of the Apple Macintosh and Microsoft Windows, the GUI has become the standard paradigm for Human Computer Interaction (HCI) today.

GUIs represent information (bits) with pixels on a bit-mapped display. Those graphical representations can be manipulated with generic remote controllers such as mice and keyboards. By decoupling representation (pixels) from control (input devices) in this way, GUIs provide the malleability to emulate a variety of media graphically. By utilizing graphical representation and "see, point and click" interaction, the GUI made a significant improvement over its predecessor, the CUI (Command User Interface) which required the user to "remember and type" characters.

However, interactions with pixels on these GUI screens are inconsistent with our interactions with the rest of the physical environment within which we live. The GUI, tied down as it is to the screen, windows, mouse and keyboard, is utterly divorced from the way interaction takes place in the physical world. When we interact with the GUI world, we can not take advantage of our dexterity or utilize our skills for manipulating various physical objects such as manipulation of building blocks or the ability to shape models out of clay.

Tangible User Interfaces (TUIs) aim to take advantage of these haptic interaction skills, which is significantly different approach from GUI. The key idea of TUIs is to give physical forms to digital information. The physical forms serve as both representations and controls for their digital counterparts. TUI makes digital information directly manipulatable with our hands, and perceptible through our peripheral senses by physically embodying it.

Tangible User Interface serves as a special purpose interface for a specific application using explicit physical forms, while GUI serves as a general purpose interface by emulating various tools using pixels on a screen.

TUI is an alternative to the current GUI paradigm, demonstrating a new way to materialize Mark Weiser's vision of Ubiquitous Computing of weaving digital technology into the fabric of a physical environment and make it invisible (Weiser, 1991). Instead of making pixels melt into an assortment of different interfaces, TUI uses tangible physical forms that can fit seamlessly into a users' physical environment.

This chapter introduces the basic concept of TUI in comparison with GUI, early prototypes of TUI that highlights the basic design principles, and discusses design challenges that TUI needs to overcome.

Urp: An Example of TUI

To illustrate basic TUI concepts, we introduce "Urp" (Urban Planning Workbench) as an example of TUI (Underkoffler and Ishii, 1999). Urp uses scaled physical models of architectural buildings to configure and control an underlying urban simulation of shadow, light reflection, wind flow, etc. (Photo1). In addition to a set of building

models, Urp also provides a variety of interactive tools for querying and controlling the parameters of the urban simulation. These tools include a clock tool to change a position of sun, a material wand to change the building surface between bricks and glass (with light reflection), a wind tool to change the wind direction, and an anemometer to measure wind speed.

Photo 1 Urp and shadow simulation

Physical building models casting digital shadows, and a clock tool to control time of the day (position of the sun).

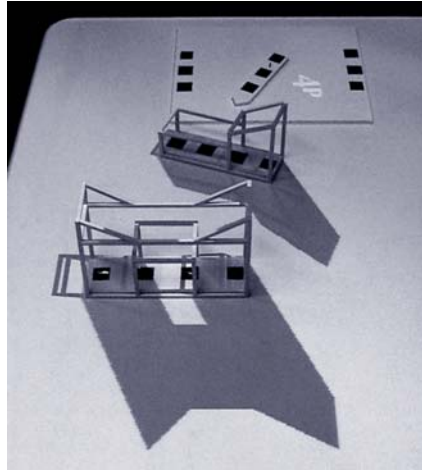
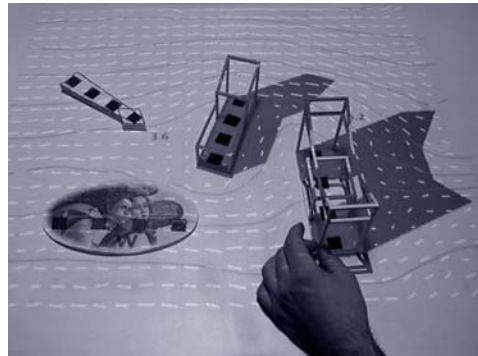


Photo 2 Urp and wind simulation

Wind flow simulation with a wind tool and an anemometer.



The physical building models in Urp cast digital shadows onto the workbench surface (via video projection), corresponding to solar shadows at a particular time of day. The time of day, representing the position of the sun, can be controlled by turning the physical hands of a "clock tool" (Photo 2). The building models can be moved and rotated, with the angle of their corresponding shadows transforming according to their position and time of day.

Correspondingly, moving the hands of the clock tool can cause Urp to simulate a day of shadow movement between the situated buildings. Urban planners can identify and isolate inter-shadowing problems (shadows cast on adjacent buildings), and reposition buildings to avoid areas that are needlessly dark areas, or maximize light between buildings.

A "material wand" alters the material surface properties of a building model. By touching the material wand to a building model, the building surface material is switched from bricks to glass, and a projected reflection of sunlight appears to bounce off the walls of the building. Moving the building allows urban designers to be aware of the relationship between the building reflection and other infrastructure. For example, the reflection off the building at sundown might result in distraction to drivers on a nearby highway. The designer can then experiment with altering the angles of the building to oncoming traffic or move the building further away from the roadway. Tapping again with the material wand changes the material back to brick, and the sunlight reflection disappears, leaving only the projected shadow.

By placing the "wind tool" on the workbench surface, a wind flow simulation is activated based on a computational fluid dynamics simulation, with field lines graphically flowing around the buildings. Changing the wind tool's physical orientation correspondingly alters the orientation of the computationally simulated wind. Urban planners can identify any potential wind problems, such as areas of high pressure that may result in hard-to-open doors or unpleasant walking environments. An "anemometer" object allows point monitoring of the wind speed (Photo 3). By placing the anemometer onto the workspace, the windspeed of that point is shown. After a few seconds, the point moves along the flow lines, to show the windspeed along that particular flow line. The interaction between the buildings and their environment allows urban planners to visualize and discuss inter-shadowing, wind, and placement problems.

In "Urp," physical models of buildings are used as tangible representations of digital models of the buildings. To change the location and orientation of buildings, users simply grab and move the physical model as opposed to pointing and dragging a graphical representation on a screen with a mouse. The physical forms of Urp's building

models, and the information associated with their position and orientation upon the workbench represent and control the state of the urban simulation.

Although standard interface devices for GUIs such as keyboards, mice, and screens are also physical in form, the role of the physical representation in TUI provides an important distinction. The physical embodiment of the buildings to represent the computation involving building dimensions and location allows a tight coupling of control of the object and manipulation of its parameters in the underlying digital simulation.

In Urp, the building models and interactive tools are both physical representations of digital information (shadow dimensions and wind speed) and computational functions (shadow interplay). The physical artifacts also serve as controls of the underlying computational simulation (specifying the locations of objects). The specific physical embodiment allows a dual use in representing the digital model and allowing control of the digital representation. In the next section, the model of TUI is introduced in comparison with GUI to illustrate this mechanism.

Basic Model of Tangible User Interface

The interface between people and digital information requires two key components; input and output, or control and representation. *Controls* enable users to manipulate the information, while *external representations* are perceived with the human senses. Fig. 1 illustrates this simple model of a user interface consisting of control, representation, and information.

In the Smalltalk-80 programming language (Burbeck, 1992; Goldberg, 1984), the relationship between these components is illustrated by the "model-view-controller" or "MVC" archetype – which has become a basic interaction model for GUIs.

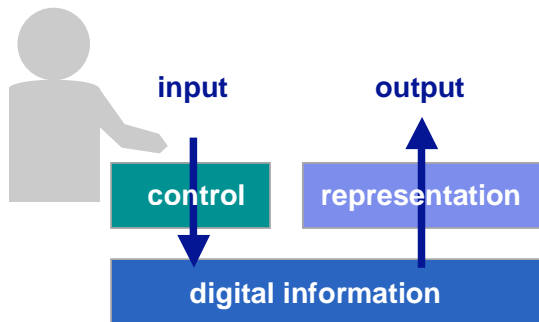


Fig. 1 User Interface

The interface between people and digital information requires two key components: 1) *external representation* (or view) that users can perceive, and 2) *control* with which users can manipulate the representation.

Drawing from the MVC approach, we have developed an interaction model for both GUI and TUI. We carry over the "control" element from MVC, while dividing the "view" element into two subcomponents: tangible and intangible representations, and renaming "model" as "digital information" to generalize this framework to illustrate the difference between GUI and TUI.

In Computer Science, the term "representation" often relates to the programs and data structures serving as the computer's internal representation (or model) of information. In this article, the meaning of "representation" centers upon external representations – the external manifestations of information in fashions directly perceivable by the human senses that include visual, hearing and tactile senses.

GUI

In 1981, the Xerox Star workstation set the stage for the first generation of GUI (Johnson, et al., 1989; Smith, 1982), establishing the "desktop metaphor" which simulates a desktop on a bit-mapped screen. The Star workstation was the first commercial system that demonstrated the power of a mouse, windows, icons, property sheets, and modeless interaction. The Star also set several important HCI design principles, such as "seeing and pointing vs. remembering and typing," and "what you see is what you get (WYSIWYG)." The Apple Macintosh brought this new style of HCI to the public's attention in 1984, creating a new trend in the personal computer industry. Now, the GUI is widespread, largely through the pervasiveness of Microsoft Windows, PDAs, and cellular phones.

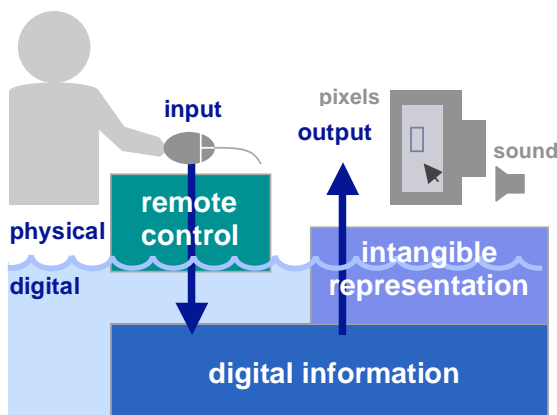


Fig. 2 Graphical User Interface

GUI represents information with intangible pixels on a bit-mapped display and sound. General purpose input

GUI uses windows, icons, and menus made of pixels on bit-mapped displays to visualize information. This is an intangible representation. GUI pixels are made interactive through general "remote controllers" such as mice, tablets, or keyboards. In the pursuit of generality, GUI introduced a deep separation between the digital (intangible) representation provided by the bit-mapped display, and the controls provided by the mouse and keyboard.

Figure 2 illustrates the current GUI paradigm in which generic input devices allow users to remotely interact with digital information. Using the metaphor of seashore that separates a sea of bits from the land of atoms, the digital information is illustrated at the bottom of the water, and mouse and screen are above sea level in the physical

domain. Users interact with the remote control, and ultimately experience an intangible external representation of digital information (display pixels and sound).

TUI

Tangible User Interface aims at a different direction from GUI by using tangible representations of information which also serve as the direct control mechanism of the digital information. By representing information in both tangible and intangible forms, users can more directly control the underlying digital representation using their hands.

Tangible Representation as Control

Figure 3 illustrates this key idea of TUI to give tangible (physical and graspable) external representation to the digital information. The tangible representation helps bridge the boundary between the physical and physical worlds. Also notice that the tangible representation is computationally coupled to the control to the underlying digital information and computational models. Urp illustrates examples of such couplings, including the binding of graphical geometries (digital data) to the physical building models, and computational simulations (operations) to the physical wind tool. Instead of using a GUI mouse to change the location and angle graphical representation of a building model by pointing, selecting handles and keying in control parameters, an Urp user can grab and move the building model to change both location and angle.

The tangible representation functions as an interactive physical control. TUI attempts to embody the digital information in physical form, maximizing the directness of information by coupling manipulation to the underlying computation. Through physically manipulating the tangible representations, the digital representation is altered. In Urp, changing the position and orientation of the building models influences the shadow simulation, and the orientation of the "wind tool" adjusts the simulated wind direction.

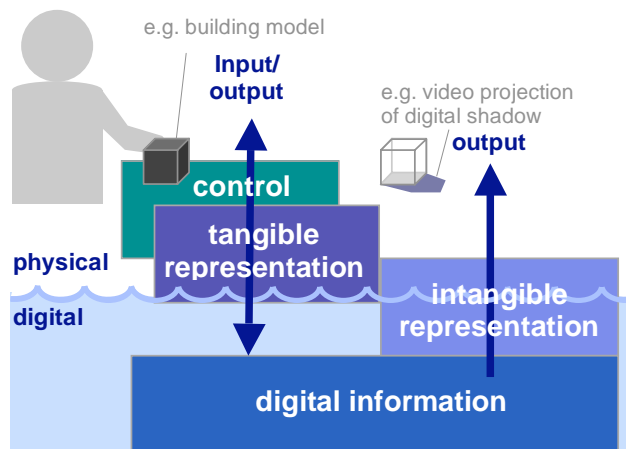


Fig. 3 Tangible User Interface

By giving tangible (physical) representation to the digital information, TUI makes information directly graspable and manipulable with haptic feedback. Intangible representation (e.g. video projection) may complement tangible representation by synchronizing with it.

Intangible Representation

Although the tangible representation allows the physical embodiment to be directly coupled to digital information, it has limited ability to represent change many material or physical properties. Unlike malleable pixels on the computer screen, it is very hard to change a physical object in its form, position, or properties (e.g. color, size) in real-time. In comparison with malleable "bits," "atoms" are extremely rigid, taking up mass and space.

To complement this limitation of rigid "atoms," TUI also utilizes malleable representations such as video projections and sounds to accompany the tangible representations in the same space to give dynamic expression of the underlying digital information and computation. In the Urp, the digital shadow that accompanies the physical building models is such an example.

The success of a TUI often relies on a balance and strong perceptual coupling between the tangible and intangible representations. It is critical that both tangible and intangible representations be perceptually coupled to achieve a seamless interface that actively mediates interaction with the underlying

digital information, and appropriately blurs the boundary between physical and digital. Coincidence of input and output spaces and realtime response are important requirements to accomplish this goal.

[note] There exist certain types of TUIs which have actuation of the tangible representation (physical objects) as the central mean of feedback. Examples are inTouch (Brave, et al., 1998), curlybot (Frei, et al., 2000a), and topobo (Raffle, et al., 2004). This type of force-feedback-TUI does not depend on "intangible" representation since active feedback through the tangible representation serves as the main display channel.

Key Properties of TUI

While Figure 2 illustrates the GUI's clear distinction between graphical representation and remote controls, the model of TUI illustrated in Figure 3 highlights TUI's integration of physical representation and control. This model provides a tool for examining the following important properties and design requirements of tangible interfaces (Ullmer and Ishii, 2000).

Computational coupling of tangible representations to underlying digital information and computation:

The central characteristic of tangible interfaces is the coupling of tangible representations to underlying digital

information and computational models. One of the challenges of TUI design is how to map physical objects and their manipulation to digital computation and feedback in a meaningful and comprehensive manner.

As illustrated by the Urp example, a range of digital couplings and interpretations are possible, such as the coupling of data to the building models, operations to the wind tool, and property modifiers to the material wand.

Deciding the embodiment and mapping of the controller is dictated by the type of application envisioned. We give examples cases in which a range of specificity of embodiment is used. In some applications, more abstract form of physical objects (such as round pucks) are used as generic controllers that are reusable to control a variety of parameters by rotating and pushing a button (Patten, et al., 2001). When a puck is used as a dial to control a simulation parameter, graphical feedback is given to complement the information such as scale of the dial.

Embodiment of mechanisms for interactive control with tangible representations:

The tangible representations of TUIs serve simultaneously as interactive physical controls. Tangibles may be physically inert, moving only as directly manipulated by a user's hands. Tangibles may also be physically actuated, whether through motor-driven force feedback approaches (e.g. inTouch, Curlybot) or magnet-driven approaches such as Actuated Workbench (Pangaro, et al., 2002).

Tangibles may be unconstrained and manipulated in free space with six degrees of freedom. They may also be weakly constrained through manipulation on a planar surface, or tightly constrained, as in the movement of the abacus beads with one degree of freedom.

In order to make interaction simple and easy to learn, TUI designers need to utilize the physical constraints of the chosen physical embodiment. Because the physical embodiment, to some extent, limits the interaction choices, a designer must design the interaction so that the actions supported by the object are based on well-understood actions related to the physical object. For example, if a bottle shape is chosen, then opening the bottle by pulling out a cork is a well-understood mechanism (Ishii, et al., 2001). This understanding of the culturally common manipulation techniques helps disambiguate the users' interpretation of how to interact with the object.

Perceptual coupling of tangible representations to dynamic intangible representations:

Tangible interfaces rely on a balance between tangible and intangible representations. Although embodied tangible elements play a central, defining role in the representation and control of a TUI, there is a supporting role for the TUI's intangible representation. A TUI's intangible representation, usually graphics and audio—often mediate much of the dynamic information provided by the underlying computation.

The realtime feedback of the intangible representation corresponding to the manipulation of the tangible representation is critical to insure perceptual coupling. The coincidence of inputs and output spaces (spatial continuity of tangible and intangible representations) is also an essential requirement to enhance perceptual coupling. For example, in Urp, the building models (tangible representation) are always accompanied by a "digital shadow" (intangible representation) without noticeable temporal or spatial gaps. That convinces users of an illusion that the shadows are cast from the building models (rather than the video projector).

Genres of TUI Applications

By giving physical form to digital information to enhance an experience, TUIs have a wide variety of application domains. This section gives an overview of seven genres for promising TUI applications. For a more exhaustive survey of TUIs in a historical context, I would encourage the readers to refer to: (Ullmer and Ishii, 2000). Holmquist (Holmquist, et al., 1999) and Fishkin (Fishkin, 2004), Zuckerman (Zuckerman, et al., 2005) also provided a useful taxonomy and frameworks to analyze the design space of TUIs.

1) Tangible Telepresence

One such genre is an inter-personal communication taking advantage of haptic interactions using mediated tangible representation and control. This genre relies on mapping haptic input to haptic representations over a distance. Also called "tangible telepresence", the underlying mechanism is the synchronization of distributed objects and the gestural simulation of "presence" artifacts, such as movement or vibration, allowing remote participants to convey their haptic manipulations of distributed physical objects. The effect is to give a remote user the sense of ghostly presence, as if an invisible person was manipulating a shared object. inTouch (Brave and Dahley, 1997), HandJive (Fogg, et al., 1998), and ComTouch (Chang, et al., 2002) are such examples.

2) Tangibles with Kinetic Memory

The use of kinesthetic gestures and movement to promote learning concepts is another promising domain. Educational toys to materialize record & play concepts have been also explored using actuation technology and taking advantage of i/o coincidence of TUI. Gestures in physical space illuminate the symmetric mathematical relationships in nature, and the kinetic motions can be used to teach children concepts relevant to programming and differential geometry as well as story telling. Curlybot (Frei, et al., 2000a) and topobo (Raffle, et al., 2004) are examples of toys which distill ideas relating gestures and form to dynamic movement, physics and storytelling.

3) Constructive Assembly

Another domain is a constructive assembly approach that draws inspiration from LEGO™ and building blocks, building upon the interconnection of modular physical elements. This domain is mainly concerned with the physical fit between objects, and the kinetic relationships between these pieces that enable larger constructions and varieties of movement.

Constructive assembly was pioneered by Aish and Frazer in the late 1970s. Aish developed BBS (Aish, 1979; Aish and Noakes, 1979) for thermal performance analysis, and Frazer developed a series of intelligent modeling kits such as "Universal Constructor (Frazer, 1994; Frazer, et al., 1980) for modeling and simulation. Recent examples include GDP (Anagnostou, et al., 1989), AlgoBlock (Suzuki and Kato, 1993), Triangles (Gorbet, et al., 1998), Blocks (Anderson, et al., 2000), ActiveCube (Kitamura, et al., 2001), and System Blocks (Zuckerman and Resnick, 2004). Topobo (Raffle, et al., 2004) is a unique instance that inherit the properties from both "constructive assemble" and "tangibles with kinetic memory."

4) Tokens and Constraints

"Tokens and constraints" is another TUI approach to operate abstract digital information using mechanical constraints (Ullmer, et al., 2005). Tokens are discrete, spatially reconfigurable physical objects that represent digital information or operations. Constraints are confining regions within which tokens can be placed. Constraints are mapped to digital operations or properties that are applied to tokens placed within their confines. Constraints are often embodied as physical structures that mechanically channel how tokens can be manipulated, often limiting their movement to a single physical dimension.

The Marble Answering Machine (Crampton Smith, 1995) is a classic example which influenced many following research. mediaBlocks (Ullmer, et al., 1998), LogJam (Cohen, et al., 1999), DataTile (Rekimoto, et al., 2001), and Tangible Query Interface (Ullmer, et al., 2003) are other recent examples of this genre of development.

5) Interactive Surfaces – table top TUI

Interactive surfaces are another promising approach to support collaborative design and simulation which has been explored by many researchers in the past years to support a variety of spatial applications (e.g. Urp). On an augmented workbench, discrete tangible objects are manipulated and their movements are sensed by the workbench. The visual feedback is provided onto the surface of the workbench keeping input/output space coincidence. This genre of TUI is also called "tabletop TUI" or "tangible workbench."

Digital Desk (Wellner, 1993) is the pioneering work in this genre, and a variety of tabletop TUIs were developed using multiple tangible artifacts within common frames of horizontal work surface. Examples are metaDesk (Ullmer and Ishii, 1997), InterSim (Arias, et al., 1997), Illuminating Light (Underkoffler and Ishii, 1998), Urp (Underkoffler and Ishii, 1999), Build-It (Rauterberg, et al., 1998), Sensetable (Patten, et al., 2001), AudioPad (Patten, et al., 2002), and IP Network Design Workbench (Kobayashi, et al., 2003).

One limitation of above systems is the computer's inability to move objects on the interactive surfaces. To address this problem, the Actuated Workbench was designed to provide a hardware and software infrastructure for a computer to smoothly move objects on a table surface in two dimensions (Pangaro, et al., 2002), providing an additional feedback loop for computer output, and helping to resolve inconsistencies that otherwise arise from the computer's inability to move objects on the table.

6) Continuous Plastic TUI

Fundamental limitation of previous TUIs was the lack of capability to change the forms of tangible representations during the interactions. Users had to use predefined finite set of fixed-form objects, changing only the spatial relationship among them but not the form of individual object itself.

Instead of using predefined discrete objects with fixed forms, the new type of TUI systems utilize continuous tangible material such as clay and sand were developed for rapid form giving and sculpting for the landscape design. Examples are Illuminating Clay (Piper, et al., 2002), and SandScape (Ishii, et al., 2004). Later this interface was applied to the browsing of 3D volume metric data in Phoxel-Space project (Ratti, et al., 2004).

7) Augmented Everyday Objects

Augmentation of familiar everyday objects is an important design approach of TUI to lower the floor and to make it easy to understand the basic concepts. Examples are the Audio Notebook (Stifelman, 1996), musicBottles (Ishii, et al., 1999), HandScape (Lee, et al., 2000), LumiTouch (Chang, et al., 2001), Designers' Outpost (Klemmer, et al., 2002) and I/O Brush (Ryokai, et al., 2004). It is a challenge for industrial designers to improve upon a product by adding some digital augmentation to an existing digital object. This genre is open to much eager interpretation by artists and designers, to have our everyday physical artifacts evolve with technology.

8) Ambient Media

In the early stages of TUI research, we were exploring ways of improving the quality of interaction between people and digital information. We employed two approaches to extending interaction techniques to the physical world:

- allowing users to "grasp & manipulate" foreground information by coupling bits with physical objects, and
- enabling users to be aware of background information at the periphery using ambient media in an augmented space.

At that time, HCI research had been focusing primarily on foreground activity on the screen and neglecting the rest of the user's computing environment (Buxton, 1995). However, in most situations, people are subconsciously receiving ambient information from their peripheral senses without attending to it explicitly. If anything unusual is noticed, it immediately comes to their attention, and they could decide to bring it to the foreground. For example, people subconsciously are aware of the weather outside their window. If they hear thunder, or a sudden rush of wind, the user can sense that a storm is on its way out of their peripheral attention. If it was convenient, they could then look outside, or continue working without distraction.

Ambient media describes the class of interfaces that is designed to smooth the transition of the users' focus of attention between background and foreground. Natalie Jeremijenko's Live Wire in 1995, at Xerox Parc, was a spinning wire that moved to indicate network traffic. Designing simple and adequate representations for ambient media using tangible objects is a key part of the challenge of Tangible Bits (Ishii and Ullmer, 1997). (Fig. 4) The ambientROOM is a project that explores the ideas of ambient media constructing a special room equipped with embedded sensors and ambient displays (Ishii, et al., 1998). This work was a preliminary investigation into background/peripheral interfaces, and led to the design of standalone ambient fixtures such as Pinwheels and Walter Lamp that make users aware of "digital wind" and "bits of rain" at their peripheral senses (Dahley, et al., 1998). Strictly speaking, ambient media is not a kind of TUI since in many cases there are no direct interactions. Rather

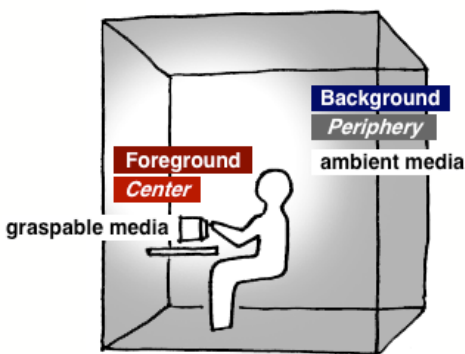


Fig. 4 Center and Periphery of User's Attention within Physical Space

ambient media serve as background information displays that complement tangible/graspable media that users manipulate in their foreground. TUI's approach to ambient media is concerned with the design of simple mappings that gives easy-to-understand form to cyberspace information and representing change in a subtle manner. We started experimenting with a variety of ambient media such as sound, light, airflow, and water movement for background interfaces for awareness of cyberspace at the periphery of human perception.

This concept of "ambient media" is now widely studied in the HCI community as a way to turn the architectural / physical spaces into an ambient and calm information environment. Another design space is low attention interfaces for interpersonal communication through ambient media, (Chang, et al. 2001). Ambient Devices further commercialized the domain of low-attention ambient media interfaces by developing the Ambient

Orb and Weather Beacon, exploring the new genre of "glanceable interfaces" (<http://www.ambientdevices.com/>).

TUI Instances

In this section, ten TUI examples are presented to illustrate the potential application domains describe in a previous section, and to highlight unique features of TUIs. However, given the limited space and rapid growth of TUI research in HCI community in recent years, the collection of examples introduced here can only cover a relatively small portion of the representative works of TUIs.

inTouch: Tangible TelePresence through Distributed Synchronized Physical Objects

inTouch is a project to explore new forms of interpersonal communication over distance through touch by preserving the physical analog movement of synchronized distributed rollers (Brave and Dahley, 1997; Brave, et al., 1998). Force-feedback is employed to create the illusion that people, separated by distance, are interacting with a shared physical object. The "shared" object provides a haptic link between geographically distributed users, opening up a channel for physical expression over distance.

Two identical mechanisms were built with three freely rotating rollers (Photo 3). Each roller is synchronized to the corresponding roller on the distant mechanism using force-feedback, so that when one roller is moved the other corresponding roller also moves. If the movement of one roller is held, then the roller transmits that resistance to the other roller. They are in a sense connected by a stiff computational spring. Two users separated by distance can then play, moving or tapping the rollers or more passively feel the other person's manipulation of the object. The presence of the other person is represented tangibly through physical interaction with the inTouch device.

Force-feedback is conventionally used to allow a user to "touch" virtual objects in the computer screen through a single point. InTouch applies this technology to realize a link for interpersonal haptic communication, instead of just touching virtual objects. InTouch allows people to feel as if they are connected through touching the rollers, to another person. Instead of touching inanimate objects, each person is touching a dynamic, moving object that is

shared.

Important features of inTouch from HCI points of view can be summarized as follows:

- 1) no boundary between "input" and "output" (i/o coincidence: the wooden rollers are force displays as well as input devices),
- 2) principal human input/output organs are hands, not eyes or ears (with the sense of touch being the primary mode),
- 3) information can be sent and received simultaneously through one's hand.

Past communication media such as video telephony set themselves the ultimate goal of reproducing the voice or the image of the human face and body as realistically as possible in order to create the illusion of "being there" for each interlocutor. inTouch takes the opposite approach by making users aware of the other person without ever rendering him or her in bodily terms and creating what we call a "tangible presence" or "ghostly presence." By seeing and feeling an object being moved in a human fashion on its own, we imagine a ghostly body. The concept of the ghostly presence provides us with a different approach to the conventional notion of telepresence.

Curlybot: A toy to record & play

Curlybot is a toy that can record and playback physical motion (Photo 4). As one plays with it, it remembers how it has been moved and can replay that movement with all the intricacies of the original gesture; every pause, acceleration, and even the shaking in the user's hand, is recorded. Curlybot then repeats that gesture indefinitely creating beautiful and expressive patterns. Children can use curlybot to gain strong intuition for advanced mathematical and computational concepts, like differential geometry, through play outside of traditional computer (Frei, et al., 2000b)

The forced-feedback technology used for real-time simultaneous communication in inTouch was employed in curlybot for the recording and playback of gestures. Two motors equipped with an optical encoder enable free rotation in addition to forward and backward movement.

When the user presses the button a red LED is illuminated to indicate the recording mode. The user then moves the curlybot around, meanwhile an encoder is recording this gesture information. Pushing the button a second time terminates recording and a green LED alights to indicate the playback mode. The microprocessor compares the current position with the stored positions and instructs the motors to retrace the steps recorded in the curlybot's memory.

This project contributes to both interface design and education. As a tangible interface it blurs the boundary between input and output as inTouch does. curlybot itself is both an input device to record gestures and a physical display device to re-enact them. By allowing the user to teach it gestures with his or her hand and body and then re-enacting those gestures in a physical space around the body, curlybot enables a strong connection between body and mind not obtainable from anything expressed on a computer screen.

From an educational standpoint curlybot allows very young children to explore "advanced" mathematical and computational concepts. Curlybot supports new ways of thinking about geometric shapes and patterns. Children can also use curlybot to explore some of the basic ideas behind computational procedures, like how complexity can be built from simple parts. This is similar to what is possible with the Logo programming language, but does not require children to read or write and thus makes advanced ideas accessible to younger children. Curlybot also draws strongly on children's intuition about their own physical actions in the world to learn – What Seymour Papert calls "body syntonic learning" (Papert, 1980). In addition, the direct input and beautifully expressive patterns that result through curlybot's repetition of the gestured keep children playing and engaged.

Topobo: 3D constructive assembly with kinetic memory

Topobo, for "topology" and "robotics," is a 3D constructive assembly system with kinetic memory, the ability to record and playback physical motion (Raffle, et al., 2004). By snapping together a combination of Passive (static) and Active (motorized) components, people can quickly assemble dynamic biomorphic forms like animals and skeletons with Topobo. Topobo allows users to animate those forms by recording the movement of pushing, pulling, and twisting them, and later observe the system play back those motions repeatedly. This record & play function was inherited from the prior curlybot project, and the constructive assembly function was inherited from the commercial toy, Zoob™.

For example, a dog can be constructed and then taught to gesture and walk by twisting its body and legs. The dog will then repeat those movements and walk repeatedly. The same way people can learn about static structures playing with regular building blocks, they can learn about dynamic structures playing with Topobo. Topobo works like an extension of the body giving one's gestural fluency. Topobo embeds computation within a dynamic building system so that gestural manipulation of the material becomes a programming language (Photo 5).

Topobo is inspired by current trends in computational media design and by artists and empiricists using visual explorations and models of natural phenomena to more deeply appreciate patterns found in the natural world. In this spirit, Topobo is designed to allow people to use experimentation, play, and self-expression to discover and explore common natural relationships between natural forms and dynamic motion. Building toys and educational manipulatives have been used for years by children to learn about the world through model making.

Unique among modeling systems is Topobo's coincident physical input and output behaviors (which is common

among inTouch, curlybot too). The system is comprised of 10 different primitives that can be snapped together in a variety of ways. Nine of these primitives are called "Passive" because they form static connections. These static connections are constraining the form and the range of motion available to the structure. One "Active" primitive is built with an embedded encoder and motor which is programmed by demonstration. These motorized components are the only ones that move, so the system is able to faithfully record and replay every dynamic manipulation to a structure.

mediaBlocks: Token and Constraint approach

The mediaBlocks system is a tangible interface for manipulating lists of on-line digital media such as video clips and images (Ullmer, et al., 1998). Whereas Urp provides a spatial interface for leveraging object arrangements consistent with real-world building configurations, the mediaBlocks system provides a relational interface for manipulating more abstract digital information.

The mediaBlocks are small, digitally tagged blocks, dynamically bound to lists of on-line media elements. The mediaBlocks support two major modes of use. First, they function as capture, transport, and playback mechanisms for moving on-line media between different media devices. In this mode, conference room cameras, digital whiteboards, wall displays, printers, and other devices are outfitted with mediaBlock slots. Inserting one of the mediaBlocks into the slot of a recording device (e.g., a camera) activates the recording of media into on-line space, and the dynamic binding of the media to the physical block.

Similarly, inserting one of the bound mediaBlocks into a playback device (e.g., video display) activates playback of the associated on-line media. Inserting mediaBlocks into slots mounted on computer monitors provides an intermediate case, allowing mediaBlock contents to be exchanged bidirectionally with traditional computer applications using the GUI drag-and-drop operation.

The second functionality of mediaBlocks uses the blocks as physical controls on a media sequencing device (Photo 6). A mediaBlock "sequence rack" (partially modeled after the tile racks of the Scrabble game) allows the media contents of multiple adjacent mediaBlocks to be dynamically bound to a new mediaBlock carrier. Similarly, a second "position rack" maps the physical position of a block to an indexing operation upon its contents. When mediaBlocks are positioned on the left edge of the position rack, the first media element of the block is selected. Intermediate physical positions on the rack provide access to later elements in the associated media list of the block.

Digital Desk: Pioneer of Tabletop TUI

Digital Desk (Wellner, 1993) is a pioneering work to demonstrated a way to integrate physical and digital documents processing on a table. Wellner brought some of the functionality we typically associate with GUIs onto the physical desktop. This table used a camera and a microphone to detect finger presses on a graphical interface displayed on a desk with a video projector. Wellner used this desk for tasks such as graphic design and spreadsheet computations on physical paper. This system also employed some physical props, such as a scanner that would scan items and place them directly on the tabletop interaction surface.

Wellner's research pointed the way toward enabling the computer to perform some of the operations we traditionally associate with GUIs in a tabletop environment. The Digital Desk also illustrates some of the compelling reasons for considering computer interfaces based on horizontal interactive surfaces. Because many work surfaces in our environment are already planar, horizontal or nearly horizontal surfaces, integrating computer interfaces into these surfaces may provide an opportunity for new types of relationships between computation and physical objects, and may help create computer systems that are more relevant to problem domains with established work practices based on tabletops.

The Digital Desk inspired many tabletop tangible interfaces including the Luminous Room project (Underkoffler, et al., 1999) from which Urp (Underkoffler and Ishii, 1999) was created. Sensetable (Patten, et al., 2001) is another example.

Sensetable and AudioPad: Tabletop TUI for realtime music performance

Sensetable (Patten, et al., 2001) is a system that wirelessly tracks the positions of multiple objects on a flat display surface. The sensetable serves as a common platform for a variety of tabletop TUI applications such as Audio Pad and IP Network Design Workbench.

Audiopad (Patten, et al., 2002) is a composition and performance instrument for electronic music which tracks the positions of objects on a tabletop surface and converts their motion into music. One can pull sounds from a giant set of samples, juxtapose archived recordings against warm synthetic melodies, cut between drum loops to create new beats, and apply digital processing all at the same time on the same table. Audiopad not only allows for spontaneous reinterpretation of musical compositions, but also creates a visual and tactile dialogue between itself, the performer, and the audience.

Audiopad is based on the Sensetable platform that has a matrix of antenna elements which track the positions of electronically tagged objects on a tabletop surface. Software translates the position information into music and graphical feedback on the tabletop. Each object represents either a musical track or a microphone (photo 8).

Experience of Audiopad with tangible user interface through a series of live performances suggests that interacting

with electromagnetically tracked objects on a tabletop surface with graphical feedback can be a powerful and satisfying tool for musical expression. The integration of input and output spaces gives the performer a great deal of flexibility in terms of the music that can be produced. At the same time, this seamless integration allows the performer to focus on making music, rather than using the interface. Spatial multiplexed inputs of TUI also supported two performers play music simultaneously and collaboratively (Photo 8).

IP Network Design Workbench: Event Driven Simulation on Senstable

The IP Network Design Workbench (IPNWDWB) is the collaborative project between NTT Comware and the Tangible Media Group. The IP Network Design Workbench supports collaborative network design and simulation by a group of experts and customers (Kobayashi, et al., 2003). This system is also based on the Senstable platform which can wirelessly detect the location and orientation of physical pucks. Simulation engine is based on the event-driven simulation model. Using Senstable system, users can directly manipulate network topologies for modeling, control simulation parameters of nodes and links using physical pucks on the sensing table, and simultaneously see the simulation results projected onto the table in real-time (Photo 9).

The goal of IPNWDWB is to make simulation tools more accessible for non-experts, so that they can join the network design process and interact with experts more easily than using traditional GUI computer. This system was commercialized and has been used for collaborative network design with customers to ensure their understanding of the performance and cost of network enhancements dealing with the increases of network traffic caused by Voice over IP and/or streaming video, for example. Because of the large tiling horizontal work surface and TUI interaction that invites all the participants to touch and manipulate pucks simultaneously, the process of decision making becomes much more democratic and more convincing than ordinary PowerPoint presentations through conventional GUI.

If we compare IPNWDWB with Urp, we notice a big difference in the nature of applications. In Urp, we used physical scale models of buildings, which humans have used for thousand years to design cities, as tangible representations of urban models. Therefore, it is very natural to apply TUIs to such domains (urban planning, landscape design) in which physical models have been used long before the birth of digital computers.

In contrast, IP Network Design is based on event-driven simulation model which are quite abstract and new. This modeling technique requires digital computers. IPNWDWB is important since it demonstrated that TUI can empower the design process even in abstract and computational application domain which does not have straight-forward mappings from abstract concepts to physical objects. There are a wide range of modeling and simulation techniques such as System Dynamics and Event-Driven Simulation that uses 2D graph representation. We learned that many of these abstract computational applications can be supported by Senstable-like TUI platforms in the collaborative design sessions. For example, simultaneously changing parameters, transferring control between different people or different hands and distributing the adjustment of simulations dynamically are interactions enabled by TUI.

Actuated Workbench: Closing a loop of computational actuation and sensing

The aforementioned tabletop TUI systems share a common weakness. While input occurs through the physical manipulation of tangible objects, output is displayed only through sound or graphical projection on and around the objects. As a result, the objects can feel like loosely coupled handles to digital information rather than physical manifestations of the information itself.

In addition, the user must sometimes compensate for inconsistencies when links between the digital data and the physical objects are broken. Such broken links can arise when a change occurs in the computer model that is not reflected in a physical change of its associated object. With the computer system unable to move the objects on the table surface, it cannot undo physical input, correct physical inconsistencies in the layouts of the objects, or guide the user in the physical manipulation of the objects. As long as this is so, the physical interaction between human and computer remains one-sided.

To address this problem, the Actuated Workbench was designed to provide a hardware and software infrastructure for a computer to smoothly move objects on a table surface in two dimensions (Pangaro, et al., 2002).

The Actuated Workbench is a new technology that uses magnetic forces to move objects on a table in two dimensions. It is intended for use with existing tabletop tangible interfaces, providing an additional feedback loop for computer output, and helping to resolve inconsistencies that otherwise arise from the computer's inability to move objects on the table.

Actuation enables a variety of new functions and applications. For example, a search and retrieve function could respond to a user query by finding matching data items and either moving them to another place on the tabletop or wiggling them to get the user's attention. A more powerful function would be one in which the computer could physically sort and arrange pucks on the table according to user-specified parameters. This could help the user organize a large number of data items before manually interacting with them. As a user makes changes to data through physical input, he or she may wish to undo some changes. A physical undo in this system could move the pucks back to their positions before the last change. It could also show the user the exact sequence of movements she had performed. In this sense, both "undo" and "rewind" commands are possible.

One advantage that table top tangible user interfaces offer is the ease with which multiple users can make simultaneous changes to the system. Users can observe each other's changes, and any user can reach out and physically change the shared layout without having to grab a mouse or other pointing device. This is not the case, however, when users are collaborating remotely. In this scenario, a mechanism for physical actuation of the pucks becomes valuable for synchronizing multiple physically separated workbench stations (Photo 9). Without such a mechanism, real-time physical synchronization of the two tables would not be possible, and inconsistencies could arise between the graphical projection and the physical state of the pucks on the table.

In addition to facilitating the simple synchronization of these models, the Actuated Workbench can recreate remote users' actual gestures with objects on the table, adding greatly to the "Ghostly Presence" (Brave, et al., 1998) sought in remote collaboration interfaces.

Actuated Workbench is helpful in teaching students about physics by demonstrating the attraction and repulsion of charged particles represented by pucks on the table. As a student moved the pucks around on the table, the system could make them rush together or fly apart to illustrate forces between the objects.

SandScape: Continuous TUI for landscape design

SandScape (Ishii, et al., 2004) is a tangible interface for designing and understanding landscapes through a variety of computational simulations using sand. Users view these simulations as they are projected on the surface of a sand model that represents the terrain. The users can choose from a variety of different simulations that highlight the height, slope, contours, shadows, drainage or aspect of the landscape model (Photo 10).

The users can alter the form of the landscape model by manipulating sand while seeing the resultant effects of computational analysis generated and projected on the surface of sand in real-time. The project demonstrates how TUI takes advantage of our natural ability to understand and manipulate physical forms while still harnessing the power of computational simulation to help in our understanding of a model representation.

The SandScape configuration is based on a box containing 1 m diameter glass beads lit from beneath with an array of 600 high-power infra-red LEDs. Four IR mirrors are placed around the LED array to compensate for the uneven radiance distribution on the boundary. A monochrome infra-red camera is mounted 2 m above the surface of the beads and captures the intensity of light passing through the volume. The intensity of transmitted light is a function of the depth of the beads and a look-up table can be used to convert surface radiance values into surface elevation values. The system has been calibrated to work with a specific bead size and the optical properties of the material used (absorption and scattering coefficients) are critical to its successful functioning. Owing to the exponential decay of the IR light passing through the glass beads (or any other material) the intensity at the top surface can vary greatly and sometimes exceed the dynamic range of the video camera. This problem can be solved by taking several images with different exposure times and combining them to recover the effective radiance of the scene. SandScape is less accurate than its predecessor Illuminating Clay which used laser range finder to capture the geometry of a clay model (Piper, et al., 2002).

SandScape and Illuminating Clay show the potential advantages of combining physical and digital representations for landscape modeling and analysis. The physical clay and sand models convey spatial relationships that can be intuitively and directly manipulated by hand. Users can also insert any found physical objects directly under the camera. This approach allows users to quickly create and understand highly complex topographies that would be difficult and time-consuming to produce with conventional CAD tools. We believe that this "Continuous TUI" approach makes better use of our natural abilities to discover solutions through the manipulation of physical objects and materials.

At the same time the projected graphics give the user realtime feedback. While tracked physical models interfaced with a computer are not a novelty, we believe that SandScape and Illuminating Clay offer a new contribution, by using the continuous surface geometry of the model itself to act as the input/output mechanism. In so doing we hope to give the projected information the same tangible immediacy as the clay/sand material itself and allow quantitative data to support the intuitive understanding of the landscape.

Landscape architecture, as well as urban and architectural design, requires the collaboration of a number of specialists. These include earth engineers, water engineers, agrarian managers, land economists, transport engineers — to name just a few. In the current process of design, the collaboration happens at different stages, and sometimes without much direct and synchronous interaction. SandScape and Illuminating Clay provide a common platform for collaboration, centered on the table workspace. Numerous representations and analyses can be combined in a single design environment, potentially offering a greater cohesion between different specialists and streamlining the process of design.

musicBottles: transparent interface based on augmented glass bottles

musicBottles introduces a tangible interface that deploys bottles as containers and controls for digital information (Photo 11). The system consists of a specially designed table and three corked bottles that "contain" the sounds of the violin, the cello and the piano in Edouard Lalo's Piano Trio in C Minor, Op. 7. Custom-designed electromagnetic tags embedded in the bottles enable each one to be wirelessly identified.

When a bottle is placed onto the stage area of the table, the system identifies each bottle, and lights up the stage to

show that the bottles have been recognized. The opening and closing of a bottle is also detected, and as the cork is removed, the corresponding instrument becomes audible. A pattern of colored light is rear-projected onto the table's translucent surface to reflect changes in pitch and volume for each instrument. The interface allows users to structure the experience of the musical composition by physically manipulating the different sound tracks.

Humans have used glass bottles for thousands of years. Through the seamless extension of physical affordances and metaphors of the bottles into the digital world, the bottles project explores the transparency of the interface (Ishii, 2004).

A wide variety of contents, including music, weather reports, poems, and stories have been designed to test the concept (Ishii, et al., 1999). The bottles lined up on a specially designed table, the feel of the glass as we open them and the music and light from the LED lamps that come out of them together create a unique aesthetic experience. This is a pleasure not to be had from the mere click of a mouse.

Potential applications are not limited to music alone. One might imagine perfume bottles filled with poetry or wine bottles that decant stories (Mazalek, et al., 2001). More practical applications might include a medicine chest full of bottles that tell the user how and when to take them and let the hospital know when they do. As an intimate part of our daily lives, glass bottle interfaces offer a simple and transparent interface.

Pinwheels: ambient interface to

Pinwheels is an example of ambient media that demonstrate a new approach to interfacing people with online digital information through subtle changes in sound and movement, which can be processed in the background of awareness. Pinwheels spins in a "bit wind" and represents a invisible flow of digital information such as network traffic as physical movement within an architectural spaces (Photo 12).

Nature is filled with subtle, beautiful and expressive ambient media that engage each of our senses. The sounds of rain and the feeling of warm wind on our cheeks help us understand and enjoy the weather even as we engage in other activities. Similarly, we are aware of the activity of neighbors through passing sounds and shadows at the periphery of our attention. Cues like an open door or lights in an office help us subconsciously understand the activities of other people and communicate our own activity and availability.

Current personal computing interfaces, however, largely ignore these rich ambient spaces, and squeeze vast amounts of digital information into small rectangular screens. Information is presented as "painted bits" (pexels) on flat screens that must be in the center (foreground) of a user's focus to be processed. In order to broaden the concept of "display" to make use of the entire physical environment as an interface, using ambient media, information can be manifested as subtle changes in form, movement, sound, color, smell, temperature, or light. We call them "ambient displays."

The Pinwheels evolved from the idea of using airflow as ambient media. However, we found that the flow of air itself was difficult to control and to convey information. As an alternative, we envisioned that a visual/physical representation of airflow based on the "spinning pinwheels" could be legible and poetic. The Pinwheels spin in the "bit wind" at different speeds based upon their input information source.

Ambient displays are envisioned as being all around and suited to the display of

- a) people's presence (awareness of remote people's status / activities),
- b) atmospheric and astronomical phenomena, or
- c) general states of large and complex systems (e.g. atomic power plant).

For instance, an atmospheric scientist might map patterns of solar wind into patterns of Pinwheel spins in a room.

There are many design challenges surrounding ambient displays. One of them is the mapping information to the physical motion and other ambient media. A designer of ambient displays must transform the digital data into a meaningful pattern of physical motion that successfully communicates the information. The threshold between foreground and background is another key issue. Ambient displays are expected to go largely unnoticed until some change in the display or user's state of attention makes it come into the foreground of attention. How to keep the level of display at the threshold of a user's attention is an open design issue.

Contributions of TUIs

TUI is generally built from systems of physical artifacts with digital coupling with computation. Taken together as ensembles, TUI has several important advantages over traditional GUI as well as limitations. This section summarizes those contributions of TUIs and required design considerations.

Double Interactions Loop – immediate tactile feedback

One important advantage of TUI is that users receive passive haptic feedback from the physical objects as they grasp and manipulate them. Without waiting for the digital feedback (mainly visual), users can complete their input actions (e.g. moving a building model to see the interrelation of shadows).

Typically there are two feedback loops in TUI, as shown in Figure 5.

1) The passive haptic feedback loop provides the user with an immediate confirmation that he or she has grasped and moved the object. This loop exists within a physical domain, and it does not require any sensing and processing by a computer. Thus, there is no computational delay. The user can begin manipulating the object as desired without having to wait for the second feedback loop, the visual confirmation from the interface. In contrast, when user uses a mouse with a GUI computer, he or she has to wait for the visual feedback (2nd loop) to complete an action.

2) The 2nd loop is a digital feedback loop that requires sensing of physical objects moved by users, computation based on the sensed data, and displaying the results as visual (and auditory) feedback. Therefore, this 2nd loop takes longer than the 1st loop.

Many of the frustrations of using current computers come from the noticeable delay of digital feedback as well as a lack of tactile confirmation of actions taken by computers. We believe the double loops of TUI give users a way to ease those frustrations.

[note] Actuation technology introduced in Actuated Workbench will contribute to add another loop, that of physical actuation. Fig. 5 illustrates the 3rd loop introduced into the TUI model by computer-controlled actuation and sensing. The 3rd loop allows the computer to give feedback on the status of the digital information as the model changes or responds to internal computation.

Persistence of tangibles

As physical artifacts, TUI tangibles are persistent. Tangibles also carry physical state, with their physical configurations tightly coupled to the digital state of the systems they represent. The physical state of tangibles embodies key aspects of the digital state of an underlying computation.

For example, the physical forms of the Urp building models, as well as their position and orientation on the workbench of the system, serve central roles in representing and controlling the state of the underlying digital simulation system. Even if the mediating computers, cameras, and projectors of Urp are turned off, many aspects of the state of the system are still concretely expressed by the configuration of its physical elements.

In contrast, the physical form of the mouse holds little representational significance because GUIs represent information almost entirely in visual form.

Coincidence of Input and Output Spaces

Another important feature (and design principle) of TUI is coincidence of input and output spaces to provide seamless information representation that spans both tangible (physical) and intangible (digital) domains.

GUI utilizes the mouse and keyboard as generic "remote" controllers (input), and the screen serves as main output medium. Thus, there is spatial discontinuity between those two spaces. There is also multimodal inconsistency, as touch is the main input while vision is the only output.

TUI tries to coincide inputs space and output space as much as possible to realize seamless coupling of physical and digital worlds (Ishii and Ullmer, 1997). An example of this seamless coupling of is Underkoffler's Urp

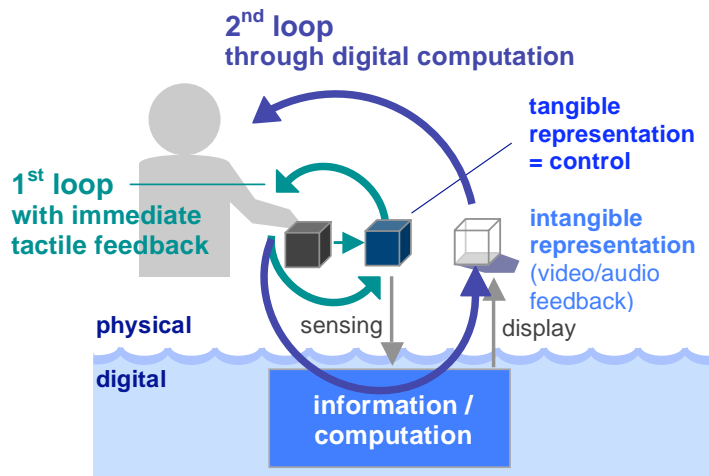


Fig. 5 TUI's Double Feedback Loops

TUI provides two feedback loops:

- 1) 1st immediate tactile feedback, and
- 2) 2nd feedback through digital processing with possible delay.

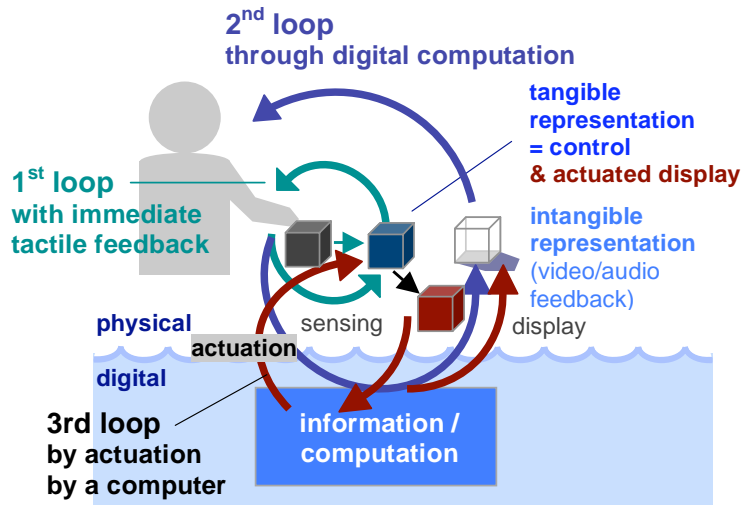


Fig. 6 TUI with Actuation (Actuated Workbench)

Computational actuation provides another loop for computer to control the position of objects (tangible representation) on the workbench.

(Underkoffler and Ishii, 1999). A series of architectural models serve as the input devices, and output in the form of a wind and shadow simulation is projected down onto the same tabletop surface, on top of and around the building models. Illuminating Clay (Piper, et al., 2002) and SandScape (Ishii, et al., 2004) demonstrates another example of i/o coincidence using continuous flexible material: sand. Curlybot and topobo demonstrate the same concept using the contact surface of the tangibles as input and output to digitize the person's physical motion.

Special Purpose vs. General Purpose

GUIs are fundamentally general purpose interfaces that are supposed to emulate a variety of applications visually using dynamic pixels on a screen and generic remote controllers such as the mouse and keyboard. On the other hand, TUIs are relatively specific interfaces tailored to certain type of applications in order to increase the directness and intuitiveness of interactions.

The selection of the correct and specific application domain is critical to apply TUI successfully to take advantage of existing skills and work practices (e.g. use of physical models in urban planning).

One notable aspect of Urp is its use of objects with very application-specific physical forms (scaled building models) as a fundamental part of the interface. Physical building models represent the buildings themselves in the interactive simulation. Thus they give the user important visual and tactile information about the computational object they represent. Indicators such as a clock and weather vane work in reverse in the Urp system. Instead of the clock hands moving to indicate the passage of time, the user can move the clock hands to change the time of day for the shadow study (Photo 1). Likewise, he or she can change the orientation of the weather vane to control the direction of the wind (Photo 2).

In the design of TUI, it is important to give an appropriate form to each tangible tool and object so that the form will give an indication of the function available to the users. For example, the clock hands allow people to automatically make the assumption that they are controlling time.

Of course, this special-purpose-ness of TUIs can be a big disadvantage if users would like to apply it to a wide variety of applications since customized physical objects tailored to certain application can not be reused for most of other applications. By making the form of objects more abstract (e.g a round puck), you lose the legibility of tangible representation and the object will become a generic handle rather than the representation of underlying digital information. It is important to attain a balance between specific/concrete vs. generic/abstract to give a form to digital information and computational function.

Space-Multiplexed Input

Another distinct feature of TUI is space-multiplexed input (Fitzmaurice, et al., 1995a). Each tangible representation serves as a dedicated controller occupying its own space, and encourages two-handed & multi-user simultaneous interaction with underlying computational model. Thus TUI is suitable for collocated collaboration allowing concurrent manipulation of information by multiple users.

GUI, in contrast, provides time-multiplexed input that allows users to use one generic device to control different computational functions at different points in time. For instance, the mouse is used for menu selection, scrolling windows, pointing and clicking buttons in a time-sequential manner.

TUI can support not only collocated collaboration, but also remote collaboration using actuation mechanism to synchronize the physical states of tangibles over distance. Actuated Workbench is an example of such a technology that extends TUI for remote collaboration (Pangaro, et al., 2002).

In the Urp scenario, applying the Actuated Workbench technology, it is possible to have two distributed Urp tables in different locations, connected and synchronized over the internet. One Urp can be in Tokyo, while the other Urp can be in Boston, and the shadows are synchronized as the urban planning team moves the buildings around the Urp space. The movement of buildings can be also synchronized by the actuation mechanism. When the building planner moves a building location, both the local and the remote shadow will update simultaneously and position and orientation of moved building is also synchronized. This synchronization of distributed workbench allows both teams to discuss changes to the situation in realtime, and provides a common reference for otherwise ethereal qualities such as wind, time, and shadows.

Conclusion

The author met a highly successful computational device called the "abacus" when he was two years old (Photo 13). He could enjoy the touch and feel of the "digits" physically represented as arrays of beads. This simple abacus was not merely a digital computational device. Because of its physical affordance, the abacus also became a musical instrument, imaginary toy train, and a back scratcher. He was captivated by the sound and tactile interaction with this simple artifact.

His childhood abacus became a medium of awareness too. When his mother kept household accounts, he was aware of her activities by the sound of her abacus, knowing he could not ask her to play with him while her abacus made its music.

This abacus suggests to us a new direction of Human-Computer Interaction (HCI) that we call Tangible User

Interfaces (TUI). First, it is important to note that the abacus makes no distinction between "input" and "output." Instead, the beads, rods, and frame serve as physical representations of numerical information and computational mechanism. They also serve as directly manipulatable physical controls to compute on numbers.

Second, the simple and transparent mechanical structure of the abacus (without any digital black boxes) provides rich physical affordances (Norman, 1999) so that even children can immediately understand what they can do with this artifact without reading a manual.

TUI pursues these features further into the digital domain by giving physical form to digital information and computation, employing physical artifacts both as representations and controls for computational media. Its design challenge is a seamless extension of the physical affordances of the objects into the digital domain.

This chapter introduced the basic concept of TUI and a variety of examples of TUI applications to address the key properties of TUI and its design challenges. TUI is still in its infancy, and extensive research is required to identify the killer applications, scalable TUI toolkits, and a set of strong design principles.

The research of TUI which gives physical forms to digital information/computation naturally crosses with the paths of industrial/product design as well as environmental/architectural design. It has also made an impact on the media arts/interactive arts community. The author hopes that TUI design will contribute to promote those interdisciplinary design research initiatives in the HCI community to bring strong design culture as well as media arts perspective to the scientific/academic world.

Mark Weiser's seminal paper on Ubiquitous Computing (Weiser, 1991) started with the following paragraph"

"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it."

I do believe that TUI is one of promising paths to his vision of invisible interface.

Acknowledgements

The author would like to thank to many of the former and current TMG students and colleagues in the MIT Media Lab for their contributions to the TUI research in the past 10 years at the MIT Media Lab. Especially thanks are due to Dr. John Underkoffler, Dr. Brygg Ullmer, Dr. Kimiko Ryokai, Dr. James Patten, Angela Chang, Hayes Raffle, Amanda Perks, Oren Zuckerman. Thanks are also due to Things That Think and Digital Life Consortia at the Media Lab for their support of our research. The author also would like to thank Prof. Bill Buxton and Dr. George Fitzmaurice for our collaboration on Bricks project in 1994 at University of Toronto (Fitzmaurice, et al., 1995b) which laid the foundation of TUI, and Prof. Mitchel Resnick for his vision of Digital Manipulatives (Resnick, et al., 1998) which influenced on our design of educational TUI applications.

References

- Aish, R. (1979). 3D input for CAAD systems. *Computer-Aided Design* 11, 2, 66-70.
- Aish, R. and Noakes, P. (1979). Architecture without numbers - CAAD based on a 3D modelling system. *Computer-Aided Design* 16, 6, 321-328, Nov. 1984. .
- Anagnostou, G., Dewey, D. and Patera., A. (1989). Geometry-defining processors for engineering design and analysis. *The Visual Computer* 5, 304-315.
- Anderson, D., Frankel, J. L., Marks, J., Agarwala, A., Beardsley, P., Hodgins, J., Leigh, D., Ryall, K., Sullivan, E. and Yedidia, J. S. (2000). Tangible interaction + graphical interpretation: a new approach to 3D modeling, *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., pp. 393-402.
- Arias, E., Eden, H. and Fisher, G. (1997). Enhancing communication, facilitating shared understanding, and creating better artifacts by integrating physical and computational media for design, *Proceedings of the conference on Designing interactive systems: processes, practices, methods, and techniques*, ACM Press, pp. 1-12.
- Brave, S. and Dahley, A. (1997). inTouch: A Medium for Haptic Interpersonal Communication (short paper). *Conference on Human Factors in Computing Systems (CHI '97) (Atlanta, March 1997)*, ACM, pp.
- Brave, S., Ishii, H. and Dahley, A. (1998). Tangible interfaces for remote collaboration and communication, *Proceedings of the 1998 ACM conference on Computer Supported Cooperative Work (CSCW 98)*, ACM Press, pp. 169-178.
- Burbeck, S. (1992). *Applications Programming in Smalltalk-80(TM): How to use Model-View-Controller (MVC)*.
- Buxton, W. (1995). Integrating the Periphery and Context: A New Model of Telematics. *Graphics Interface '95*, 239-246.
- Chang, A., O'Modhrain, S., Jacob, R., Gunther, E. and Ishii, H. (2002). ComTouch: design of a vibrotactile communication device, *Proceedings of the conference on Designing interactive systems: processes, practices, methods, and techniques*, ACM Press, pp. 312-320.
- Chang, A., Resner, B., Koerner, B., Wang, X. and Ishii, H. (2001). LumiTouch: an emotional communication device, *CHI '01 extended abstracts on Human factors in computing systems*, ACM Press, pp. 313-314.
- Cohen, J., Withgott, M. and Piernot, P. (1999). LogJam: a tangible multi-person interface for video logging. *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit (Pittsburgh, Pennsylvania, United States)*, ACM Press pp. 128-135.

- Crampton Smith, G. (1995). *The Hand That Rocks the Cradle*, I.D., pp. 60-65.
- Dahley, A., Wisneski, C. and Ishii, H. (1998). *Water Lamp and Pinwheels: Ambient Projection of Digital Information into Architectural Space* (short paper). Conference on Human Factors in Computing Systems (CHI '98) (Los Angeles, April 1998), ACM, pp. Conference Summary of CHI '98.
- Fishkin, K. P. (2004). A taxonomy for and analysis of tangible interfaces. *Personal Ubiquitous Comput.* 8, 347-358.
- Fitzmaurice, G. W., Ishii, H. and Buxton, W. (1995a). *Bricks: Laying the Foundations for Graspable User Interfaces*. Conference on Human Factors in Computing Systems (CHI '95) (Denver, Colorado), ACM, pp. 442-449.
- Fitzmaurice, G. W., Ishii, H. and Buxton, W. A. S. (1995b). *Bricks: laying the foundations for graspable user interfaces*, Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co., pp. 442-449.
- Fogg, B., Cutler, L. D., Arnold, P. and Eisbach, C. (1998). *HandJive: a device for interpersonal haptic entertainment*, Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co., pp. 57-64.
- Frazer, J. (1994). *An Evolutionary Architecture* Architectural Association, London.
- Frazer, J., Frazer, J. and Frazer, P. (1980). *Intelligent physical three-dimensional modelling system*, *Computer Graphics* 80, North Holland, pp. 359-370.
- Frei, P., Su, V., Mikhak, B. and Ishii, H. (2000a). *curlybot: designing a new class of computational toys*, Proceedings of the SIGCHI conference on Human factors in computing systems (CHI 2000), ACM Press, pp. 129-136.
- Frei, P., Su, V., Mikhak, B. and Ishii, H. (2000b). *curlybot: designing a new class of computational toys*. Proceedings of the SIGCHI conference on Human factors in computing systems (CHI 2000) (The Hague, The Netherlands, April 01 - 06, 2000), ACM Press, pp. 129-136.
- Goldberg, A. (1984). *Smalltalk-80: The Interactive Programming Environment*. Addison-Wesley.
- Gorbet, M., Orth, M. and Ishii, H. (1998). *Triangles: Tangible Interface for Manipulation and Exploration of Digital Information Topography*, Conference on Human Factors in Computing Systems (CHI '98), ACM.
- Holmquist, L. E., Redstr, J. and Ljungstrand, P. (1999). *Token-Based Acces to Digital Information*, Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing, Springer-Verlag, pp. 234-245.
- Ishii, H. (2004). *Bottles: A Transparent Interface as a Tribute to Mark Weiser*. *IEICE Transactions on Information and Systems* E87-D, 6, 1299-1311.
- Ishii, H., Fletcher, H. R., Lee, J., Choo, S., Berzowska, J., Wisneski, C., Cano, C., Hernandez, A. and Bulthaupt, C. (1999). *musicBottles*. ACM SIGGRAPH 99 Conference abstracts and applications (Los Angeles, California, United States), ACM Press, pp. 174.
- Ishii, H., Mazalek, A. and Lee, J. (2001). *Bottles as a minimal interface to access digital information*, CHI '01 extended abstracts on Human factors in computing systems, ACM Press, pp. 187-188.
- Ishii, H., Ratti, C., Piper, B., Wang, Y., Biderman, A. and Ben-Joseph, E. (2004). *Bringing clay and sand into digital design - continuous tangible user interfaces*. *BT Technology Journal* 22, 4, 287-299.
- Ishii, H. and Ullmer, B. (1997). *Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms*. Conference on Human Factors in Computing Systems (CHI '97) (Atlanta, March 1997), ACM, pp. 234-241.
- Ishii, H., Wisneski, C., Brave, S., Dahley, A., Gorbet, M., Ullmer, B. and Yarin, P. (1998). *ambientROOM: Integrating Ambient Media with Architectural Space* (video), Conference on Human Factors in Computing Systems (CHI '98), ACM, pp. Conference Summary of CHI '98.
- Johnson, J., Roberts, T. L., Verplank, W., Smith, D. C., Irby, C. H., Beard, M. and Mackey, K. (1989). *The Xerox Star: a retrospective*. *IEEE Computer* 22, 9, 11-26, 28-29.
- Kitamura, Y., Itoh, Y. and Kishino, F. (2001). *Real-time 3D interaction with ActiveCube*, CHI '01 extended abstracts on Human factors in computing systems, ACM Press, pp. 355-356.
- Klemmer, S. R., Thomsen, M., Phelps-Goodman, E., Lee, R. and Landay, J. A. (2002). *Where do web sites come from?: capturing and interacting with design history*, Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, ACM Press, pp. 1-8.
- Kobayashi, K., Hirano, M., Narita, A. and Ishii, H. (2003). *A tangible interface for IP network simulation*, CHI '03 extended abstracts on Human factors in computing systems, ACM Press, pp. 800-801.
- Lee, J., Su, V., Ren, S. and Ishii, H. (2000). *HandSCAPE: a vectorizing tape measure for on-site measuring applications*, Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press, pp. 137-144.
- Mazalek, A., Wood, A. and Ishii, H. (2001). *genieBottles: An Interactive Narrative in Bottles*, Conference Abstracts and Applications of SIGGRAPH '01, ACM Press, pp. 189.
- Norman, D. A. (1999). *Affordance, conventions, and design, interactions*, ACM Press, pp. 38-43.
- Pangaro, G., Maynes-Aminzade, D. and Ishii, H. (2002). *The actuated workbench: computer-controlled actuation in tabletop tangible interfaces*, Proceedings of the 15th annual ACM symposium on User Interface Software and Technology (UIST 2002), ACM Press, pp. 181-190.
- Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books.
- Patten, J., Ishii, H., Hines, J. and Pangaro, G. (2001). *Sensetable: a wireless object tracking platform for tangible user interfaces*, Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press, pp. 253-260.
- Patten, J., Recht, B. and Ishii, H. (2002). *Audiopad: A Tag-based Interface for Musical Performance New Interface for Musical Expression*.

- Piper, B., Ratti, C. and Ishii, H. (2002). Illuminating clay: a 3-D tangible interface for landscape analysis, Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, ACM Press, pp. 355-362.
- Raffle, H. S., Parkes, A. J. and Ishii, H. (2004). Topobo: a constructive assembly system with kinetic memory, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2004), ACM Press, pp. 647-654.
- Ratti, C., Wang, Y., Piper, B., Ishii, H. and Biderman, A. (2004). PHOXEL-SPACE: an interface for exploring volumetric data with physical voxels, Proceedings of the 2004 conference on Designing interactive systems: processes, practices, methods, and techniques, ACM Press, pp. 289-296.
- Rauterberg, M., Fjeld, M., Krueger, H., Bichsel, M., Leonhardt, U. and Meier, M. (1998). BUILD-IT: a planning tool for construction and design, CHI 98 conference summary on Human factors in computing systems, ACM Press, pp. 177-178.
- Rekimoto, J., Ullmer, B. and Oba, H. (2001). DataTiles: a modular platform for mixed physical and graphical interactions, Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press, pp. 269-276.
- Resnick, M., Martin, F., Berg, R., Borovoy, R., Colella, V., Kramer, K. and Silverman, B. (1998). Digital manipulatives: new toys to think with, Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co., pp. 281-287.
- Ryokai, K., Marti, S. and Ishii, H. (2004). I/O brush: drawing with everyday objects as ink, Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press, pp. 303-310.
- Smith, D. (1982). Designing the Star User Interface, Byte, pp. 242-282.
- Stifelman, L. J. (1996). Augmenting real-world objects: a paper-based audio notebook, Conference companion on Human factors in computing systems: common ground, ACM Press, pp. 199-200.
- Suzuki, H. and Kato, H. (1993). AlgoBlock: A tangible programming language -- a tool for collaborative learning, the 4th European Logo Conference, pp. 297-303.
- Ullmer, B. and Ishii, H. (1997). The metaDESK: Models and Prototypes for Tangible User Interfaces, Symposium on User Interface Software and Technology (UIST '97), ACM Press, pp. 223-232.
- Ullmer, B. and Ishii, H. (2000). Emerging frameworks for tangible user interfaces. IBM Systems Journal 39, 3&4, 915-931.
- Ullmer, B., Ishii, H. and Glas, D. (1998). mediaBlocks: physical containers, transports, and controls for online media, Proceedings of the 25th annual conference on Computer graphics and interactive techniques, ACM Press, pp. 379-386.
- Ullmer, B., Ishii, H. and Jacob, R. J. K. (2003). Tangible Query Interfaces: Physically Constrained Tokens for Manipulating Database Queries, INTERACT 2003 Conference, IFIP.
- Ullmer, B., Ishii, H. and Jacob, R. J. K. (2005). Token+constraint systems for tangible interaction with digital information. 12, 81-118.
- Underkoffler, J. and Ishii, H. (1998). Illuminating Light: An Optical Design Tool with a Luminous-Tangible Interface, Conference on Human Factors in Computing Systems (CHI '98), ACM Press/Addison-Wesley Publishing Co., pp. 542-549.
- Underkoffler, J. and Ishii, H. (1999). Urp: a luminous-tangible workbench for urban planning and design, Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, ACM Press, pp. 386-393.
- Underkoffler, J., Ullmer, B. and Ishii, H. (1999). Emancipated pixels: real-world graphics in the luminous room, Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co., pp. 385-392.
- Weiser, M. (1991). The computer for the 21st Century. Scientific American 265, 3, 94-104.
- Wellner, P. (1993). Interacting with Paper on the DigitalDesk. Communications of the ACM 36, 7, 87-96.
- Zuckerman, O., Arida, S. and Resnick, M. (2005). Extending tangible interfaces for education: digital montessori-inspired manipulatives, Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press, pp. 859-868.
- Zuckerman, O. and Resnick, M. (2004). Hands-on modeling and simulation of systems, Proceeding of the 2004 conference on Interaction design and children: building a community, ACM Press, pp. 157-158.

----- end of the text

An Encompassing View on Tangible Interaction: A Framework

Eva Hornecker

Interact Lab, University of Sussex, Falmer, Brighton BN19PF, UK
eva@ehornecker.de

ABSTRACT

Our current understanding of human interaction with hybrid or augmented environments is very limited. We here focus on ‘tangible interaction’, denoting systems relying on embodied interaction, tangible manipulation, physical representation of data, and embeddedness in real space. This synthesis of prior ‘tangible’ definitions allows us to address a larger design space integrating approaches from different disciplines. We introduce a framework that contributes to understanding the (social) user experience of tangible interaction. This understanding lays the ground for evolving knowledge on collaboration-sensitive design.

INTRODUCTION

Tangible User Interfaces (TUIs) and Tangible Interaction are terms increasingly gaining currency within HCI. Through embedding computing in the everyday environment and supporting intuitive use these approaches share goals with other novel approaches to HCI. Variations have been pursued over the last two decades as ‘graspable user interfaces’ [7], ‘tangible user interfaces’ [19], or ‘tangible interaction’ [3, 5]. Design in this domain requires not only designing the digital but also the physical, as well as designing new types of interaction: these are new challenges for design and HCI. Through various effects these systems lend themselves to the support of face-to-face social interaction, reflected in a considerable number of systems aimed at cooperative scenarios [1, 18, and see 19].

Research until recently focused on developing new systems. A move towards concepts and theory can be detected from a special issue on ‘tangible interfaces in perspective’ [10]. However, attempts to develop frameworks have concentrated mainly on defining terms or on categorizing and characterizing systems (e.g. [17, 19]). While supporting structural analysis, mapping out the design space and detecting uncharted territory, these offer little advice when designing for real world situations and seldom address users’ interaction experience. There is still a need for conceptual frameworks, that unpack why ‘tangible

interaction’ works so well for users [6]. Equally there is a need for principled approaches supporting research and design of these new hybrid environments.

We have chosen to use ‘tangible interaction’ as an umbrella term, drawing together several fields of research and disciplinary communities. This view encompasses a broad scope of systems relying on embodied interaction, tangible manipulation and physical representation of data, being embedded in real space and digitally augmenting physical space. It covers approaches from HCI, computing, product design and interactive arts. The proliferation of computing into everyday appliances draws product designers towards IT product design [3, 5]. Artists and museums experiment with hybrid interactives. Increasingly systems are developed by users e.g. in architecture or biology. This becomes even more prominent with computing moving beyond the desktop and ‘intelligent’ devices spreading into all fields of life and work. Thus a conceptual understanding of this new interface type and knowledge supporting design becomes even more important.

In this position paper we can only give a short overview of our framework that focuses on the user experience of interaction and aims to unpack the interweaving of the material/physical and the social aspects of interaction. It is described in more detail in [13]. The framework contributes to the larger research agenda of Embodied Interaction [6, 15], offering four themes and a set of concepts. It builds upon results from a PhD project [11] and recent studies. One theme is described in detail in [12].

A BROAD VIEW ON TANGIBLE INTERACTION

We now give an overview of the dominant views and approaches on ‘tangible interaction’ and propose a deliberately broad, encompassing view. A look at the above mentioned approaches from other disciplines reveals that the ‘tangible interface’ definition frequently used in HCI is too narrow to encompass these. From the characterizations found, we can distinguish a data-centered view, pursued in Computer Science and HCI; an expressive-movement-centered view from Industrial and Product Design; and a space-centered view from Arts and Architecture:

- *Data-centered view*: [6, 10, 19] define ‘tangible user interfaces’ as utilizing physical representation and manipulation of digital data, offering interactive couplings of physical artifacts with “computationally

This position paper is based on the first pages of a full paper accepted for CHI 2006, being shortened and slightly altered.

CHI 2006, April 22–28, 2006, Montréal, Québec, Canada.
Copyright 2006 ACM 1-59593-178-3/06/0004...\$5.00.

mediated digital information” [10]. This characterization of TUIs is dominant in HCI publications. Conceptual research from HCI and computer science tends to explore types of coupling and representations [19].

- *Expressive-Movement-centered view*: An emerging ‘school’ in product design aims to go beyond form and appearance and to design the interaction itself. This view emphasizes bodily interaction with objects, exploiting the “sensory richness and action potential of physical objects”, so that “meaning is created in the interaction” [5]. Design takes account of embodied skills, focuses on expressive movement and ‘rich’ interaction with ‘strong specific’ products tailored to a domain [3, 14]. The design community prefers the term ‘tangible interaction’.
- *Space-centered view*: Interactive arts and architecture increasingly talk about ‘interactive spaces’. These rely on combining real space and real objects with digital displays or sound installations [2, 16]. “Interactive systems, physically embedded within real spaces, offer opportunities for interacting with tangible devices”, and “trigger display of digital content or reactive behaviors” [4]. Full-body interaction and use of the body as interaction device or display are typical for this approach.

Tangible interaction, as we understand it, encompasses a broad scope of systems, building upon and synthesizing these approaches from different disciplinary backgrounds. These share the characteristics of tangibility/ materiality, physical embodiment of data, embodied interaction and bodily movement as an essential part of interaction, and embeddedness in real space [2, 3, 4, 5, 6, 10, 19].

This concept of *tangible interaction* has a broader scope than Ullmer and Ishii’s [19] description of tangible interfaces: “giving physical form to digital information” and its subsequent physical control, which is often used to define TUIs. Tangible interaction includes tangible appliances or remote control of the real world [14]. This approach focuses on designing the interaction itself (instead of the interface) and exploiting the richness of bodily movement [3]. Interaction with ‘interactive spaces’ by walking on sensorized floors or moving in space [2, 16] further extends our perspective, the body itself being used as input ‘device’. Instead of using a restrictive definition, it seems more productive to address this larger design space and to interpret these attempts at conceptualization as emphasizing different facets of a related set of systems.

RELATED WORK ON ‘TANGIBLE’ FRAMEWORKS

Previous attempts to develop frameworks have focused mainly on defining terms, categorizing and characterizing systems, on types of coupling. Most frameworks take a structural approach, systematically mapping out an abstract design space, but seldom address the human interaction experience. The most notable push towards a theory of tangible interaction and an understanding of the interaction experience, was provided by Dourish [6]. He emphasizes how social action is embedded in settings, focusing on the

social construction of meaning. Thus materiality itself, and its relation to the social has been less discussed.

Support of social interaction and collaboration might be the most important, domain-independent feature of tangible interaction, but this issue has attracted little explicit attention. The pioneering work by [1, 18] of analyzing social use of TUIs and identifying social affordances found few followers. Even though many researchers agree that TUIs are especially suited for co-located collaboration, conceptual work often only briefly mentions visibility of actions and distributed loci of control as collaborative affordances. Evaluations often assess individual use, or give primarily anecdotal accounts of field use.

The research community therefore lacks concepts for analyzing and understanding the social aspects of tangible interaction and design knowledge on how to design so as to support social interaction and collaboration. This has motivated the development of our framework, which focuses on the (social) interaction experience, addressing the broader design space of ‘tangible interaction’.

OUR FRAMEWORK ON TANGIBLE INTERACTION

The framework is structured around four themes (figure 1) that are not mutually exclusive, but interrelated, offering different perspectives on tangible interaction. A set of concepts elaborates each theme, providing more concrete handles for understanding their implications. Themes are:

- *Tangible Manipulation* • *Spatial Interaction*
- *Embodied Facilitation* • *Expressive Representation*

We now (briefly) present the four themes, explaining each theme’s relevance for tangible interaction and presenting the related concepts, characterized with a short question in colloquial language. A more detailed description of themes and concepts can be found in the authors’ CHI paper [13].

Theme: Tangible Manipulation (TM)

Tangible Manipulation refers to the reliance on material representations with distinct tactile qualities that is typical for tangible interaction. Tangible Manipulation is bodily interaction with physical objects. These objects are coupled with computational resources [19] to control computation. The main concepts, colloquially phrased, are:

Haptic Direct Manipulation: Can users grab, feel and move ‘the important elements’?

Lightweight Interaction: Can users proceed in small, experimental steps? Is there rapid feedback during interacting?

Isomorph Effects: How easy is it to understand the relation between actions and their effects? Does the system provide powerful representations that transform the problem?

Theme: Spatial Interaction (SI)

Spatial Interaction refers to the fact that tangible interaction is embedded in real space and interaction therefore

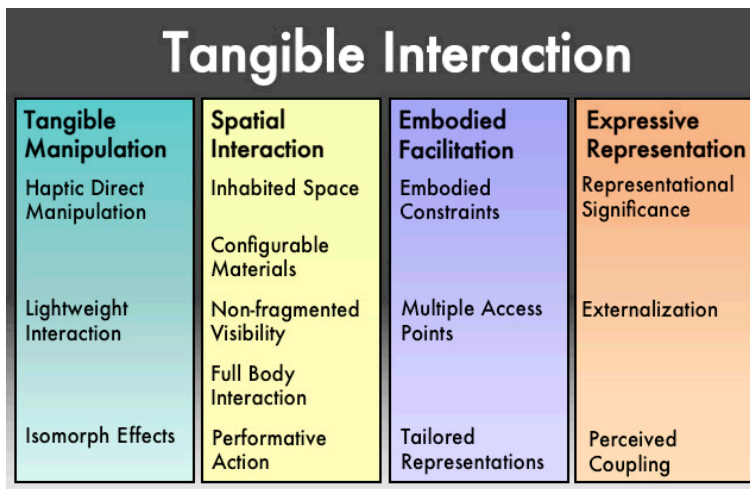


Figure 1. Tangible Interaction Framework with themes and concepts

occurring by movement in space. The interfaces take up space and they are situated in places. Interaction with spatial installations or interactive spaces can be interpreted as a form of tangible interaction that is not restricted to moving objects in space, but relies on moving one's body. The main concepts for Spatial Interaction are:

Inhabited Space: Do people and objects meet? Is it a meaningful place?

Configurable Materials: Does shifting stuff (or your own body) around have meaning? Can we configure the space at all and appropriate it by doing so?

Non-fragmented Visibility: Can everybody see what's happening and follow the visual references?

Full-Body Interaction: Can you use your whole body?

Performative Action: Can you communicate something through your body movement while doing what you do?

Theme: Embodied Facilitation (EF)

Embodied Facilitation highlights how the configuration of material objects and space affects and directs emerging group behavior. We literally move in physical space and metaphorically in software space. Tangible interaction embodies structure and thereby styles, methods and means of facilitation. We can learn from facilitation methods how to shape physical and procedural structure so as to support and subtly direct group processes (for details see [12]). The main concepts are:

Embodied Constraints: Does the physical set-up lead users to collaborate by subtly constraining their behavior?

Multiple Access Points: Can all users see what's going on and get their hands on the central objects of interest?

Tailored Representation: Does the representation build on users' experience? Does it connect with their experience and skills and invite them into interaction?

Theme: Expressive Representation (ER)

Expressive Representation focuses on the material and digital representations employed by tangible interaction systems, their expressiveness and legibility. Often hybrid representations combine material and digital elements, each with distinct representational qualities. In interaction we 'read' and interpret representations, act on and modify them. Here the main concepts are:

Representational significance: Are representations meaningful and have long-lasting importance? Are physical and digital representations of the same strength and salience?

Externalization: Can users think and talk with or through objects, using them as props to act with? Do they give discussions a focus and provide a record of decisions?

Perceived Coupling: Is there a clear link between what you do and what happens? Are physical and digital representations seemingly naturally coupled?

On the Framework

Frameworks focus our view, providing us with concepts that systematize our thinking. We feel that our approach is distinct from other frameworks by not offering taxonomies, but perspectives and themes for analysis and as conceptual guidance for design. Perspectives allow for systematic shifts of focus, highlighting different aspects of one object. The themes and concepts summarize our experiences from system assessments and reflections on design, in combination with a literature review on the use of material artifacts in social situations, distilling a set of social affordances [11]. The overall framework thus is the result of a synthesis of previous works and concepts developed by us. Recurrent themes or insights from literature have been integrated and fused into a larger framework focusing on the (social) use experience of tangible interaction.

Figure 1 can be read from left to right as referring to the design space of tangible interaction from the specific to the general. *Tangible Manipulation* is the most specific theme, relying on the use of material objects. It applies best to systems usually referred to as tangible interfaces [19] and tangible appliances. *Spatial Interaction* and *Embodied Facilitation* provide insights relevant for the broader research area of 'embodied interaction' [6], where movement in space and physical configuration of computing resources are central characteristic, e.g. mobile interaction and ubiquitous computing. *Expressive representation*, insofar as it concerns tangible representations, is specific to tangible interaction, but can be generalized to mixed reality representations. The *Embodied Facilitation* and *Spatial Interaction* themes are the most concerned with understanding and supporting social interaction. The other two themes address aspects of the user experience that support social interaction in

indirect ways, e.g. lowering participation thresholds, making action publicly available, or providing shared references, while being important for single users as well.

The framework is organized on three levels of abstraction. The themes offer perspectives at an abstract level and define broad research issues such as the role of space. Themes are each elaborated by a set of concepts that provide analytical tools, summarize generic issues, help to pinpoint design mistakes and successes or to guide design on a conceptual level. A level of more directly applicable design 'guidelines' is in development for practical purposes. These are not meant to be strict rules, they rather act as 'design sensibilities' [4, 8], inspiring and thought-provoking suggestions.

CONCLUSION

Several previous frameworks have aimed at design for social interaction (e.g. [8]) or at tangible interfaces. Few have combined both fields of interest. Our framework contributes to the larger research agenda of *Embodied Interaction* [9], providing insight into the relation of embodied and social interaction. It integrates and fuses relevant recurrent themes and concepts from previous attempts at conceptualizing tangible interaction. For example the seminal work of Fitzmaurice [7] addressed issues strongly related to the tangible manipulation theme, albeit focusing on the usability and effectiveness of haptic directness. In recent years more emphasis has been directed to the aesthetic and expressive aspects of manual interaction with objects [3, 5]. Yet these attempts have mostly investigated the individual user experience. While all of these are important contributions that have inspired us, they often considered isolated aspects. Our aim has been to integrate these into a wider framework that focuses on the overall (social) use experience. Our aim has been to develop a better understanding of the user experience of tangible interaction and concepts for analyzing its social aspects along with knowledge aiding collaboration-sensitive design.

ACKNOWLEDGMENTS

Geraldine Fitzpatrick, John Halloran, Paul Marshall, Mark Stringer and all other members of the Interact Lab. Jacob Buur encouraged and mentored the development of this framework. This work was financed during the last months by the Equator IRC GR/N15986/01.

REFERENCES

1. Arias, E., Eden, H. and Fischer, G. Enhancing Communication, Facilitating Shared Understanding, and Creating Better Artifacts by Integrating Physical and Computational Media for Design'. *Proc. of DIS '97*, ACM (1997), 1-12.
2. Bongers, B. Interactivating Spaces. *Proc. Symposium on Systems Research in the Arts, Informatics and Cybernetics* (2002).
3. Buur, J., Jensen, M.V. and Djajadiningrat, T. Hands-only scenarios and video action walls: novel methods for tangible user interaction design. *Proc. of DIS'04*. ACM (2004), 185-192.
4. Ciolfi, L. *Situating 'Place' in Interaction Design: Enhancing the User Experience in Interactive Environments*. Ph.D. Thesis, University of Limerick (2004)
5. Djajadiningrat, T., Overbeeke, K. and Wensveen, S. But how, Donald, tell us how? *Proc. of DIS'02*, ACM (2002), 285-291.
6. Dourish P. *Where the Action Is. The Foundations of Embodied Interaction*. MIT Press (2001).
7. Fitzmaurice G. W. *Graspable User Interfaces*. PhD thesis, University of Toronto, Canada (1996).
8. Fitzpatrick, G. *The Locales Framework: Understanding and designing for Wicked Problems*. Kluwer (2003)
9. Holmquist L. E., Redström J. and Ljungstrand P. Token-based access to digital information. *Proc. of HUC'99*, Springer (1999). 234-245.
10. Holmquist, L., Schmidt, A. and Ullmer, B. Tangible interfaces in perspective: Guest editors' introduction. *Personal & Ubiquitous Computing 8(5)* (2004) 291-293.
11. Hornecker, E. *Tangible User Interfaces als kooperationsunterstützendes Medium*. PhD-thesis. University of Bremen (2004)
12. Hornecker, E. A Design Theme for Tangible Interaction: Embodied Facilitation. *Proc. of ECSCW'05*, Springer (2005). 23-43
13. Hornecker, E. and Buur, J. Getting a Grip on Tangible Interaction: a Framework on Physical Space and Social Interaction. Accepted Paper for CHI 2006
14. Jensen, M.V., Buur, J. and Djajadiningrat, T. Designing the user actions in tangible interaction. *Proc. of Critical Computing Aarhus 2005*. ACM (2005). 9-18
15. Robertson T. Cooperative Work and Lived Cognition. A Taxonomy of Embodied Actions. *Proc. of E-CSCW'97*, Kluwer (1997), 205-220.
16. Rubidge, S. and MacDonald, A. Sensuous Geographies: a multi-user interactive/responsive installation. *Digital Creativity Vol 15, No. 4*, 2004, 245-252
17. Sharlin, E., et al. On tangible user interfaces, humans and spatiality. *Personal and Ubiquitous Computing 8(5)* (2004), 338-346.
18. Suzuki H. and Kato H. Interaction-level support for collaborative learning: Algoblocks - an open programming language. *Proc. of CSCL (1995)*, 349-355.
19. Ullmer B. and Ishii H. Emerging frameworks for tangible user interfaces. *IBM Systems Journal 39(3-4)* (2000), 915-931.

From “Human-Computer Interaction” to “The Use and Design of Digital Artefacts”

Submission to the CHI 2006 Workshop “What is the Next Generation of Human-Computer Interaction?”

Lars Erik Holmquist
Future Applications Lab, Viktoria Institute
leh@viktoria.se
www.viktoria.se/fal

When seeing the title of this workshop, I realize that in academia, we are often fixated on the computer itself rather than on what it is supposed to accomplish. Our conferences are called things like “Human-Computer Interaction”, “Ubiquitous Computing”, or “Computer Graphics”. Even supposedly alternative paradigms like “ambient” or “pervasive” computing places too much emphasis on *how* things are done, rather than *what* we should be doing.

In fact, the notion of “human-computer interaction” is a remnant of an increasingly outmoded way of thinking about digital technology. The term “computer” itself is now so loaded with meaning, that looking for new ways to “interact with computers” can be genuinely counterproductive. Instead of looking for new *human-computer interfaces*, what we should think about is *interaction models* and *applications* where humans and computing (rather than computers) co-exist. I propose that rather than “human-computer interaction”, we should talk about “the use and design of digital artefacts.”

Let’s take a few steps back. During the last decade, HCI researchers have been developing a multitude of alternatives to the traditional desktop computer. In the hope of improving how we interact with digital information, we have seen a number of new paradigms, such as virtual reality, graspable interfaces, tangible interfaces, augmented reality, mixed reality and so on.

But at the same time, re-thinking the desktop computer has become a largely irrelevant pursuit. It is a battle that has already been both won and lost. Won, in the sense that the computer we have today is as successful an incarnation of the original vision of “intuitive interfaces” as we could hope for. Lost, in the sense that progress has already bypassed the “computer” – and a large part of the HCI community with it.

The modern computer interface came together at Xerox PARC in the mid-70’s, building on research in the 60’s on interactive graphics, input devices, networking and other fundamental technologies. After Xerox had failed to capitalize on their inventions, Apple licensed some of the patents and in 1984 launched the Macintosh. The rest is history, as windows, icons, mouse and pointers eventually became the standard for interacting with computers. So successful was this paradigm that the gigabyte and gigahertz workstations we all use today are – on the surface at least – hard to distinguish from a Xerox terminal of almost 30 years ago. But if capabilities like memory and speed have increased many thousand times, why are we still using the same tired setup of screen, keyboard and pointer to get data into and out of the computer?

The main reason is that while the idea was excellent to start with, technology took a while to catch up. The Xerox Star, launched in 1981, had just about everything we expect of a computer today; WYSIWYG editing, intuitive graphical interfaces, networking with laser printers and servers, and so on. It was also practically unusable. The graphics-intensive software and flexible networking capabilities coupled with a slow processor and limited memory led to a system that had no chance of running at reasonable speed. The engineers had been so concerned with making a “perfect” system that they failed to take into account if it was actually feasible given the technological limitations of the day.

Around the same time, engineers at a rival company were working within strict technical limitations on a project that turned out to be much more successful: the IBM Personal Computer, or PC. Compared to the Star, the PC was laughably primitive – no networking, no graphical interface, no WYSIWYG editing, not even a mouse. But it worked. And the PC standard turned out to be remarkably flexible, so that over the years it has been able to accommodate all the Star’s features and more – turning it into the computer that Xerox tried, and failed, to make.

Today’s desktop computer is very close to what the engineers at Xerox tried to create, but couldn’t with the technology of the day. The interface is by no means perfect, but for what it does it is pretty damn good – and it is not likely to change. For information processing, document creation, network management, and the hundreds of other tasks we perform at the desktop, it works just fine, thank you very much. And desktop computers with graphical user interfaces are used by hundreds of millions of people around the world. The battle has been won – the intuitive computer interface is here.

This remarkable success story is perhaps the best confirmation possible that a sufficiently strong vision of interactive technology can have a huge – if in this case substantially delayed – impact. The interface that the engineers and designers at Xerox envisioned just needed a few decades of increasing computing power to become viable, but the basic idea was good enough to still be in use 30 years later.

Today, there is no need to re-think the fundamental way in which we interact with desktop computers. Instead, if we want to progress we should be looking for applications that are completely *orthogonal* to the desktop. This means abandoning not only the idea of a “computer”, but even that of an “interface” altogether!

The problem is more deeply rooted than many perhaps realize. The traditional (and technically correct) view of a computer is a programmable central processor with which humans communicate using a set of input and output devices. This view has been relevant ever since the first digital computers right up until the last decade or so, and is still correct when it comes to the desktop computer. But it is not the right way to think for future digital applications. As computing becomes diffused into the world, it is not meaningful anymore to separate “computer” and “interface” into distinct entities.

Yet, this is exactly the thinking behind many popular new interface paradigms. One example is “tangible interfaces”. The vision described by Ishii and Ullmer (1997) was to turn “painted bits” (graphical user interfaces) into “tangible bits” (tangible user interfaces). This involved creating

new physical input and output devices, such as Phicons, Ambient displays, etc. These input and output devices were all clearly different from the mouse, keyboard and screen paradigm; yet, at heart they were fundamentally the same. Rather than being a new model for interacting with computation, tangible interfaces keep the same mental model of a computer – a central processor coupled with a set of inputs and outputs. This model is still obviously reflected in popular toolkits for building tangible interfaces, such as Phidgets (Greenberg and Fitchett 2001), where new interaction devices are always controlled by a central (desktop) computer.

A problem with this way of thinking is that it decouples the computation from the actual activity. The “computer” is on one side of a barrier, with the human on the other, and the interface sits in-between. It can lead to the mistake of thinking that it is enough to re-design the *interface*, whereas in fact the *whole activity* must be considered as a unit if a real change is to be accomplished. But the “computer” (or rather the program running on it) is not an abstract entity isolated from the user – because if we change what the computer does, we also change the interaction.

Therefore, creating novel input and output devices is simply not enough, if we keep the same mental model of a human and a computer separated by an interface. “Re-thinking the way we interact with computers” to find the interactive applications of the future is about as fruitful as making a new running shoe by “re-thinking the car”.

If we continue the analogue of a car, we do not think of the gas pedal as a “human-motor interface”, or the steering wheel as a “human-wheel interface” even though that is technically correct (in the same way that the mouse or GUI is a “human-computer interface”). Instead, we think of them as controlling the things a car should do – taking us from one place to the other. In the same way, we do not “interact with computers” anymore, we “do things that are supported by computations” such as write documents listen to music, etc. – just like we do not “interact with the motor and the wheels” in a car, we “drive from A to B”.

An alternative way of thinking that I find attractive, but still not completely satisfactory, is the “augmented artefacts” approach (Beigl et al 2001). This was also the way of thinking that motivated the Smart-Its project, a system for digitally augmenting everyday artefacts with sensing and computation (Holmquist et al 2004). In this approach, the starting point is an everyday artefact that is augmented with sensing and computation. The original function of the artefact is unaffected by this augmentation but additional functionality is added.

In the case of the *MediaCup*, for instance, an ordinary coffee was augmented by adding a processor, heat and movement sensors, and an infrared transceiver. The original function remains – the cup is still good for drinking from – but the cup is also context-aware, and can for instance give a report on its position and usage. While the *MediaCup* still required a central computer for data processing (as did most, but not all, Smart-Its-based demonstrations) it was still a break from the human-computer interface mental model in that users never consciously “interacted” with a computer – they just used a coffee cup as they would have done anyway.

Of course, I am not the first one to try to drop the “computer” from the field of human-computer interaction. The terms “interaction design” and, more recently “user experience design” are commendable alternatives. “Information appliances” have been introduced to describe smaller,

more specialized digital devices. And the original vision of “ubiquitous computing” as proposed by Mark Weiser captured much of what I am looking for, in particular by explicitly substituting “computers” with “computing”, thus emphasising what computers do rather than the actual computer (a fine point that is often lost!)

However, for the sake of this workshop, I wanted to find a term that focuses on the specific artefacts that will usher in the next era, not of human-computer interaction, but of human activities augmented by digital technology. Therefore, I propose to replace “human-computer interaction” with “the use of digital artefacts”. This captures both the activity aspect (“use”), that we are doing something new (“design”), the reliance on computation (“digital”), and finally, that we do not consider interface and computer separate but the entire object as a whole (“artefact”). It is a slightly unwieldy term, but it will do for now.

I look forward to discussing this further at CHI 2006 – no matter what the workshop happens to be called!

References

Beigl, M., HW Gellersen, A Schmidt (2001). Mediacups: experience with design and use of computer-augmented everyday artefacts. *Computer Networks*, 2001.

Greenberg, S. and Fitchett, C. (2001) Phidgets: Easy development of physical interfaces through physical widgets. Proceedings of ACM UIST 2001 User Interface Software and Technology, 2001.

Holmquist, L.E. et. al. Building Intelligent Environments with Smart-Its. *IEEE Computer Graphics & Applications*, January/February 2004, pp 56-64.

Ishii, H. and Ullmer, B. (1997). Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms, in *Proceedings of Conference on Human Factors in Computing Systems (CHI '97)*, (Atlanta, March 1997), ACM Press, pp. 234-241.

**What is the Next Generation of Human-Computer Interaction?
CHI 2006 WORKSHOP**

George Fitzmaurice
Autodesk
210 King Street East, Toronto, Ontario, Canada
fitzmaurice@acm.org

To answer what is the next generation of HCI, we need to look back a bit and determine what generation we are on now and what constitutes a generational shift. Everyone can agree that command line and GUI are two distinct generations. So, how do we know the boundaries of one generation from the next? For example, what about the cell phone and PDA interfaces? What about the internet revolution? One could argue that these systems still use the WIMP (windows, icons, menus, pointer device) elements and so belong to the GUI generation, yet I think they introduced some significant changes and usage models.

GUIs are composed with WIMP (Windows, Icons, Menus, Pointing device) elements. There should be a corresponding set of elementals for this next generation UI. However, do we have to start from scratch? Are we trying to reinvent the wheel? It's like saying the caveman invented the wheel so "what's the next generation of the wheel?" Yet, the wheel will always be valuable. It just continues to get refined and augmented.

In fact, perhaps we have recently reverted back to command line interfaces since we type in text for specifying web addresses and conducting web searches. But let's look forward.

Future Trends

In general I believe the domain of human computer interaction has expanded from focusing primarily on "office worker" activities to consumer activities (e.g., entertainment, content delivery and commerce). Here are some thoughts for the future HCI generational changes:

Less representational and more actual. We will expand our interactions to real world situations and I think "Augmented Reality" is a good encompassing phrase. Currently, GUIs deal with representations of things (e.g., a file icon represents a document). The next generation will deal with the manipulation of real world artifacts. Within the GUI desktop, interface elements will continue to look more realistic (e.g., like VCR transport controls for playing DVDs)

Provide transparent transactions. Our transactions with the computer are very explicit and everyone around us knows when we are interacting with a computer. In the future, we should have the capability to provide transparent transactions. That is, other people observing us are unaware that we are performing transactions (e.g., looking up a person's

name while having a conversation with the person). In terms of augmented reality, it is also possible to have transparent transactions with a computer while manipulating physical artifacts (e.g., an RFID tag embedded in our clothes signals when we are in a store).

No more storage management. Why do I have to keep managing my data (moving it between computers and devices, conducting backups, getting more storage, moving my data when I get a new system)? Perhaps there can be a central repository on the net (say iron mountain) that manages, stores and caches my data with an unlimited storage capacity. While this would be handy for a GUI based system, with an augmented reality system, I doubt we want to have to deal with storage issues. Instead the reality-based UI concept will focus on issues of *ownership* of an object and if I currently have an object *present* with me.

More pleasurable rather than efficient articulation. For explicit, human-computer dialogs, next generation interfaces should focus on defining pleasurable instead of solely efficient designs. For example, much of the research effort has been in designing efficient interaction techniques (e.g., reduce pointing time, reduce number of button clicks). This may be akin to the progression of design and overall experience for the automobile. The original automobile was purely functional. Slowly it evolved into offering luxury finishes, smooth manual gear shifting with high performance suspension systems.

These last three are more targeted towards existing GUI based systems:

GUI: Move from ugly aesthetics towards cinematic visual experiences. On traditional desktop systems, most of the visual experience is quite poor. Shifting towards a more cinematic visual experience is important when dealing with people who are more casual users. The Apple Mac OS X operating system is a good example here of shifting towards more cinematically rich visual experiences. For the augmented reality systems, the analogy does not hold up as strongly – it means shifting from real world objects and physical environments to highly stylized environments (e.g., theatrical stages or Disney world) and desirable artifacts (e.g., iPod nano).

GUI: Tackle the huge learnability of complex software systems. Many applications are very difficult to learn from an initial “walk-up and use” situation as well as providing continual learning and improvement for the user. How can we provide better ways of finding and presenting this knowledge? One way may be to use video-based examples.

GUI: video friendly interactions. For GUI based systems, our future designs must handle video as a core element. We need to be able to go beyond the idea of creating, editing, or just watching a video. I think the time has come where video can be used as an input stream such that the content is analyzed, objects are recognized, tracked and selectable with embedded links to augmented content.

Target a task or activity

In all of this analysis, we first must be guided by determine what people want to do. Next we need to determine if and how we can improve upon their desired activity by introducing computational artifacts. Only then can we begin to design and determine the proper input/output devices, computer infrastructure and preferred interaction model.

Here are my favorite samples of tasks or activity for the next generation:

- As we may think (“memex”) by Vannevar Bush. His vision was to augment intelligence by providing a memory extender.
- Retrieve and display a person’s name when you meet them.
- Commerce. The ultimate killer application. Shopping occurs in the real world with real artifacts and is ubiquitous.

Real World Interaction (RWI)

With digitally augmented real world interactions, there are a number of determinations users will conduct when interacting with the digitally augmented objects. These properties include the :

- (1) *Scope of operability*. What environments does this object operate in? For example, it only works when it is on a sensing table or above ground.
- (2) *Location sensing*. The object has an awareness of it’s physical location (such as GPS) and reacts to events based on locations.
- (3) *Sharability*. Will the information or state of the digital object be transmitted to other elements or users in the interactive system for collaborative use?
- (4) *Online or Offline states*. Does the digital object have access to the internet, cell phone network or a personal PC network? Secondly, if the digital object has network capabilities, is it currently connected or disconnected.
- (5) *Data flow*. Does the digital object represent a “push” technology or a “pull”? A push technology is one that if an event happens with the object, the results are manifested at a different location. A pull technology is one that if an event happens with the object, the object reacts and has added information or value immediately available (often manifested with visual display or feedback).
- (6) *Anatomy of a transaction*. What signals the start and end of a transaction? What is the duration of a transaction? Some transaction could last days or months, etc.
- (7) *Persistence of transaction*. Users need to determine how lasting the effect of the interaction will be – how long with this information be stored and relevant for future interactions.
- (8) *Privacy of transaction*. Who can access information retained by the digital object or the transaction log of interactions the digital object has participated in and possibly transmitted to other devices on a network?

All of these descriptions have an implicit assumption that a physical artifact serves as a container or conduit for some digital artifacts. Consequently, the physical characteristics of the physical artifact will determine the “chunking” and phrasing of the combined physical and digital interaction.

Sample technology for discussion - Electronic Shelf Labels

1) NCR RealPrice (<http://www.ncr.com/en/products/hardware/esl.htm>)

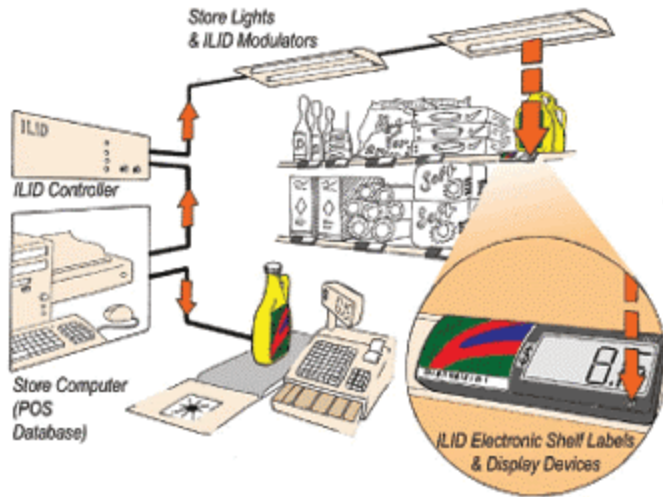
These electronic shelf labels allow prices to be changed dynamically and almost instantly throughout the store by communicating with a wireless base station. They are a good example of augmenting the real world with digital computation and displays.



2) ILID Fujitsu¹

“ILID is a unique technology that uses a building’s existing lighting system to carry information to electronic display devices such as electronic shelf price labels. The technology provides a simple low-cost means of transmitting data from a central computer to a large population of receiving devices.

¹ Excerpt taken from ILID Fujitsu web page:
(<http://www.fujitsu.com/au/services/industries/retail/solutions/esl/>)



ILID Electronic Shelf Labels are fully portable, self-powered and operate in adverse environments such as freezers and fresh produce areas. Because light is used to transmit price changes and other information, there is no need for radio frequency or infrared transmission hardware. The devices themselves can be located almost anywhere – they do not need to be in a direct line of sight as they will work with reflected light or even light transmitted through a translucent object.

The ILID system revolutionizes the way both retailers and customers view product pricing. For retailers it becomes an effective pricing tool by eliminating time-consuming price-marking and by simplifying product promotion. For customers it guarantees that the price on the shelf will be identical to that paid at the checkout. Also, other information such as savings on discounted prices can be quickly and easily displayed.”

Interaction Design for Active Bodies: Two Themes

Scott R. Klemmer, Björn Hartmann

Stanford University HCI Group
Computer Science Department
Stanford, CA 94305-9035, USA
{srk, bjoern}@cs.stanford.edu

Leila Takayama

Stanford University CHIME Lab
Communication Department
Stanford, CA 94305-2050, USA
takayama@stanford.edu

ABSTRACT

Our physical bodies play a central role in shaping human *experience* in the world, *understanding* of the world, and *interactions* in the world. Reality-based interfaces promise to better leverage the capabilities of our bodies in human-computer interaction. In two themes, *thinking through doing* and *thick practice*, we introduce aspects of human bodily engagement with the world that designers of interactive systems should attend to.

INTRODUCTION

One of the most sweeping (and unintended) transformations that the desktop computing paradigm has brought about is the extent to which the physical performance of work has homogenized. With a keyboard and mouse interface, the use of our bodies for writing a paper is the same as for editing photographs. And playing music. And communicating with friends and family. Reality-based interfaces such as tangible UIs and augmented reality applications promise to reintroduce our bodies' capabilities for rich and subtle expression into human-computer interaction. For a combination of computation with real-world actions and artifacts to be successful though, a thorough understanding of what each can offer is needed first.

In our current work, we are shifting focus from the technologies of embodied interaction to the participating human bodies themselves. New design considerations emerge when our bodies are understood as more than just "Baby Bubbleheads" (*i.e.*, the Model Human Processor [3]). By drawing on psychology, sociology, and philosophy, we seek to elucidate how human bodies shape *experience* in the world, *understanding* of the world, and *interactions* in the world. We hope to contribute to the discussion about the future of HCI a set of interaction design considerations for active bodies and a language that can both stimulate ideation and give guidelines for evaluation of reality-based interfaces.

We have identified *five themes* that we believe are particularly salient for designing and evaluating interactive systems. The first, *thinking through doing*, describes how thought (mind) and action (body) are deeply integrated and how they co-produce learning and reasoning. The second, *performance*, describes the rich actions our bodies are capable of, and how physical action can be both faster and

more nuanced than symbolic cognition. *Visibility* describes the role that observable actions on physical artifacts play in collaboration and cooperation. *Risk* explores how the uncertainty and risk of physical co-presence shapes interpersonal and human-computer interactions. The final theme, *thickness of practice*, argues that because the pursuit of digital verisimilitude is more difficult than it might seem, embodied interaction that gives primacy to the physical world may be a more prudent path. In this paper, we provide an overview of two of the five themes and show how both are concerned with the tradeoffs between physical *presence* and virtual *representation*.

THINKING THROUGH DOING

Direct physical interaction with the world is a key constituting factor of cognitive development during childhood. The importance of physical action as an active component of cognition extends throughout life. In this section, we review the connection between thinking and doing as uncovered by educational theorists, gesture researchers, and cognitive scientists. Cumulatively, their empirical work point towards a common nexus of perception, cognition, and action.

Learning through doing

Being able to move around in the world and interact with pieces of the world enables learning in ways that reading books and listening to words do not. Jean Piaget [18] posited that cognitive structuring requires both physical and mental activity. Particularly for infants in the sensorimotor stage of development, *physical interaction in the world facilitates cognitive development* [11]. In this very basic sense, humans learn about the world and its properties by interacting within it.

Pedagogies such as the Montessori method [17] employ bodily engagement with physical objects to facilitate active learning. The use of tangible manipulatives has been shown to improve elementary school student understanding of mathematical concepts. Such educational methods nicely leverage the bodily basis of mathematical concepts for learning [14]. Physical reasoning can also play an important role in professional and higher education. An example is the Illuminating Light interface [23], which enables students to create light reflection simulations by placing tangible reflectors on a tabletop surface with virtual projection (see Figure 1).

The Role of Gesture

Just as moving about in the world helps learning, gesture plays a role in pre-linguistic communication for babies [9] and aids cognition and fully linguistic communication for adults. From studies of gesturing in face-to-face interactions, we know that people use gesture to conceptually plan speech production [2] and to communicate thoughts that are not easily verbalized [4].

While gesturing is normally thought of as having a purely communicative function, many studies suggest that gesture may also play a helpful role for the speaker; for example, even congenitally blind children gesture [10]. For both adults and children, gesturing has been shown to lighten their cognitive workload [6].

Beyond continuing work in gesture recognition and synthesis, a less obvious but no less important point is that systems that constrain gesturing (*e.g.*, by having your hands stuck on a keyboard) are likely to hinder the user's thinking and communication abilities. Consider telephones: we have seen shifts from corded phones to cordless phones to mobile phones and mobile phone headsets. Experimental studies demonstrated that more physical mobility increased user creativity and disclosure of personal information in microphone use [24]. The increasingly freeing form factors of telephone technologies help users think and communicate by getting out of the their way.

Epistemic Action

Bodily engagement with physical and virtual *environments* constitutes another important aspect of cognitive work. We are familiar with people leaving keys or notes for themselves in strategic locations so that they serve as reminders in the future.

Distinguishing *pragmatic* action—manipulating artifacts to directly accomplish a task — from *epistemic* action — manipulating artifacts to better understand the task's context [12], provides an interpretation for such behavior. A laboratory studies of Tetris gameplay [15] illustrates the importance of epistemic action. One might expect that the predominant task in Tetris is piece movement with the *pragmatic* effect of aligning the piece with the optimal available space. However, contrary to intuitions, the proportion of shape rotation that are later undone by backtracking increases (not decreases) with increasing Tetris-playing skill levels: players manipulate pieces to understand what they are and how different options would work.

An analogous example of how a user's environment may be appropriated to facilitate cognitive work that would otherwise have to be done entirely in the head is how people move lettered tiles around in the game of Scrabble to help explore possible word formations [16].

Thinking through prototyping

Iterative design practices provide yet another perspective on the importance of concrete, artifact-centered action in the

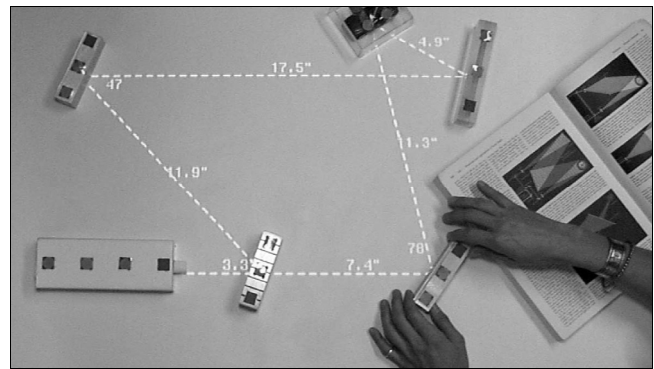


Figure 1 The tangible Illuminating Light workbench lets students learn about optical systems by designing them.

world to aid thought. *Reflective practice*, the framing and evaluation of a design challenge by *working it through*, rather than just *thinking it through*, points out that physical action and cognition are interconnected [19]. The production of concrete prototypes provides the crucial element of surprise, unexpected realizations that the designer could not have arrived at without producing a concrete manifestation of her ideas.

The *backtalk* that artifacts provide helps uncover problems or generate suggestions for new designs. Prototypes thus become the “essential medium for information, interaction, integration, and collaboration” [20]. Beyond backtalk, creating intermediate tangible artifacts allows for expression of tacit knowledge. It also facilitates communication within a design team, with clients, or users, by providing a concrete anchor around which discussion can occur. Prototypes then present us with a different kind of embodiment: they themselves embody design ideas or specifications, render them concrete and, in doing so, inform the designer's thinking.

THICK PRACTICE

It may seem a platitude, but it is worth repeating that, “if technology is to provide an advantage, the correspondence to the real world must break down at some point” [7]. Interaction design is simultaneously drawn in two directions. First, the promise of new technology is that it provides previously unavailable functionality. Second, in designing almost any new technology, one is drawing on existing human understanding of the world. In the creation of the new, much technology formalizes some aspects of a work practice. However, a complete formalization of practice is an elusive goal — important aspects of real-world work often remain invisible to system designers and are subsequently denied by the introduction of new technology (*cf.* [21]).

This section argues that interfaces that *are the real world* can obviate many of the difficulties of attempting to *model all of the salient characteristics of a work process as practiced*. This argument builds on Weiser's exhortation to design for “embodied virtuality” rather than virtual reality [25]. Designing interactions that *are the real world* instead

of ones that simulate or replicate it hedges against simulacra that have neglected an important practice.

A system that respects the primacy of physical practice is Final Scratch [1], which provides access to digital music for DJs through specially encoded vinyl records (see Figure 2). These vinyl records contain a time code instead of an audio signal. The system interposes a soundcard into the signal path between turntable and mixer to pick up the time code, link it to playback of digital music files on a laptop computer, and return that audio signal to the inputs of the mixing console. FinalScratch affords continuity of practice—skills acquired over years of practice still apply since the physical interface has not changed. DJs regard it as superior to competing digital control products (such as CD players with jog dials) because digital controls do not provide the sensory richness or the nuance of manipulation offered by the “real thing.”

Books with Voices [13] augments paper transcripts of oral histories with barcodes printed alongside the text. These can be scanned by a PDA to access original audio recordings. In retaining the printed paper page as the primary artifact around which interaction is structured, the system embraces existing reading practices, grafting digital media onto them.

The project of technology is the creation of increasingly malleable materials, and computation is perhaps the most malleable created so far. Given the techno-utopian ideology of computer science, it can seem heretical to suggest that one should undertake a project other than replacing the physical world. Clearly, the digital world can provide advantages. To temper that, we argue that *because there is so much benefit in the physical world*, we should take great care before unreflectively replacing it. More precisely, from a design perspective, solutions that carefully integrate the physical and digital worlds—leaving the physical world alone to the extent possible—are likely to be more successful by admitting the improvisations of practice that the physical world offers.

REPRESENTATION AND PRESENCE

A shared problematic underlies the discussion of both preceding themes: which parts of a reality-based interface should be assigned to the virtual realm and which to the physical realm in order to best combine the affordances of both? Is it more advantageous to start with a symbolic model and reify it by providing physical handles that *represent* objects in the virtual world, as in the Illuminating Light workbench? Or should one instead focus on preserving the *presence* of existing artifacts and enhance them with virtual functionality, as in the FinalScratch system?

The promise and challenge of reality-based interfaces lies in the fact that they simultaneously represent an underlying digital model (they stand for something else), while they are also present (in immediate existence) in the actor's world.



Figure 2 Final Scratch: encoded vinyl for digital music.

The crux for interaction design then is deciding how to tie properties arising out of the presence of the interface to properties of the digital information being represented and vice versa.

Tangibility offers both direct familiarity and a set of common metaphors to leverage in interaction. But some mappings between the physical and the virtual work, while others clearly don't. An example of an interactive system that successfully leverages our familiarity with everyday physics is the automotive drive-by-wire system that uses force feedback to alter driver perceptions of the road [22]. It discourages lane drifting by exerting forces on the wheel such that the driver has the impression that the driving lane is shaped like a shallow bathtub. On the other hand, a professor of computational logic demonstrates to his class how propositional inference can be performed in physical space using stackable boxes—however, in this case a purely symbolic representation of the problem on paper will yield a solution much more quickly.

So what guidelines can we give to designers? Djajadiningrat *et al.* [5] cite that many tangible user interface researchers have gravitated towards “natural” mapping, which are intuitively understandable because of their straightforward identification of virtual attributes with physical ones. *I.e.*, objects are positioned in virtual space by moving physical handles in physical space. We share their concern that these identifications are only possible for a restricted domain of systems—how do we interact with symbolic information that does not have an obvious physical equivalent? While we cannot provide ready operationalizable guidelines, we believe the problem of developing successful representation strategies is a fruitful area for further research.

CONCLUSION

Hollan and Stornetta argue that the impact of electronic media should not be measured by how well they can approximate the affordances of face-to-face interaction, but rather how they can surpass the constraints of co-presence and co-location to offer value that motivate their use even if face-to-face communication is available [8]. Similarly, we should not just strive to approach the affordances of

tangibility in our interfaces and interactions, but to go beyond what mere form offers. A better understanding of the affordances of physicality and concreteness for the design of interactive systems is a first step in this direction. We believe this work will be of value both *generatively*—helping designers come up with new solutions—and for *evaluation*—providing a rich set of axes for analyzing the benefits of systems.

REFERENCES

- 1 Final Scratch, 2005. Stanton Magnetics: Hollywood, FL. <http://www.finalscratch.com>
- 2 Alibali, M.W., S. Kita, and A. Young. Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes* 15. pp. 593-613, 2000.
- 3 Card, Stuart K., Thomas P. Moran, and Allen Newell, *The Psychology of Human-Computer Interaction*. Hillsdale: Lawrence Erlbaum. 469 pp. 1983.
- 4 Church, R. B. and S. Goldin-Meadow. The mismatch between gesture and speech as an index of transitional knowledge. *Cognition* 23. pp. 43-71, 1986.
- 5 Djajadiningrat, Tom, Stephan Wensveen, Joep Frens, and Kees Overbeeke. Tangible products: redressing the balance between appearance and action. *Pers Ubiquit Comp* 8: Springer-Verlag. pp. 294-309, 2004.
- 6 Goldin-Meadow, S., H. Nusbaum, S. D. Delly, and S. Wagner. Explaining math: Gesturing lightens the load. *Psychological Science* 12(6). pp. 516-22, 1991.
- 7 Grudin, Jonathan. The case against user interface consistency. *Commun. ACM* 32(10). pp. 1164-73, 1989.
- 8 Hollan, Jim and Scott Stornetta. Beyond being there. In *Proceedings of Human factors in computing systems*. pp. 119-25, 1992.
- 9 Iverson, J. and S. Goldin-Meadow. Gesture paves the way for language development. *Psychological Science* 16(5). pp. 367-71, 2005.
- 10 Iverson, J. and S. Goldin-Meadow. What's communication got to do with it: Gesture in blind children. *Developmental Psychology* 33. pp. 453-67, 1997.
- 11 Kermoian, R. and J. J. Campos. Locomotor experience: A facilitator of spatial cognitive development. *Child Development* 59(4). pp. 908-17, 1988.
- 12 Kirsh, D. and P. Maglio. On distinguishing epistemic from pragmatic action. *Cognitive Science* 18. pp. 513-49, 1994.
- 13 Klemmer, Scott R., Jamey Graham, Gregory J. Wolff, and James A. Landay. *Books with Voices: Paper Transcripts as a Tangible Interface to Oral Histories*. CHI: ACM Conference on Human Factors in Computing Systems, CHI Letters 5(1): ACM Press. pp. 89-96, 2003.
- 14 Lakoff, George and Rafael E. Núñez, *Where Mathematics Comes From*. New York, NY: Basic Bookspp. 2000.
- 15 Maglio, Paul and David Kirsh. Epistemic action increases with skill. In *Proceedings of Cognitive Science Society*: LEA, 1996.
- 16 Maglio, Paul, Teenie Matlock, D. Raphaely, B. Chernicky, and David Kirsh. Interactive skill in Scrabble. In *Proceedings of Twenty-first annual conference of the cognitive science society*: Lawrence Erlbaum, 1999.
- 17 Montessori, Maria, *The Montessori method*. New York: Frederick A. Stokes Co. 376 pp. 1964.
- 18 Piaget, Jean, *The origins of intelligence in children*. New York: International University Press. 419 pp. 1952.
- 19 Schön, Donald A., *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books. 1983.
- 20 Schrage, Michael, *Cultures of Prototyping*, in *Bringing Design to Software*, T. Winograd, Editor. ACM Press: New York, 1996.
- 21 Suchman, Lucy. Do Categories Have Politics? The Language/Action Perspective Reconsidered. *Computer Supported Collaborative Work* (2). pp. 177-90, 1994.
- 22 Switkes, Joshua P., Eric J. Rossetter, Ian A. Coe, and J. Christian Gerdes. Handwheel force feedback for lanekeeping assistance: Combined dynamics and stability. In *Proceedings of International symposium on advanced vehicle control*. Arnhem, Netherlands. pp. 47-52, 2004.
- 23 Underkoffler, John and Hiroshi Ishii, *Illuminating light: an optical design tool with a luminous-tangible interface*, in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1998, ACM Press/Addison-Wesley Publishing Co.: Los Angeles, California, United States.
- 24 Wang, Qianying and Clifford Nass. Less Visible and Wireless: Two experiments on the effects of microphone type on users' performance and perception. In *Proceedings of Human factors in computing systems*. Portland, Oregon, 2005.
- 25 Weiser, Mark. The Computer for the 21st Century. *Scientific American* 265. pp. 94-104, 1991.

B. Broader Perspectives, Psychological Considerations

B.1 Opportunities and Challenges for HCI Design and Research

Jim Hollan University of California, San Diego

B.2 Matching Human Abilities and Limitations: The Basis for Better Future User Interfaces

David Kieras University of Michigan

B.3 Coping with Information Overload in the New Interface Era

Steven M. Drucker Microsoft Research, Redmond, WA

B.4 Some Musings On Building and Showing Future UI Prototypes

Andrew D. Wilson Microsoft Research, Redmond, WA

Opportunities and Challenges for HCI Design and Research*

Jim Hollan

Distributed Cognition and Human-Computer Interaction Lab
Department of Cognitive Science, University of California, San Diego
(Email: hollan@cogsci.ucsd.edu Web: <http://hci.ucsd.edu/hollan>)

The decreasing cost, increasing power, and miniaturization of commodity computing devices continues to enable ever wider application of computation. The history of computing hardware can be summarized as a progression from a focus on low-level components towards integration on larger and larger scales, from vacuum tubes and transistors to LSI, VLSI, chipsets, personal computers, LANs, WANs, and now the global internet. Today there is also a move in the opposite direction as the monolithic “computer” is being unbundled into fragmentary components. Currently these components are reemerging coalesced in a multitude of forms, ranging from the rapidly evolving cell phone to novel embeddings in an expanding array of everyday objects.

But this machine-oriented view is far too narrow, because progression on the human side has been at least as dramatic and important, from isolated single users, to timesharing, to groupware and support for community activities, to the frontier where ubiquitous, wireless, context-aware, multi-modal, mobile computing enables currently unknown social possibilities. These radical changes and associated further imbuing of our professional and personal activities with computation present enormous new challenges for HCI design and research.

These challenges make this a most propitious time to convene a workshop to discuss the next generation of HCI and, I would argue, to take stock of where the field is, how we got here, and how best to proceed. Understanding computationally-based systems and ensuring their future design respects human needs and abilities are intellectual challenges of singular importance. These challenges involve not only complex theoretical and methodological issues of how to design effective computationally-based representations and mechanisms of interaction and communication but also confront complex social, cultural, and

political issues such as those of privacy, control of attention, and ownership of information.

As with many challenges there is also opportunity. For example, the same forces leading to the unbundling of the monolithic computer and ubiquitous computing are also changing the nature and richness of data we can collect about human activities. In the history of science, the appearance of new technologies for collecting or analyzing data has frequently spawned rapid scientific advancement. The human genome project, for example, would have been unfathomably complex without automatic DNA sequencing technology. In the present case, a new generation of inexpensive digital recording devices is revolutionizing data collection for studying human activity, extending it to situations that have not typically been accessible and enabling examination of the fine detail of action captured in meaningful settings.

There are myriad important issues that would benefit from workshop discussion. Topics range from research strategy questions (e.g., tradeoffs between small-science and big-science approaches, the importance of open source for HCI, etc.) to how to better bridge the cultural chasms between the disciplines involved in HCI to support the interdisciplinary graduate programs so urgently needed to train the next generation of human-computer interaction researchers.

In this position paper I briefly comment on the proposed discussion starting point for the workshop, describe opportunities arising from new facilities to capture and share detailed records of activity, and discuss the importance of capitalizing on these opportunities as one component in elaborating a research agenda for next generation HCI research.

Proposed Starting Point: Reality-Based Interaction

The proposed starting point for the workshop is the notion of “natural or realistic interfaces or reality-based interfaces” that “increasingly draw their strength from ex-

*Position paper for CHI 2006 Workshop: What is the Next Generation of Human-Computer Interaction?

plotting users' pre-existing skills and expectations from the real world rather than trained computer skills."

It is argued that

A unifying characteristic for much next generation HCI may thus be that it increasingly taps into the users' abilities and pre-existing knowledge. Direct manipulation moved user interfaces toward more realistic interaction with the computer; next generation reality-based interfaces push further in this direction, increasing the realism of the interface objects and allowing the user to interact even more directly with them. ...

Finally, work that helped define the GUI generation is a model for us. Shneiderman took a set of disparate new user interfaces and unified them through their common characteristics [12]. Hutchins, Hollan, and Norman then explained their power and success of these interfaces with a theoretical framework [9]. Our hope is to take a first step in that direction for the emerging generation.

Although our paper is now over two decades old, I still think the distinction it makes between two fundamentally different metaphors (conversation versus model world) for interface design has sufficient currency to warrant further discussion. I would also note that as we argued then, and as Rob reiterates in the call, simple mimicking of the real world is obviously insufficient. It would be valuable to discuss how to draw on users' knowledge of interacting with the world without being constrained to merely mimicking those interactions. This was certainly a central motivation for our early Pad++ zoomable interface work [2] and continues to motivate our current Dynapad work [1]. The goal has been to exploit people's wonderful ability to imbue space with meaning but remove or relax some of the restrictions of physical space.

A Permeable Boundary between Physical and Digital

A related idea worthy of workshop attention is that of making the boundary between the physical and the digital permeable. One example is Rekimoto's work on augmented surfaces [11] in which one can move windows from the virtual world of the display onto a physical table or from the physical world into the virtual world. Another particularly compelling example is Guimbretièrè's innovative work on Paper Augmented Digital Documents (PADD) [6]. In this approach, the digital and paper world are put on equal footing: paper and computers are simply

different ways to interact with documents during their lifecycle. When paper affordances are needed, a document is retrieved from the database and printed. The printer acts as a normal printer but adds a pen-readable pattern to each document. Using a digital pen, the document can now be marked like a normal paper document. The strokes collected by the pen are combined with the digital version of the document. The resulting augmented document can then be edited, shared, archived, or participate in further cycles between the paper and digital worlds.

Distributed Cognition and Embodied Interaction

In terms of theory, it is important to note that there is currently a shift in cognitive science away from the notion of cognition as a property of isolated individuals and toward a view of cognition as a property of larger social and technical systems [8, 3, 7]. This extends the reach of cognition to encompass interactions between people as well as interactions with resources in the environment. As a consequence, the human body and the material world take on central rather than peripheral roles. As Andy Clark put it, "To thus take the body and world seriously is to invite an emergentist perspective on many key phenomena - to see adaptive success as inhering as much in the complex interactions among body, world, and brain as in the inner processes bounded by the skin and skull." [3] This new perspective on cognition is emerging from the fields of distributed cognition [8, 5] and embodied interaction [3, 10, 4]. I think it promises to provide an intellectual basis for a paradigm shift in thinking about the dynamics of interaction; one that takes material and social structures to be elements of cognitive systems and views on-going activity as a continually renegotiated emergent product of interaction. This paradigm shift promises to be vitally important for the next generation of HCI.

A Critical Opportunity

In addition to developments in theory, recent advances in digital technology present unprecedented opportunities for the capture, storage, and sharing of activity data. This is important because to understand the dynamics of human activity in sufficient detail to be able to make informed design choices, we must first understand the full context of those activities and this can only be accomplished by recording and analyzing detailed data of real-world behavior. I argue that the ability to capture and share such data has created a critical moment in the practice and scope of research. I think our field has an exciting opportunity to capitalize on this by developing a shared infrastructure to assist in integrating the theoretical and analytic frameworks required to build a stronger scientific base for HCI research.

While detailed activity data are certainly needed to advance science, more data cannot be the whole answer, since many researchers already feel that they are drowning in data. Data without appropriate theoretical and analytical frameworks do not lead to scientific advances. Another obstacle to fully capitalizing on the opportunity provided by digital recording devices is the huge time investment required for analysis using current methods. In addition, we need to understand how to coordinate analyses focused on different levels so as to profit from the theoretical perspectives of multiple disciplines.

Reducing the Cost of Analysis

Today the high labor cost of analyzing rich activity data leads to haphazard and incomplete analyses or, all too commonly, to no analysis at all of much of the data. Even dataset navigation is cumbersome. Data records are chosen for analysis because of recording quality, interesting phenomena, and interaction density—producing a haphazard sampling of the recorded set. Good researchers have a nose for good data, but with a tendency to focus on small segments of the record that contain “interesting” behavior, analyze them intensively, and then move on to the next project.

When analysis is so costly, few analyses can be done—so datasets are severely underutilized—and researchers come to have a large investment in the chosen data segments. Since each analysis may appear as an isolated case study, it can be difficult to know how common the observed phenomena may be. Larger patterns and contradictory cases can easily go unnoticed. Well-known human confirmation biases can affect the quality of the science when each analysis requires so much effort.

Increasing the Power of Analysis

Researchers studying activity in each of the disciplines involved in HCI are beginning to appreciate the importance of understanding patterns that emerge from the interactions of multiple dynamically linked elements. Such interactive patterns may be invisible to approaches that decompose activity into the more or less independent components created by historical distinctions among disciplines. This is why a multidisciplinary approach is necessary. But tools that match this multidisciplinary vision are also needed. Without them we cannot address questions such as: How shall we come to see patterns in interactions among the many modes of multimodal activity? How can we approach the long-standing problem of the relations between patterns that emerge on larger time scales with the short time scale mechanisms that give rise to them?

The richly multimodal nature of real-world activity makes analysis difficult. A common strategy has been to

focus on a single aspect of behavior or a single modality, and to look for patterns there. However, the causal factors that explain the patterns seen in any one modality may lie in the patterns of other modalities. In fact, recent theorizing suggests that activity unfolds in a complex system of mutual causality. Analysis may still be based on decomposition of the activity, as long as there is a way to put the pieces back together again. That is, as long as there is a way to visualize the relations among the many components of multimodal activity.

In addition, as activity unfolds in time, describable patterns that take place on the scale of milliseconds are located in the context of other describable patterns that display regularities on the scale of seconds. Those patterns in turn are typically embedded in culturally meaningful activities whose structure is described on the scale of minutes or hours. Patterns at larger time scales are created by and form the context for patterns at shorter time scales.

The high cost of performing analyses on data that represent real-world activity means not only that too few analyses are conducted, but that analyses tend not to be shared. Most often when results of an analysis are published, neither the activity of doing the analysis, nor the procedure that was used are shared. This creates a situation in which most analyses are idiosyncratic and possibly non-replicable.

I will argue that a primary component needed to advance HCI research is an open software infrastructure to support capture of increasingly rich activity data, speed and improve analysis, and facilitate sharing with a wider research community.

A Final Thought

The current state of HCI owes a deep intellectual debt to research at Xerox Parc and the development of the Alto. Researchers in the early days at Parc designed a new computational environment with crucial attention to the importance of the interface. Is it not time to do this again? Might not a major limitation in building the next generation of HCI be that we are building additional layers on top of operating systems themselves designed with virtually no conception of the kinds of services needed today?

References

- [1] BAUER, D., FASTREZ, P., AND HOLLAN, J. Computationally-enriched piles for managing digital photo collections. In *Proceedings IEEE Symposium on Visual Languages and Human-Centric Computing* (2004).

- [2] BEDERSON, B. B., AND HOLLAN, J. D. Pad++: A zooming graphical interface for exploring alternate interface physics. In *Proceedings of the ACM Symposium on User Interface Software and Technology* (1994), pp. 17–26.
- [3] CLARK, A. *Being There: Putting Brain, Body and World Together Again*. MIT Press, 1997.
- [4] DOURISH, P. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, 2001.
- [5] GOODWIN, C. Gesture, aphasia, and interaction. In *Language and Gesture* (2000), D. McNeill, Ed., Cambridge University Press.
- [6] GUIMBRETIERE, F. Paper augmented digital documents. In *Proceedings of the ACM Symposium on User Interface Software Technology* (2003), pp. 51–60.
- [7] HOLLAN, J., HUTCHINS, E., AND KIRSH, D. Distributed cognition: Toward a new theoretical foundation for human-computer interaction research. *ACM Transactions on Human-Computer Interaction* (2000), 174–196.
- [8] HUTCHINS, E. *Cognition in the Wild*. MIT Press, Cambridge, 1995.
- [9] HUTCHINS, E. L., HOLLAN, J. D., AND NORMAN, D. A. Direct manipulation interfaces. *Human-Computer Interaction 1* (1985), 311–338.
- [10] NUNEZ, R. Could the future taste purple? reclaiming mind, body and cognition. *Journal of Consciousness Studies* 6 (1999), 41–60.
- [11] REKIMOTO, J., AND SAITOH, M. Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 1999), ACM Press, pp. 378–385.
- [12] SHNEIDERMAN, B. Direct manipulation: A step beyond programming languages. *IEEE Computer* 16, 8 (1983), 57–69.

Matching Human Abilities and Limitations: The Basis for Better Future User Interfaces

David Kieras

University of Michigan

What would make good future user interfaces? Ones that work well with human abilities and limitations.

This approach is not the same as reality-based interfaces - these may or may not help because first, non-computer tasks do not necessarily relate to computer tasks, and second, not all current real-world tasks or situations are well-suited for humans, making them poor models for interfaces.

To elaborate, the reality we evolved for was hunter-gatherer culture. This implies certain powerful abilities, but they may not be manifest in an arbitrarily selected current real task. Manually controlling a glass and steel box moving in a crowd of similar boxes at 70 mph without getting killed is a remarkable skill that most of us can perform pretty well, But this does not mean that it is "natural" and thus a basis for a new interface technology - for example, it is very hard to learn and requires (or should require) great concentration.

Another problem with reality-based interfaces is that the reality of tasks is constantly changing with technology changes. For example, "cut and paste" used to refer to an actual physical activity in document preparation, but not any more. Changing a cursor position on a vertical display by changing the relative position of a small box on a nearby horizontal surface was not a real world activity before, but it is now. Basing future design on the changing nature of "reality" is not obviously a good idea.

A better way to see what could be done better in the future is to examine what made GUIs better at least for time than command language interfaces (CLIs), and then extrapolate some of the lessons learned to the future. The qualities of GUIs have not been fully appreciated; in fact, it would be better to say that GUIs have been misappreciated; it is time for a re-interpretation of the GUI concept.

As a first step, a brief survey of some relevant aspects of human perception, cognition and action will set the stage. One perceptual-motor ability is that visual search is extremely efficient; people are able to use peripherally-visible properties of objects to guide visual search. In addition, humans have good memory for stable visual locations.

Another human ability is the ease of aimed movements to visually identified objects. This has recently been shown to be immune to Hick's law effects, at least as long as the stimuli and

responses are spatially compatible. But even keystrokes are heavily influenced by spatial compatibility factors; classic research shows that in a simple choice reaction task using a row of response buttons, the human first identifies what response to make, then determines where to make it; then the keystroke is made.

Another set of perceptual abilities is the omnidirectionality, parallelizability, and spatiality of the auditory channel. Localized sound can be used to designate spatial locations, but the auditory system does not require a particular orientation to perceive a sound. Sound can be responded to simultaneously with other channels. However, perception of external speech can be interfered with by simultaneous or subvocalized speech and vice-versa.

A familiar aspect of human memory often cited in the interface area is that people can recognize information better than they can recall it.

People have several forms of working memory; the best explored is verbal working memory, which is usually represented in speech-based forms. A little-appreciated problem with speech-based interfaces is that speech input or speech production interferes with the use of verbal working memory.

A frequently overlooked, but well-documented form of working memory is task working memory: people have an extremely large and reliable working memory for information in tasks in which they are highly skilled. This means that systems and interfaces in complex task domain have rather different memory support requirements depending on the skill of the user.

A final area concerns human ability to acquire procedural knowledge. Humans have a powerful ability to easily learn procedures as long as they are simple and consistent, and these aspects of learnability can be quantified in terms of procedural knowledge representations. These results lead to a direct characterization of interface consistency: Interfaces that have consistent procedures can be learned much more easily because the learner can simply re-use the shared procedural knowledge. This is the classic common elements theory of transfer of learning in modern theoretical garb.

However, it is known that transfer from previous skills can be weak unless the common elements are quite similar. A common result is that transfer of training fails to appear in many educational contexts - e.g. learning Latin will not help you learn arithmetic. This argues that expecting skills to be transportable from one domain to another is dubious, especially if the relationship is not obvious to the learner.

In spite of this limitation, people can do a good job of learning how to use an unfamiliar system if their knowledge of the system (or "mental model") supports inferring the needed procedures - which is possible only if the interface has a non-arbitrary relationship to the system structure or knowledge domain. For example, we can usually figure out how to operate an unfamiliar digital clock because such devices have similar mechanisms and shared economic constraints on the technology cause them to have controls and displays that interface to these mechanisms in

certain typical ways. At the other extreme, we can usually figure out how to navigate through menus or web pages based on meaning of the words present in the interface and our knowledge of the domain.

The good and bad points of current GUI methodology in these terms can now be briefly presented. First, the good points are:

1. The visual presentation and point-and-click actions on visible objects takes advantage of the great ease and efficiency of human perceptual-motor abilities.
2. Heavy reliance on superior recognition memory.
3. Conventions leading to stable spatial positions of commonly accessed objects, eliminating the need for visual search.
4. Conventions leading to a standard set of menu items for common actions across applications, based on common functions and features, greatly reducing the amount of procedural knowledge to be learned. In contrast, CLIs in practice are inconsistent across applications.
5. Pressure to use a small, simple set of procedures since the interface operations are so limited at the perceptual-motor level, consisting only of various combination of mouse points and button/keystroke operations. Although some GUIs have managed to be otherwise, in general this pressure has resulted in simple and consistent procedures for picking commands or options once the correct display is present. In contrast, CLIs typically display little or no such consistency; each command and its options are typically a special case.
6. Using words and graphical presentations often directly related to the task domain.

The bad aspects of current GUIs in these terms are:

1. An inadequate use of visual guidance in icon search. It is difficult to find objects whose locations are not already known. Most serious is that overall color coding is rendered almost useless by the tendency to use detailed multicolored icons.
2. Serious under-utilization of the auditory channel. Except for an occasional “beep”, the auditory channel of almost all computer systems is not used to facilitate task performance.
3. Under-use of speech I/O. Despite repeated attempts, perhaps this is due to incompatibility with verbal working memory or other speech activity.
- 4.. Extreme featuritis of many current GUI applications. This means that the basic procedural simplicity of the GUI is obscured by arbitrary and difficult-to-learn menu

transversals to access features often arranged without regard to their frequency or utility, or even the domain semantics, that as a result can not support inference of the complex procedures. Another aspect is that instead of recognizing on-screen items as relevant, the user must recall a sequence of menu picks, thus relying on a less powerful form of memory. The overall benefit compared to CLIs is remarkably reduced.

As a side note, traditional CLIs were generally far less usable than they could have been. A CLI could also have relatively simple and consistent procedures. If other aspects of the interface were carefully designed, the ease of learning and use of a CLI would be competitive with that of a GUI.

Thus future interfaces can be considerably improved by taking more advantage of what humans are good at, and relying less on what people have trouble with. In particular, future interfaces should

1. Exploit human visual search and spatial abilities better by re-thinking how the interface objects are presented.
2. Use the auditory and speech channels as parallel and possibility spatial I/O channels, while ensuring that interference with verbal memory is minimized.
3. Use existing analysis and modeling techniques to ensure procedural simplicity and consistency.
4. Ensure that the semantics of the domain can support inference of procedures, through the development of techniques for mapping domain semantics to/from an interface (e.g. a combination of cognitive walkthrough and latent semantic analysis).
5. Support the differing memory requirements of tasks by designing interfaces to either provide intensive memory support for the task state, or not interfere with expert use where it is not needed.

Coping with Information Overload in the New Interface Era

Position paper: Steven M. Drucker, Next Media Research Group, Microsoft Research, 12/15/2005

It has been said that there is more information in a single, Sunday, New York Times newspaper than people saw in their entire lifetimes during the Medieval Ages. Newspapers themselves are giving way to new forms of media including television and the Internet. Between instant messaging, email, blogs, RSS feeds, web sites, we have more information instantly available to us at our desktops than ever before. This incredible inundation with information is often an opportunity, but at the same time, finding ways to effectively cope with the information overflow can be a tremendous challenge.

Widespread communication, in the forms of telephony and the internet, account for some of this increase, but in the past this has been primarily in a fixed location (desktop) due to the ‘wires’ necessary for connectivity. With the increase in wireless networks and mobile telephony and the advent of ubiquitous sensing, mobile interaction, and augmented environments, the amount of information available at *any moment* challenges us even further. Under these circumstances, we can’t necessarily use the same standard (often not quite effective) techniques that we’ve evolved for managing that information on the desktop (such as standard sized displays, keyboards and pointing devices). We need to find new and different ways to cope with (and hopefully take advantage of) this surfeit of information.

While it is difficult to anticipate the technological innovations that will occur, there are some areas that we *can* predict with reasonable accuracy. Human capabilities such as sensory modalities, memory, and cognition will *not* improve on their own. Coping with the information will have to come from either automatic methods of filtering the data (I’m lumping these into ‘AI’ techniques) or more effective ways of representing and interacting with the representations (information visualization). Presumably some combination of both of these areas will define the kinds of interfaces we will use in the future.

Furthermore, it is not enough only to look at progress in the technology, but as practitioners in the field of HCI, we also need to clearly understand the tasks that we will be performing in the future. As such, this position paper will start a taxonomy of interaction, devices, human abilities, *and* tasks and how they might be altered by next generation UI technologies.

Human capabilities:

Human capabilities do not change to a large extent, especially in comparison to the rapid progress in technology. Our sensor modalities, our motor systems, and our cognitive systems are essentially the same for the last 50,000 (potentially 1 million) years. With the exception of direct implantation to or recording from the brain, we will likely use these same modalities to perceive information or affect change (Table 1), while this table is not exhaustive, it hopefully presents the kinds of thinking that will promote discussion in the workshop:

| Human sensor modalities | Devices for mobile interaction | Strengths | Weaknesses |
|-------------------------|--------------------------------|-------------------------------------|------------------------|
| Visual: | Mobile displays | Always available, portable, private | Small, power consuming |
| | Expandable displays | (same as above) | Not available yet |

| | | | |
|-----------------------|---------------------------------------|--|---|
| | Projected displays | Potential problems for projection surface | Privacy, power |
| | Head mounted displays | Good resolution, potential for augmented reality | Obtrusive, tracking difficulties for alignment, power, can separate from others |
| | Displays embedded in the environment | Can use all the power necessary, can be large, multiperson | Privacy, potentially expensive |
| Auditory: | Synthetic speech | Can attend to other tasks | Not expressive enough, slow and linear feedback |
| Verbal | | | |
| Sound effects | General synthesizer | Easy to do, can be general purpose | need to learn 'vocabulary' of interaction |
| Tactile/haptic | Computer controlled Braille, pushpins | good potential in combination with others | not high in information bandwidth |
| Vestibular | | | not generalized |
| Taste/Smell | | evocative of memories | difficult to do well, not well suited for high bandwidth communication |
| Other? | | | |

Human control modalities

Human control modalities have also not changed significantly, though new sensing devices have enabled new kinds of input and control (Table 2).

| Human control modalities | Devices for mobile interaction | Strengths | Weaknesses |
|---------------------------------|---|--|--|
| Gestural: | Hands, free space or touch pad | Expressive, good control, can use simultaneous input | Fatiguing without appropriate supports, need to learn a vocabulary |
| | Whole body | Can have added benefits (exercise) | Same as above, privacy |
| | Device manipulation (pens, spaceball, etc.) | Mature form of input | Information bandwidth may be low |

| | | | |
|-------------------------|--|--|--|
| Verbal: | Speech recognition | General (perhaps too general) | Need to know context for commands, privacy |
| Controller | Keyboards | Potentially high in information bandwidth, familiar | Small keyboards can be difficult to use, even harder to be fast, difficult to use in the air |
| | Projected keyboards | Can be larger in size but still portable | No tactile feedback |
| | Alternate keyboards (chorded) | Potentially fast, portable | Learning curve |
| | Other controllers (remote controllers, game controllers, musical instrument controllers) | Comfortable to use, portable, can be expressive or relative high information bandwidth with learning | Limited controls, need to know/learn mapping |
| Tangible devices | | Natural mapping between task and object | not generalized. May need many devices to perform task |
| Other? | | | |

Cognitive Limitations:

While not sensory modalities, there are also perceptual and cognitive limitations for people that limit the style of the human computer interaction. Limited attention, limited memory, limited visual acuity, limited audio resolution, all have influences on the kind, number, and combinations of output devices that will be useful.

Information tasks:

Finally, in a one last table, we have high level information tasks that we need to be addressed by interfaces.

| | |
|---------------------|--|
| Notification | Bringing relevant information to the attention of the user. (Context can be very important for this, both spatial and temporal, ie. Bring information about where the user might be, or what the user might need to do at the moment.) |
| Acquisition | Learning new information |
| Search | Finding new (or old) relevant information |
| Organization | Keeping track of relevant information for synthesizing or communication (or to facilitate search). |
| Synthesis | Creating new information for self or others |
| Sharing | Communicating information to others |

We have ways of coping with all these tasks but new interaction techniques and devices can make these tasks easier, can make users more powerful for performing current tasks (allow faster completion, make results better, and allow more tasks to be addressed at the same time). Furthermore, communication between individuals can be facilitated which might either be a task unto itself or enhance other tasks. And finally, this can all potentially be done in new and disparate locales.

Challenges in the design of new interfaces for information management:

Many practitioners in the field of HCI believe that clear task models are crucial for defining effective interfaces. User centered design, working with people facing real problems, is an important factor in coming up with polished interaction systems. But it is also clear that users seldom anticipate fundamental variations on existing methodologies. When asked how to steer the very first automobiles, the overall response from people was 'reins', since that was what people of the time were familiar with. It is appropriate to identify problems with currently performing tasks, especially given new, mobile domains, but solutions will need to be designed, prototyped, and tested with an open mind.

Conventions may be adapted from existing techniques for interacting with large amounts of information, but significant changes will need to occur as well. For example, direct manipulation has had an incredible influence on the style and types of interaction that we have on our desktop. Direct Manipulation interfaces have the advantage of showing affordances for interaction. This is one of the most significant advantages since occasional or novice users can often infer what needs to be done by the reflected state of the system. By contrast, command driven UI's need to be learned and context appropriate commands can be easily forgotten. However, in a mobile environment, natural pointing and interaction with abstract data objects can be difficult (unless using an auxiliary display and appropriate gestural recognition). It could be much more natural to use a speech command driven UI since we nearly always can simply speak, though, as mentioned before, command driven displays have their own set of problems. Furthermore, direct manipulation interfaces often have problems dealing effectively with large numbers of information. Repetitive tasks are not easily specified and transformations on groups can be difficult.

To date, I've been examining novel interfaces that use information visualization techniques, but aimed at casual or novice users to help them cope with large amounts of information. In particular, I've been focusing on enhanced display capabilities (powerful GPUs, and large screens), though I am interested in finding new ways to interact with these visualizations in mobile and other contexts. In particular, four interfaces (MediaFrame, TimeQuilt, Movie Variations, Remote Photo Browser) all represent novel ways of dealing with lots of media (photos and video). See <http://research.microsoft.com/~sdrucker/interest.htm> for videos and references. Finding ways to generalize these techniques to other kinds of information, as well as new interaction techniques will bring on exciting new opportunities.

Conclusions:

In thinking about the interfaces of the future, it is not enough to think only of the technology, or only of the users, but a combination of the challenges that will be facing us in the future, user abilities and limitations, as well as the technological directions that we are currently taking.

Some Musings On Building and Showing Future UI Prototypes

Andrew D. Wilson
Microsoft Research

Introduction

For better or worse, I have spent a great deal of time demonstrating prototypes of “future UI” systems to people of all kinds. In this paper I would like to lay out some observations I have made regarding people’s reactions to these demo systems. These are very all preliminary thoughts, half-cocked, naïve or even obvious, but I think they can at least serve as interesting talking points.

People are wowed by the darndest things.

Now that large flat panel displays and projectors are relatively commonplace, it is refreshing to give a demo where people don’t make comments on how stunning your display is. Even so, people will often completely miss the point of your demo, and will be completely snowed by the overall whiz-bang nature of it. At one level it is a lot of fun to show systems that deeply impress people, but at another level it is frustrating that in the process they can miss the point of your little innovation.

One common phenomenon is the absolutely robust fascination with *action at a distance*. That is, people tend to be impressed by almost any form of action initiated by some control that is removed from the site of action. This effect is strongest when the mechanism is hidden. Of course, most people are not impressed by the lights coming on by flipping a switch on the wall. A key component then must be the level of surprise, or degree to which expectations are violated when they move a knob here and something changes there. People’s expectations are by and large cast by their experience with personal computers—the most amazing thing they will ever see is the optical mouse.

Magicians are probably operating on many of the same principles. There are probably some things we can learn from magicians, though magicians may be more intent on redirecting attention. Presumably we are less interested in that?

People love the Minority Report interface.

Most people’s conception of future UI is tightly bound to how they are portrayed in sci-fi movies. Sci-fi movie interfaces are a great source of imagination for everyone, but they do not always give a balanced view of what’s really possible or even desirable. There’s a great variety of demos before and after the movie *Minority Report* that are gesture-based, and use interesting sensing in ways that trigger people’s memories of the movie. Interestingly, more than half of the people that I demo to can cite the movie but cannot remember its name (really)—that at least speaks to the power of the particular scene.

Most people don’t fully grasp some of the problems with such an interface. Fatigue is significant problem with these kinds of interactive systems. It is interesting to note, for

example, that before the mouse there was the light pen (a proto-Minority Report interface?), and that one of the draws of the mouse was that you didn't have to hold your hand up for long periods. John Atherton uses the transparent display so that the director can get nice shots of his expensive actor (and it's cool, besides). I believe there *are* real advantages of these kinds of interfaces, but what are they exactly? "I'm too busy having fun moving this thing around the screen."

People expect one-size-fits-all computing.

As mentioned above, people carry a lot baggage when it comes to how they expect computers should behave, i.e. the WIMP interface. Unfortunately, these expectations have an adverse effect in proposing a UI that behaves radically differently. It is shocking how many people will seriously ask whether they can read their email or surf web pages on just about any device you can imagine.

The future-UI designer is thus faced with a difficult choice in how to present their work: try to get full-up Windows running on their device or show their device on some whacky new barely-working application framework. The approach of emulating Windows input is limited by the fact that it usually emulates the mouse very poorly and therefore provides a pretty miserable experience. Inventing a completely new UI suffers from the fact that the UI is probably not very well developed and very narrow, and maybe even doesn't do anything useful at all.

One way to talk around these issues is to sketch the vision of *ubiquitous computing*. For the present purposes this vision predicts that the future is populated with *many* interfaces, each perhaps specialized to a given application or style of interaction. For example, it's not reasonable to expect to code up a C++ program while driving your car. Living room media PCs probably shouldn't present dense fields of text and widget, and kitchen computers are probably fine with coarse levels of input. Thus the future is not about one-size-fits-all computing, but rather smart *specialization*, where some interfaces are free to employ modes of interaction that make sense for a given range of applications. It will be playground of interactive techniques in which its perfectly okay if the wall computer doesn't provide fantastic mouse emulation.

The challenge then is how to design a variety of interactions, without falling back to the WIMP interface, forcing it onto devices where it can't take advantage of a device's special advantages. How will users learn and understand such a variety of interfaces? Are there primitives of interaction more basic than WIMP that we should be using?

Kids have far fewer preconceptions.

It is interesting to watch young children use an interactive table system and compare their usage pattern with that of typical adults. Kids will use possibly both hands, putting their whole (small) hands on the surface, while most adults will form a proper extended index finger and fully expect a cursor-based interaction. This is probably due to the fact that most kids have never used an ATM touch screen and have minimal experience with PC computing, and so do not have many preconceptions about computing interfaces. In

many ways, they expect the direct manipulation interface and are not surprised by it in the least, while adults are quite the opposite.

The differences don't stop there. Kids tend to move much faster than adults, which can be a problem for many prototypes. Kids seem to be *a lot* harder to impress than adults. Perhaps this is due in part to increasingly early experience with video games or the relative technical sophistication of today's toys.

It will be interesting to see how today's generation of children will question today's interfaces and re-invent them. Instead of waiting until they are out of college (at least) to have access to computers powerful enough to build interesting prototypes, they will be able pursue their wildest dreams right away.

Building prototypes is fun. Evaluating them is hard.

Many of us have been having *a lot* of fun building interesting prototype systems that demo well. However, evaluating these prototypes to show that users can understand them, or showing scientifically that they improve some aspect compared to more conventional approaches, can be very difficult.

Chief among the difficulties is that because the systems tend not to be incremental improvements on what we already have but more a reinvention of the interface, typical users are wowed by the novelty. Most will say they like it or enjoy it but when you press for specifics you find they like it just because it is new and interesting. Subjective evaluations that include questions such as "Did you like it?" are therefore not helpful. Comments solicited during brief demonstration-length encounters are interesting and can be valuable feedback, but they tend to miss important considerations such as whether users could really live with such an interface day in, day out. Longitudinal studies are expensive and usually require a fairly refined (robust) prototype.

Even when the user study is truly informative and well executed, the effort put in the study is usually a fraction of the effort spent to get the system working in the first place, and once the paper is done, we're onto the next thing. Most of us are engineers and inventors at heart after all. It is therefore difficult to find answers to some of the deeper questions raised by these new interfaces. *Tangible computing*, for example, satisfies an intuition that input and output should be embodied in the real world, yet this very powerful intuition is supported by very few scientific studies. Does tangible computing scale? In order to make it real must we spoil it with layers of modal interactions?

One can argue that the field (whatever it is) is very young and we are still in an exploratory mode, where we are at once picking the low-hanging fruit and exploring the outer boundaries of what is even possible.

Traditional HCI evaluation metrics capture only a part of what's going on.

When it comes to the application of traditional HCI evaluation metrics such as time to task completion, reaction time, Fitts index of difficulty and so on, as applied to future UI

concepts, one is reminded of the adage of the drunk looking for his keys under the lamp-post because that is where there is good light. Traditional HCI metrics seem ill-suited to address many aspects of future UI prototypes that generate a lot of excitement, such as the interface's *novelty*, the sheer *pleasure* in using a device, and the quality of "flow" experience [2] enabled by the interface. Don Norman [3] argues that there are a variety of dimensions along which a device or interface can be evaluated, some of which have nothing to do with *performance per se*, such as the *reflective* qualities that relate to the user's self image and satisfaction in using the device. Norman further argues that in order for a device to be a success it must be successful on multiple levels. It is interesting to note how this dove-tails nicely with the observation that the future will lead to a larger variety of interactions—not everything must be built to do word processing.

The novelty of a UI prototype presents an interesting dilemma. On the one hand, scientists studying usability often spend a great deal of effort to remove novelty effects (longitudinal studies being the most relevant method). On the other hand, novelty is a quality that can drive users to initially engage in a device in the first place, and can provide some much needed variety in one's life. In some sense, as consumers we crave novelty and actively seek it out. The question of whether novelty is a "good" or "bad" thing has probably more to do with its staying power and whether it gets in the way of performing the main task. Unfortunately there are few good ways to talk about the impact of novelty. The quality of pleasure in using the device, and the abstract quality of "flow" have similar stories.

Video games provide a wide variety of user interfaces (at least, in comparison with most other computing), but very little is known or written about what specifically makes a game fun or engaging. The new field of study that considers the design of games may provide some directions on non-traditional ways to evaluate user interface prototypes [1].

Conclusion

This short paper presents a number of thoughts related to how prototypes of future UIs are shown and how they are positioned within the HCI field. While many of the comments are of a critical nature, they should also be viewed as opportunities for areas of future work. Many of these problems are very difficult, and it is probably unreasonable to expect clean solutions for any one, particularly in light of the overall arc of *interface diversity* presented: the *future* of the future UI is likely to be messy but full of interesting worthwhile diversions.

References

1. Blythe, M.A., Overbeeke, K., Monk, A.F. and Wright, P.C. (eds.). *Funology: From Usability to Enjoyment*. Kluwer Academic Publishers, 2004.
2. Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience*. Harper and Row, 1990.
3. Norman, D.A. *Emotional Design: Why We Love (Or Hate) Everyday Things*. Basic Books, 2004.

C. New Interaction Styles

C.1 Moving Towards Naturalistic Interactions With Invisible Interfaces: Can "Digital Simplicity" Be Achieved?

Beverly L. Harrison Intel Research Seattle

C.2 Egocentric Interaction

Thomas Pederson Umea University, Sweden

C.3 Core Tangibles and Tangible Visualizations: prospects for tangible convergence and divergence

Brygg Ullmer Louisiana State University

C.4 Computational Corporeality: Computers, Materials, and the Body

Amanda Parkes MIT Media Laboratory

C.5 The Future of Interface Design, Through a Child's Eyes

Hayes Raffle MIT Media Laboratory

C.6 Frames of Reference for Portable Device Interaction

Sile O'Modhrain Queen's University, Belfast

C.7 Embodied Interfaces: The Next Generation of HCI?

Chris Creed University of Birmingham, UK
Russell Beale University of Birmingham, UK

Moving Towards Naturalistic Interactions With Invisible Interfaces: Can “Digital Simplicity” Be Achieved?

Beverly L. Harrison

Intel Research Seattle

1100 NE 45th Street, Seattle, WA 98105

beverly.harrison@intel.com

(206)545-2503

THREE CHALLENGES OF INVISIBLE INTERFACES

Past user interface design leveraged real world metaphors and the naturalistic associations of these to create what was hoped to be more easy to use technologies (e.g., file folders, the desktop, the classic trash can, now subtly renamed the more environmentally friendly “recycle bin”). Pictorial representations (icons) were intended to facilitate understanding, learning, and the direct manipulation of objects of interest. These waves of prior innovation enabled technologies to become accessible to a broader community of users in ways that are now considered quite pervasive.

Technology is now equally pervasive in mobile or handheld devices, household appliances, and is often embedded invisibly in the environment around us (e.g., sensors, cameras, wireless networking). Where formerly users had explicit interactions with particular technologies they deliberately selected (by virtue of using a particular computer or device, a particular input mechanism, and/or a particular application), they may now be interacting in implicit ways with imprecisely selected technology at almost any moment in time. These interactions can occur whether users are aware of them or not, and whether users intended them or not.

Early visions of the future presented by Wellner’s DigitalDesk [38, 39] and by Weiser’s ubiquitous computing [36, 37] have been extended upon and are reflected in substantial research over the last 10 years [e.g., 8, 9, 11, 17, 26, 28, 29, 30, 31, 33] (including my own work [e.g., 6, 7, 13, 20, 21, 35]). A goal of these emerging projects is to seamlessly blend the affordances and strengths of physically manipulatable objects with virtual environments, functionality or artifacts thereby leveraging the particular strengths of each. This seems like a natural progression towards making the next UI “metaphor” the real world itself: real objects having real properties that are linked to or embedded with the virtual artifacts that they control. This reflects a powerful, only slightly understood user interface paradigm of “invisible interfaces”; interfaces that are always on, embedded in our everyday objects or environment, and subtly invoked by our naturalistic interactions. However, as with any technology, getting the design right is crucial if our aspiration is for widespread accessibility, predictability, ease of use and ubiquity. A first grand challenge for invisible interfaces is better *articulating what the user interaction model is* and how the associated design principles need to evolve to adapt to this model.

While the explicit manipulation of an everyday object may influence or trigger embedded technologies, research has simultaneously extended to examine the broader context of this interaction. (This focus on context of use appears to be particularly emphasized in ubiquitous computing.) Deliberate user interactions may produce implicit and unintended consequences because of the contextual assumptions made in designing and embedding these technologies (e.g., picking up a medicine bottle to read the code on the label for a refill vs. picking it up to take the medication [34, 25]). Furthermore, some environmentally embedded technologies are activated simply by virtue of the user being in a particular location without any expressed interactions at all (e.g., in view of a camera system, in range of a particular wireless network, in range of a motion sensor [e.g., 3, 10]). The idea of “situated action” is not new [32] however, it seems that the importance of understanding context is of increasing importance in disambiguating user intent in this newer interaction space [e.g., 5, 15, 22, 25]. A second grand challenge for invisible interfaces is *understanding or correctly inferring user context* and how this impacts design.

As a direct consequence both of the changing nature of what it means to interact with technologies in this invisible interface paradigm and due to the increasingly important role context plays, there is a crucial transformation needed for evolving evaluation techniques and methodologies. These invisible interfaces are used in dynamic and often highly mobile contexts and locales. They often involve a complex mesh of infrastructure, devices and sensors all of which must work as a coherent whole and thus must be assessed as a system. Evaluation methods that might have worked well for single technologies or specific interaction methods do not seem well suited to this more challenging problem domain. Modified techniques are being adapted and tested to try and address these unique attributes and challenges [e.g., 4, 12, 16, 24]. However, research on the methodologies and evaluation tools is in the early stages with much more promising results still ahead. A third grand challenge for invisible interfaces is *creating or evolving evaluation techniques and methodologies*.

INTERACTION MODELS FOR “INVISIBLE INTERFACES”

My past work has investigated three fundamentally different mechanisms of interacting with technology in the ubiquitous computing domain: 1) physically manipulative interfaces through object handling and deformation via

embedded sensors [e.g., 6, 7, 13]; 2) embedded “inferred” interfaces through object location/presence sensing via RFID and computer vision [e.g., 20, 21, 35]; and 3) interaction resulting from embedding technology in the environment such as cameras, sensors, or wireless networking. These may be deployed independently or in concert to create an overall user experience typically characterized by the absence of visible user interface controls. Independently each of these creates a different model of interaction and different design constraints and possibilities.

Physically Manipulative Interaction

These interfaces typically rely on pressure sensors, accelerometers, motion sensors, light sensors, sound input and output, motors, and wireless communication to enable users to physically press, push, tilt, or juxtapose handheld devices or objects in order to convey commands. (I would like to differentiate interactions where object “deformation” is used from those where merely picking up an object is the interaction and hence command, discussed next).

There are already some compelling prototype examples reported in the current HCI literature:

- Scrolling a menu on a PDA by tilting the PDA [26]
- Zooming text by pushing/pulling the device towards/away from the user [8]
- Rolling or turning knobs [2]
- Squeezing/physically interacting with plush toys [18, 19] or physically deforming object surfaces [23]
- Or in my own prior work [6, 7, 13], exploring user interface techniques for navigating and displaying documents in portable reading devices/ebooks by squeezing, stroking, flicking, and tilting.

We find that designing these interactions confronts us with a host of subtle design issues, and that there are no articulated design principles to guide us through them. We believe that the user interfaces above are the forerunners of a new paradigm for user interface design, user interfaces in which a computing device is directly and physically manipulated, which we have called *Manipulative User Interfaces* (MUIs).

The MUI paradigm extends the GUI paradigm in 2 fundamental ways: 1) GUIs are somewhat indirect, in that the mouse controls a remote cursor behaving as interaction intermediary, whereas in MUIs, users use their bodies to directly manipulate devices, objects and/or their displays to specify commands. 2) The MUI paradigm opens up a much richer space of interactions than just pointing. Consider, for example, the Tamagotchi [1] “virtual pet” toy, which you interact with by means of a number of buttons. The toy requires its owner to provide it with food (by pressing the feeding button), to give it attention (by pressing the petting button), etc. But from a MUI perspective you could interact with the Tamagotchi in much more affective ways. You could soothe the pet by stroking it, play with it by tossing it in the air, ease its hunger by rubbing its belly, and so forth.

These manipulations can be divided into three categories: manipulations that change the *spatial position* of the device (translation, rotation, orientation, tilting), manipulations that change the *structural properties* of the device (deformations, pressing, squeezing, flicking, shaking), and manipulations that change the *inter-device relationship* (juxtaposition, stacking, tiling, proximate association or dissociation). All of these can be temporary or permanent in effect, can be parameterized in various ways, and can be performed alone, simultaneously, or in sequence.

In MUIs the *extent of embodiment* can vary. In some cases there is a tight coupling between the manipulation, the perceived content this affects, and feedback (both tactile and visual) indicating the resultant effect. All are embodied within a single object to be manipulated. In other instances the manipulation frequently affects content/data via a controller intermediary and visual feedback of the effect are represented in an external system distinct from the controller. Examples of this include recent graspable-object interactions, such as “phicons” [11, 28], “bricks” [9], and “doll's head” [14] techniques, where the object being manipulated acts as a remote control for actions taking place outside the object. The boundary between these two approaches can be somewhat blurred.

Embedded “Inferred” Interaction

This class of interaction is one where the act of selecting an object or moving it triggers a technological response thereby communicating an implied or inferred “request” from the user. This is typically achieved by computer vision, RFID, or location detection where technological modifications to an object are invisible to the user (barcodes and glyphs being the exception) yet handling these instrumented objects has an effect. A number of prototypes have been built to demonstrate a variety of applications:

- Moving physical blocks representing buildings for urban planning [33]
- The pick-and-drop work that attaches virtual content to a physical transport mechanism or physically selecting a projections of objects to move items between surfaces or devices using gestures [26]
- And some of my own work [20, 21, 35] looking at augmenting books, staples, business cards, posters, and augmented in/out boards and whiteboards with phicons (Collaborage).

Unlike environmentally sensed interactions, in embedded inferred interactions the participation of both a user and an embodied object are necessary. While the technologies deployed in both situations have similarities (wireless communication, detecting changes in location across wireless networks, detecting presence/absence of an object or person of interest), embodied inferred interactions are instantiated in particular devices or objects that users hold, carry and interact with. Whereas in environmentally sensed interactions, the environment infers interactions from the

user him/herself regardless of the presence of absence of any particular object (e.g., the user's presence triggers a motion sensor, floor pressure pad, camera system).

As with most taxonomies or categorizations, there are situations where the distinction between an object-based inferred interaction and an environmentally aware application is somewhat blurred, for instance, in cases where an object has a tight affiliation with a user's identity (e.g., a cell phone the user habitually carries even when not in use) but the object's presence is virtually forgotten. The object (or device or sensor) is used to transmit information to environmental sensors as a back channel (e.g., cell phone transmits GPS location data) while the user is not deliberately interacting with this object and may be unaware of this active background transmission. One could argue that there is an interaction mechanism embodied in an object co-present with a user and thus this scenario is an embedded inferred interaction. However, the user has not participated actively (or perhaps even knowingly) in this interaction, the object's role is passive (from the user's perspective) and the application is crucially more dependent upon the environmental infrastructure sensing the user's presence (rather than the object's presence) thus it could equally easily be considered an environmentally sensed interaction. Perhaps the value of categorizing scenarios and applications will best be determined as we evolve design guidelines for these categories.

Environmentally Sensed Interaction

The proliferation and availability of largely invisible communication, camera and sensor-based technologies have created new possibilities for environmentally sensed interactions: global cellular networks, city-wide and building-wide wireless networks, city-wide and building-wide camera coverage, room or house scale sensors, etc. It is becoming more commonplace to see applications that utilize this infrastructure to sense and infer things about individuals, groups, behavior patterns, and trends (including divergences from norms). As describe above, individuals may be participants (knowingly or unknowingly) by virtue of habitually carrying devices that afford sensing or by themselves occupying or moving through a space that is instrumented.

There have been a number of implementations of "smart home" and "smart kitchen" prototypes that used sensors to detect and even identify individuals and their activities [e.g., 3, 25]. One could argue that home alarm systems are a simple but early instance of these. City wide camera-based technologies are deployed in the UK and in some US cities for security and traffic monitoring. Any number of location sensing applications have recently been built to track people, the places they frequent, the routes they take, and the activities they are doing [e.g., 3, 10, 15, 16, 25].

While all three of the above interaction categories have implications for inadvertent use, uninformed use, and privacy, environmentally sensed interaction is perhaps the

most problematic and challenging [e.g., 10 presents an excellent overview of issues and studies]. There are no "objects" channeling users' intentions and express interactions. Nor can we leverage such objects to inform users of system activity. Deriving interaction models and design guidelines for this class of application must incorporate these considerations.

Conscious, semi-conscious, and unconscious interactions

We characterize the interaction models as being *conscious* (fully aware of interactions with an object), *unconscious* (fully unaware of interactions by virtue of having no explicit instantiation), and *semi-conscious* (sporadic awareness and forgetfulness of objects that afford interactions whether or not the user is aware).

We deliberately wish to use these as metrics to categorize 5 dimensions of interaction: awareness, intent, ubiquity, ease of use/design and skill. In fact, one might argue that as interaction mechanisms and technologies become more pervasive and ubiquitous (if well designed), they may migrate from being consciously invoked, manipulated, or monitored to semi-consciously or sporadically manipulated/monitored to unconsciously embedded in habitual routines without much explicit planning or thought about use. This raises interesting and difficult questions. How does the acquisition of skill and expertise relate to technology consciousness? How is design related to this and does bad/good design support or hinder consciousness? Is the migration of technology from consciously aware usage to semi- or unconscious use a positive progression?

UNDERSTANDING AND INFERRING USER CONTEXT

Many interactions with invisible interfaces rely on sensors to help make educated guesses about the user's context. If these sensors and interaction mechanisms are embodied within a device or object, parameters scoping the user intention may sometimes be inferred. In fact, stronger assumptions might be possible when considering sequences of object interactions [e.g., 25], for instance, activity inferencing that characterizes high level activities like "making a cup of tea" based on seeing interaction with the water tap, kettle, cup, tea box, and milk. One significant research problem is reliably and accurately abstracting these lower level actions into higher level activities, when users vary, lower level actions vary, sensors vary, and the data is noisy and ambiguous. However, surprisingly good results have come from targeting particular activity types that seem more amenable to prediction, especially if combined with supervised learning [e.g., 5, 16, 22, 25]. For instance, it is easy to detect someone in motion versus someone who is stationary or a change in location for a particular object. It is more difficult (and potentially intrusive) to detect contexts that have few physical cues (e.g., changes in mental state or switches in cognitive tasks). Determining task attributes that make certain activities most amenable to inferencing, evaluating training and inferencing algorithms, and assessing how reliable the

inferencing needs to be in order to be useful are important areas of ongoing and further research. There is a significant body of work going on in machine learning, computer vision, speech and signal processing, and generally in context-aware computing to address many of these issues.

Knowing something about a user's context can greatly enhance the ability of invisible interfaces to behave in predictable or anticipated ways by providing data to help disambiguate user intentions. A combination of sensor-based data for activity inferencing and user supplied training data (used to establish ground truth) are proving to be interesting and useful techniques. This combination of system log data and user supplied field data are also a crucial component of more general evaluation strategies.

EVALUATION TECHNIQUES AND METHODOLOGIES

Ubiquitous computing and invisible interfaces pose particular challenges for evaluating whether or not technologies and applications are usable or successful. Traditional empirical studies can assess whether or not individual technologies work reliably and predictably in controlled situations. However ubiquitous computing typically requires a collection of technologies to work in concert and thus isolating, modeling, and evaluating individual components does not provide good indicators for whether or not the more complex system will be usable. Ubicomp technologies are used in dynamic contexts and in changing locations, where demands on the users' visual and mental attention varies dramatically and somewhat unpredictably. These complexities are compounded by the lack of design guidelines and interaction models to guide in developing usable and predictable "invisible interfaces" (if there is no interface per se, how do you "design" it?).

Evaluation methodologies have evolved to combine controlled laboratory studies and Wizard of Oz prototypes with in-situ field methods. In addition to ethnographic studies (observed user behavior) and diary or journal studies (self-reported user behavior), new techniques and metrics are being tested and applied [e.g., 4, 15, 16]. Ideally, these capture data that are in-situ, involve multiple participants, take place over (sometimes extended) periods of time, and are quantitative and qualitative in nature. Most notably, experience sampling methods (ESM) are eliciting user responses from the field by using sensors, inferred events, or contextual inferences to time prompting for users to answer questions delivered over mobile devices. These questions can thus be tuned to fit the nature of the inferred situation and to increase the likelihood of responses (because they are timed for less disruptive moments).

We are further interested in exploring methods for quantifying and measuring not only usability/predictability but also effort. While still in its early stages, we are investigating the application of subjective and objective measures of mental and physical workload such as those typically applied in traditional Human Factors Engineering, most notably the NASA Task Load Index, SWAT

(Subjective Workload Assessment Technique), and SWORD (Subjective Workload Dominance). We are hoping that the metrics captured by these techniques may usefully quantify key factors in ubicomp applications and technologies that enable us to compare, contrast and systematically assess new approaches in design.

SUMMARY

I wish to summarize by highlighting some questions and issues that I believe the research community needs to address. Does categorizing the extent to which an interface is embodied in an object help us in formulating design principles, interaction models and evaluation metrics? How do we define and go about designing "invisible interfaces" if interaction mechanisms aren't visible? What is the interaction model? What is the role of user's context and how do we best make use of that? How accurately do we need to infer context? What do we communicate to the user about what we are inferring, when we infer, where we infer, and whether it is correct? What kinds of evaluation methodologies will most help us in assessing new usage patterns, new technologies, and invisible interactions? Will this evolution of technology result in outcomes that are "digitally simplistic"? Should this be an aspiration? If not, what are the measures for success?

ACKNOWLEDGMENTS

Much of my prior work was done at Xerox PARC in collaboration with Ken Fishkin, Roy Want, Anuj Gujar, and Tom Moran. Some of my work on activity centered design and mobile devices was done at IBM Almaden Research Center in collaboration with Tom Moran, Shumin Zhai, and John Barton. Most recently I have been working with researchers at Intel Research Seattle, including James Landay, Anthony Lamarca, Sunny Consolvo, Ian Smith and Matthai Philipose. The efforts and influences of these collaborations are gratefully acknowledged.

REFERENCES

1. Bandai Co. Ltd, the Tamagotchi Page. <http://www.urban.or.jp/home/jun/tama/>
2. Brave, S. Ishii, H. and Dahley, A. Tangible Interfaces for Remote Collaboration and Communication. *Proc. of CSCW'98*, ACM Press, 1998, pp. 169-178.
3. Brooks, R. A., "The Intelligent Room Project," *Proc. of the 2nd International Cognitive Technology Conference*, IEEE CS Press, 1997, pp.217-8.
4. Consolvo, S. and Walker, M. Using the Experience Sampling Method to Evaluate UbiComp Applications, *IEEE Pervasive Computing*, April-June 2003, pp. 24 – 31.
5. Dey, A. K., Kortuem, G., Morse, D. R., Schmidt, A. (Eds) Special Issue on Situated Interaction and Context Aware Computing, *Personal and Ubiquitous Computing*, 2001, Vol 5, Springer-Verlag.
6. Fishkin, K. P., Gujar, A., Harrison, B. L., and Want, R. Embodied User Interfaces for Really Direct Manipulation. *ACM*, Sept. 2000, Vol.43 No.9, pp. 75-80.

7. Fishkin, K. P., Moran, T., and Harrison, B. L. Design Principles for Manipulative User Interfaces. *Proc.s of Engineering for Human-Computer Interaction*, Heraklion, Crete, 1998.
8. Fitzmaurice, G. Situated Information Spaces and Spatially Aware Palmtop Computers, *CACM*, 36(7), July 1993, pp.38-49.
9. Fitzmaurice, G., Ishii, H., and Buxton, W. A. S. Bricks: Laying the Foundations for Graspable User Interfaces. *Proc.s of CHI'95*, ACM Press, 1995, pp. 422-449.
10. Friedman, B., Kahn, P. H., Hagman, J., and Severson, R. L., The Watcher and the Watched: Social Judgements about Privacy in a Public Place, In press. 2005. To appear in the *Human Computer Interaction*. 2005.
11. Gorbet, M. G., Orth, M., and Ishii, H. Triangles: Tangible Interface for Manipulation and Exploration of Digital Information Topography. *Proc. of CHI'98*, ACM Press, 1998, pp. 49-56.
12. Harrison, B. L., Cozzi, A., Moran, T. P. Roles and Relationships for Unified Activity Management, *Proc. of Group 2005*, ACM Press, 2005, pp. 236-245.
13. Harrison, B. L., Fishkin, K. P., Gujar, A., Mochon, C., and Want, R. Squeeze Me, Hold Me, Tilt Me! An Exploration of Manipulative User Interfaces, *Proc. of CHI'98*, ACM Press, 1998, pp. 17-24.
14. Hinckley, K., Pausch, R., Goble, J. and Kassel, N. Passive Real-World Interface Props for Neurosurgical Visualization, *Proc of CHI'94*, ACM Press, 1994, pp. 452-458.
15. Intille, S. S., Bao, L., Tapia, E. M., Rondoni, J. Acquiring In Situ Training Data for Context Aware Ubiquitous Computing Applications, *Proc of CHI 2004*, ACM Press, 2004, pp. 1-8.
16. Intille, S. S., Rondoni, J., Kukla, C., Ancona, I., and Bao, L. A Context-Aware Experience Sampling Tool. Posters in *Proc. of CHI 2003*, ACM Press, 2003, pp. 972-973
17. Ishii, H. and Ullmer, B. Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms. *Proceedings of CHI'97*, ACM Press, 1997, pp. 234-241.
18. Kaminsky, , Dourish, P., Edwards, K., Lamarca, A., Salisbury, M., Smith, I. SWEETPEA: Software tools for programmable embodied agents. *Proc. of CHI'99*, ACM Press, 1999, pp. 144-151.
19. Marti, S. and Schmandt, C. Physical embodiments for mible communication agents. *Proc. of UIST 2005*, ACM Press, 2005, pp. 231-240.
20. Moran, T. P., Saund, E., van Melle, W., Bryll, R., Gujar, A., Fishkin, K., Harrison, B. L. The Ins and Outs of Collaborative Walls: Demonstrating the Collaborage Concept. Short paper in *Proc.s of CHI'99*. ACM Press. 1999, pp. 192-193.
21. Moran, T. P., Saund, E., van Melle, W., Gujar, A., Fishkin, K., Harrison, B. L. Design and technology for collaborage: collaborative collages of information on physical walls. *Proc. of UIST'99*. ACM Press, 1999, pp. 197-206.
22. Moran, T.P., and Dourish, P. (Eds). Special Issue on Context Aware Computing, *Human Computer Interaction*, 16, 2-4 (2001). ACM Press. NY, NY.
23. Murakami, T., Hayashi, K., Oikawa, K., and Nakajima, N., DO-IT: Deformable Objects as Input Tools, *SIGCHI '95 Conference Companion*, ACM Press, 1995, pp. 87-88.
24. Mynatt, E. D., Ugarashi, T., Edward, K., and Lamarca, A. Flatland: New dimensions in office whiteboards. *Proc of CHI99*. ACM Press, 1999, pp. 346-353.
25. Philipose, M., Fishkin, K. P., Perkowitz, M., Patterson, D., Fox, D, Kautz, H., and Hahnel, D. Inferring Activities from Interactions with Objects. *IEEE Pervasive Computing*, Issue: Oct-Dec 2004, pp. 10-17.
26. Rekimoto, J. and Saitoh, M. Augmented Surfaces: A Spatially Continuous Workspace for Hybrid Computing Environments. *Proc of CHI99*, ACM Press, 1999. pp.378-385.
27. Rekimoto, J., Tilting Operations for Small Screen Interfaces. *Proc. of UIST '96*, ACM Press, 1996, pp.167-168.
28. Small, D., and Ishii, H. Design of Spatially Aware Graspable Displays. *Extended Abstracts of CHI'97*, ACM Press, 1997, pp. 367-368.
29. Streitz, N. A. and Russell, D. M. Basics of Integrated Information and Physical Spaces: The State of the Art. *Proc. of CHI'98*, ACM Press, 1998, pp. 273-274.
30. Streitz, N. A. Integrated Design of Real Architectural Spaces and Virtual Information Spaces. *Proc. of CHI'98*. ACM Press, 1998, pp. 263-264.
31. Streitz, N. A., Konomi, S., and Burkhardt, H.-J. Cooperative Buildings: Integrating Information, Organization, and Structure. *Proc. from the 1st International Workshop CoBuild'98*, Springer-Verlag. 1998.
32. Suchman, L. A. *Plans and Situated Action: The Problem of Human Machine Communication*. Cambridge University Press. 1987.
33. Underkoffler, J. and Ishii, H. URP: A Luminous Tangible Workbench for Urban Planning and Design. *Proc.s of CHI'99*, ACM Press, 1999, pp. 386-393.
34. Wan, D. Magic Medicine Cabinet: A Situated Portal for Consumer Healthcare, *Proc. Of 1st Int'l Symp. Handhel and Ubiquitous Computing (HUC99)*, Springer-Verlag, pp. 352-355.
35. Want, R., Fishkin K., Gujar, A., Harrison, B. L., Bridging Physical and Virtual Worlds with Electronic Tags. *Proc. of CHI '99*, ACM Press, 1999, pp. 370-377.
36. Want, R., Schilit, B. N., Adams, N. I., Gold, R., Petersen, K., Goldberg, D., Ellis, J. R., and Weiser, M. An Overview of the ParcTab Ubiquitous Computing Experiment. *IEEE Personal Communications*, December 1995, pp. 28-43.
37. Weiser, M. The Computer for the 21st Century. *Scientific America*, 265(3), 1991, pp. 94-104.
38. Wellner, P. Interacting with paper on the Digital Desk, *CACM*, 36(7), July 1993, pp. 86-96.
39. Wellner, P. tactile manipulation on the DigitalDesk. Video in CHI'92 Special Video Program, *ACM SIGGRAPH Video Review 79*, 1992.

Egocentric Interaction

Thomas Pederson

Department of Computing Science

Umeå University

SE-90187 Umeå

+46 (0)90 786 6548

top@cs.umu.se

ABSTRACT

This position paper for the CHI2006 workshop “What is the Next Generation of Human-Computer Interaction?” presents our emerging conceptual framework for modelling and designing everyday activity support systems. The proposed framework is based on the assumption that a suitable wearable computing and sensing platform will be feasible in 5-10 years from now.

INTRODUCTION

Looking back on the HCI work made during the period 1950-1975 and comparing it to the work performed during the 25 years after that it seems that while the first period was about finding and defining a working interaction paradigm for non-specialist computer use, the past 25 years has mainly been about polishing it. Why it is so is of course an open question. From a more balanced and less cynical viewpoint one might say that the reason that PC/WIMP is still the ruling computing paradigm is because all other tried alternatives (although few have actually been proposed) have been less suitable for the tasks at hand. The PC/WIMP paradigm was simply a great design from the start. Another reason might be that there has been no series of inventions in the field of electronics (corresponding to the ones in the early days) that could inspire new ways of interaction. Another important factor is the enormous “paradigm polishing” industry that has evolved during the mentioned time period and doing everything it can in order to ensure that the (for them) very lucrative paradigm stays.

The end of the PC/WIMP polishing era?

Although the PC/WIMP interaction paradigm probably will continue to be the focus for the major part of the HCI community for another few years, recent advances in hardware technology and standardisation (in particular electronics miniaturisation, battery improvements, new sensor technology, and ubiquitous wireless network access) might soon trigger a paradigm shift just similar to how in the past the electronics development played an important role in determining the interaction paradigm of the early days of HCI research.

In fact, as portable/wearable computers, wireless networks and the new possibilities for making computers aware of events in the physical world emerged during the 1990’s, a growing number of academics within the HCI community left their “paradigm polishing” activities and began searching for an interaction paradigm that could better embrace the interaction possibilities enabled by this new hardware. Meanwhile, the established PC industry of course has continued to push for the use of the PC/WIMP paradigm also in this new technological context.

If the state-of-the-art hardware technology of today is enough for driving the paradigm shift, or if one or two more electronics inventions have to arrive first, is an open question. The increased interest in alternative interaction devices is an indication that the shift is at least approaching. There is of course also a slight possibility that the existing WIMP paradigm could be adapted and work well enough also for controlling computing in the new technological and social context as well. It is however unlikely, considering the basic assumptions of the WIMP paradigm such as

- the assumption that the human agent can dedicate all attention to the interaction with the virtual environment provided by the computer (e.g. does not bike or drive a car)
- the assumption that the real world environment in which the interaction takes place is always the same (quiet, in the shadow, etc.)
- the assumption that input and output devices are few, and the same, at all times (i.e. screen, keyboard, and mouse)

As acknowledged widely by researchers in proactive HCI areas (e.g. Augmented/Mixed Reality, Ubiquitous/Pervasive Computing, Graspable/Tangible User Interfaces, Wearable Computing, and Context Awareness) these assumptions do in general not hold, or severely limit interaction possibilities, in non-WIMP paradigm contexts.

EGOCENTRIC INTERACTION CORNER STONES

Our general aim is to design computer systems that facilitate everyday activities no matter if they take place in the physical world, the virtual world (mediated by some interactive computing device) or — which we believe will become increasingly common in the future — somewhere inbetween (Pederson, 2003). The technology advancements and increased interest in sensing and actuation technologies throughout the past 15 years makes this aim not only an

interesting academical challenge but also a realistic one from a more pragmatcal perspective.

The egocentric view on interaction we propose differs from more classical HCI models by building on the following corner stones:

1. **A physical-virtual design perspective.** Objects of interest to a particular human agent are modelled uniformly no matter if manifested in physical space, virtual space, or physical-virtual space.
2. **One human, no user interface.** Focus is on interaction between a single human agent and objects of interest to that human agent, specifically ignoring infrastructure that mediates virtual objects, such as personal computers.
3. **Strengthening connections between physical and virtual artefacts.** Focus is on the role of potentially automatic interplay between a) physical and virtual objects [residing in] b) physical and virtual environments [making up] c) the physical and the virtual world.
4. **“Applications” emerging by doing.** Modelled and infrastructurally supported physical-virtual environments (“applications”) are just as likely to emerge through everyday egocentric interaction by specific human agents, as being pre-defined by external physical-virtual environment designers.
5. **Support for living rather than just work activities.** The aim is to support personal everyday activity 24 hours, 7 days a week, without drawing the classical border between work and leasure, home and out.

The term ‘egocentric’ has been chosen to signal that it is the human body and mind of a specific human individual that (sometimes literally) acts as a centre of reference to which all interaction modelling and activity support is anchored.

Motivation for Corner Stones 1 & 2

What matters for human agents when manipulating objects for the purpose of performing well-known activities is the changes made to objects of interest (henceforth called *domain* objects), whether it is about calculating next years project budget in a spreadsheet application or knitting a pullower. What does *not* matter is the way the domain objects are accessed. E.g., expert PC users do not have to pay attention to how the physical mouse corresponds to the virtual arrow. Instead, they pay attention to the effects of their virtual activities as if they were inside the virtual environment themselves. Dragging a virtual document to the virtual trash can is as natural as throwing the corresponding paper document in the physical trash can below the desk. Based on observations such as these, the physical-virtual design perspective [2] suggests that physical and virtual domain objects should be as central for physical-virtual designers as they are for the users of those designs and that many classical HCI user interface concerns are of less importance if not completely ignorable.

Motivation for 3

One dimension of the physical-virtual gap has to do with the lack of causality crossing the divide [2]. State changes of an object in the physical world typically does not have any effect on a corresponding object in the virtual world, or vice

versa. Infrastructure keeping corresponding physical and virtual artefacts in synch would bring the two worlds closer to each other.

Motivation for 4 & 5

The space of possible activities is much larger in the physical (real) world compared to in the virtual world. It is harder as environment designer to have detailed control over human activity (and the way it is performed) in the physical world compared to the almost complete control offered when designing environments and behaviour in the virtual world.

The dynamics of “physical-virtual applications” caused by the extreme mobility (in physical space) of virtual environments, compared to the immobility of physical environments. The extreme mobility of virtual environments, paired with ever cheaper virtual environment providers (computing devices) makes it technically possible to provide virtual environments in almost any physical situation, not just the typical office setting. Thus, the design space for activity support more or less automatically expands from classical work settings to include just about any kind of human activity thinkable.

THE PHYSICAL-VIRTUAL DESIGN PERSPECTIVE

Computers, embedded in the “background” as well as more obtrusive artefacts (e.g. PCs, PDAs, cellular phones), play an increasingly important role in human activity. However, there are still things that most people would prefer to do “off-screen” in the physical (real) world, such as having parties, reading long text documents, or spending vacation. I argue that there exists a class of activities that are neither physical or virtual, but “physical-virtual” [2]. People frequently do parts of an activity in the physical world (e.g. proof-reading a text document under construction) and parts in the virtual world (e.g. adjusting paragraphs within “the same” document in a word processing environment). This behaviour is likely to become more common. Hence, future environments should be designed with such *physical-virtual activities* in mind.

The Classical User Interface Can Be Ignored

The proposed physical-virtual perspective is a way to deal with the gap between the physical and the virtual world¹, and to facilitate the exploration of designing information technology for helping human agents bridging it. The assumption is that a reduced physical-virtual gap means less “friction” for physical-virtual activities. Physical and virtual space is modelled together, and automatic mechanisms for synchronising related phenomena in both worlds are imagined to be offered by systems that have been developed with the physical-virtual perspective in mind. By viewing the physical and virtual worlds as one, we believe the chance to *make* them one increases.

Adopting the physical-virtual perspective involves abstracting away the classical HCI concepts of input and output devices, giving them a background role as Inter-World Event Mediators (IWEMs). Fig. 1 and 2 illustrate the basic

1. Dimensions of the gap between the physical and virtual worlds are explored in [2].

cases of physical and virtual human action (object manipulation). IWEMs are shown as white squares.

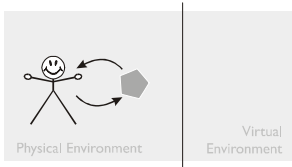


Fig. 1. Physical action [2]

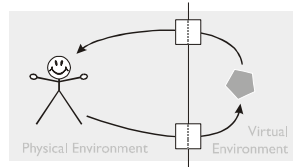


Fig. 2. Virtual action [2]

In order to arrive at a definition of physical-virtual activity I have found it useful to define human border-bridging activity on a lower level of abstraction first:

Definition 1: A *physical-virtual action pair* consists of two actions belonging to the same activity and often time-wise adjacent, where the first action is constrained (by lack of action support in the current environment) or chosen (e.g. based on individual preferences) to be performed in the physical world and the other action is constrained/chosen to be performed in the virtual world, or vice versa. [2]

Physical-Virtual Artefacts

Among physical-virtual action pairs we can sometimes identify one or several information-mediating objects that are subject to indirect or direct human manipulation in *both* actions, objects that transcend the physical-virtual border by being present in both worlds. Such objects are referred to as *Physical-Virtual Artefacts (PVAs)* and for denoting the presentations of them in the two different worlds, the term *PVA manifestation* is used. A text document presented in both the physical (e.g. printed on paper) and the virtual world (e.g. within a word processing environment) would serve as a good example of a PVA, where each manifestation affords different kinds of manipulation.

Definition 2: A *physical-virtual action* is an action on a PVA where both the physical and virtual manifestations are directly controlled and/or monitored by the agent. [2]

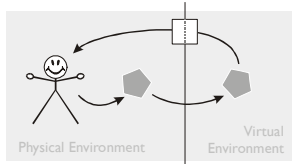


Fig. 3. Physical → virtual action [2]

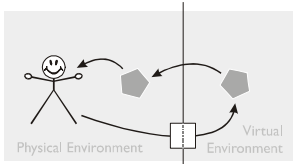


Fig. 4. Virtual → physical action [2]

Fig. 3 and 4 illustrate two possible kinds of physical-virtual actions. Finally, the concept of physical-virtual activity is defined as follows:

Definition 3: A *physical-virtual activity* is an activity consisting of a sequence of actions containing a) at least one physical-virtual action pair or b) at least one physical-virtual action. [2]

One Space, One Magnifying Glass

By viewing the physical and the virtual worlds as equally important for human activity, the proposed physical-virtual perspective makes terms tied to an implicit virtual-world bias such as “context awareness” obsolete. It also expands the meaning of “location tracking” (currently having an implicit physical-world bias) to include also space and place in the virtual world. It invites the viewing of the relationship between physical and virtual environments from

unconventional angles. For instance, why should not the current (local) state of the virtual world influence how activity in the physical world is interpreted? Could it not be of use for physical-world “applications” to be aware of their virtual-world context? And why is virtual-world location tracking (e.g. web pages browsed by a human agent) not considered when designing interactive environments?

EGOCENTRIC INTERACTION

The egocentric interaction perspective is based on a situative model of what a specific human agent can see and not see, reach and not reach at any given moment in time (Fig. 5). The model is based on the physical-virtual design per-

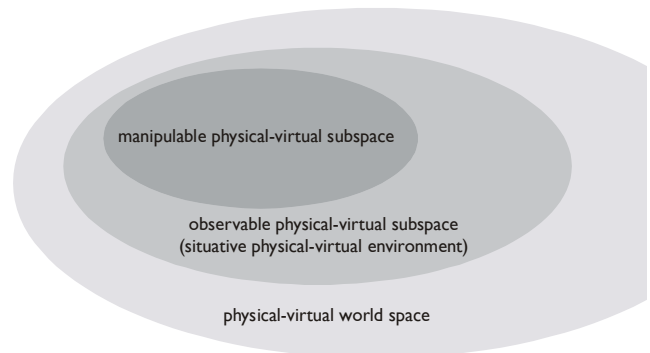


Fig. 5. A situative physical-virtual space model [2].

spective briefly outlined in the previous section. Thus, physical and virtual domain objects are treated as being located in the same space. As a specific human agent changes physical and/or virtual location, objects come into and leave the observable physical-virtual subspace in a dynamic fashion. Somewhat simplified, one can say that it is the borders of the observable subspace which defines the set of objects that can possibly be part of a physical-virtual “application” at any given time-instant for the specific human agent.

The idea of using computers for assisting individual human agents in everyday life is not new but has gotten increased relevance in the last 5-10 years because of increased capacity of mobile and wearable devices. One example is the research performed at Georgia Tech investigating the possibilities in creating an always present, context-aware “digital assistant” [6]. The egocentric view differs from their and most other similar “intelligent assistant” approaches, by focusing on detecting presence of physical (and virtual) objects rather than places or persons, for detecting and contextualizing human activity. The approaches are, of course, complementary in this respect. However, as mentioned earlier, by taking a world-neutral physical-virtual design stance, the egocentric view on interaction differs from most existing HCI modelling approaches by not seeing the state of the real world as merely *context* to virtual activity but an *inseparable part* of it.

Computing Infrastructure for Egocentric Interaction

The egocentric approach follows the current HCI trend, breaking with classical Task Analysis that assume human agents to perform all actions based on rational decisions for reaching well-defined goals most efficiently. Egocentric

computing systems do not necessarily have to actually *know* what the modelled activity is about but rather what the human agent seems to need (in time and space) in order to perform it, mainly based on historical data of object use. Thus, *emerging* individualised physical-virtual “applications” rather than traditional *pre-defined* general-purpose ditto designed by application designers.

Being a relatively abstract perspective on future HCI, the egocentric view does not assume the use of any particular kind of technology for supporting interaction between the human agent and the physical-virtual world. For instance, computing and sensing technology for tracking physical activity of a specific human agent could be imagined to be either worn by the human agent herself, or be situated in the surrounding physical environment. The same goes for virtual environment providers (computing devices providing access to the virtual world) which could be both worn by their user or ubiquitously distributed throughout the physical environment like in Mark Weiser’s vision [7].

For reasons of privacy, efficiency, design complexity, feasibility, and cost, we have found an implementation approach based on wearable sensing and computing power most attractive. The basic idea is to make the wearable egocentric computing system as self-sufficient as possible, reducing the problem of “uneven conditioning” [4]. Information about activities performed using devices external to the wearable egocentric interaction system (e.g. in the case when the human agent is using a desktop PC) need to be transmitted through some standard wireless communication protocol to the wearable egocentric interaction device for analysis. Complemented with real-world object manipulation information from sensors, the egocentric interaction system would (at least in theory) be able to gracefully model activities across the physical-virtual gap.

As an example of a useful application, wearable egocentric interaction technology has the potential of ensuring that human agents always have the necessary physical and virtual objects at hand for successfully performing the activities they like to. Such systems can act in the conscious “foreground”, reminding the wearer to bring this or that physical object along when changing physical location, or in the “background” by setting up physical-virtual environments prior to the human agent’s arrival, making the emerging physical-virtual world a slightly smoother place to be in.

APPLYING THE FRAMEWORK

As is common in the explorative branches of HCI, the interaction theory presented in this paper has been and will continue to be developed hand in hand with prototype systems. The aspects of the egocentric interaction framework related to the physical-virtual design perspective have inspired and been inspired by the development of the Magic Touch system [3; 1] which among other things provided limited support for end-user design of Physical-Virtual Artefacts based on wearable position and object identification sensors.

easyADL

A more extensive application of the egocentric interaction framework is currently undertaken within the easyADL project where the focus lies on recognition and support of everyday activities based on the situative physical-virtual space model (Fig. 5). easyADL is a two-year research project started in June 2005 investigating new forms of computer technology for facilitating everyday life of individuals suffering dementia disease¹. The goal is to develop a wearable computer-based “cognitive prosthesis” which to some extent could compensate for mild cognitive impairments when performing everyday tasks such as getting dressed, preparing dinner, and going to the bathroom.

In order to speed up the design process, and in order to compensate for limitations with currently available sensing technologies, Virtual Reality models are used to simulate wearable real-world sensors, feeding “clean” data to the activity modelling algorithms under investigation. The method also facilitates the experimentation and comparison between different wearable sensor configurations for successful recognition of activities based on a set of objects nearby and/or manipulated by the human agent within a given time span.

CONCLUSION

This position paper has presented an everyday activity modelling framework (under development) based on automatic recognition of basic object manipulation performed by a specific human agent in a joint physical-virtual space.

REFERENCES

1. Pederson, T. (2004). Magic Touch. In Video Proceedings of UBICOMP2004, Nottingham, UK, ACM Press.
2. Pederson, T.: From Conceptual Links to Causal Relations — Physical-Virtual Artefacts in Mixed-Reality Space. PhD thesis, Dept. of Computing Science, Umeå university, report UMINF-03.14, ISSN 0348-0542, ISBN 91-7305-556-5 (2003)
Permanent URL: <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-137>
3. Pederson, T.: Magic Touch: A Simple Object Location Tracking System Enabling the Development of Physical-Virtual Artefacts in Office Environments. Journal of Personal and Ubiquitous Computing 5 54-57, Springer Verlag (2001)
4. Satyanarayanan, M. (2001). Pervasive computing: vision and challenges. *IEEE Personal Communications* 8(4):10-17, 2001.
5. Sellen, A. J., & Harper, R. H. R.: The myth of the paperless office. MIT Press (2002)
6. Starner, T. (2001). The Challenges of Wearable Computing: Part 1 & 2. *IEEE Micro* 21(4), 44-52 & 54-67.
7. Weiser, M. (1991). The Computer for the 21st Century. *Scientific American*, 265(3), 66-75.

1. <http://www.cs.umu.se/research/easyadl>

Core Tangibles and Tangible Visualizations: prospects for tangible convergence and divergence

Brygg Ullmer, Louisiana State University
Department of Computer Science and
Center for Computation and Technology (ullmer@lsu.edu)

ABSTRACT

We describe efforts toward “core tangibles” and “tangible visualizations.” These combine both convergent and divergent approaches toward the advance of tangible interfaces as a practical, useful approach for computationally-mediated interaction. We also describe several key enabling electronic, programmatic, and web-based tools.

INTRODUCTION

Both academics and the popular press have long celebrated prospects for technological “convergence.” In 2006, whether one views set-top boxes, cell phone/mp3/camera/PDAs, or beyond, convergence and functional integration are now mass-market realities. Further progressions toward “smaller, faster, cheaper, and more of them” seem inevitable [2].

In parallel, prospects for *divergence* and *dis-integration* (functional decomposition) in the forms and interfaces of computationally-mediated systems are also gaining traction. Buxton makes the comparison between multi-purpose Swiss Army knives and “simpler specialized devices, each of which may well be kept in a different location,” noting that the latter is often preferred in daily practice [2]. The site [7] presents more than 300 appliances, vehicles, wearables, and beyond which are based upon embedded Linux systems, but externalize their interfaces in forms which often cannot be recognized as computer consoles or other WIMP GUIs.

User interfaces like DataTiles [11] have begun to illustrate concrete approaches by which both functional decomposition, and grammatic composition of the resulting subelements, can be realized. Also, efforts such as [1, 3–5] consider prospects of divergent, decomposed interfaces for the critical usage context of colocated collaboration. We believe such examples illustrate the emergence of rich, diverse, complementary ecosystems of both convergent and divergent computationally-mediated devices and systems.

Our research centers on applications of tangible interfaces for visualization, simulation, presentation, and education, often toward collaborative use by scientist end-users. Both

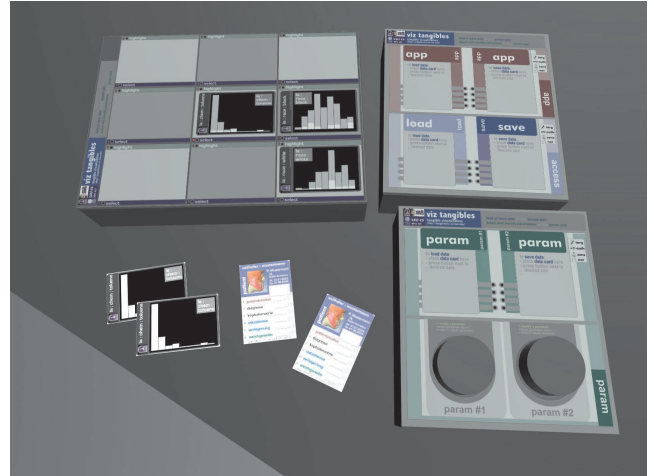


Figure 1. Core tangibles: illustration of several interaction devices used in combination with RFID-tagged “data cards” to collaboratively interact with a visualization (Figure 2). The right devices are “pages” composing two “interaction pads” (now replaced with four 10x10cm “interaction tiles”).

within these contexts and more broadly, we see convergence and divergence as critical, interdependent themes for tangible interfaces and other physically-situated interaction techniques. In response, we are developing two complementary approaches: *core tangibles* and *tangible visualizations* – complemented by several key enabling technologies.

Core Tangibles

Most tangible interfaces have existed as isolated islands. Inspired by Durrell Bishop [10], we see open-ended, ad-hoc, and exploratory combinations of diverse interoperating digital artifacts as among the most promising prospects of tangible interfaces. To date, this has rarely been possible.

Moreover, many common, critical operations within GUIs – e.g., opening and closing applications; loading, saving, and navigating data; and adjusting parameters – have generally remained inaccessible from tangible interfaces. Instead, users have been required to “leave” the TUI to interact with GUI consoles or command shells; to rely upon a supporting human operator; or to settle for highly limited applications.

In response, we are developing “core tangibles” [14]. Core tangibles support “core operations” which are common across a broad spectrum of applications, including data + ap-

plication selection and invocation; parameter manipulation; discrete and continuous navigation; and spatial placement. We currently embody core tangibles as modular, composable tiles, building on an approach introduced by DataTiles [11].

We anticipate core tangibles serving several major roles:

- enabling modular, composable, open-ended re-use of core operations across diverse applications;
- providing “in-band” access to core operations, rather than requiring dependencies on GUIs or human operators.
- supporting collaboration among multiple colocated and/or distributed participants, by decomposing the interface into multiple discrete, semi-independent display + control components, which are coordinated by shared graphical, audible, and haptic displays.

Core tangibles illustrate a kind of “convergence” in the area of tangible interface design. On the one hand, individual core tiles are specialized, both in their faceplate controls and underlying function. Simultaneously, they embody “convergent” functionality that is common across many applications, and potentially can be used in combination with many different kinds of interfaces (including GUIs, AR, VR, speech/gesture UIs, et al.).

Tangible Visualization

While core tangibles address a kind of “tangible convergence,” we are even more strongly invested in “divergent” tangible interfaces that are more specialized and physically representational in form. Specifically, we are developing a class of highly specialized, divergent interactive systems that we call “tangible visualizations.”

Tangible visualizations can be seen as “physical descriptions of complex systems.” Our approach seeks to realize systems of specialized interaction devices which physically represent specific data, operations, and application domains. These are often intended for collaborative use (both colocated and distributed); and often in service of visualization, simulation, presentation, and education.

We continue by describing two usage scenarios for core tangibles and tangible visualizations. The first is an approach for collaborative visualization; the second is an interactive physical description of a major network infrastructure and its associated resources. We briefly discuss related work, and introduce several key electronic, programmatic, and graphical enabling technologies.

SCENARIO 1: COLLABORATIVE VISUALIZATION

One of our main driving applications is support of interactive, collaborative visualization of scientific content. Computation had made a major impact on the pursuit of science. Data acquisition/digitization, simulation, and computational analysis are routine across many domains. In each case, visualization takes on a critical role for meaningfully engaging with large quantities of data.

Collaboration is also critical to many aspects of science. It may take place between peers with aligned or complementa-



Figure 2. AccessGrid-linked meeting; core tangibles in use (co-located and distributed users collaboratively manipulate a visualization application. Five sets of tangibles (three local, two remote) are used together with data cards to load and save data, manipulate parameters, etc. Most functionality is operational, but image is composited.)

ry expertise; between instructors and students; and in outreach with the general public, among other contexts.

We are working with a team of diverse scientists (physicists, chemists, biologists, ecologists, et al.) who are actively engaged in collaboration and dependent upon visualization. Their collaborations are both colocated and distributed (spanning multiple campuses and continents). They are heavy users of roughly a dozen “AccessGrid” video conferencing rooms spanning as many different sites. Each room has two or three wall-projections, 6-10 pan-tilt-zoom cameras, and network connections ranging from 100Mbps to 10Gbps (Figure 2a).

In meetings within AccessGrid rooms, “interaction” with visualizations is almost exclusively conducted using two software platforms: distributed PowerPoint and movie players. Generally, all interaction is controlled directly by the dedicated AccessGrid operators (who typically must simultaneously manage ~5 computers, 5-10 video links, and numerous applications). Thus, these visualizations not only fall short of collaborative control; often, they are not even under direct control of the presenters. While this is the most sophisticated and heavily-used video conferencing setup we have encountered, no one is satisfied with current visualization support.

We have received funding to develop a new approach for collaborative visualization, and to deploy our system and bring it into daily use in more than a dozen AccessGrid (AG) rooms. We have made three successful tests of our early hardware and software in meetings spanning three sites – Baton Rouge; Seattle/San Diego; and Brno (Czech Republic). These were linked by hi-def, low latency video conferencing (three channels video, one of shared 3D visualization) over dedicated 10Gbps network connections; all interaction with visualizations was controlled by our tangibles.

We anticipate interaction with core tangibles in this context to take roughly the following form. Each participating site will project one or (usually) more camera-fed remote video links; one or more shared visualizations; and a viz/interaction status display (Figure 2a,b). The AG operator will continue (at least initially) to maintain control of the video links, while control of the visualizations will shift to the conference participants (we expect to deploy 2-4 sets of tangibles within each AG room).

All conference participants will have the ability to load shared visualizations; alter both visual (rotate/zoom/...) and functional (time step/slice #/...) parameters; capture screen shots; and perform other operations via the core tangibles. In a conference with three sites, each with three sets of tangibles, this means that up to ~nine different participants can potentially take interactive control of the shared visualizations. (We expect the actual number of involved participants will be more, as multiple can share a single set of tangibles.)

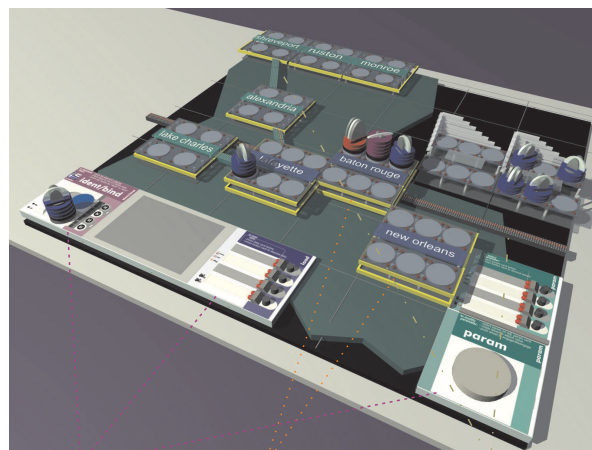
We anticipated and have observed that maintaining awareness and harmonizing collaborative interactions is both a challenge and potential strength of our approach. We are addressing collaborative control in several ways, including automatic software “clutching” to mediate handoff; visual feedback, both on shared projection displays and on the interaction devices themselves; audible feedback; and force feedback.

SCENARIO 2: LONIWORKS

The above scenario describes the use of “convergent,” widely-applicable interaction devices. In contrast, our “LoniWorks” scenario centers upon “tangible visualizations:” special-purpose, “divergent” interaction devices which take the form of physical models specific to particular problem domains. Simultaneously, LoniWorks itself depends upon core tangibles, illustrating interdependence between convergent and divergent interaction approaches.

LoniWorks’ target domain is the Louisiana Optical Network Initiative (LONI). LONI is a \$40M, 10-year project to deploy a 40Gbps network backbone and compute infrastructure across Louisiana, linked with the national NLR infrastructure. LONI has many research, commercial, and public service implications. E.g., terabyte-sized Hurricane Katrina recovery datasets which took roughly a week to transfer could be transmitted with LONI in roughly 4 minutes.

However, LONI is a complex infrastructure. Its capabilities and behavior are unlike existing infrastructure; e.g., in contrast with mainstream Internet carriers, it is based on band-



core tangibles: basic operations such as load, identify, bind, modify params
embedded linux pda: allows limited graphical interaction with several core operations (e.g. ID and bind tokens to machines, jobs, etc.)
network nodes: nodes for each LONI terminus. each node includes array of pads for placing
resource tokens (representing super-computers, vidconf, ...)
network paths: LED matrix (4 wide, one LED per network lambda); red = LONI, blue = NLR

Figure 3. LoniWorks: tangible visualization of network. (high-resolution; PDF of image can be viewed full-screen)

width reservations and dedicated channels. This makes it potentially difficult to understand and control by its end users: research, commercial, and public service users who are not themselves network specialists.

We have begun developing a tangible visualization of LONI, providing a physical, manipulable model of the infrastructure, its network traffic, compute jobs, videoconferencing, and other activities (Figure 3). We believe this will be valuable in helping a diverse audience to collaboratively understand, use, plan, and manage this critical resource.

For example, consider Figure 3. Distinctively shaped tokens represent major LONI-linked resources. Specifically, major compute clusters are represented. The number of layers represent the number of CPUs \log_4 (16 processors = two layers, 1024 processors = five layers, etc.), allowing diverse resources to be physically compared. Similarly, the height and width of arced token tops represents node and system RAM (which can be more important than CPUs for computations such as bioinformatics). Colors code for different vendors. Resource tokens are embedded with RFID tags, allowing their placement on different network nodes to be monitored.

While the forms of these tokens are still undergoing refinement, and important supporting textual annotations are not shown, we believe they will allow both technologists and the broader public to qualitatively distinguish, compare, and interact with diverse compute resources. (We are also developing tokens representing video conferencing facilities and other major LONI resources.)

Many tasks (e.g., hurricane and coastal erosion simulations) involve coordinating multiple computation resources and multiple teams spanning the state. Manipulating these resource tokens on the LONIworks interface allows users to marshal and configure resources; monitor network traffic on the LED-illuminated network paths; query and browse logs of past LONI utilization and scheduled reservations; and perform other key tasks. While supporting only a fraction of the operations accessible to a network professional at a traditional computer console, the interface should be far more accessible for education and end-user interactions.

RELATED WORK

The background for core tangibles is described in [14].

Broadly viewed, the tangible visualization approach extends into prehistory. From before the era of digital computing, we find precedent in board and card games (with a 5,000 year history); the memory boards and Tjuringa of Africa and Australia; the early mechanical controls for the Panama Canal; and mimic boards and map boards (once used widely in industrial and transportation control centers, later in tiled forms).

The term “tangible graphics” has long been used in the blind community, as a descriptor for diagrams in physical relief which support tactile inspection. Within HCI, the term “tangible visualization” was perhaps first used in 2001, in conjunction with three related papers. Strata/ICC [12] and preceding work [13] lead directly into the present work. Patten et al. [9] used the term “tangible visualization” to describe the use of a general-purpose tangible interface to interactively manipulate graphical visualizations. Also, Marti et al.’s [8] “WeatherTank” created a “tangible visualization” as a kind of vivarium which potentially could be mapped to illustrate the dynamics of diverse information feeds.

ENABLING TECHNOLOGIES

Our work is closely linked with the development of several key enabling technologies. While mostly beyond the scope of this short paper, we introduce these briefly below.

Electronics

The core keyboard/pointer/display interaction model underlying the GUI/WIMP paradigm has remained relatively stable for roughly 30 years. In contrast, tangible interfaces typically force developers into an intimate relationship with electronics, mechanics, and firmware. Given the central role of physical form in TUIs, we believe the need for their designers to possess rudimentary electronic and mechanical fluency is likely to remain. However, relatively few HCI research labs have the human resources to deal with major electronics and mechanics development.

Moreover, the need to build up each TUI from the lowest level of abstraction constrains the complexity of what can reasonably be designed. In particular, we believe the “tangible visualization” approach ranges from impractical to implausible, if each system must be developed from the microcontroller and resistor level.

In response, we are developing “function blades,” which expose sensing, display, communications, and computation capabilities in a consistent, modular, extensible, easily employed fashion. These blades are each 10x2x1cm in size, corresponding to the 10cm modal length defined by human’s palmar hand grasp posture. Each blade has two 14-pin (7x2) DIP connectors. One connector is identical across all blades, and provides power and communications (both USB and I²C). The second connector’s function is entirely specific to the individual blade.

We use these blades extensively in both the tiled forms of core tangibles (which each accommodate 5-10 underlying function blades), and within our tangible visualizations. Non-blade electronics (e.g., Phidgets and Particles) can also be embedded within tiles using their standard USB and I²C buses. We also use tiles themselves as a kind of functional aggregation and scope (from a software and functional perspective) in our design of tangible visualizations (Figure 3). In order to support the evolution of tangible interfaces, we are working to release a suite of blades, core tangibles tiles, and associated software as open source hardware, firmware, and software.

Software APIs

While we believe blades and bladed tiles could become an important technology tool for the development of tangible interfaces, diversity is a virtue, and tangibles using widely varying hardware/software toolkits should be expected. As we feel interoperability should remain a major TUI goal, resolving some form of hardware, software, and toolkit-independent device description seems highly important.

Toward this, we have begun working on several forms of XML-based device descriptions. One of these (in collaboration with Dr. Albrecht Schmidt) is at the hardware protocol level. This will allow diverse software to interact with underlying functions implemented by specific blades, Phidgets, Smart-Its, Particles, etc.

A second level of XML description relates to the composite functional roles of individual interaction devices. Here, we intend to expose an API which is largely independent of the underlying technical implementation. Thus, ideally an interaction device could be reimplemented with blades, Phidgets, etc., while leaving its external functional API unmodified. We are interested in finding ways to combine these efforts with others such as [6]. We are also exploring different kinds of registries and matching software for pairing interacting interaction devices together, including (e.g.) Prolog interactions with Grid-enabled SQL databases.

Finally, while we hope end-users will wish to interact with our tangible interfaces, we believe graphical interfaces and other interaction techniques will sometimes remain preferable. We also wish to map the same set of tangibles across various kinds of software implementations (e.g., visualizations using Amira, VTK, OpenDX, AVS, TechPlot, etc.). To support both of these ends, we have begun developing a “core operations API” (COA) and “InterViz API” (IVA) to descri-

be some of the high-level software functionalities we wish to control with tangibles and other interfaces.

Web-based interaction

We expect tangible interfaces will in time fill certain useful functional roles, but many forms of HCI will continue to be carried forth using graphical interfaces. Correspondingly, we feel it is important to support “interaction life cycles” that can gracefully interlink tangible and graphical interaction.

We have begun to develop graphical, web-based “invitations to interaction” and “interaction take-aways” which are closely coupled with our tangible interfaces. As one example, we have created software which creates “viz reports.” These are web-based compilations of annotated images which are automatically generated as the product of computer simulations of various physical processes.

In addition to providing a visual overview, these viz reports serve as a launch-point and terminus for interaction. We are developing web-based means by which users can quickly transfer selected images (corresponding to “checkpoints” of manipulable, fully interactive visualizations) into “data cards” for manipulations with our core tangibles. Similarly, we are developing software for our core tangibles which allows them to (e.g.) “screen shot” visualizations together with associated parameters; and automatically synthesize viz reports composed of this content.

ACKNOWLEDGEMENTS

A number of individuals and funding sources have contributed to this work. Graduate students Rajesh Sankaran, Karun Kallakuri, Srikanth Jandhyala, and Cornelius Toole have been critical co-developers. The core tangibles research was begun at the Zuse Institute Berlin in the group of Hans-Christian Hege, with collaboration by Werner Bengler and Stefan Zachow and support by the EC IST GridLab project. The Max Plank Institute funded the first deployment of core tangibles at MPG AEI; Christa Hausman, Ed Seidel, and Werner Bengler were critical toward this. The current deployment of core tangibles is funded by the NSF MRI (0521559) and NIH INBRE programs, through leadership by Drs. Stephen D. Beck, Harold Silverman, William Wischusen, and John Quebedeaux.

REFERENCES

1. Binder, T., De Michelis, G., Gervautz, M., Jacuci, G., Matkovic, K., Psik, T., and Wagner, I. (2004). Supporting configurability in a mixed-media environment for design students. In *Personal and Ubiquitous Computing*, v8n5, pp. 310-325.
2. Buxton, W. (2000). L'infographie en pleine Evolution. *Pixel*, 53 Septembre/Octobre, 14-15. In English at <http://www.billbuxton.com/Trends.html>
3. Hornecker, E., and Buur, J. (2006). Getting a grip on tangible interaction: a framework on physical space and social interaction. Accepted to *CHI'06*.

4. Inkpen, K., and Mandryk, R. (2003). Multi-Display Environments for Co-located Collaboration. In *Proc. of ECSCW'03*.
5. Ishii, H., Underkoffler, J., Chak, D., Piper, B., Ben-Joseph, E., Yeung, L., and Kanji, Z. (2002). Augmented Urban Planning Workbench: Overlaying Drawings, Physical Models and Digital Simulation. In *Proc. of ISMAR'02*, pp. 203-212.
6. Leland, N., Shaer, N, Jacob, R. (2004). TUIMS: Laying the Foundations for a tangible user interface Management System In *Proc. of Pervasive'04*.
7. The Linux Devices Showcase. (2005). <http://www.linuxdevices.com/articles/AT4936596231.html>. Visited Dec. 2005.
8. Marti, S., Seetharam, D., and Ishii, H. (2001). A Tangible Interfaces using Weather Metaphors. http://web.media.mit.edu/~stefanm/tangible/WeatherTank_short.20
9. Patten, J., Ishii, H., Hines, J., and Pangaro, G. (2001). Sensetable: a wireless object tracking platform for tangible user interfaces. In *Proc. of CHI'01*, pp. 253-260.
10. Polynor, R. (1995). The Hand That Rocks the Cradle. *I.D.*, May/June 1995, pp. 60-65.
11. Rekimoto, J., Ullmer, B., and Oba, H. (2001). DataTiles: A Modular Platform for Mixed Physical and Graphical Interactions. In *Proc. of CHI'01*, pp. 269-276.
12. Ullmer, B., Kim, E., Kilian, A., Gray, S. and Ishii, H. (2001). Strata/ICC: Physical Models as Computational Interfaces. In *Extended Abstracts of CHI '01*, pp.373-374.
13. Ullmer, B. (2002). Tangible Interfaces for Manipulating Aggregates of Digital Information. Ph.D. dissertation, MIT Media Laboratory.
14. Ullmer, B., Jandhyala, S., Sankaran, R., Kallakuri, K., and Toole, C. (2006). Core Tangibles: Composable Physical Artifacts for Common Operations across Diverse Applications. Under review by *DIS'06*; online at <http://csc.lsu.edu/~ullmer/papers/ullmer-dis06sub-coretangibles.pdf>

Computational Corporeality: *Computers, Materials, and the Body*

Amanda Parkes
Tangible Media Group
MIT Media Lab
77 Massachusetts Ave.
Cambridge, MA 02139, USA
amanda@media.mit.edu

INTRODUCTION

At their core, computers are design tools, they allow us to expand our ability to create, think, organize, and learn. The systems which we build to interface with the digital data they provide can be considered in the the same paradigm of all tools which we use for non-digital tasks in our daily lives. However, the complication of a human-computer interface lies in the fact that it provides both the raw material and the functionality of use in the same space and of the same material. There is no longer the distinction between organic materiality and the discrete function of the tool with which we manipulate it. This can allow for remarkable advantages in ease of use but also negates many experiences of bodily intuition and feedback.

As Tangible User Interfaces [1] become more refined and sophisticated, the future of human computer interaction lies in finding a bridge between the organic structures in the natural world which we seek to emulate and manipulate and the rigidity and exactness of the digital systems which provide the tools for manipulation. As Ellen Lupton comments, “organic forms and materials provide designers with a humanist vocabulary that affirm society’s place within the natural world.” [2] The constructed and the organic are converging, and the digital materials and tools in development should address this phenomenon by providing an organic material means to engage the tactile senses in the act of creating and modeling.

The rise of ubiquitous computing has brought about the development of innovative systems involving a multiplicity of small computers embedded in everyday objects and the surrounding environment, we are no longer constrained to think of computers as a box on a table. By combining the notion of ubiquitous computing with the approach of direct manipulation, to improve the ‘directness’ and ‘manipulability’ of an interface [3], the possibility for a new class of interaction tools and materials emerges. This new class of materials has its basis in the interaction techniques and tools of Tangible User Interfaces, designed to give physical form to digital information. In order for human-

computer interfaces to reach a more sophisticated state, they must perform, respond and react in ways that mimic the body and human behavior, not just, or necessarily in their intelligence, but in their materiality.

I have identified four areas which are important for transitioning computer interactions to a more embodied state.

INTERFACING WITH THE BODY

Design has taught us that material choice changes the experience of any interaction. However, interactions with computational systems remains largely disembodied, constraining the digital data to be part of an ‘other’ or external object. The incorporation of different materials that mimic organic, softer bodily qualities (eg. silicones, textiles) as well as actuation technologies that mimic natural motion (eg. artificial muscle and shape memory alloy deformation) can expand how we interact with digital data and the kind of functionality which computers can offer us.

DISCRETE VS. CONTINUOUS: MODULARITY AND PRIMITIVES

All natural materials are in essence created of discrete components which combine to create continuous systems. The importance of modularity in creating a material language is evident in our environment from the microscopic level, such as building blocks of biological systems (cells) or chemistry (atoms) to the architectural level, such as refabricated panels. Modularity leads to systems of physical primitives, grammars forming the basis of constructive assembly systems. Our physical interface design should be informed and guided by this model.

PHYSICALITY IN THE FEEDBACK LOOP

The materials of our keyboards, the haptic feedback of our buttons, for example, define our computer experience as much as the pixels on the screen which we control. However, this phenomenon is rarely addressed in interface design. The notion of extending the feedback loop of our interfaces to include an input from the physical behavior

of the materials we control and manipulate. can tie our interactions to the natural world and the intuition of our bodily experience.

FORM FOLLOWS FUNCTION FOLLOWS INTERACTION

Physical form can no longer just be considered just to prescribe the function of an object. With multilayered interfaces, and multifunctional devices, the demands of the form are complex and often times at odds with each other. A camera has very different ergonomic considerations than a phone, however they are increasingly incorporated into the same device. The challenge for designers is to look first at the interaction and ascertain how this affects the function which will in turn determine the form. Digital and physical modularity can offer some options to these paradoxical problems, but we must first put interaction first in the design process and have functionality follow.

REFERENCES

1. Ishii, H. and Ullmer, B. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. *Proceedings of CHI 1997*, ACM Press, (1997), 234-241.
2. Lupton, Ellen. Skin: New Design Organics. in Skin: Surface, Substance and Design. New York: Princeton Architectural Press, 2002.
3. Fitzmaurice, G. Ishii, H. and Buxton, W., Bricks: Laying the Foundations for Grapable User Interfaces. *Proceedings of CHI 1995*, ACM Press (1995).

The future of interface design, through a child's eyes

Hayes Raffle

Tangible Media Group

MIT Media Lab

77 Massachusetts Ave.

Cambridge, MA 02139, USA

hayes@media.mit.edu

Abstract

One way to try to predict the future of UI design is to look at the new ways children are doing design today. There is strong evidence that education in the 19th century shaped the arts and design of the 20th century, and I believe that today's pedagogy that uses computers will be a cornerstone in future applications of interactive systems. When we consider UI design, we are thinking of the computer as a design tool, and I will approach the future of UI design with a brief overview some trends that present computerized systems and tools to encourage children to be designers.

Introduction

Kindergarten was invented by Frederick Froebel in the 1830's, a crystallographer heavily influenced by Pestalozzi and Goethe. His educational philosophy emphasized unity in the natural world and fostered an understanding of one's role in society and nature. His pedagogy used activities and specialized objects called "gifts and occupations" (figure 2) to explore relationships between divergent natural phenomena [2].

Many of the greatest artists and designers of the 20th century were among the original Kindergartners. Klee, Kandinsky, Mondrian, Braque, Frank Lloyd Wright, Buckminster Fuller, Le Corbusier, Mies van der Rohe and Gropius all either taught kindergarten or were students of kindergarten. We find that some of their mature work was almost identical and untransformed from the exercises, strong evidence that their early learning shaped later aesthetic practices. The Bauhaus and the Modernist movement, cornerstones in 20th century art and design, were directly informed by Kindergarten. These movements later helped inform the development of UI design when researchers sought to make interfaces that lay people and professionals could use for art and design practices.

Our next generation of designers will be influenced by the technologies they have today, and by design-oriented interfaces that are currently being developed in both research and industry.

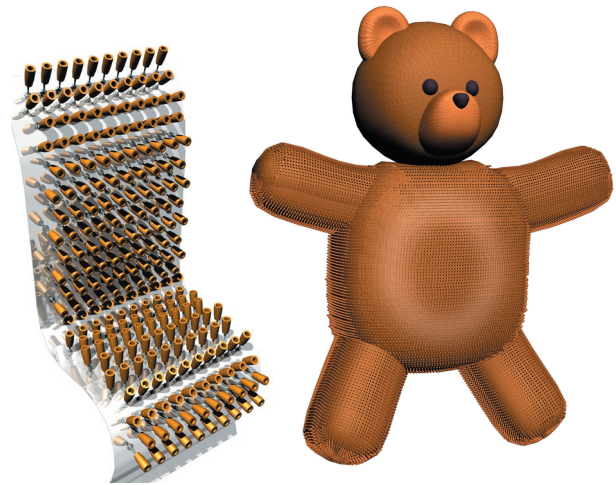


Figure 1. Super Cilia Skin conceptual rendering. A material with kinetic memory could someday be used to design children's toys to engage emotions and support learning.

Child as programmer

In the 1970's Seymour Papert, a student of Piaget, argued that children learned best when they were actively designing things. He viewed the computer as a tool for children to understand how they think and how the world around them works, and invented the LOGO programming language for kids to create computer programs to do things like play music and draw pictures [8]. His work influenced mainstream education and led to the LEGO Mindstorms products, which were based on LEGO/LOGO research at the MIT Media Lab. This work generally used the dominant UI of the day, from command line interfaces to contemporary GUIs with drag and drop iconic programming languages. One goal is to help kids think in terms of systems concepts, and to test ideas through active construction of models. In the case of LEGO/LOGO, physical models are created with LEGO and computational models are created with a GUI and downloaded into the toy [14].

This work was part of a larger effort to create "digital manipulatives" that embed computation into familiar children's toys. The idea is that computers can make ideas about feedback and emergence more salient to kids, and the UI shifted towards making the physical toy a greater part of the "interface" [14].

Tangibles and learning

Tangible Interfaces assume that the physical object should be the primary interface to the program structure and behavior [6]. Where digital manipulative and tangibles intersect we find specialized programming paradigms that are developed to make a specific set of ideas accessible to children. Many of these projects build on classic play patterns, and use interactivity to allow children to design different kinds of dynamic systems.

For example, “flow blocks” and “system blocks” can be physically connected and adjusted to create different dynamic patterns of light and sound. They are somewhere between a marble chute building toy and real dynamic modeling, where kids can explore some of the effects of feedback and cyclic behavior in flow-based systems. Physical models are built, parameters are adjusted and the behavior is observed to change in different ways [17].

Other projects have used organization of blocks to represent program structure. For example, Wyeth’s blocks allowed young children to construct logical models and see their models respond to conditional and feedback behaviors [16].

Record and play has been explored in different ways to explore different modalities, including sight, touch, and sound. Since this programming model has been argued to facilitate computational design for young children, I will briefly overview some projects that use record and play.

Sound: Tell Tale was a caterpillar whose body segments could record different audio clips. When arranged in different sequences, a story could be played different ways [1]. Storymat [3] recorded children’s spoken stories as



Figure 2. The original kindergarten “gifts” emphasized design using geometry and craft techniques.

dolls were moved around on a special mat that could record their position. When dolls revisited those areas later, the stories would play back. Where one project used physical construction, the other used spatial mapping for recordings.



Figure 3. With IO Brush, children can paint with the color and movement of any object.

Light and sound: IO Brush (figure 3) is a paintbrush with an embedded video camera that can record still or video images and sound, and paint the images and sound onto a large plasma display. Children can create paintings with familiar colors from the environment, make animations, or hide stories in a painting that are released when a child touches different parts of the painting [15].

Movement: Curlybot is a small object that can record a playback its movements on a table top. Children used curlybot to tell stories, or explore geometry with the toy’s looping playback [5].

Movement and form: Topobo (figure 4) is a constructive assembly system, also with kinetic memory, that children can use to build and animate different kinds of creatures. For instance, a child can make a moose and teach it to walk by twisting its body around. The moose will then walk on its own. The same way kids learn about buildings by playing with blocks, they can learn about animal locomotion by playing with Topobo [12]. The creators of Topobo also developed different tangible interfaces to modulate recordings, as well as to sample and sequence playback motions.

Touch and sight: Super Cilia Skin (figure 1) is a textural interface that can record and playback the movements of an array of hair-like actuators. The designers proposed that a kinesthetic fabric could be used in children’s plush toys as an interface to computational behavior. The argument was that the subtle, organic qualities of textiles were the qualities that helped children form personal emotional

connections to the objects that are an important part of their development, and that an interface designed to emphasize these qualities could allow young children to form such connections with an interactive system [13].

Sight: Moovl is a web applet that allows children to use a GUI to draw and animate their drawings. Animation uses a record and play paradigm, where drawings are animated based on gestural paths that the user inputs. Moovl was designed to be used with a tablet PC, and although it is not a tangible UI, in this way is more like drawing than “using a computer” (GUI) [9].

While there seems to be a “record and play” trend among some design tools, there may also be potential for this approach to enhance computational sports. Exertion Interfaces argued that UI designs that promote physical activity support social relationships [10]. AR approaches to games like “Can You See Me Now,” where people simultaneously play a game either at web terminals, or running around with handhelds and GPS units, also use physical activity as input [4]. Similarly, Dance Dance Revolution requires and promotes dancing as input and play pattern. Although all of these projects have different “interface designs,” the trends towards physicality and tangible UIs may lead towards more specialized multi-modal interfaces that blend ideas from gestural, pen, tangible and graphical interfaces.

Consumer electronics and commercial toys

Consumer electronics like cell phones, mp3 players and video game systems present the dominant UIs that children use today. These devices do not generally encourage children to be designers, with the exception of things like text messaging and digital photography. Although the toy industry has not adjusted to this shift in children’s play patterns, designers have introduced several educational toys that use less expensive sensing and embedded technologies to build on more traditional play and learning patterns.

Leap Frog is one of the most creative mass-market educational toy companies, who made a name for themselves creating talking toys that could teach young kids phonics. On a talking bus, pressing a letter A on the side of the bus would sound out the letter A, make an animal in the bus sing and dance, and use the letter A in a word. Later, Leap Frog developed a talking book called the Leap Pad with interchangeable paper booklets and ROM cartridges. When a child touches different parts of the printed page (either with a special pen or finger), the book will sing or talk to the user. It can sound out words, identify letters, or teach you geography. Most of their products are display only (i.e. a “touch and hear” paradigm), although the recently introduced the “Fly pen computer” uses Anoto pen technology to encourage children to create their own content. This limited form of pen computing seems most compelling with the pre-scripted games, or in more open-



Figure 4. This walking Topobo moose was designed by two eighth grade girls.

ended writing exercises that encourage creative writing and give some feedback and structure to the activity [7].

Neurosmith [11] has introduced a number of good toys that use a variety of interfaces design approaches including gestural and tangible. Most of their toys encourage physical activity, use music as content and display, and some use tangible manipulations of objects to emphasize sequencing and concepts of parallel/serial organization and nesting. They use sound as a primary display, often coupled with physical organization of toys, or gesture. One thing their toys make clear is that limited interfaces are best for children; when a child cannot be successful if they forget a piece of hidden information, the toy will be frustrating. UI designs are best kept simple and tailored for specific activities and types of design. Although the GUI (with its all-purpose keyboard and mouse) challenges this idea, simplicity and specificity are fundamental in disciplines like product design, may become a dominant design goal in tomorrow’s UI designs.

Looking forward:

In today’s toys, we find a variety of interfaces, including gestural, graphical, tangible, voice, record and play. They all seem to work well in different situations, and don’t conflict when different techniques are used to engage different senses, leading to multi-modal interaction. As computers are reaching out into the environment, balancing multi-modal interactions will become a more common design problem in the future.

I feel there are two important questions regarding the future of children’s design tools, and the future of design in general. First, how will currently disparate ideas about designing with interactive systems be contextualized? To a certain extent, different paradigms will work best

for different modalities, for example gestural interfaces make sense for certain kinds of physical games and pen computing makes sense for writing exercises. However, there may be potential in relating the world of computational processes to the natural world through carefully designed materials and activities.

In creating Kindergarten, Froebel used the craft and design techniques of his day to help children model and understand the natural world. Today researchers in UI design present new design techniques, and due to the field of interactive material design, are beginning to present new craft materials like color-changing fabrics and materials with gestural memory. With materials, the material itself is a central part of the interface. How can these new craft and design techniques be distilled and organized to help children understand our modern interpretation of nature?

Where Froebel wove clay, ink, colored paper and cloth into a circle of pedagogical pattern and activity, we may soon be able to approach the next design revolution by considering the expressive capabilities of materials. With materials, pre-programmed material behaviors may actually become a dominant part of an interface. Already, our building blocks can come to life, and fabric can replay touch and color. A new pedagogy that relates objects, materials, modalities, and behaviors to nature may give people an intuitive way to create meaning through design in our increasingly technologized culture.

REFERENCES

1. Ananny, M. *Telling Tales: A new way to encourage written literacy through oral language*. PhD Thesis, Media Lab, MIT: available <http://web.media.mit.edu/~ananny/thesis.html> [28th February, 2003], 2001, 165.
2. Brosterman, N. 1997. *Inventing Kindergarten*. New York, Harry N. Adams, Inc.
3. Cassell, J. and Ryokai, K. 2001. "Making Space for Voice: Technologies to Support Children's Fantasy and Storytelling." *Personal Technologies* 5(3): 203-224.
4. Crabtree, A, et al. "Orchestrating a mixed reality game 'on the ground'." *Proceedings of CHI 2004*, ACM Press, (2004), 391-398.
5. Frei, P., Su, V., Mikhak, B., and Ishii, H. 2000. "curlybot: Designing a New Class of Computational Toys." *Proceedings of Conference on Human Factors in Computing Systems (CHI) '00*, 129-136. ACM Press.
6. Ishii, H. and Ullmer, B. *Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms*. *Proceedings of CHI 1997*, ACM Press, (1997), 234-241.
7. Leap Frog toys. <http://www.leapfrog.com>.
8. Papert, S. *Mindstorms: Children Computers and Powerful Ideas*. Cambridge, Massachusetts: Perseus Publishing, 1980.
9. Moovl. <http://www.moovl.com>.
10. Mueller, F et al. *Exertion Interfaces: Sports over a Distance for Social Bonding and Fun*. *Proceedings of CHI 2003*. ACM Press, (2003), 561-568.
11. Neurosmith toys. <http://www.neurosmith.com>.
12. Raffle, H. Parkes, A. Ishii, H. *Topobo: A Constructive Assembly System with Kinetic Memory*. *Proceedings of CHI 04*. ACM Press, (2004), 869-877.
13. Raffle, H., Tichenor, J. and Joachim, M. 2003. "Super Cilia Skin, an Interactive Membrane." *Extended proceedings on Human Factors in Computing Systems (CHI) '03*, 529-530. ACM Press.
14. Resnick, Martin, Berg, et al. *Digital Manipulatives: New Toys to Think With*. Paper Session, *Proceedings of CHI 1998*, ACM Press, (1998) 281-287.
15. Ryokai, K., Marti, S., Ishii, H. *IO Brush: Drawing with Everyday Objects as Ink*, in *Proceedings of CHI 04*.
16. Wyeth, P. and Purchase, H. *Tangible Programming Elements for Young Children*. *Proceedings of CHI 2002*, ACM Press, (2002) 774-775.
17. Zuckerman O., Arida, S., and Resnick M. (2005). *Extending Tangible Interfaces for Education: Digital Montessori-inspired Manipulatives*. *Proceedings of CHI 2005*.

Frames of Reference for Portable Device Interaction

Sile O'Modhrain

Sonic Arts Research Centre
Queens University, Belfast

sile@qub.ac.uk

ABSTRACT

Increasingly, our mobile devices are acquiring the ability to be aware of their surroundings, both their physical environments and their digital neighbourhoods. Alongside this awareness of the outside world, these devices are acquiring the ability to sense what is happening to them - how they are being held and moved. The coincidence of connectedness, awareness and richly multimodal input and output capabilities brings into the hand a device capable of supporting an entirely new class of haptic or touch-based interactions, where gestures can be captured and reactions to these gestures conveyed as haptic feedback directly into the hand. Thus one can literally shake the hand of a friend, toss a file off ones PDA, or be lead by the hand to a desired location in a strange city. While this new interaction paradigm opens up a vast array of potential application domains, it also poses a number of challenges. In particular, how can such devices support interactions that will have consequences in environments with different spatial frames of reference – the world-centred frame of reference of the location-aware application, the body-centred frame of reference of the gestural interface, and the device-centred frame of reference of a screen-based application. This paper presents some prototype applications for handheld devices that explore the implications of different frames of reference for actions in the mobile context.

Author Keywords

Portable devices, Haptic Interfaces, Gestural Control, Frames of Reference.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI)

INTRODUCTION

Portable devices are becoming more and more powerful and the types of applications they can support are becoming more sophisticated. Increased processing power, memory and the addition of multiple forms of motion and location sensing bring into the hand a device capable not only of supporting more demanding applications such as video capture and editing, but also of supporting entirely new forms of interaction. However, the development of many applications for the portable context has often been seen as an adjunct, or at most a parallel to that of their desktop relatives. This approach has resulted in a one-size-fits-all solution for the design of interaction with such applications in situations where screen real-estate is limited and the range of motion supported by pointing devices such as joysticks on phones is at an entirely different scale to that being engaged in by the user in the mobile context. Crossan et al [2], for example, have shown that people are less accurate at selecting targets on a PDA when walking versus sitting but that, when walking, people are more likely to tap items in synchrony with the downward phase of the walking gait, suggesting that the motion of walking does effect their interaction with the portable device.

Alongside the scale of a gesture or action, a second factor to be considered, and the factor which is the primary focus of this paper, is the frame of reference of the interaction. In other words, where is the interaction embodied? Is it entirely contained within the hand-held device (e.g. an address list) or is it embodied in the world (a location-aware application such as a map). Moreover, how do both of these spaces relate to the user's own body space, the fundamental frame of reference for their own motion and action?

THE ROLE OF THE BODY IN EMBODIED INTERACTION

The term 'embodiment', defined by Dourish as 'the transition from the realm of ideas to the realm of everyday experience' [9], encompasses not only physical embodiment (of objects such as tables and chairs), but also embodied actions such as speech and gesture. For Dourish, the notion of embodiment is related to Heidegger's phenomenological approach to the world and the way we act upon it. Heidegger distinguished between two

categories of interaction -those where the world is present ("vorhanden") and where the world is acted upon ("zuhanden"). Dourish translates these concepts as "present-at-hand" and "ready-to-hand", and suggests that embodiment is equivalent to Heidegger's concept of "zuhanden". Dourish argues that a Human-computer interface is, when acted upon, "ready-to-hand".

The concept of body-mediated or embodied interaction, of the coupling of interface and actor, has become increasingly relevant within the domain of HCI in general and presents particular opportunities for the domain of mobile device interaction in particular. With the reduced size and cost of a wide variety of sensor technologies and the ease with which they can be wirelessly deployed, on the body, in devices we carry with us and in the environment, comes the opportunity to use a wide range of human motion as an integral part of the interaction with all sorts of applications. As Fishkin et al have pointed out [5], There are many directions in which HCI design is developing beyond the GUI, all of which move toward a greater integration of the body's motion and its sense of its own motion in interaction design. Virtual Reality (VR) approaches a situation where the user is drawn into a high-quality, animated 3D world on the display. In its extreme, the display migrates onto the user's body as goggles, headphones and even touch-feedback-enabled clothing. The second approach that of Augmented Reality (AR) [9], recognises that computation is embodied in physical devices that exist as elements in the physical world and that the physical configuration of these computational devices is a major determinant of their usability. A third approach, that of Enactive Interface design [4], places the body at the centre of the interaction.

How does the notion of embodiment relate to enaction? Varela, Thompson and Rosch define the relationship thus:

“By the term embodied we mean to highlight two points: first, that cognition depends upon the kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological, and cultural context. By using the term action we mean to emphasize once again that sensory and motor processes, perception and action, are fundamentally inseparable in live cognition.” [8]

The enactive approach, then, consists of two elements: (1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided. Thus, enaction can be considered to be a specific case of embodiment, where the 'body' that provides structure and meaning for the interaction is that of the user. This subtle shift of the seat of embodiment from the world to the body provides the opportunity to reconsider the role of motion and action as elements of interaction, because it allows for notions of embodied knowledge to inform

design. Enaction does not preclude either VR or AR approaches, but merely introduces the notion of the body as the meaning-making mechanism in the increasingly physically encompassing world of the human-computer interface.

ENACTION AND TOUCH

The consequence of adopting an enactive approach to the design of interactions with portable devices is a shift of emphasis that acknowledges that the body is far more capable of discovering meaning in the context of the physical world than any artificially constructed metaphor. Here the sense of touch is key - without a sense of touch we cannot move and without moving we cannot act. As Mine et al point out [6], without touch, a user can no longer feel their surroundings to tell where they are or use the felt collision with stationary objects to refine spatial perception. Further, they cannot Use the inertial properties of wielded objects to refine tool manipulation. In short the combined feedback from tactile and haptic proprioceptive systems provides a myriad of cues that help us move and act in the world: The tactile sense, mediated by receptors in the skin, relies on movement between the skin and an object's surface in order for any sensation to be perceived - without movement, objects disappear from our tactile view.

The kinaesthetic sense, mediated by receptors in muscles and joints, must support the tactile sense by providing information about motion and self-motion. The proprioceptive sense, in turn, orients the entire system with respect to gravity and the outside world.

If we are to take on board the implications of the enactive approach to interaction design within the context of mobile device interaction, then we cannot overlook the importance of touch in the acquisition of the kind of embodied knowledge we rely on for physical interactions with the environment. However, the complexity and interdependence of the touch senses requires an approach to the design of applications that use touch where its various roles in sensing, motion and action are well understood and supported by appropriate sensory cues.

FRAMING ACTION – SOME DESIGN EXAMPLES

Over the past three years, the Palpable Machines group at Media Lab Europe developed a handheld platform for prototyping applications to test the hypothesis that tightly coupling motion (of the user) with touch feedback (from the handheld device) could unlock some of the potential for acquiring and using enactive knowledge in the context of portable device interaction. That is not to say one can't acquire tacit knowledge of the interface to an existing PDA or phone. What concerned us was the appropriateness of the scale of motion and the coherence of frames of reference for actions to be performed.



Figure 1. Body Mnemonics

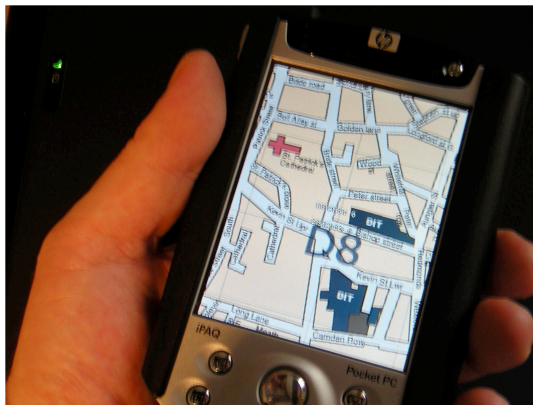


Figure 2. Topographic Torch

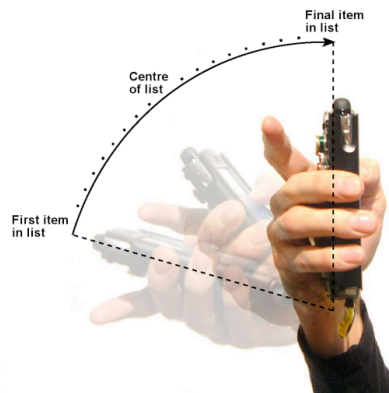


Figure 3. Tilting Lists

Broadly speaking, our prototypes fall into three categories that are distinguished by the frames of reference for their interaction. The first category is that of device-centred interactions, where the world of the application is contained within the device – i.e. the device embodies the set of tasks it supports. Here we were interested to discover whether adding haptic feedback would really result in improved performance in a tilt-to-scroll task because the action of tilting would have an associated haptic reaction from the device. We tested two scrolling paradigms, absolute versus relative scrolling with respect to tilt, and found that the presence of haptic feedback to indicate the movement of the cursor from one item in the list to the next significantly improved performance, particularly at the boundaries of the list where the angle of the screen with respect to the user was most extreme [7]. One important factor to note here is that while the world of the application was certainly contained within the device, it was the user's knowledge of the position of their hand with respect to gravity, their proprioceptive knowledge, that provided meaning for the tilt-based interaction.

The second category of applications used the body itself as the frame of reference for interaction with the portable device. For 'Body Mnemonics', the user's body becomes the reference frame for storing and retrieving information in their portable device [1]. While the application in its final form has not been empirically tested yet, results of a conceptual study where participants placed information in locations around an image of a body and were asked to recall where they had put their data a week later, indicate that the concept of placing and retrieving information in this way is potentially very powerful. What remains to be tested is whether in practice such an interface improves people's ability to access their information while they are moving.

The third category of interest, that of applications where the physical world provides the frame of reference for actions, is perhaps the most interesting. Here we have developed an egocentric map based on the metaphor of a torch so that. As the person holding the device turns to view their surroundings, the map also rotates so that it is always overlaid precisely on the physical environment. They can then tilt the device to zoom their map view, much as if they were casting the beam of a torch onto the world around them. Here it is hoped that the automatic alignment of the

world and map views, combined with the egocentric perspective, will support the user in building a physically grounded understanding of the mapped space around them.

It is worth noting here that, though not strictly related, there are some interesting similarities between the three categories of applications we chose to develop and the paradigms for interaction with VEs described by Mine et al [6].

Their 'Direct Manipulation' paradigm, where they looked at ways to use haptic proprioception to control manipulation in VEs, loosely maps to our device-centred approach. Their second category of 'physical mnemonics' directly corresponds to our body-centred approach and their third paradigm, 'Gestural Actions' (i.e. ways of using body-centred actions to issue commands), encompasses all three approaches, but is particularly relevant for the Topographic Torch.

With respect to frames of reference for action, an important question which must still be addressed for all interaction paradigms is what happens if the frame of reference for an interaction is not constant, either in scale or locus. As we have already noted, the scale of motion appears to effect the way people interact with mobile devices while walking. Is there a similar cost associated with moving between different spatial frames of reference for their actions or does the body's sense of its own space act to mediate any apparent disparities.

SUMMARY

The increased availability of low-cost sensors and actuators provides us as designers with an unprecedented opportunity to reconsider the way in which we design interfaces for portable devices. However,

the use of motion as an input modality is likely to require that the locus of an action and any associated feedback, be situated in a clearly defined reference frame. As mobile devices become more powerful, the designer of applications for the mobile user is likely to have to negotiate a design space with multiple frames of reference for a user's actions. Here a better understanding of the senses of touch, of the body's motion and its sense of its own motion, may be the

key to providing a meaningful bridge between multiple, interleaved and interdependent spaces.

ACKNOWLEDGMENTS

This work was carried out by the members of the Palpable Machines group at Media Lab Europe, in particular Jussi Angesleva, Stephen Hughes, Dr. Ian Oakley and Mike Bennett. Algorithms for gesture recognition were developed by Stephen Strachen and Dr. Rod Murray-Smith of the Hamilton Institute, Maynooth University. Finally the Topographic Torch project grew out of collaboration with colleagues in the CommonSense group at MLE.

REFERENCES

1. Ängeslevä, J., Oakley, I., Hughes, S. & O'Modhrain, S., Body Mnemonics: Portable device interaction design concept. In Proceedings of UIST'03 November 2-5 2003
2. Crossan, A., Murray-Smith, R., Brewster, S., Kelly, J., and Musizza, B. Gait Phase Effects in Mobile Interaction. In Proceedings of CHI '05, Portland 2005
3. Dourish, P. 'Where the Action Is', MIT Press, 2000.
4. Enactive Network <http://www.enactivenetwork.org>
5. Fishkin, K. P., Gujar, A., Harrison, R. L., Moran, T. P., and Want, R. Embodied User Interfaces for really Direct Manipulation. Communications of the ACM 43, 9, September 2000
6. Mine, M., Brooks, F., Séquin, C., "Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction," Proc. of SIGGRAPH '97, 1997.
7. Oakley, I. and O'Modhrain, S. Tilt to Scroll: Evaluating a Motion Based Vibrotactile Mobile Interface In IEEE Proceedings of the World Haptics Conference, March 2005.
8. Varela, F., Thompson, E., Rosch, E. (1991). The Embodied Mind. Cambridge, MA: MIT Press.
9. Weiser, M. The computer for the 21st century. Scientific American 265, 3 (Mar. 1991), 94-104.

'Embodied Interfaces': The Next Generation of HCI?

Chris Creed and Russell Beale
School of Computer Science
University of Birmingham
Birmingham, B15 2TT
[cpc,r.beale]@cs.bham.ac.uk

Abstract

Advocates of embodied agents often assume that such agents will enhance human-computer interaction (HCI) as they take advantage of our pre-existing social skills and provide an interface that is natural and engaging to use. But this is not guaranteed. A discussion is provided detailing some of the key technical and social issues that will need to be overcome for embodied interfaces to be of use in the next generation of HCI, along with an overview of related experiments that are to be conducted over the coming months. Final conclusions regarding embodied interfaces are then provided.

1. Introduction

The use of (human-like) embodied agents in human-computer interaction (HCI) is a research area which has received increasing interest over recent years. Advances in computer graphics have made it possible to efficiently create realistic three dimensional (3D) graphics that can be incorporated into interfaces. Embodied characters often exhibit a variety of human-like behaviours, including emotional expression, speech (synthetic and recorded), gestures and other facial expressions. There has been a wide debate regarding the use of embodied agents in interfaces with arguments both for (e.g. Lester et al., 1997) and against their use (e.g. Shneiderman and Maes, 1997). Supporters of embodied agents believe that they will take advantage of our pre-existing social skills (e.g. analysing tone of voice, facial expressions and body posture) thus making the interaction more natural and engaging. Additionally, it is believed that embodied interfaces will make the system more enjoyable to interact with and thus increase users' motivation to interact with the system, which it is believed will promote cognitive processes such as learning and problem solving.

Opponents of embodied agents argue that humanisation of an interface might have derogatory impact on HCI as it may produce false mental models of the agent's capabilities (Norman, 1997). For example, agents which exhibit human-like behaviour may be perceived as more intelligent than they actually are, which can lead to incorrect expectations about the system's abilities. Another argument is that embodied agents can have an impact on vital cognitive resources such as attention and can lead the user to be easily distracted from the task they are performing. Moreover, it has been suggested that there is no need to embody interface agents as users seem to treat computers as social entities even when they are not represented in a human form (Reeves and Nass, 1996).

Over recent years there have been a variety of studies which have empirically tested supporters' assumptions regarding embodied agents. However, these studies often seem to provide conflicting results regarding the helpfulness of such agents and this can make it difficult to assess the impact they may have in the next generation of HCI (Dehn and Van Mulken, 2000). Moreover, many of the agents that have been developed to date are of poor quality and cannot interact naturally and effectively with users. This inevitably causes people to question the potential role that embodied agents will play in future interfaces.

This position paper will start by discussing the main technical and social issues that will need to be overcome for embodied interfaces to become widespread. Related

experiments that are to be conducted over the coming months will then be outlined along with general conclusions about the future role of embodied agents in HCI.

2. Future of Embodied Agents in HCI

High-quality embodied agents have the potential to effectively play a variety of roles in HCI including sales representatives, trainers, customer support advisors, employee mentors, teachers, site guides, marketers, motivational coaches, comedians and a wide range of others. However, despite the amount of research that has been conducted over the last decade, this potential is yet to be fulfilled as very few interfaces today make use of embodied agents.

2.1 Technical Issues

Several technical issues need to be resolved in order for embodied agents to be of use in HCI. Initially, an agent's ability to make use of human social cues (e.g. eye gaze and emotional expressions) must improve significantly. Studies which have suggested that we treat computers as social actors (Reeves and Nass, 1996) add strength to the possibility that we may respond to embodied agents as though they are human, but if, for example, an emotional expression of empathy is poorly animated and a synthetic voice is used, it is unlikely to have the beneficial psychological impact that has been documented in human-human interaction. Only recently, developing a realistic 3D virtual human face was a huge undertaking in terms of both time and expense (Plantec, 2004). However, it is now becoming possible for people without extensive artistic and animation experience to create realistic virtual humans in a fraction of the time using affordable software that automates much of the process. This technology is likely to improve over the coming years increasing the potential for creating realistic-looking virtual humans.

While it is likely that realistic looking virtual humans will be available over the coming years, a major technical issue that will need to be overcome is an agent's ability to have a *natural* conversation with a human user. For this to be possible, embodied agents will need to combine the work of other large research areas such as affective computing, context-aware interfaces, natural language processing and other sub-fields of artificial intelligence. Only when these research areas have matured sufficiently will embodied agents be able to fulfil their potential. Kurzweil (2005) predicts that this will happen in the near future. With computer processing speeds doubling every year, Kurzweil believes that we will have virtual environments with realistic virtual humans (which will still be *unable* to pass the Turing Test) by 2010. By 2030, he predicts that it will be difficult to distinguish between virtual and real humans.

2.2 Social Issues

Should these predictions turn out to be correct, it will be possible to create embodied agents that are indistinguishable from humans within a few decades. The implications of this would be huge and would raise a number of social issues. How will people respond to agents that act like humans? Will we embrace or reject the technology? How will we feel about taking advice from virtual humans? Can synthetic emotional expressions have the same psychological impact as human emotion? Even if virtual humans are technologically viable in the future, it is not guaranteed that we will feel comfortable interacting with them.

If these predictions are not correct, then we will not be able to produce completely human-like agents. But this simply sets up another set of problems: what forms of representation usefully benefit interaction – equally, what domains benefit from embodiment, and why?

Changes in the nature of the interaction are also caused by changes in the users and their perceptions and expectations, as well as in the technologies employed. For

example, internet shopping has grown strongly over the past few years, well after the dot com boom era, as more and more users become familiar with the technologies and systems. As more advanced games infiltrate more and more homes for entertainment purposes, so users become more familiar interacting with embodied entities on the screen, and hence this more familiar interaction metaphor could drive a widespread acceptance of, and demand for, embodied interaction in general.

3. Experiments

In order to understand how people might respond to future embodied agents, it is necessary to simulate how that interaction might take place. Over the coming months, we plan to conduct a number of experiments that will investigate our responses to embodied agents and are currently in the process of building a nutritional coach that will attempt to help motivate people to eat more healthily. The agent will make use of psychological behaviour change models and will use a number of the strategies and techniques that human therapists use when helping clients change problematic behaviour.

3.1 Research Questions

A number of recent studies have suggested that users tend to like and trust emotional agents more than unemotional agents (Brave et al., 2005; Bickmore and Picard, 2004). In human-human interaction, people are more easily influenced by people they like and trust and are more likely to act on their advice. Does the same apply in HCI? If emotional agents are perceived as more likeable and trustworthy than unemotional agents can they potentially motivate people more effectively? An agent's representation may also effect the strength of a user's response to synthetic emotion and consequently might have an impact on an agent's ability to influence user attitudes and behaviour. The main research questions that we are going to address include: -

1. How do users respond to synthetic displays of emotion? Can emotional agents motivate people to change problematic behaviour more effectively than unemotional agents?
2. What impact does the type of representation (e.g. textual content, synthetic speech, multimedia video) used by an agent to express emotion have upon a user's perceptions and behaviour?

3.2 Experiment Overview

The initial experiment to be conducted will investigate how people respond to synthetic displays of emotion. Subjects will be divided into two conditions: emotion and no emotion (control). An interaction will then take place with a (embodied) nutritional coach which will ask questions related to the subject's current diet and motivation to change their eating behaviour. To reduce the possibility of the agent being perceived as lacking intelligence through poor dialogue, subjects will respond to the agent's questions by selecting from pre-scripted responses. After the interaction with the agent is complete, subjects will be able to view educational material about maintaining a healthy lifestyle for as long as they desire. Once they have finished looking at the material provided they will be asked to complete an online questionnaire which will be used to assess their perceptions of the interaction and the agent. Subjects will also be debriefed to elicit qualitative information about their perceptions of the interaction.

3.3 Further Experiments

We intend to conduct a similar experiment again over an extended period of time (i.e. around four weeks), which will test the impact of an agent's representation on a user's reactions to synthetic displays of emotion. Subjects will be asked to interact

with the coach on multiple occasions to measure how our perceptions of emotional and embodied agents change with time and to examine the long-term effects that interacting with a computational agent has on peoples' eating behaviour. Further experiments will be conducted with other problematic behaviours (e.g. smoking) to test the consistency of any effects found.

4. Conclusion

It remains difficult to predict the role that embodied interfaces will play in the next generation of HCI. When considering what this role might be, it can be easy to look at the standard of agents developed to date and conclude that they will not play a significant role in our future interaction with computers. The graphics and animation are often poor, a synthetic and monotone voice or static text is usually used for communication purposes, and the agents themselves have limited use of language. However, it is now becoming easier to create human-looking agents with affordable software and with computer processing speeds accelerating at an exponential rate, it is likely that agents developed over the coming years will be far superior to the ones currently available.

While it seems almost inevitable that embodied agents will become more advanced in their capabilities, it does not guarantee that they will be adopted by users. People may find it strange interacting with a virtual human and could reject the technology outright. Research which has focused on how we respond to embodied agents has failed (to date) to clarify whether embodied agents will enhance or hamper HCI, with results from related studies often conflicting with each other. More experiments of the type outlined above which simulate how we might interact with embodied agents in the future are essential in helping us understand the potential benefits of using such agents in HCI.

The future of embodied agents in HCI depends on whether the main technical and social issues outlined can be resolved. If they can, embodied interfaces have the potential to enhance HCI in a wide variety of domains including entertainment, teaching, sales, the helping professions, and product support. However, if these issues cannot be resolved, embodied interfaces (in many cases) will become a frustrating interaction experience and are likely to be of limited appeal.

References

- Bickmore, T. & Picard, R. (2004) Towards caring machines. In *CHI '04: Extended abstracts on Human factors in computing systems*, New York, USA, pp. 396-403. ACM Press.
- Brave, S., Nass, C. & Hutchinson, K. (2005) Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies* 62(2), 161-178.
- Dehn, D. & Van Mulken, S. (2000) The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies* 52(1), 1-22.
- Kurzweil, R. (2005) *The Singularity is Near*. New York: Penguin.
- Lester, J., Converse, S., Kahler, S., Barlow, T., Stone, B. & Bhogal, R. (1997) The persona effect: affective impact of animated pedagogical agents. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, USA, pp. 359-366. ACM Press.
- Norman, D. (1997) How People Might Interact With Agents. In Bradshaw, J. (Ed.) *Software Agents*, pp. 49-55. Cambridge: MIT Press.
- Plantec, P. (2004) *Virtual Humans*. New York: Amacon.
- Reeves, B. & Nass, C. (1996) *The media equation: How people treat computers, televisions, and new media like real people and places*. New York: Cambridge University Press.
- Shneiderman, B. & Maes, P. (1997) Direct Manipulation vs. Interface Agents: Excerpts from debates at IUI 97 and CHI 97. *Interactions* 4(6), 42-61.

D. New Interface Designs and Systems

D.1 Here Or There Is Where? Haptic Egocentric Interaction With Topographic Torch

Mike Bennett UCD Dublin
Sile O'Modhrain Queen's University, Belfast
Aaron Quigley UCD Dublin

D.2 Personalized Interactions with Reality Based Interfaces

Angela Chang MIT Media Laboratory
Hiroshi Ishii MIT Media Laboratory

D.3 Reality-Based Object Movement Techniques for 3D

Darius Dadgari York University, Toronto
Wolfgang Stuerzlinger York University, Toronto
Ji-Young Oh University of Arizona

D.4 Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction

Desney S. Tan Microsoft Research, Redmond, WA

D.5 The Future of Mobile Device Research in HCI

Jacob O. Wobbrock Carnegie Mellon University

D.6 Federated Devices: Augmenting Mobility with Ambient Resources

Frankie James SAP Research, Palo Alto, CA

D.7 Reflecting on the Potential of VR for the Future of HCI

Timo Ropinski University of Munster, Germany
Frank Steinicke University of Munster, Germany
Klaus Hinrichs University of Munster, Germany

D.8 Pervasive Interaction - Using movement and speech to interact with computers

Thomas Riisgaard Hansen University of Aarhus, Denmark
Jakob E. Bardram University of Aarhus, Denmark
Mads Sogaard University of Aarhus, Denmark

D.9 Managing Extrinsic Costs via
Multimodal Natural Interaction
Systems

Rebecca Lunsford Natural Interaction Systems,
Portland, OR

Ed Kaiser Natural Interaction Systems,
Portland, OR

Paulo Barthelmess Natural Interaction Systems,
Portland, OR

Xiao Huang Natural Interaction Systems,
Portland, OR

David McGee Natural Interaction Systems

Here Or There Is Where? Haptic Egocentric Interaction With Topographic Torch

Mike Bennett¹

Sile O'Modhrain²

Aaron Quigley¹

Systems Research Group¹
School of Computer Science and Informatics,
UCD Dublin, Belfield,
Dublin 4, Ireland
{mike.bennett, aaron.quigley}@ucd.ie

Sonic Arts Research Centre²
Queen's University of Belfast,
Belfast BT7 1NN,
Northern Ireland
sile@qub.ac.uk

ABSTRACT

In this paper we present and describe Topographic Torch; a handheld digital mapping tool featuring a novel egocentric interaction model. Topographic Torch was designed to encourage and enable people to explore spatial relationships of the world around them in a “natural” manner. Users of Topographic Torch can physically point at objects of interest in the world to automatically see those objects on a map. Enabling people to physically point at objects of interest forces them to use their location in the world as an egocentric frame of reference. An egocentric frame of reference may enhance people’s ability to understand the relationships between where they are in the world and where other objects of interest are in relation to them. We describe Topographic Torch’s interaction model and elaborate on how it functions, along with an outline of a preliminary user study.

Author Keywords

Egocentric Interaction, Embodied Interaction, Haptics, Spatial Cognition, Wayfinding, Maps, Zoomable Displays, Handheld Devices

ACM Classification Keywords

H5.2. Information Interfaces and Presentation: Input Devices and Strategies, Graphical User Interfaces, Interaction Styles.

INTRODUCTION

In this paper we present Topographic Torch; a handheld digital mapping tool featuring a novel egocentric interaction



Figure 1. Example of what a Topographic Torch user of sees on-screen. The red dot represents the user’s location in the world.

model. Our design motivation with Topographic Torch was to enable people to explore geographical spatial relationships of the world around them in a “natural” manner. Specifically we wanted to enhance people’s ability to understand the relationships between where they are in the world and where other objects of interest are in relation to them. For example, when visiting an unfamiliar city it is not uncommon to:

1. locate where you are on a map
2. rotate the map so it is aligned with the world, and,
3. figure out the direction and location of objects of interest in the world and on the map with reference to your location.

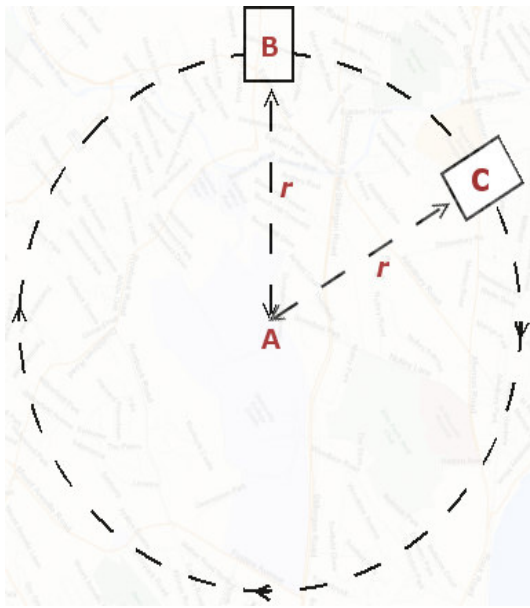


Figure 2. Representation of how the viewport rotates around the user in Topographic Torch. Point A is the location of the user in the world, Point B is the region of the map shown on-screen, Point C is a region of the map that is shown after user rotation. r is the radius of the circle around which B moves when the user rotates on the spot.

When using Topographic Torch people can physically point at objects of interest in the world, such as distant buildings, and see where those objects are on a map in relation to where they are.

Using Topographic Torch can be thought of as similar to using a handheld torch in a dark room. With a handheld torch you can point in various directions to shed light on areas of interest, with limitations on how far the beam of light will travel. A second way of thinking about Topographic Torch is as a tool for carrying out “Point-and-Query” interactions with the world and data overlaid on the world. An important aspect of Topographic Torch is that actions, e.g. pointing, are carried out with reference to the location of the person in the world (embodied interaction).

BACKGROUND & RELATED WORK

Maps displayed on handheld devices are increasingly used in variety of contexts for wayfinding tasks, e.g. where is the nearest supermarket, what is the shortest route from location A to location B, etc. Yet there are a number of serious challenges with using these maps [1, 2, 4]. Limited screen sizes of handheld devices restrict how much of a map can be displayed to the user at any one time. Small display sizes make it harder to understand the spatial relationship between two points on a map, i.e. if the two points are far enough apart they cannot be simultaneously displayed on screen [6]. An inability to see different parts of the map at once, especially multiple points of interest, restricts a person's ability to understand spatial relationships.

Previous work, particular in psychology, has shown that there are a number of ways of navigating and understanding spaces, i.e. survey, procedural and landmark knowledge [9, 10, 11]. Most of these variations can be viewed as differences in the frames of reference used for navigating spaces, i.e. relative, egocentric, and intrinsic frames of reference [4, 8].

Levine [6, 7] explored the implications of these differences for map reading and design. The results of this were a number of experiments that lead to the following principles about map design [2]:

1. Alignment Principle, maps should be aligned with the world they represent.
2. Forward-Up Principle, the upward direction of the map should correspond to what is in front of the person using the map.
3. Two-Point Theorem, a person should be able to relate two points in the world to two points on the map.

In Topographic Torch each of Levine’s principles are employed in the context of an egocentric frame of reference.

TOPOGRAPHIC TORCH

Interaction Model

Topographic Torch is a handheld tool, consisting of an iPaq PDA (Personal Digital Assistant) running custom developed software.

Users are presented with a map on the PDA screen (Figure 1). The map is centered on the location where the user is standing in the world (Figure 2, Point A). The map is initially displayed at a level of zoom such that streets and street names can be readily identified. A user does not press buttons or adjust sliders to interact with the map. Movement around and explorations of the map are controlled in two separate though related ways. The two ways of moving around the map are by pointing and tilting. By moving around the map the user is able to see different parts of the map in the viewport, i.e. on the PDA screen.

Pointing

To use Topographic Torch the user holds the PDA so that the screen is reasonably parallel with the surface of the Earth. Then the user points the PDA in any direction, e.g. North, South, East or West. This causes the map to automatically rotate around the point where the user is located in the world (Figure 2, Point A). Rotation stops when the map is properly aligned with the world. For example when a user physically points North the on-screen map will update to show what is North of the user (Figure 2, anywhere along line r).

Topographic Torch’s automatic alignment means users do not have to physically or mentally rotate maps. All

alignment is done with regards to the user's physical location, i.e. an egocentric frame of reference. Therefore Levine's Alignment Principle automatically occurs as part of the fundamental design of Topographic Torch.

Tilting

Tilting the PDA scrolls the on-screen map backwards and forwards along the direction the user is pointing in (Figure 2, line r). To scroll forward the user tilts the device forward, and to scroll backwards the user tilts the device backwards. When tilted forward the PDA should be angled such that the front of the device becomes closer to the ground, and the back of the device becomes further away from the ground. It is not possible for a user to scroll so far back they begin to see what is behind them. A user can only scroll back to where they are located on the map. If they wish to see what is behind them they must turn around and point in that direction. Unlike other digital maps pressing left or right buttons, or tilting left or right in the case of Topographic Torch, has no effect – it does not cause the map to scroll left or right. Tilting enables users to explore parts of the map that can be physically distant and off-screen.

Tilting, as implemented in Topographic Torch, automatically fulfills Levine's Forward-Up principle. What the user sees on-screen is always in front of the direction the user is pointing in.

Egocentric Scrolling

With the interaction model described so far an important question is: What would happen if a user rotates/points in different directions when the region of the map at Point B in Figure 2 is displayed on-screen? In existing mapping tools a user would expect to see what is directly to the left or right of Point B, i.e. the on-screen map would scroll left or right. This is not the case with Topographic Torch because scrolling and movement around the map are tightly integrated with where the user is located in the world.

Instead user rotation causes the viewport, which is initially at Point B, to traverse the circumference of a circle. The centre of the circle is Point A, where the user is located in the world, and the radius of the circle is the distance between Point A and Point B. As a user tilts forward and back the radius r increases or decreases. For example in Figure 2 if the user is facing North (Point B), and then rotates approximately 50 degrees right they would see the region of the map at Point C.

It should be noted that the viewport rotates as well. The viewport maintains a tangent to the circle while traversing the circumference of the circle. By maintaining the viewport at a tangent to the circle scrolling will always occur from the top to the bottom or from the bottom to the top of the screen. This top to bottom scrolling is important to do because it maintains Levine's Forward-Up Principle for user interactions with maps.

Distant-Dependent Automatic Zooming (DDAZ)

As the user looks at regions of the map that are further and further away from them the map is automatically zoomed out. They can see less detail but more overview. As they look at regions closer and closer to their location the map zooms in more and more. They can see more detail but less overview. This is done for a number of reasons. Distant-Dependent Automatic Zooming can be viewed as a variation of Speed-dependent Automatic Zooming [5]

Firstly it is done to try and build on how we see objects in the world. We are unable to see distant objects as clearly as we can see close objects, e.g. buildings, and distant objects can appear smaller than close objects. By dynamically altering the scale and level of detail as a function of distance (length of r) we are attempting to influence a user's sense of the distance between where they are and where what they are looking at is.

Secondly, as a user looks at parts of the map that are further and further away the sensitivity of pointing increases. As r increases in length then the distance traveled per degree around the circle circumference increases. A one degree change in the direction a user is pointing, when looking at regions of the map that are close by, does not traverse a large amount of the map. A one degree change when r is large causes large amounts of the map to be traversed. Therefore when r is large there are two potential negative effects:

1. Slight changes in the direction the user is pointing cause very large changes in what is shown on-screen, thus any kind of small physical jitter or movement by the user leads to a constantly updating unreadable screen display.
2. If the user is trying to explore the area around a point, then a small change in pointing angle leads to a large amount of the map getting traversed. This makes it hard to understand the relationship between regions of the map, because there is no visual scrolling continuation between the regions.

By using DDAZ in Topographic Torch we prevent these two potential issues. By zooming out as r increases the rate of traversal around the circumference of the circle can be maintained as a fixed rate of movement.

IMPLEMENTATION DETAILS

Topographic Torch runs on an iPaq plugged into a MESH [3]. MESH is a hardware platform for the iPaq that captures a wide range of haptic information. MESH has X, Y, and Z axis magnetic compasses, gyroscopes and accelerometers, and a number of other very useful features, e.g. vibro-tactile feedback, GPS, etc. The magnetic compass and the accelerometer are used for carrying out tilt compensation to establish what direction Topographic Torch is being pointed in. Tilt compensation adjusts the magnetic compass data so that the direction of magnetic North is not lost when Topographic Torch is tilted. For example if you tilt a

traditional magnetic compass too much the compass needle gets stuck and provides incorrect directional information.

Low pass filtering is carried out on the data captured from MESH. This reduces the jitter introduced by the users' kinesthetic system and contributes to making display updates smoother. Maps are stored as bitmaps, though vector based maps would be better for the map transforms, i.e. zooms and rotations.

PRELIMINARY STUDY

We have carried out a preliminary study of Topographic Torch. The purpose of the study was to provisionally examine whether Topographic Torch helped users understand the relationship between where they are in the world and where various target locations were. Is angular error greater or less with Topographic Torch when a user had to understand the relationship between two points in the world and on a map? Angular error is defined as the difference in degrees between the direction a user thinks a location is in and what direction it actually is in. A secondary purpose of the study was to establish how to experimentally examine Topographic Torch.

There were two groups in the experiment, with four subjects taking part. Group 1 carried out the tasks using a paper based map, and Group 2 carried out the tasks using Topographic Torch. Both maps were the same. In both Groups the subjects stood in the same fixed location. There were two main tasks. Task 1 was a timed task which consisted of pointing in a specific direction and asking users to find a specific location in that direction. In Task 2 users were given a target on a map, and then asked to indicate the direction of the target in the real world relative to their location. Though the number of subjects is not large enough to draw meaningful conclusions it would seem that Topographic Torch subjects fared better at Task 2, while paper based map subjects were faster at Task 1.

After the experiment we informally exposed all subjects, along with number of others, to Topographic Torch. In general users appeared to quickly understand how pointing Topographic Torch would automatically align the map, and how tilting backward and forwards scrolled the viewport along the direction they were facing. A number of these users initially kept on tilting Topographic Torch left and right to scroll left and right. Even though they did understand how Topographic Torch worked they took time to adjust to the idea of rotating their whole body to point in directions of interest. This may indicate the egocentric frame of reference in Topographic Torch is not something everyone immediately adapts to.

CONCLUSION

In this paper we have presented a novel egocentric mapping reading tool. Motivations for the interactive design

decisions behind Topographic Torch have been covered while touching on how these decisions meet Levine's map design principles. The impact of maintaining an egocentric frame of reference in each of the interactions is covered. The impact of the egocentric frame of reference for interactive design can be especially seen in the Distant-Dependent Automatic Zooming and the Egocentric Scrolling.

ACKNOWLEDGMENTS

Thanks to Ken Haase and Stephen Hughes for their suggestions and help. A portion of this research was carried out while the first two authors worked at Media Lab Europe.

REFERENCES

1. Baudisch, P., and Rosenholtz, R. Halo: A Technique for Visualizing Off-Screen Locations. In *Proc. CHI 2003*, 481-488, 2003.
2. Darken, R. P., and Sibert, J. L. Navigating Large Virtual Spaces. In *International Journal of Human-Computer Interaction*, 8(1), 49-72, 1996.
3. Hughes, S., Oakley, I., and O'Modhrain, S. MESH: Supporting Mobile Multi-modal Interfaces. Presented at *ACM UIST 2004*, Sante Fe.
4. Hunt, E., and Waller, D. Orientation and wayfinding: A review. ONR technical report N00014-96-0380, Arlington, VA: Office of Naval Research, 1999.
5. Igarashi, T., and Hinckley, K. Speed-dependent automatic zooming for browsing large documents. In *Proc. UIST 2000*, 139-148.
6. Levine, M., Jankovic, I. N., and Palij, M. Principles of Spatial Problem Solving. *Journal of Experimental Psychology: General*, 111(2), 157-175, 1982.
7. Levine, M., Marchon, I., and Hanley, G. The Placement and Misplacement of You-Are-Here Maps. *Environment and Behavior*, 16(2), 139-157, 1984.
8. Mou, W., and McNamara, T. P. Intrinsic Frames of Reference in Spatial Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 162-170, 2002.
9. Thorndyke, P. W., and Goldin, S. E. Spatial Learning and Reasoning Skill. In *Spatial Orientation: Theory, Research and Application*. 195-217, 1983, New York: Plenum Press.
10. Thorndyke, P. W., and Hayes-Roth, B. Differences in Spatial Knowledge Acquired from Maps and Navigation. *Cognitive Psychology*, 14, 560-589, 1982.
11. Thorndyke, P. W., and Stasz, C. Individual Differences in Procedures for Knowledge Acquisition from Maps. *Cognitive Psychology*, 12, 137-175, 1980.

Personalized Interactions with Reality Based Interfaces

Angela Chang

Tangible Media Group
MIT Media Lab
Cambridge, MA 02139
anjchang@media.mit.edu

Hiroshi Ishii

Tangible Media Group
MIT Media Lab
Cambridge, MA 02139
ishii@media.mit.edu

ABSTRACT

In this paper, we describe a perspective on the evolution of HCI toward *personalized* interfaces, moving toward unique customized interfaces that possess expressive qualities defined by their end-users. The increasing availability of personal portals and configurable skins, coupled with the means to distribute personalizations, allow a wealth of novel interface mappings to coexist. We reflect on potential social implications of personalization.

Keywords

Reality-based interfaces, tangible interfaces, sensorialism, portals, customization

INTRODUCTION

This paper aims to draw attention to the increasing ability for a user to have *personalization*, or *individual customization of sensorial expression*. The trend we are reporting is highly relevant to the discussion of reality-based interfaces [10], as we believe that the personalization aspect will give incentive for people to enhance their personal affinity and value for their interfaces. We describe examples of personalization, and discuss important features for promoting personalization: ease of customization and distribution.

We also propose some metrics for describing these interfaces. This paper is prompted by significant development of the next generation of interfaces which expand on the traditional desktop metaphor by enhancing sensory integration. Some of these initiatives revolve around augmenting graphics with touch and spatial orientation (such as augmented reality and virtual reality) [7], while others aim to emphasize the information available in the physical world (e.g. tangible interfaces and reality-based interfaces)[5].

These research fronts, however, have largely been focused on the *development of customized* platforms that support a wide range of specialized scenarios (for example, Figure 1 displays a information interface called metaDesk, which was only demoed in research contexts [5]). These research



Figure 1. Tangible Media Group's metaDesk allows how users manipulate and display information simultaneously, demoed at the Media lab in 1997.

tools have, with rare exception, not been commercialized for general use. There has been little incentive or ability for a large group of consumers to acquire non-WIMP interfaces such as tangible, augmented reality or virtual reality interfaces. In order for the next generation of HCI to become pervasive, we feel there must be a way for users to learn about and access new interfaces. These interfaces must also be easily reproduced or mass-produced somehow. The rest of this paper discusses some features of personalization, and how new, personalized interfaces can be made accessible to non-research users.

BACKGROUND

We note that ergonomics and cognitive science have received much focus in prior interface design research. These research approaches are based on performance metrics, such as have led to enormous achievements in understanding how to design more efficient and intuitive interfaces, and improve accessibility for a general populace. For example, the desktop metaphor has allowed users to understand how to navigate a file system, and organize documents [15]. Meanwhile, ergonomics is assessed by means of Fitts law-type usability experiments, for example, target acquisition times for comparing menu layouts [14].

Rather than enhancing performance or increasing functionality, there is some evidence that there are other, features of interest for new interfaces. For example, one study found that although mixed initiative voice menus were efficient, users preferred system-initiative menus [17]. Other research has been done on more subjective measures, like pleasure and playability of an interface[2,3,4,8,11].



Figure 2. This Ambient Orb (left foreground) allows a user to customize colors associated with trading activity.

Along this line of thought, we believe personalization is another research area that is equally important.

PERSONALIZATION

Personalization is the ability of an interface to be customized, by allowing the user to alter the sensory experience. (Examples of personalization are devices which support interchangeable faceplates.) Personalization allows the underlying utility of a device or application to remain the same, yet, the interface between the person and the sensory experiences are alterable. Some devices allow flexibility in the input/output method, like the Ambient Orb[1], allows users to map financial performance to color (figure 2.)

Personalized interfaces are interfaces that can be customized, allowing users some degree of flexibility in input/output modality, display and control. For example, a user may choose to increase the size of the buttons on their browser, or change the look of the buttons altogether. Instead of using a ring tone on a mobile phone, a user can decide to receive vibrations alerts. Sensorial, aesthetic mappings can be changed at the user's whim. Parts of interfaces can be pared down or augmented.

Personalization serves to enhance self-expression and identity, rather than optimizing the interface function. As objects are increasingly differentiated based on design, they begin to take on their own identity while reflecting the self-expression of their users. Consumers are increasingly more sensitive to branding, using design to differentiate themselves rather than being content with mass-produced functional products. Norman presents five different levels of customization, from no customization to creating a whole new product, and suggests that users inherently want to customize ("we are all designers", p.224)[16]. In any case, the two essential characteristics for successful personalization are *ease of customization* and *ease of distribution* for sharing the customized interfaces.

Ease of User Customization

What makes personalization so interesting is the variety of product categories where consumers can now participate in creating the look and feel of the end product. Potential buyers are given many levels of choices, so the product

they take home is unique, designed around their personal preference. Instead of only choosing from among few possibilities, such as faceplates, buyers have the option of creating custom products that no other person may have an identical copy of. This flexibility in look and feel was a design approach that had been applied in fashion and accessories, and has finally migrated to other sectors.

For example, the Scion car model allows a potential buyer can specify what type of external body shape and internal audio features they would like (www.scion.com). People accessorize their phones with custom faceplates covers, ringtones, backgrounds and tassels. The user can then differentiate their objects from those of other users.

Another example of personalization we'll discuss in personal portals. For example, Mozilla's firefox (<http://www.mozilla.org/themes/download/>), allows users to download themes and extensions that reflect the user's preference (figure 3). Themes (or skins) are mainly graphical alternatives to the basic interface (see figure 3 right), while extensions allow custom functionality specific to the user (such as the ability to have an Amazon or Google search bar). The ease of skinning a Firefox browser requires two clicks-- the user first clicks on the available skins to download and clicks again to select the theme to use. A restart is required, but the customization is executed within seconds.

These individualized portals reflect the preferences and priorities of a specific user. Another user might find another selection of sensory modalities, themes, and extensions more useful. Once changed, the sensorial experience of using the interface is unique to the user. For example, if interfaces were swapped between a business man's portal and a sixteen-year-old girl's browser, they might find the other person's interface unintuitive and inefficient. The businessman's phone might emphasize sleek lines, large screen display areas, and minimal audio and tactile display, while a youth might prefer bold colors, ornamental details, and loud audio and tactile stimulation. In response to an alert, the professional's device might vibrate while the teenager's device blinks.

In order to support personalization, designers should account for *how* personalization can occur, by creating a consistent mechanism for altering the experience of the product. DeMichelis and Susani describe consistency in their discussion of multifunction vs. general purpose interfaces[6]. The design team needs to specify the hardware or software interface available for customization and make an effort to simplify the process of customization. A user should be able to browse available options and select the ones they want, while time to configure the customization is minimal. In embedded devices, forcing a user to go to a website to configure their mobile phone, for example, may be too complex and involve too much effort.

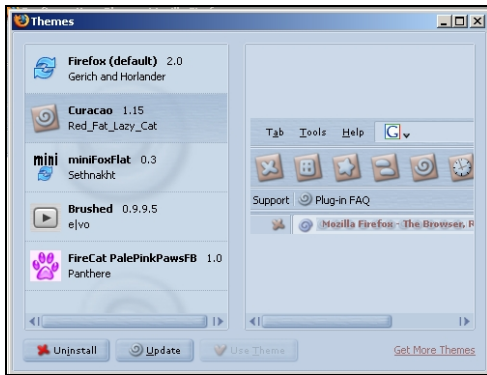


Figure 3. Firefox themes are selected by clicking on the button next to the graphic.

Ease of Distribution

A distribution channel for customizations is another key aspect to enabling users to successfully personalize their devices, software, or hardware. We see a vast array of internet tutorial sites for downloading custom versions of operating systems and applications (for example, Winamp, FireFox, Messenger). Many freely available and open-source software initiatives support downloading extensions and themes. Meanwhile, a mobile phone user can go into a convenience store and find accessories and faceplates. Most of all, the internet, combined with the established product distribution patterns help consumers access information and select among many personalization options.

DISCUSSION

Personal Fabrication

At the MIT Media lab, there is ongoing research in tools for developing nations, particularly in the context of Fab Labs (<http://fab.cba.mit.edu>), a project headed by Professor Neil Gershenfeld. Gershenfeld sees a revolution in personal computing called, personal fabrication. With the right types of tools and sharing of information, remote villagers can create their own workshops and make custom tools and share knowledge about their solutions through the internet[9].

Rather than just providing packaged solutions, visitors to Fab Labs can create and use their own interfaces to answer a need, or just for fun. Any code or examples are documented and shared online. As the ability to create customize interfaces becomes more widespread, we expect a vast array of different interfaces to become commodities, with exchanges happening and designs being adapted and reused. We cite evidence of solutions to problems specific to remote villages, such as testing water quality (Gershenfeld, p.167), or harvesting power (Gershenfeld, p. 90). As users are empowered to create interfaces to solve specific needs, they can share them, and new designs can proliferate throughout a community.

Similarly, popular websites like Hack-A-Day (www.hackaday.com) and “make” magazine (www.makezine.com) allow people to learn about how to

create customized products from old electronics. Sites where customizations are downloaded often offer toolkits for creating and uploading new customizations, adding to the pool of selections

Cultural diversity and Innovation

With the increasing availability of personalized solutions, we envision that diverse populations can participate in using new technology. For example, elderly web surfers might download themes with large buttons, or simplified menus and expect the same web surfing functionality as their grandchildren. Instead of relying on manufacturers to create and distribute solutions to interface problems, there might be grass-roots creation and adoption of unique solutions. Essentially, we expect that although there will be a proliferation of new interfaces, and that the interface grammars may get diluted; however, we expect that there will always be value added to the user.

The Proliferation of Personalization

Most people are not creators, and in general, most people are consumers of content. With the advent of the internet and electronic publishing, it has become a lot easier for custom creations to be passed on and adapted. Likewise, the cost of manufacturing has decreased, and mass production of interchangeable parts is now designed into products at little cost to the consumer. As a result, we expect these personalized interfaces to be culturally rich and useful to many diverse and scattered populations. Instead of neglecting groups of users for the 95 percentile of a population, we envision that increased personalization includes previously untargeted user populations. Personalization could embody universal usability.

Personalization is the opposite of “context awareness”. There has long been a view of personalized interfaces relating to content that adjusts based on tracking the user’s behavior, such as in Amazon’s 1-Click or recommender systems. Context awareness is concerned with technology that can automatically adapt behavior. This is a misuse of personalization, with much potential for insensitive or miscues [12]. Personalization, as described in this paper, is directly controlled by the user, and relies on a user specifying behaviors for the interface. For example, a phone might be augmented with a custom accessory or bracelet. Instead of relying on the manufacturer to create an identity for the user, the user creates her own mappings so that when their friend calls the accessory lights up or vibrates.

Humans Co-evolving with Technology

In future work, it might be helpful to have metrics for personalization, in order to better compare different products. As a starting point, possible metrics are: *variety*, *time to install*, and the *number of distinctive styles* the user is able to identify with. Affective characteristics, such as emotional tones or signals can also be used. Although these metrics might not help measure efficiency, they can certainly allow people to begin a dialogue on

personalization. It remains to be proven whether personalizations can be compared or how their effects can be assessed. McLuhan's the "medium is the message" urges researchers to examine the psychological and social impact of the adaptation of current interfaces [13]. Through personalization, better self-expression and identity may allow societies to communicate in new ways that change the nature of interactions. For example, in the case of Apple products or Saturn cars, brand affinity often becomes a social tie. Likewise, video gamers and hackers speak different languages and concern themselves with different social issues pertaining to their passions. Research can certainly be done to assess the different ideologies and viewpoints available to users of personalized interfaces.

Convergence or Divergence?

For this workshop, there is a question about whether interfaces are converging or diverging. For the most part, we have discussed how products are converging to enable better interface personalization. While the actual interfaces are diverging in form and function, the ability to modify any interface according to a user's preferences is becoming standardized through common use or communal decision-making (e.g. the W3C consortium <http://www.w3.org/>). In essence, customization becomes a commodity, based on stable, but functional platforms.

WORKSHOP PROPOSAL

At the workshop, the authors will demonstrate a mobile application with user interface that supports personalization. This augmented device will showcase some applications which might be useful in research on personalized information displays. While these implementations are very basic, we hope to provoke discussion on personalization applications.

CONCLUSIONS

We have presented our alternate vision of the evolution of reality based interfaces. We hope that this paper has raised some thought-provoking issues about how society may adapt to increasing personalization of their interfaces. We also hope that research in personalization and development of supporting features are worthy of further research.

ACKNOWLEDGMENTS

We thank Richard Whitney for his discussions on the evolution of human computer interfaces and many other researchers from the Tangible Media Group. We especially thank Neil Gershenfeld from the MIT Media Lab's Fabrication Labs, who generously allowed me to sit in on his 'How to Make Almost Anything' Class in Fall 2005.

REFERENCES

1. Ambient Orb, Ambient Devices, Jan. 05, 2006 http://www.ambientdevices.com/cat/orb/Stock_Orb.pdf

2. Azuma, R., *A Survey of Augmented Reality*, Presence: Teleoperators and Virtual Environments:6(4), 355-385.
3. Blythe, M. et.al., *Funology: From Usability to Enjoyment*, Kluwer Academic Publishers, 2003.
4. Bødker, S., Christiansen, E., Designing for Ephemerality and Prototypicality. Proceedings of the 2004 conference on Designing Interactive Systems (DIS '04), ACM Press, 255-260.
5. Brave, S., Ishii, H., Dahley, A. Tangible interfaces for remote collaboration and communication. Proceedings of Computer Supported Cooperative work (CSCW '98), ACM Press, New York, NY, 1998, 169-178,.
6. De Michelis G and Susani, M. (2003) The Swiss Pattada: designing the ultimate tool (with original drawings by Marco Susani). *Interactions* 10.3:44-53.
7. Feiner, S., MacIntyre, B. and Seligmann, D. (1993), Knowledge-based augmented reality, *Communications of the ACM*, 36(7):53-62.
8. Galloway, A. et.al, Panel: Design for hackability. Proceedings of the 2004 conference on Designing Interactive Systems (DIS 04), ACM Press, 363-366.
9. Gershenfeld, N. *FAB: The Coming Revolution on Your Desktop--From Personal Computers to Personal Fabrication*. Basic Book, New York, NY, 2005.
10. Jacobs, R. Reality Based Interfaces class, taught in fall at Tufts University, Medford, MA 2005. Class website <http://www.cs.tufts.edu/courses/description/COMP/025-0-RBI>.
11. Jordan, P.W. *Designing Pleasurable Products: An Introduction to the New Human Factors*, Taylor & Francis, Philadelphia, PA, 2002.
12. Kabel, Marcus. *Wal-Mart: Error to blame for MLK link*, Associated Press, Jan 5, 2006, http://hosted.ap.org/dynamic/stories/W/WAL_MART_WEB_SITE?SITE=SCGRE&SECTION=HOME&TEMPLATE=DEFAULT
13. McLuhan, M. *Understanding Media: The Extensions of Man*, The MIT Press, Cambridge, MA 1994.
14. J. Nielsen. *Usability Engineering*. AP Professional, Cambridge, MA 1993.
15. Norman, D.A., *The Psychology of Everyday Things*, Basic Books, New York, NY, 1988.
16. Norman, D.A., *Emotional Design: Why We Love (Or Hate) Everyday Things*, Basic Books, New York, NY, 2004.
17. Suhm, B., et.al. A comparative study of speech in the call center: natural language call routing vs. touch-tone menus, Proceedings of CHI 02, ACM Press, 283-290.

Reality-Based Object Movement Techniques for 3D

Wolfgang Stuerzlinger, Darius Dadgari

York University

Toronto, Canada

<http://www.cs.yorku.ca/~wolfgang|~dariusd>

Ji-Young Oh

University of Arizona

Tucson, AZ, USA

<http://3dvis.optics.arizona.edu>

ABSTRACT

Scene layout and part assembly are basic tasks in 3D object manipulation. While movement methods based on 3D or 6D input devices exist, the most efficient 3D movement techniques are based on utilizing only two degrees of freedom. This poses the problem of mapping the motion of the 2D input device to efficient and predictable object motion in 3D. We present a new method to map 2D input to 3D motion in this paper. The object position follows the mouse cursor position closely, while the object always stays in contact with other surfaces in the scene. In contrast to existing techniques, the movement surface and the relative object position is determined using the whole *area* of overlap of the moving object with the static scene. The resulting object movement is visually smooth and predictable, while avoiding undesirable collisions. The technique also utilizes the fact that people easily recognize the depth-order of shapes based on occlusions. The proposed technique runs in real-time. Finally, the evaluation of the new technique with a user study shows that it compares very favorably to conventional techniques.

INTRODUCTION

Moving objects is one of the most basic tasks of 3D scene construction. When people design a scene with multiple objects, they repeatedly realign or adjust different parts, to explore the design space. Our goal is to provide an efficient and smooth object motion technique aimed at facilitating this explorative process in 3D manipulation systems such as Computer Aided Design (CAD), Virtual Reality (VR), and Augmented Reality (AR) systems.

In general, object motion is performed either with a mouse in CAD, with desktop-VR/AR systems, or with 3D or 6D input devices in VR and AR systems. We briefly discuss previous work by reviewing related work with input devices with two DOF as well as input devices with more DOFs.

Related Work – 2D Input Devices

Strauss categorized possible solutions to the problem of mapping 2D input to 3D movement [18]. He enumerates the idea of using handles, moving objects parallel to the view plane, using obvious structures to determine the plane of motion, and heuristics. The authors also claim that there is no “perfect” solution, as there is no approach that is both easy-to-use and robust at the same time. Consequently, users need to frequently check if the object is the desired position, which can become tedious.

Many commercial CAD systems utilize some form of handles to provide for 3D object motion. While this solution allows no room for failure or unexpected results, the task of moving an object may become tedious, as the user has to mentally separate the desired 3D movement into 1D or 2D components. Moreover, if objects are in contact, these objects may occlude handles and may make it difficult or even impossible to manipulate an object. The second category of movement techniques, moving objects parallel to the view plane, is rarely employed as the results of object movement depends strongly on the current view direction, which is undesirable. The third approach, which utilizes other structures in the scene, typically uses a ray from the eye point through the current pixel to identify the first intersection point with the scene. This intersection is then used to compute the position of the 3D object. E.g. Bier [1] used this approach in his snap-dragging technique, which snaps to the closest visual feature in a wire-frame display. However, this approach suffers from severe problems in complex scenes. As an example for a heuristic approach we list the idea of using a library of predefined objects with predefined movement behaviors. These behaviors are then used to constrain objects to particular places in a scene (such as an “on-wall” constraint). A ray along the current mouse position is then used to find the places in the scene where the constraints are fulfilled and the object is close to the cursor position. In the work on the MIVE system [17], we evaluated an implementation of this idea with several user studies and showed that such a technique allows naïve users to quickly populate 3D scenes with predefined objects.

Related Work – 3D Input Devices

A number of authors have investigated the performance of object manipulation with 3D input devices, such as a spaceball or a six degree-of-freedom tracker. Such devices enable direct interaction with a 3D scene and are typically used in VR/AR systems. One of the first researchers to present a system was Bolt in 1980 [2]. Subsequently many other researchers studied the creation and manipulation of 3D environments in VR. Bowman [3], Mine [8], and Pierce [13] proposed different 3D manipulation methods that can be applied in a variety of settings. For a more complete overview over work in this area, we refer the reader to a recent book about 3D user interfaces [4].

Several of these systems use collision detection to prevent interpenetration of objects. However, few utilize constraints for object manipulation and these support only the simplest

geometric constraints (e.g. “on-plane”). A general constraint-based system was presented by Multigen-Paradigm in the ‘SmartScene’ technology [16], which provides for arbitrary scripting of object behavior. Users can interact with the scene using tracked pinch-gloves, and the pre-defined object behaviors facilitate 3D object manipulation.

Observations about Real-World Object Manipulation

The following observations are based on data collected in a series of user studies on 3D object movement techniques targeted at naïve users. In these studies (see e.g. [10, 12, 17]) we experimented with many different alternative strategies, including the use of 2D, 3D, and 6D input devices as well as many forms of widgets to move objects in a 3D scene. Based on the results of these studies, we believe that the following set of rules captures the most important design decisions for 3D object movement techniques.

1. In the real world, (almost) all objects are attached or connected to other objects.
2. Objects do not interpenetrate each other.
3. Bringing objects in contact with each other is the natural way to position them relative to each other.
4. Only visible objects can be manipulated.
5. The most important cue for judging 3D position in real scenes is occlusion.
6. Many standard computer graphics techniques such as “handles”, “wireframe”, “3 orthogonal views”, etc. tend to confuse users.
7. In general, 3D or 6D input devices provide less precision than 2D input devices.
8. Users seem to consider the entire area of visual overlap of the (moving) foreground object with the (static) background scene when deciding where the object is.

Please note that the above observations are based on experiments with naïve users, i.e. users who have no 3D computer graphics education or significant gaming experience.

In the following, we briefly discuss why each of the above points is important when dealing with naïve users.

1. Floating objects are exceptional in the real world and our observations in the user studies confirm that most users are surprised when an object starts to float when moved. That means that the correct *default* for any manipulation technique for 3D object motion is that objects should stay in contact with the rest of the world!
2. Most naïve users get confused if objects interpenetrate each other. This is easily solved with collision detection, something that can be computed nowadays in real-time even for complex scenes [5, 7].
3. The paradigm of sliding an object on the surface of another until it reaches the desired position is the most natural way to position objects – and in fact makes any

task much easier. This is easily demonstrated by watching a child position toy blocks. The technical way to implement this is to choose a movement surface from the set of surfaces of the static scene and to displace the object on that surface. Often this is realized via the definition of constraints to object movement, see the section on previous work.

4. Naïve users don’t try to manipulate objects that are not visible. They typically rotate the view/scene so that the part in question is visible. One indication for this is that a comparison of different interaction methods for 3D input devices found that the best techniques are based on the notion of ray casting [14]. Ray casting identifies the first object that is visible along an infinite ray from the manipulation device into the scene (much like a laser pointer). Hence, it is sufficient to allow the user to select all objects from a 2D image! And indeed, in [14], the authors hypothesize that all ray casting techniques can be approximated as 2D techniques.
5. As documented by research into visual perception, people judge 3D position based on several cues. Besides perspective, the most important cue for 3D position is occlusion [21]. In our studies, we found that for scenes without floating objects (see above), perspective and occlusion combined with the ability to quickly move the viewpoint are usually sufficient to allow humans to understand the 3D position of an object in relation to other objects. Finally, it is interesting to note that recent research confirmed that from an end-users point of view, most stereo technologies are not very mature and are tiresome and/or problematic to use on a daily basis (e.g. [6, 20]).
6. The idea that one has to use “handles” to move an object in 2D is an instance of an indirect manipulation technique. It is sufficient to point out that in the domain of (2D) desktop environments, this idea was very rapidly eclipsed by the idea of direct manipulation [15], as this paradigm proved to be much simpler to understand. Similarly, most naïve users can’t readily interpret a wire frame view or three simultaneous views.
7. A human hand held in free space will “jitter” more than a hand that is supported by a real surface. That means that an input device that is limited to 2DOF provides more precision and hence usually affords also more efficient manipulation. In VR/AR research, this has been already realized through the adoption of techniques such as the Personal Interaction Panel [19], which effectively transforms a 3DOF input device into a 2DOF input device.
8. Practically all techniques for 3D object motion use the current position of the pointing device to compute the (new) 3D position. This effectively reduces the computation to a point mapping problem, and all current 3D object motion techniques are based on this idea. However, research into vision in primates discovered that

the perceptive field for an object that is being held in the hand covers the whole object [9]. In other words, there is strong evidence that the whole visual area of an object is used to judge 3D position. And indeed, we observed in our user studies that point-based techniques do not work as well as area-based techniques.

The presented list is based on our observations of naïve users. However, we would like to point out that while it may be possible that expert users can achieve higher performance by ignoring some of the above observations, we believe that for many kinds of routine scene modifications even expert users will greatly benefit from techniques that follow these guidelines.

REALITY-BASED 3D OBJECT MOVEMENT

Based on the above discussion, we designed a new 3D object movement technique that fulfills these criteria. Instead of using widgets, we allow the user to simply “grab” any object and slide it across the scene by dragging it to the desired position to visually assess the impact of a change. One of the main ideas behind this is that this will greatly facilitate exploration.

As the user moves an object, he/she utilizes his/her knowledge of the other surface(s) hidden by an area covered by the moving object. We implement this by always moving the object on one of the surfaces that it occludes. This implicitly guarantees that the object always stays attached to other objects. More precisely, we look for the closest visible surface behind the moving object and move the manipulated object onto it. Finding visible surfaces can be done very efficiently with graphics hardware, as can the detection of collisions [5, 7].

The attractiveness of the first alternative is its simplicity, as we only need to determine the foremost surface of the

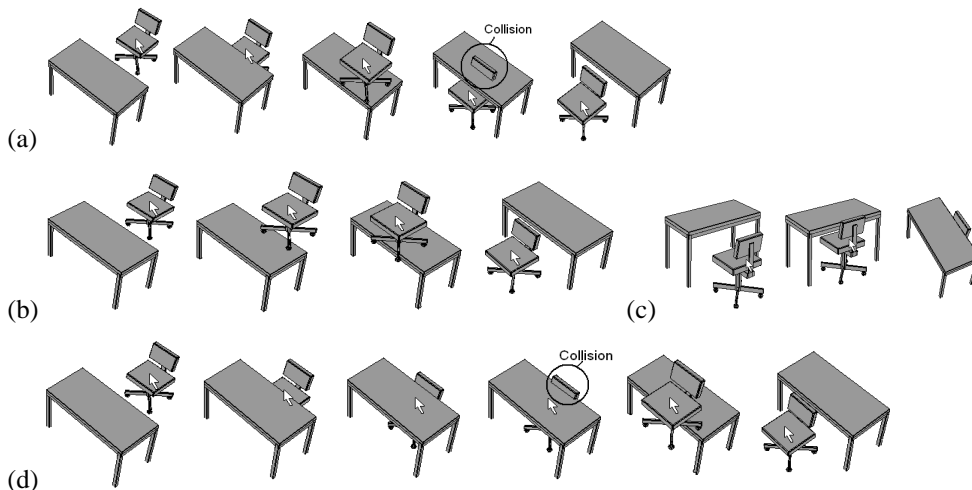


Figure 2. Image sequences illustrating object movement based on (a) the mouse position technique, (b) & (c) a technique based on the foremost visible surface, and (d) our new technique based on the foremost surface behind the moving object. For a detailed explanation please refer to the text.

(static) scene, regardless of the current position of the moving object. Furthermore, this method ensures that the moving object is always closer to the viewer than the rest of the scene. This implicitly guarantees that there are no collisions with other objects. The downside of this method is that when a scene is cluttered with many objects, and there are consequently many surfaces, then the moving object will jump frequently in depth, and positioning the object requires more attention from the user. A pilot study of an implementation of this idea showed that this is indeed a problem that users encounter in practice.

The second alternative again identifies the first surface behind the moving object, but ignores any surface closer to the viewer than the moving object. In this method, the moving object does not immediately pop out to the surface in front of the user, unless the moving object becomes the one closest to the viewer. Usually, this conforms better to the intentions of the user. The limitation of this alternative is that when a small object moves forward, it may penetrate another object in front. To address this problem, we employ a collision detection method. Once a collision is found, the object jumps also in front of the colliding object, as with the first alternative.

Figure 1 depicts several movement sequences with the original mouse-ray techniques and the two new techniques mentioned in this section. Here the goal is to slide the chair under the table. Figure 1(a) shows the movement based on the foremost surface behind the mouse position. As soon as the mouse pointer overlaps with the surface of the table, the chair moves on top of it. However, when the mouse pointer moves off the table again, the chair drops immediately to the floor and ends up in a position where it collides with the table (fourth image). There are two ways to get the chair under the table: One is to change the viewpoint - Figure

1(c). The other alternative is to “grab” the chair by the top part of the backrest (an area that is visually quite small) and to move it towards the table while avoiding overlap with the table itself. However, this is very non-intuitive, and most users do not realize that this is possible – especially since the position that needs to be “grabbed” is not at all obvious.

The image sequence depicted in Figure 1(b) illustrates the technique that utilizes the foremost surface behind the image of the moving object. As soon as the image of the chair overlaps with the

table (second image), the chair starts to slide on the table surface. Only when the image of the chair does not overlap with the table anymore, does the chair drop down to the floor. Note that no collision occurs with this technique, but the only way to drag the chair under the table is again to change the viewpoint, as in Figure 1(c).

Finally, Figure 1(d) illustrates the new technique that utilizes the first surface behind the moving object itself. As the chair slides on the floor it continues to move underneath the table, because the first surface visible *behind the chair* is the floor. In the third image, the chair is clearly in the desired position and the user is finished. For illustration purposes, we continue this sequence with the fourth image from the left, where a collision occurs, which prompts the technique to move the chair onto the table.

Implementation

This algorithm can be efficiently implemented by using the features of modern graphics hardware. In particular, we can detect the foremost hidden surface by rendering the background scene. Rendering the moving object into a separate area and using the depth information computed by the hardware for both images aids in the identification of how much the object has to be moved in 3D. To avoid problems due to discretization, the implementation also uses an item buffer, which encodes where each object is visible on the screen, to compute a more precise answer. Further details of the implementation are described in [11]. Finally, our technique allows objects to move freely in 3D space, when an object is seen over the background (i.e. there is no visual overlap with other objects). In this case, the movement surface is chosen to be the axis-aligned plane that is most orthogonal to the viewing direction. This provides users with the option to quickly create “floating” objects, if necessary.

Evaluation

The presented technique works well for general shapes, even for objects that have large concavities or curved surfaces. For curved surfaces, the API’s of current graphics cards necessitates an approximation of the curved surface into many small planar surfaces, which allows our algorithm to work without issue.

We evaluated the described scheme in several user studies that asked the participants to assemble various objects with a mouse in a desktop 3D modeling system [12]. We found that users had little problem in understanding how the new technique works, were significantly faster with it, and were quickly able to utilize the technique to its full potential.

CONCLUSION AND FUTURE WORK

Based on the results of a series of users studies, we presented a list of guidelines for techniques to move objects in 3D scenes. Then we presented a new reality-based technique to move objects in 3D scenes that provides intuitive 3D object motion, which naïve users found easy to understand and easy to utilize effectively. As this technique uses graphics hardware, it can handle even complex scenes effi-

ciently. In the near future, we plan to evaluate this technique in IVY, our six-sided fully immersive VR system as well as other VR/AR setups.

REFERENCES

1. E. Bier, Snap-dragging in three dimensions. *SIGGRAPH 1990*, 193-204.
2. R. Bolt, Put-that-there, *SIGGRAPH 1980*, 262-270.
3. D. Bowman, et. al, An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. *Symp. on Interactive 3D Graphics*, 1997, 35-38.
4. D. Bowman, E. Kruijff, J. LaViola, I. Poupyrev, *3D User Interfaces: Theory and Practice*, Addison-Wesley, 2004.
5. K. Dave, K. P. Dinesh, CInDeR: Collision and Interference Detection in Real-time using graphics hardware, *Graphics Interface*, 2003.
6. D. Diner, D. Fender, *Human Engineering in Stereoscopic Viewing Devices*. Plenum Press, 1993.
7. N. K. Govindaraju, et al, CULLIDE: interactive collision detection between complex models in large environments using graphics hardware. *SIGGRAPH Workshop on Graphics Hardware 2003*, 25-32.
8. M. Mine, F. Brooks, C. Sequin, Moving Objects in Space: Exploiting proprioception in virtual environments *SIGGRAPH 97*, pp. 19-26.
9. S. Obayashi, et al., Functional brain mapping of monkey tool use, *NeuroImage 14*: 853-861, 2001.
10. J.-Y. Oh, W. Stuerzlinger, A system for desktop conceptual 3D design, *Virtual Reality*, 2004 7: 198-211.
11. J.-Y. Oh, W. Stuerzlinger, Moving objects with 2D input devices in CAD Systems and Desktop Virtual Environments, *Graphics Interface 2005*, 141-149.
12. J.-Y. Oh, Desktop 3D conceptual design systems, Ph.D Thesis, 2005. York University.
13. J. Pierce, et al., Image plane interaction techniques in 3D immersive environments. *Symp. on Interactive 3D Graphics*. 1997. 39-43.
14. I. Poupyrev, et al., Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. *Eurographics 1998*.
15. B. Shneiderman, Direct manipulation: A step beyond programming languages, *IEEE Computer 16*, 8, 1983, 57-69.
16. SmartScene, promotional material, Multigen-Paradigm, 1999.
17. G. Smith, et. al., 3D Scene Manipulation with 2D Devices and Constraints, *Graphics Interface 2001*, 135-142.
18. P. S. Strauss, et al., The design and implementation of direct manipulation in 3D. *SIGGRAPH 2002 Course Notes*, 2002.
19. Z. Szalavári, M. Gervautz, The Personal Interaction Panel, A Two-Handed Interface for Augmented Reality, *Eurographics 1997*, 335-346.
20. Z. Wartell, L. F. Hodges, W. Ribarsky, A geometric comparison of algorithms for fusion control in stereoscopic HTDs, *IEEE TVCG*, 8, 129-143, 2002.
21. C. Wickins, J. Hollands, Spatial displays, in *Engineering psychology and human performance*, Prentice-Hall, 1999.

Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction

Desney S Tan[†]

Microsoft Research

One Microsoft Way, Redmond, WA 98052, USA

desney@microsoft.com

ABSTRACT

Recent advances in cognitive neuroscience and brain imaging technologies provide us with the unprecedented ability to interface directly with activity in the human brain. Beyond traditional neuroscience work, researchers have begun to explore brain imaging as a novel input mechanism. This work has largely been targeted at allowing users to bypass the need for motor movement and to directly control computers by explicitly manipulating their brain signals. In our work, we use brain imaging to passively sense and model the user's state as they perform their tasks. In this workshop, we will discuss how we are using brain imaging to explore human cognition in the real world, to evaluate interface design, and to build interfaces that adapt based on user state. We will ground our discussions in user studies we have conducted to perform task classification while users perform various tasks. We will also provide lessons learned as well as a general methodology so that other HCI researchers can utilize these technologies in their work.

INTRODUCTION

Human-computer interaction researchers continually work to increase the communication bandwidth and quality between humans and computers. We have explored visualizations and multimodal presentations so that computers may use as many sensory channels as possible to send information to a human. Similarly, we have devised hardware and software innovations to increase the information a human can quickly input into the computer. Since we have traditionally interacted with the external world only through our physical bodies, these input mechanisms have all required performing some motor action, be it moving a mouse, hitting buttons, or speaking.

Recent advances in cognitive neuroscience and brain imaging technologies provide us with the unprecedented ability to interface directly with activity in the brain. Driven by the growing societal recognition for the needs of people with physical disabilities, researchers have begun to use these technologies to build brain-computer interfaces, communication systems that do not depend on the brain's normal output pathways of peripheral nerves and muscles [9]. In these systems, users explicitly manipulate their brain activity instead of using motor movements in order to produce signals that are measured and used to control computers.

While removing the need for motor movements in computer interfaces is challenging and rewarding, we believe that the full potential of brain imaging as an input mechanism lies in the extremely rich information it provides about the user. In our work, we will use brain imaging to passively monitor users while they perform their tasks in order to acquire a more direct measure of their state. We will use this state either as feedback to the user, as awareness information for other users, or as supplementary input to the computer so that it can mediate its interactions accordingly. This communication channel is unlike any we have had and has the potential not only to change the way we view ourselves acting within the environment and to augment traditional interaction processes, but also to transform the field of human-computer interaction.

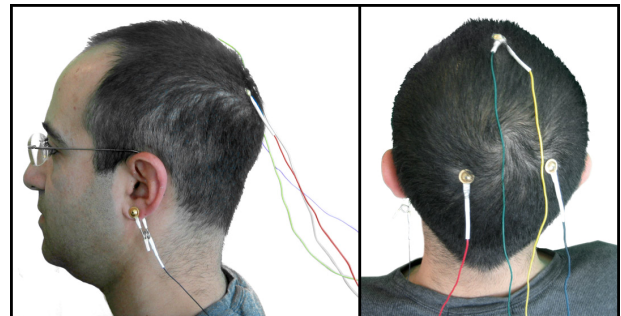


Figure 1. EEG Electrode placements in our experimental setup.

BACKGROUND

The human brain is a dense network of about 100 billion nerves cells called neurons. Each neuron communicates with thousands of others in order to regulate physical processes and to produce thought. Neurons communicate either by sending electrical signals to other neurons through physical connections or by exchanging chemicals called neurotransmitters (for a more detailed discussion of brain function, see [1]). Advances in brain sensing technologies allow us to observe the electrical, chemical, and physical changes that occur as the brain processes information or responds to various stimuli.

With our current understanding, brain imaging allows us only to sense general processes and not the full semantics of our thoughts. Brain imaging is not mind reading. For exam-

[†] Parts of this work performed in collaboration with Johnny Chung Lee, Carnegie Mellon University.

ple, although we can probably tell if a user is processing language, we cannot determine the semantics of the content. We hope that the resolution at which we are able to decipher thoughts grows as we increase our understanding of the human brain and abstract thought, but none of our work is predicated on these improvements happening.

Brain Imaging Technologies

While there are a myriad of brain imaging technologies, which we will review briefly in the workshop, we believe that only two, Functional Near Infrared (fNIR) imaging and Electroencephalography (EEG), present opportunities for inexpensive, portable, and safe devices, properties that are important for brain-computer interface applications within HCI research.

fNIR technology works by projecting near infrared light into the brain from the surface of the scalp and measuring optical scattering that represent localized blood volume and oxygenation changes (for a detailed discussion of fNIR, see [4]). These changes can be used to build detailed functional maps of brain activity. While we have begun exploring this technology, most of our current work has utilized EEG.

EEG uses electrodes placed directly on the scalp to measure the weak (5-100 μ V) electrical potentials generated by activity in the brain (for a detailed discussion of EEG, see [10]). Because of the fluid, bone, and skin that separate the electrodes from the actual electrical activity, signals tend to be smoothed and rather noisy. Additionally, EEG is also susceptible to non-cognitive signals, or “noise”, which have been problematic in traditional EEG work. However, we show how we can actually exploit this noise for accurate task classification within HCI applications.

USING BRAIN-SENSING SIGNALS

User State as an Evaluation Metric

One use of user state derived from brain imaging could be as an evaluation metric for either the user or for computer systems. Since we might be able to measure the intensity of cognitive activity as a user performs certain tasks, we could potentially use brain imaging to assess cognitive aptitude based on how hard someone has to work on specific tasks. With proper task and cognitive models, we could potentially use these results to generalize performance predictions in a much broader range of tasks and scenarios.

Rather than evaluating the human, a large part of human-computer interaction research is centered on the ability to evaluate computer hardware or software interfaces. This allows us not only to measure the effectiveness of these interfaces, but more importantly to understand how users and computers interact so that we can improve our computing systems. Thus far, we have been relatively successful in learning from performance metrics such as task completion times and error rates. We have also used behavioral and physiological measures to infer cognitive processes, such as mouse movement and eye gaze as measures of attention, or

heart rate and galvanic skin response as measures of arousal and fatigue.

However, there remain cognitive phenomena such as cognitive workloads or particular cognitive strategies, which are hard to measure externally. For these, we typically resort to clever experimental design or subjective questionnaires which give us indirect metrics for specific cognitive phenomena. In our work, we explore brain imaging as a measure that more directly quantifies the cognitive utility of our interfaces. This could potentially provide powerful measures that either corroborate external measures, or more interestingly, shed light on the interactions that we would have never derived from external measures alone.

Adaptive Interfaces based on User State

If we take this idea to the limit and tighten the iteration between measurement, evaluation, and redesign, we could design interfaces that automatically adapt depending on the cognitive state of the user. Interfaces that adapt themselves to available resources in order to provide pleasant and optimal user experiences are not a new concept. In fact, we have put quite a bit of thought into dynamically adapting interfaces to best utilize such things as display space, available input mechanisms, device processing capabilities, and even user task or context.

In our work, we assert that adapting to users’ limited cognitive resources is at least as important as adapting to specific computing affordances. One simple way in which interfaces may adapt based on user state is to adjust information flow. For example, verbal and spatial tasks are processed by different areas of the brain, and cognitive psychologists have shown that processing capabilities in each of these areas is largely independent [1]. Hence, even though a person may be verbally overloaded and not able to attend to more verbal information, their spatial modules might be capable of processing more data. Sensory processes such as hearing and seeing, have similar loosely independent capabilities.

Using brain imaging, the system could know approximately how the user’s attentional and cognitive resources are allocated, and could tailor information presentation to attain the largest communication bandwidth possible. For example, if the user is verbally overloaded, additional information could be transformed and presented in a spatial modality, and vice versa. Alternatively, if the user is completely cognitively overloaded while they work on a task or tasks, the system could present less information or choose not to interrupt the user with irrelevant content until the user has free brain cycles to better deal with more information. This is true even if the user is staring blankly at the wall and there are no external cues that allow the system to easily differentiate between deep thought and no thought.

Finally, if we can sense higher level cognitive events like confusion and frustration or satisfaction and realization (the “aha” moment), we could tailor interfaces that provide feedback or guidance on task focus and strategy usage in

training scenarios. This could lead to interfaces that drastically increase information understanding and retention.

INITIAL RESULTS

For a BCI technology to be useful as a communication device, the system must be capable of discriminating at least two different states within the user. A computer can then translate the transitions between states or the persistence of a state into a form that is appropriate for a particular application [6].

Our initial (as yet unpublished) results represent our first steps in exploring how BCI technology can be applied to HCI research. The contributions of this work are twofold. First, we demonstrate that entry into the field does not necessarily require expensive high-end equipment, as is commonly assumed. Results from a user study we conducted show that we were able to attain 84.0% accuracy classifying three different cognitive tasks using relatively low cost (\$1500) off-the-shelf EEG equipment (see Figure 1 for setup). Second, we present a novel approach to performing task classification utilizing both cognitive and non-cognitive artifacts measured by our brain sensing device as features for our classification algorithms. We present a second user study showing 92.4% accuracy classifying three tasks within a more ecologically valid setting, determining various user states while playing a computer game.

User Study 1

In our first study, we used machine learning approaches in an attempt to take only EEG data and classify which mental task a user was performing. We adopted the general experimental design presented by Keirn and Aunon [6] in an effort to extend their results using a low-cost EEG device.

Based on pilot studies with our system, we chose three tasks to classify: 1) *Rest*, in which we instructed participants to relax and to try not to focus on anything in particular; 2) *Mental Arithmetic*, in which we instructed participants to perform a mental multiplication of a single digit number by a three digit number, such as “ 7×836 ”; 3) *Mental Rotation*, in which we cued participants with a particular object, such as “peacock”, and instructed them to imagine the object in as much detail as possible while rotating it.

For data collection, we used the Brainmaster AT W2.5, a PC-based 2-channel EEG system. This device retails for approximately USD\$1500, comparable to the cost of a laptop computer. Eight (3 female) users, ranging from 29 to 58 years of age, volunteered for this study. All were cognitively and neurologically healthy, and all were right handed, except for one participant who had a slight nerve injury in his right hand and who had trained himself to depend more on his left hand.

In order to classify the signals measured using our EEG device, we first performed some basic processing to transform the signal into a time-independent data set. We then computed simple base features, which we mathematically

Mean Classification Accuracy vs. Averaging Scenarios (Mental Tasks)

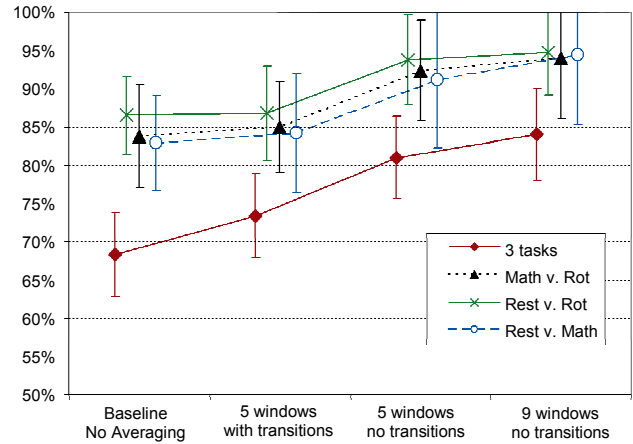


Figure 2. Plot of overall classification accuracies for three mental tasks in study 1 under various averaging scenarios. Error bars represent standard deviation.

combined to generate a much larger set of features. Next, we ran these features through a selection process, pruning the set and keeping only those that added the most useful information to the classification. This pruning step is important as it eliminates features not useful for classification and prevents a statistical phenomenon known as overfitting. Our feature generation and selection process is similar to that used by Fogarty et al. in their work on modeling task engagement to predict interruptability [8]. We then used the pruned set of features to construct Bayesian Network models for task classification.

The Bayesian Network classifiers for these three mental tasks yielded classification accuracies of between 59.3 and 77.6% ($\mu=68.3$, $\sigma=5.5$), depending on the user. The prior for these classifications, or the expected result of a random classifier, was 33.3%. It should be noted that we would expect a human observer to perform only as well as a random classifier. The pair-wise classifiers had a prior of 50% and yielded accuracies of between 68.5 and 93.8% ($\mu=84.4$, $\sigma=6.0$). Using various averaging schemes, we attained improvements of up to 21% (going from 67.9% to 88.9% for participant 1), with average improvements ranging between 5.1% and 15.7% for the 3 task classifier. See Figure 2 for a summary of these results.

We will present this entire methodology in more detail at the workshop as well as discuss how averaging may be used to enhance the classification accuracies, leading us to our final results. Interestingly, this methodology can be generalized for use with time sequence data coming from other sources, such as physiological or environmental sensors.

User Study 2

The classification accuracies we achieve in the first study suggest that we can indeed reliably measure and classify performance of our three mental tasks. However, we cannot be certain that the phenomena providing the classification power is entirely generated by neuronal firings in the brain.

We believe that various mental tasks are involuntarily coupled with physiological responses [7] and that it is difficult, if not impossible, to isolate cognitive activity using EEG in neurologically healthy individuals. This is problematic for researchers ultimately aiming to apply the technology to disabled individuals, as they have to guarantee that the features of interest are generated solely by the brain. For this reason, many researchers have conducted extensive work to remove the ‘confounds’ introduced by physical artifacts before classification [e.g. 5] or have limited their data collection to include only participants who suffer from the same disabilities as those as their target users.

However, since we are aiming to apply this to a generally healthy population, we only need to determine the reliability of the features in predicting the task. This concept was briefly explored by Chen and Vertegaal for modeling mental load in their physiologically attentive user interfaces [3]. In fact, if non-cognitive artifacts are highly correlated with different types of tasks or engagement, we should fully exploit those artifacts to improve our classification power, even though the neuroscience community has spent large efforts to reduce and remove them in their recordings.

We conducted a second user study to explore using both cognitive and non-cognitive artifacts to classify tasks in a more realistic setting. The tasks we chose involved playing a PC-based video game. This experiment serves as a demonstration of a general approach that could be applied to a much broader range of potential applications.

The game we selected for this experiment was Halo, a PC-based first person shooter game produced by Microsoft Game Studios. The tasks we tested within the game were: 1) *Rest*, in which participants were instructed to relax and fixate their eyes on the crosshairs located at the center of the screen; 2) *Solo*, in which participants navigated the environment, shooting static objects or collecting ammunition scattered throughout the scene; 3) *Play*, in which participants navigated the environment and engaged an enemy controlled by an expert Halo player.

Using the same processing methodology as before, the baseline classification accuracies were between 65.2 and 92.7% ($\mu=78.2$, $\sigma=8.4$) for the 3-task classifiers and 68.9 and 100% ($\mu=90.2$, $\sigma=8.5$) for the pair-wise classifiers. After averaging, we were able to achieve 83.3 to 100% accuracies ($\mu=92.4$, $\sigma=6.4$) for 3-task and 83.3 to 100% ($\mu=97.6$, $\sigma=5.1$) for pair-wise comparisons. While shifted slightly higher, the graph of the results looks similar to that of user study 1.

While it would have been possible to achieve reasonable classification accuracy by hooking into the game state, or even by monitoring the keyboard and mouse, our goal in running this experiment was not to show the most ideal way of discriminating these tasks but to demonstrate the impact of non-cognitive artifacts on EEG-based task classification in a realistic computing scenario. We assert that EEG shows interesting potential as a general physiological input sensor

for distinguishing between tasks in a wide variety of computing applications.

CONCLUSION

The idea of using brain imaging in HCI work has been previously proposed [11] and we believe that the technology is now ripe for us to re-articulate this unexplored area of inquiry as well as the challenges that have to be addressed in order for this research project to be a success. We believe that our initial methodology and results will quickly get other researchers caught up with the topic and look forward to inspiring discussion that follows.

ACKNOWLEDGMENTS

We thank Johnny Lee, Sumit Basu, Patrick Baudisch, Ed Cutrell, Mary Czerwinski, James Fogarty, Ken Hinckley, Eric Horvitz, Scott Hudson, Kevin Larson, Nuria Oliver, Dan Olsen, John Platt, Pradeep Shenoy, Andy Wilson, and members of the Visualization and Interaction Group at Microsoft Research for invaluable discussions and assistance.

REFERENCES

1. Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.
2. Carey, J (ed.). (2002). *Brain facts: A primer on the brain and nervous system* (4th edition). Washington DC, USA: Society for Neuroscience.
3. Chen, D., & Vertegaal, R. (2004). Using mental load for managing interruptions in physiologically attentive user interfaces. *SIGCHI 2004*, 1513-1516.
4. Coyle, S., Ward, T., Markham, C., & McDarby, G. (2004). On the suitability of near-infrared (NIR) systems for next-generation brain-computer interfaces. *Physiological Measurement*, 25, 815-822.
5. Gevins, A., Leong, H., Du, R., Smith, M.E., Le, J., DuRousseau, D., Zhang, J., & Libove, J. (1995). Towards measurement of brain function in operational environments. *Biological Physiology*, 40, 169-186.
6. Keirn, Z.A., & Aunon, J.I. (1990). A new mode of communication between man and his surroundings. *IEEE Transactions on Biomedical Engineering*, 37(12), 1209-1214.
7. Kramer, A.F. (1991). Physiological metrics of mental workload: A review of recent progress. In *Multiple Task Performance* (ed. Damos, D.L.), 279-328.
8. Fogarty, J., Ko, A.J., Aung, H.H., Golden, E., Tang, K.P., & Hudson, S.E. (2005). Examining task engagement in sensor-based statistical models of human interruptibility. *SIGCHI 2005*, 331-340.
9. Millan, J. Adaptive brain interfaces. *Communications of the ACM*, 46(3), 74-80.
10. Smith, R.C. (2004). Electroencephalograph based brain computer interfaces. *Thesis for Master of Engineering Science*, University College Dublin.
11. Velichkovsky, B., & Hansen, J.P. (1996). New technological windows into mind: There is more in eyes and brains for human-computer interaction. *SIGCHI 1996*, 496-503.

The Future of Mobile Device Research in HCI

Jacob O. Wobbrock

Human-Computer Interaction Institute
 School of Computer Science
 Carnegie Mellon University
 Pittsburgh, PA 15213 USA
 jrock@cs.cmu.edu

ABSTRACT

In this paper, I present four trends that point toward the increasing importance of mobile device research in Human-Computer Interaction (HCI). These trends indicate a future in which the gap between the user and the desktop is even wider than today, mobile devices have proliferated even further in society, and computer use is more intermittent and in a greater variety of contexts than our current user models accommodate. The implications are that mobile devices must be made more accessible to an aging population, they must be designed for “situational impairments” incurred from on-the-go use, they must adopt flexible and rapid input mechanisms, and the models that describe them must be revamped to accommodate mobile use and behavior. There are also opportunities for using mobile devices in computer education and medical care in developing nations, where mobile devices are more common than desktop PCs.

Keywords

Mobile devices, handheld devices, mobile phones, input techniques, interaction techniques, situational impairments, physical impairments, user models, computer education, medical care, developing nations.

INTRODUCTION

The last decade has seen a surge in off-desktop human-computer interaction due to the prolific spread of mobile technologies. Such technologies include PDAs, handheld communicators, pocket music players, two-way pagers, digital cameras, smart watches, GPS units, medical and factory devices, and mobile phones. There are also numerous hybrid devices that combine two or more of these devices into a single unit. Along with the advent of these devices has come a flurry of HCI research on topics including mobile input techniques [10], handheld web browsing [9,20], adaptive mobile interfaces [14], interfaces that span from devices to desktops [13], sensing devices [4], and many new mobile applications [7]. Thus far, mobile HCI research has focused mainly on the devices

themselves: how to accommodate small screens, how to make devices smarter, how to design faster input mechanisms, how to establish more reliable communications, etc. In ten years, we have quickly reached the point where we can no longer afford to consider devices in isolation, but must take into account the larger social and contextual factors surrounding mobile device use. Like the early '90s that saw desktop-based HCI go beyond the GUI, we must take mobile HCI research beyond the device itself.

I see four important trends in society and technology that have direct consequences for mobile HCI: (1) the overall aging of the population; (2) the increasing amount of personal computing done away from the desktop; (3) the increasing capabilities of ever-smaller devices; and, perhaps most importantly, (4) the convergence of computing capabilities onto the mobile phone. Taken together, these trends require that the future of mobile HCI research be one which considers *context* as much as *capability*.

In the sections that follow, I discuss these trends and their implications for research, giving examples of projects that would capitalize on them. Then I discuss my own research in related areas, and finally draw some conclusions.

IMPROVING MOBILE DEVICE ACCESSIBILITY

Trend #1: The overall aging of the population.

The current population of the United States is 296.5 million people. By 2050, this number is projected to swell to 419.9 million [15], an increase of nearly 42% in only 45 years. Of the current population, 12% are aged 65 or over,¹ and this number is projected to reach 20% by just 2030 [8]. Europe is also aging. The percent of people aged 65 or over is projected to reach 23.5% by 2030, up from just 14.7% in 2000 [8]. Also consider that the average life expectancy for American males is 75 years, and for females it is 80 years. Life expectancy in Canada is even slightly higher. Clearly, the elderly are fast becoming a crucial demographic to consider, and one for whom current mobile interfaces may not be suitable.

Copyright is held by the author/owner(s).

CHI 2006, April 22-27, 2006, Montréal, Québec, Canada.

“What is the Next Generation of Human-Computer Interaction?”

¹For comparison, Canada currently has 13% of its population aged 65+. Japan is the world’s highest at 20%. Europe as a whole is 16%, whereas Africa is just 3% [8].

In particular, the rapid aging of the population has implications for mobile device accessibility. As people age, they incur various impairments including loss of visual acuity, strength, fine motor control and coordination, and pain in the fingers, wrists, or other joints (e.g. due to arthritis). These impairments are particularly problematic for operating mobile devices, since devices' shrunken form factors, miniature buttons, small fonts, postage-stamp screens, and low pixel counts make even able-bodied use difficult. As the baby boomers give way to Generation X, people who grew up with mobile devices and are accustomed to using them will be increasingly challenged to do so unless devices are made more accessible. Although there have been numerous efforts at making desktop computers accessible, there have been almost no efforts to improve mobile device accessibility. (Recent exceptions are [1,22].) The infamous Digital Divide, a sad but real concept that loomed large in conversations of the Internet in the '90s, now threatens to be prominent again unless device manufacturers, interaction designers, and assistive technologists can make mobile devices accessible to people with physical disabilities.

A major challenge, however, is that the approaches traditionally taken to improve desktop accessibility are not likely to work on mobile devices. A main difference, of course, is that interaction with a desktop PC is mediated by the mouse and the keyboard, neither of which exist on a mobile device. Thus, the approach of *emulating* input devices with accessible hardware or software, like mouse or keyboard emulators do, will probably not work on mobile devices. Since mobile device interaction is with the fingers or a pen, it is more direct than interaction with a desktop PC, and therefore more difficult to emulate. Thus, new approaches to providing accessibility are required.

A promising aspect of research in this area is that it stands to improve the design of mobile devices for *everyone*, not just for those with physical impairments [19]. This is particularly important for mobile devices because, as I will argue in the next section, all users incur impairments of some kind when using mobile devices on-the-go.

RESPONDING TO SITUATIONAL IMPAIRMENTS

Trend #2: The increasing amount of personal computing done away from the desktop.

As mobile devices permeate our lives, greater opportunities exist for interacting with computers away from the desktop. But the contexts of mobile device use are far more varied, and potentially compromised, than the contexts in which we interact with desktop computers. For example, a person using a mobile device on the beach in San Diego may struggle to read the device's screen due to glare caused by bright sunlight, while a user on an icy sidewalk in Pittsburgh may have gloves on and be unable to accurately press keys or extract a stylus. By comparison, the differences between these people's desktop experiences would almost certainly not be so dramatic.

These observations suggest that we need to better understand *situationally-induced impairments and disabilities (SIIDs)* [17,18], or "situational impairments" for short. For example, although situational impairments have been mentioned in the literature, to my knowledge no studies have been conducted to find out how input techniques are affected by the act of walking. (A related exception was a study of text *reading* while walking [12].) Our models of users and empirical user tests are mainly geared toward unimpaired desktop use in a laboratory setting. Thus, our understanding of situational impairments is in its infancy, and more research in this area is needed.

Once situational impairments are better understood, it would be useful and interesting to discover whether physical impairments and situational impairments affect users in similar ways. For example, does a user with poor vision face similar challenges as a user dealing with glare? Is a user whose fingers are inflamed due to arthritis similar to a user whose fingers are very cold? Is a user with tremor similar to a user who is trying to write while walking? Depending on the answers to these questions, certain designs may be able to benefit large numbers of users, and designs successful for one user group may transfer with minimal refinement to another.

Ultimately, it should be feasible to construct devices and interfaces that automatically adjust themselves to better accommodate situational impairments. This kind of sensing and adaptation is in the spirit of the work begun by Hinckley et al. [4] but is taken further to specifically address situational impairments. A device could sense environmental factors like glare, light levels, temperature, walking speed, gloves, ambient noise—perhaps even user attention and distraction—and adjust its displays and input mechanisms accordingly. For example, imagine a device that is aware of cold temperatures, low light levels, and a user who is walking and wearing gloves. The device could automatically adjust its contrast, turn on its backlight, and enlarge its font and soft buttons so as to make the use of a stylus unnecessary. If it detects street noise it could raise the volume of its speakers or go into vibration mode. In short, understanding situational impairments presents us with opportunities for better user models, improved accessibility, and adaptive user interfaces.

UBI-INPUT: LEARN ONCE, WRITE ANYWHERE

Trend #3: The increasing capabilities of ever-smaller devices.

With the advent of so many new devices, it can be frustrating to learn new input techniques when encountering each new device. For instance, on the Palm PDA we had Graffiti. Then it became Graffiti 2. PocketPC devices use Jot. Then there are two-way pagers, like the *Glenayre Access Link II* (<http://www.glenayre.com>), which uses four directional arrows and a SELECT key. On the iPod, we have a scroll ring but no text entry, a feature wished for by many "power users" who have tens of thousands of mp3s and

want to search them or at least jump quickly down long lists using “type ahead.” Mobile phones offer a variety of methods, most commonly *Multitap* and *T9*. Users familiar with the former often find it awkward at first to use the latter. The input techniques on digital cameras are lacking even more, and tagging photos for later organization, search, and retrieval is a nightmare at best and impossible at worst.

The input landscape only threatens to grow more varied as new devices emerge. Furthermore, formerly non-computerized devices are now being imbued with computing power. Many of them will need at least minimal input capabilities. For example, wrist watches, which were commonly mechanical devices, may now be full-fledged PDAs, like the Fossil Wrist PDA (<http://www.fossil.com>). In the future, even simple items like credit cards may be able to accept a password, display a balance or credit limit, or even read our thumbprints. Such devices may therefore require some basic form of input.

But few people want to learn a new input technique for every new device they acquire. We therefore need input techniques that are capable of being used on multiple devices and with a variety of input mechanisms or sensors: so-called “ubi-input.” These techniques must remain consistent in the conceptual models they present to the user so that people can “learn once, write anywhere.” Examples of text entry methods that can be used on a variety of devices include MDITIM [6] and EdgeWrite [21]. If we are to take advantage of the trend of growing device capabilities, we will have to design more powerful input techniques for humans to utilize on virtually any device or off-desktop platform.

EDUCATION & MEDICINE ON MOBILE PHONES

Trend #4: The convergence of computing capabilities onto the mobile phone.

Mobile phones are by far the dominant mobile platform. More than 15 billion SMS messages were sent every month in Europe in 2004 [3]. And Africa is now the world’s fastest growing mobile phone market at 65% per year [2]. In fact, there are more people in Africa using mobile phones than using conventional landline phones.

The explosive growth of mobile phone use in both industrialized and developing nations has yet to be fully exploited by HCI researchers. Mobile phones still suffer from tedious input techniques, poor form factors [5], low resolution, unreadable fonts, and confusing user interfaces. Besides improving these problems, however, is the opportunity for HCI researchers to rethink computing on an entirely new platform apart from the desktop.

A venue for which this “rethinking” may be particularly appropriate is Africa. With the proliferation of mobile phones, African students may have opportunities for computerized education for the first time in their lives. How can mobile phones be used for education in classrooms

whose students have never seen a desktop PC? What kinds of phone-based applications could be developed to help kids learn math, science, writing, or reading? How might the teacher interact with a classroom full of students with mobile phones? The opportunities for voice I/O and for short-range networks might enable new educational opportunities in classrooms across Africa.

Another potential benefit of mobile phones in Africa is for the delivery of medical and health information, particularly to rural areas. Volunteer physicians could use phones to store and retrieve medical histories without needing to have bulky, power-hungry laptops on hand. Also, Africans in rural areas whose phones have connectivity may be able to gain information about symptoms and their treatments, enabling better informed diagnosis and treatment.

There are undoubtedly myriad ways in which mobile phones could be used as the primary computing platforms in 21st century Africa. Microsoft has highlighted this opportunity with a \$1.2 million funding offer for related projects [11]. But research will have to be conducted that involves experts across the HCI spectrum, from anthropology to interaction designers to usability engineers. The social, economic, educational, and medical issues will have to be understood before software can be written or user interfaces designed. This is a large task but the potential benefits to Africans and the computing disciplines could be enormous.

MY OWN WORK ON MOBILE INPUT & INTERACTION

I have worked on two projects that are relevant to the issues raised in this paper. First, I noticed how abysmal handheld web browsing could be on a PocketPC device, and decided to redesign the handheld web browser with new interaction techniques. These techniques were embodied in a prototype called *WebThumb* [20], which featured techniques such as picking apart page elements for retention and reuse, using the directional pad to “break apart” dense columns of links, and using rapid serial visual presentation (RSVP) [16] to intelligently “play” text in place for reading. This paper was early in a line of papers on improving handheld web browsing through interaction techniques (e.g. [9]).

My most recent work has focused on making handheld input more accessible through the development of the *EdgeWrite* input technique [22]. *EdgeWrite* uses a plastic template with a square hole to provide stability for a stylus as it moves along the edges and into the corners of the square in patterns reminiscent of hand-printed letters. *EdgeWrite* was over 18% more accurate than *Graffiti* for able-bodied novices, and 200-400% more accurate for people with moderate tremor caused by Cerebral Palsy, Muscular Dystrophy, and Parkinson’s. In addition, *EdgeWrite* has proved versatile enough to be adapted to a variety of input devices while presenting the same alphabet and conceptual model to its users [21]. Versions of *EdgeWrite* exist for PDAs, joysticks, touchpads, trackballs, buttons, watches, and mobile phones.

CONCLUSION

Important trends are underway concerning the proliferation and use of mobile devices. Although we will have the desktop computer with us for many years to come, mobile devices represent an even larger portion of the future of HCI. New research opportunities exist for improving mobile device accessibility; understanding, sensing, and responding to situational impairments; inventing new input techniques that can be used across multiple devices; and deploying new applications for education and medicine in developing nations. These exciting efforts await researchers skilled in mobile HCI and in meeting the needs of real users.

ACKNOWLEDGEMENTS

I would like to thank Brad A. Myers for his comments on this paper.

REFERENCES

- [1] Blair, P. (2005) A customizable hardware input interface for wireless, mobile devices. *Proceedings of RESNA 2005*. Arlington, Virginia: RESNA Press.
- [2] Ferrett, G. (2004) Africans rush for mobile phones. BBC News. Published May 5, 2004. <http://news.bbc.co.uk/1/hi/world/africa/3686463.stm>
- [3] GSM World. (2004) *The Netsize Guide*. <http://www.gsmworld.com/>
- [4] Hinckley, K., Pierce, J., Sinclair, M. and Horvitz, E. (2000) Sensing techniques for mobile interaction. *Proceedings of UIST 2000*. New York: ACM Press, pp. 91-100.
- [5] Hirota, N. (2003) Reassessing current cell phone designs: Using thumb input effectively. *Extended Abstracts of CHI 2003*. New York: ACM Press, pp. 938-939.
- [6] Isokoski, P. and Raisamo, R. (2000) Device independent text input: A rationale and an example. *Proceedings of AVI 2000*. New York: ACM Press, pp. 76-83.
- [7] Karlson, A.K., Bederson, B.B. and SanGiovanni, J. (2005) AppLens and LaunchTile: Two designs for one-handed thumb use on small devices. *Proceedings of CHI 2005*. New York: ACM Press, pp. 201-210.
- [8] Kinsella, K. and Phillips, D.R. (2005) Global aging: The challenge of success. *Population Bulletin 60* (1). Washington, D.C.: Population Reference Bureau.
- [9] Lam, H. and Baudisch, P. (2005) Summary Thumbnails: Readable overviews for small screen web browsers. *Proceedings of CHI 2005*. New York: ACM Press, pp. 681-690.
- [10] MacKenzie, I.S. and Soukoreff, R.W. (2002) Text entry for mobile computing: Models and methods, theory and practice. *Human Computer Interaction 17* (2), pp. 147-198.
- [11] Microsoft Corporation. (2005) *Digital Inclusion Through Mobile and Wireless Technologies Research Funding Initiative*. <http://research.microsoft.com/>
- [12] Mustonen, T., Olkkonen, M. and Häkkinen, J. (2004) Examining mobile phone text legibility while walking. *Proceedings of CHI 2004*. New York: ACM Press, pp. 1243-1246.
- [13] Myers, B.A. (2001) Using hand-held devices and PCs together. *Communications of the ACM 44* (11), pp. 34-41.
- [14] Nichols, J., Myers, B.A. and Rothrock, B. (2006) UNIFORM: Automatically generating consistent remote control user interfaces. *Proceedings of CHI 2006*. New York: ACM Press, in press.
- [15] Population Reference Bureau. *2005 Word Population Data Sheet*. <http://www.prb.org/>
- [16] Potter, M.C. (1984) Rapid serial visual presentation (RSVP): A method for studying language processing. In *New Methods in Reading Comprehension Research*, D.E. Kieras and M.A. Just (eds). Hillsdale, New Jersey: Lawrence Erlbaum, pp. 91-118.
- [17] Sears, A., Lin, M., Jacko, J. and Xiao, Y. (2003) When computers fade... Pervasive computing and situationally-induced impairments and disabilities. *Proceedings of HCI International 2003*. Amsterdam: Elsevier Science, pp. 1298-1302.
- [18] Sears, A. and Young, M. (2003) Physical disabilities and computing technologies: An analysis of impairments. In *The Human-Computer Interaction Handbook*, J. Jacko and A. Sears (eds). Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 482-503.
- [19] Steinfeld, E. (1994) The concept of universal design. *Proceedings of the Sixth Ibero-American Conference on Accessibility*. Rio de Janeiro (June 19, 1994).
- [20] Wobbrock, J.O., Forlizzi, J., Hudson, S.E. and Myers, B.A. (2002) WebThumb: Interaction techniques for small-screen browsers. *Proceedings of UIST 2002*. New York: ACM Press, pp. 205-208.
- [21] Wobbrock, J.O. and Myers, B.A. (2005) Gestural text entry on multiple devices. *Proceedings of ASSETS 2005*. New York: ACM Press, pp. 184-185.
- [22] Wobbrock, J.O., Myers, B.A. and Kembel, J.A. (2003) EdgeWrite: A stylus-based text entry method designed for high accuracy and stability of motion. *Proceedings of UIST 2003*. New York: ACM Press, pp. 61-70.

Federated Devices: Augmenting Mobility with Ambient Resources

Frankie James

SAP Research
3410 Hillview Avenue
Palo Alto, CA 94304 USA
frankie.james@sap.com

Abstract

Mobile devices are limited by their very nature in terms of the user experience they can provide: to remain portable, mobile devices require a small form factor. With increased demand for powerful mobile software, small screens and clumsy keypads quickly become liabilities for users. The Federated Devices project balances portability and usability in mobile computing by augmenting portable devices with ambient resources, such as large displays, desktop computers, and kiosks. Federation can also give user interfaces to very small devices (e.g., sensors), allowing users to manipulate otherwise inaccessible data.

Keywords

mobile, ambient computing

ACM Classification Keywords

H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces; H.5.1 [**Information Interfaces and Presentation (e.g., HCI)**]: Multimedia Information Systems – Audio input/output

Introduction

Today's computer users are becoming increasingly more mobile. Market data indicates that an ever-larger number of computer users are mobile while working, including not only traditional mobile workers (e.g., "traveling salesmen"), but also workers who spend

most of their time in an office without being at their own desk, blue-collar field workers, etc. [1]

However, mobile devices are limited by their very nature in terms of the user experience they can provide. In order to remain portable, mobile devices are constrained to a small form factor, but small screens and clumsy keypads quickly become liabilities for users. Speech input and output capabilities can enhance the usability of mobile devices, but many applications are not well suited for audio interaction, for example, applications requiring rich visual output (maps or analytics) or sensitive input (password entry).

The Federated Devices project seeks to extend the usability and functionality of devices through connection with other resources. Portable, personal devices (e.g., smart phones) can be augmented with ambient resources, such as large displays, desktop computers, and keyboards. Personal devices establish the user's identity, run applications, or connect to back-end systems, while ambient resources provide input and output. Conversely, ambient or mobile devices connected to extremely small devices with little or no user interfaces of their own can allow users to view or manipulate otherwise inaccessible data. This paper describes the types of federation, indicating where current work may be used to prototype the ideas.

Input and Output Federation

Input and output federation is the combination of devices to augment the I/O capabilities of one or the other. In particular, we focus on the augmentation of small, portable (personal) devices with ambient resources to improve mobile interaction.

Output Federation

Mobile devices often have screens that are too small to adequately display information, especially if the information involves complicated visualizations, graphics, or large quantities of text. Output federation can solve this problem by connecting with a nearby display to present the information. Federated displays can also be used for sharing data, by federating a mobile device with a projector or wall display.

The technology for display federation is already being investigated. The "Personal Server," from Intel's Ubiquity project, is a "small lightweight computer with high-density data storage capability."¹ Its primary purpose is to store data and provide computing power that can be leveraged by other devices using a wireless network, and acts as a web server to push content to the connected device through a web browser. [2] WinCuts from Microsoft Research allows users to push shared areas of live screen content to other machines, which are controlled locally. [3] PdaReach by June Fabrics² displays the contents of a PDA screen on a computer connected via the PDA synchronization cable.

What remains to be investigated are usage scenarios and design for display federation. For example, should *all* content automatically be routed to a federated large display, or should users explicitly control the portions of the output to be displayed? Such considerations likely depend on the applications involved, the security or privacy of the environment, and other factors.

¹http://www.intel.com/research/exploratory/personal_server.htm

²<http://www.junefabrics.com/index.php>

Mobile device output can be augmented using other modalities, as well. Mobile providers are already exploring multimodality to prevent competition over the visual channel between primary mobility tasks (e.g., driving) and secondary computing tasks. Federation with external audio devices would allow users to share the audio stream or improve the quality of the output. MP3 players (in particular, the Apple iPod³) are creating demand for audio federation, so that users can balance portability of their music with high-quality output.

Input Federation

Device federation also has the potential to give users the ability to choose which input devices to use with various computing resources. This idea was initially proposed almost a decade ago as a solution to the accessibility problem. [4] The basic concept was to provide users with a portable device that exactly met their needs in terms of input capabilities: low-mobility users could use single-switch or voice entry, low-vision users could use Braille entry devices, etc.

Within more general use cases, we foresee a need for users to federate input devices to enhance productivity. For example, mobile e-mail devices allow increased connectivity and productivity, but typically at the expense of awkward input mechanisms. Federation of keyboards or other input devices could be used to overcome these limitations.

At the other end of the spectrum, input mechanisms to large screen displays have been investigated by many research groups (see [5] for a recent example).

Federation offers another alternative, by allowing input through a mobile device to control a large-screen display. Our group has sponsored an initial study on input strategies for mobile devices federated with large displays, to explore whether input should come from the mobile device or via a touch-screen on the display. Initial results indicate the answer is task-dependent; further investigation is required to define the properties of tasks that lend themselves to input via one device over the other, independently of modality.

For input federation, the Apple iPod is again a good example. The iPod is portable and allows self-contained functionality, but its input and output functionality increase when federated with other devices such as computers and car stereos. Federation with car stereos is especially interesting in regards to the distribution of I/O: some products federate the iPod to a car stereo's *speakers*, handling input via the iPod itself, while others use the car's output *and* input. User satisfaction with the different products can provide insight into federation for other consumer or business applications.

Data Source Federation

Data sources, in the form of sensors, RFID tags, and other "smart items" are becoming more prevalent in today's environment. Sensors and tags can be used to track objects and provide real-time information in ways unheard of even 5 or 10 years ago. However, the devices are so tiny and so numerous that it is difficult for users to interact with them directly. Federated devices can provide new opportunities to access these nearly invisible information sources by federating them with PDAs and laptop computers that can display their state, or even modify their behavior.

³ <http://www.apple.com/ipod/ipod.html>

Research projects that investigate software updates for embedded and networked devices point the way for data source federation. The OSGi Alliance is creating an open service delivery and management platform so that "software components can be installed, updated, or removed on the fly without having to disrupt the operation of the device."⁴ Similarly, Sun Labs is researching "Small Programmable Object Technology," or Sun SPOT⁵, which allows running Java programs to be moved between sensors. The flexibility inherent in these projects opens the possibility of federating devices to access to data from nearby sensors or tags.

In data source federation, it will be important to investigate the process of connecting and disconnecting devices. This topic of course also arises in input and output federation scenarios, but the sheer volume of sensors and tags available within a local area increases the magnitude of the problem. If the user is required to search for nearby devices and explicitly grant pairing rights, selecting a few sensors (or a nearby display) from a list of thousands of possible devices becomes overwhelming. On the other hand, automatically pairing with thousands of nearby devices will generally be undesirable, as the user is unlikely to *want* to interact with all of them and may be uncomfortable not knowing which devices are connecting at any given time.

Summary

Federated devices offers many opportunities for HCI research. The different types of federation, along with

the required infrastructure, hardware, and software, also bring into play related research in computer security, data privacy, and trust.

Our approach will be to constrain initial investigations and scenarios to users within trusted environments, for example, enterprise workers whose work does not involve sitting at a desk, but who do need to access and update computer systems. Such intra-enterprise scenarios lessen the security and privacy risks (as well as the infrastructure requirements), allowing us to focus on interaction design and user scenarios. Wider-reaching scenarios in unconstrained environments can follow on later, adding rich new layers of investigation.

References

- [1] S. D. Drake, T. Glasson, M. Sandler, R. Boggs, K. Burden, and R. Giusto, "U.S. Mobile Worker Population Forecast and Analysis, 2002 - 2006," IDC 27374, June 2002.
- [2] R. Want, T. Pering, G. Borriello, and K. Farkas, "Disappearing hardware," *IEEE Pervasive Computing*, vol. 1, pp. 36 - 47, 2002.
- [3] D. S. Tan, B. Meyers, and M. Czerwinski, "WinCuts: Manipulating Arbitrary Window Regions for More Effective Use of Screen Space," presented at *CHI 2004*, Vienna, Austria, 2004.
- [4] J. Perry, E. Macken, N. Scott, and J. L. McKinley, "Disability, inability and cyberspace," in *Human Values and the Design of Computer Technology, CSLI Lecture Notes*, B. Friedman, ed. Stanford, CA: CSLI Publications, 1997, pp. 65-89.
- [5] A. Bezerianos and R. Balakrishnan, "The Vacuum: Facilitating the Manipulation of Distant Objects," presented at *CHI 2005*, Portland, OR, 2005.

⁴ <http://www.osgi.org/>

⁵ <http://research.sun.com/projects/dashboard.php?id=145>

Reflecting on the Potential of VR for the Future of HCI

Timo Ropinski, Frank Steinicke, Klaus Hinrichs
Department of Computer Science, University of Münster

1 Introduction

Burdea and Coiffet ([1]) have denoted virtual reality (VR) environments as high-end human-computer interfaces which involve real-time simulation and interaction through multiple sensorial channels. Despite this very promising facts, VR is still not mature and desktop-based computer systems are used for many tasks, which could benefit from the potential of VR interfaces. Some interaction researches even postulate the death of VR interfaces, since they have not been widely accepted so far. Currently VR systems are mainly used in research and in a few highly specialized industries as automotive design as well as gas and oil industry. The fact that these industries have permanently incorporated VR in their production cycle shows, that there are ways to unleash the potential of this promising form of human-computer interaction.

During our work with different VR environments we have identified three major shortcomings, from which we believe the wide acceptance of VR systems suffers from. First of all, the instrumentation of the user burdens certain problems and does decrease the usability of VR systems. It is usually a higher effort to take on tracked gloves or stereo glasses, than just grabbing the mouse to perform an interaction. In this paper we are not going to describe this problem, since we believe it is just a matter of time until these problems can be solved by using appropriate technologies. For instance, when observing the evolution of tracking technologies, it is very likely that in the next years user tracking without any instrumentation is possible. The second shortcoming of VR interfaces is a problem faced very often, when dealing with new technologies; it is a lack of standardization. Since VR still evolves very rapidly, there is currently not *the* standard VR system. In reality it is very unlikely, that two institutions have the same configuration of their VR system, since there is a great variety of IO devices as well as tracking systems available and all these components can be freely combined. Thus, when a VR application has been developed for a certain VR hardware setup it will not work with another setup, which may use a different stereoscopic display or another tracking system. The third major shortcoming we have identified during our work with VR interfaces is the cognitive overload the user is facing when he has to deal with 3D interactions.

When performing 3D interactions in reality the user gets different cues from the environment, e.g., the weight of an object he is moving or an obstacle blocking the movement path. Since these cues are not always provided by VR interfaces, the user has to interact in an unnatural environment often leading to confusion. Of course there have been already solutions proposed for the mentioned shortcomings. Specialized hardware devices, e.g., the phantom device can be used to simulate the weight or the texture of virtual objects. However, an application developed for this VR setup cannot be used within a VR environment not supporting haptic feedback. In this paper we will propose a solution for the two last mentioned shortcomings of VR interfaces. We describe our approach for standardizing the development of VR applications in Section 2 and propose concepts for easing 3D interactions in Section 3.

2 Interface-Independent VR Applications

In this subsection the contribution for VR²S as approach of a generic VR software are pointed out. Application specific demands on a VR software system include especially hardware-oriented issues such as support for multisensory output, various input paradigms, or device independence. Also multi-user support for collaborative interaction must be applicable within the VR system. VR²S meets these requirements and is used in research application and in cooperations with authorities.

Since the design of the VR²S is based on a generic rendering approach of a 3D graphics library, which has been implemented with the purpose to enable rapid development of interactive applications, developers are able to focus on rapid prototyping of interaction and system behavior, which enables a cost-effective implementation of 3D applications. In addition an adequate abstraction for VEs is implemented as well as the synchronization of highly independent modules by a multi-layered application programming interface. The platform independent and modular implementation of VR²S eases both extensions as well usability and allows an easy porting to other VR systems. Since VR²S supports all standard formats exchange of content is ensured. Furthermore, VR²S provides a rendering system independence API, which allows to design VR applications once and to render it with several rendering systems without the need of recoding. In interior design prototyping, for instance, a once modelled scene can be rendered for interactive exploration and sound propagation using OpenGL and OpenAL (see Figure 1).

A generic user interface concept allows to develop applications with VR²S which are not constrained to a specific user interface. The once implemented VR application can be easily used with any existing standard user interface. Thus, the borders between desktop-based and VR applications blur, since there is only one application, which can be used in different environments using different IO devices.

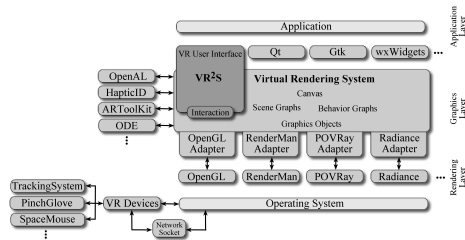


Figure 1: The system architecture of VRS ([2]) and its VR-based extension VR²S consisting of application, graphics and rendering layers.

3 Reducing the Cognitive Overload

The enhancement of human-computer interaction in VR requires the availability of intuitive and natural interaction metaphors with which fundamental interaction tasks can be accomplished. Thus, we have developed several multimodal interaction metaphors, which underline the feasibility of the described interaction concepts within VR²S. For instance, pointing is a gesture fulfilled very easily in everyday life (see Figure 2 (left)). This leads to a wide adaption of pointer metaphors in VR applications. In order to advance the selection and manipulation process of virtual pointer metaphors, we have introduced the improved virtual pointer (IVP) metaphor ([3]). The IVP metaphor allows the user to select a desired object without requiring an exact hit. A straight ray is used to indicate the direction of the virtual pointer, while an additionally visualized bendable ray points to the closest selectable object (see Figure 2 (right)). Thus, the selection process is accelerated since a user can simply perform direct object selection by roughly pointing at desired objects. To further support the user during the interaction we add multimodal feedback. When the curve bends to another object the user perceives a smooth vibration signal, whereby the signal strength increases the nearer the distance between user and object is. In addition, a gentle sound disperses from the position of that object and results in a better spatial cognition about the position of the desired object. This is just one interaction metaphor incorporated into our framework as an easy to use building block.

4 Conclusion

In this paper we have stressed three major shortcomings of VR interfaces, which as we believe are preventing VR from being the state of the art human computer interfaces. We expect that the user instrumentation problem is solved in the next years by researchers of other communities, e.g., by inventing new computer vision algorithms, allowing to track a user without the need for markers. But the other two shortcomings standardization of VR interfaces and reduction of

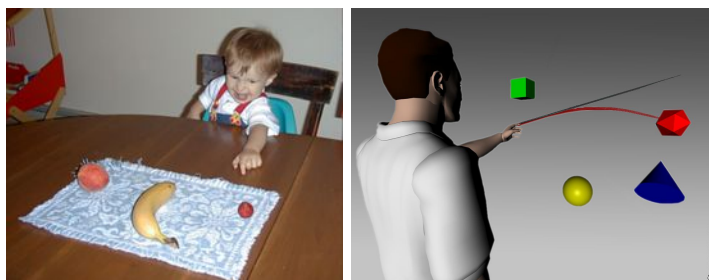


Figure 2: Gesture-based selection in everyday life (left), the IVP selection metaphor (right).

cognitive overload when performing 3D interactions are tasks which should be faced by the HCI community. The fact that automotive as well as oil and gas industry exploit VR hardware shows, that VR has the potential to be the next generation HCI. However, since there is no standard for VR interfaces the effort of time and money invested by these industries is not easily transferable to other domains. We have presented a possible solution by providing an easy to use VR framework called VR²S, which incorporates 3D interaction metaphors, which can be used as building blocks. VR²S fullfills the major demands of VR developers and it has shown its usefulness in industrial and research applications such as city and landscape planing and exploration tools. We do not proclaim VR²S as a standard but we believe, that this framework could be a step towards a standardization for the development of VR interfaces. We hope that this paper helps to initiate an investigation to solve the described problems to make VR interfaces more usable and maybe becoming the next generation human computer interfaces.

References

- [1] G.C. Burdea and P. Coiffet. *Virtual Reality Technology*. Wiley-IEEE Press, 2003.
- [2] J. Döllner and Klaus Hinrichs. A generic rendering system. In *Transactions on Visualization and Computer Graphics*, 8(2), pages 99–118, 2002.
- [3] F. Steinicke, T. Ropinski, and K. Hinrichs. Object selection in virtual environments with an improved virtual pointer metaphor. In *International Conference on Computer Vision and Graphics (ICCVG)*, pages 320–326, 2004.

Pervasive Interaction – Using movement and speech to interact with computers

Thomas Riisgaard Hansen, Jakob E. Bardram, Mads Søgaard

Pervasive Healthcare

Department of Computer Science, University of Aarhus

8200 Aarhus N, Denmark

{thomasr, madss, bardram}@daimi.au.dk

ABSTRACT

In this paper we will focus on pervasive interaction that allows the users to interact with computers when needed and not just when they are sitting in front of a computer. Inspired by the work done in hospitals we describe several projects and frameworks that address pervasive interaction. We focus on two types of interaction: Movement based interaction and speech based interaction. Finally, we suggest that the role of *context* and *uncertainty* would be relevant to elaborate further on at the workshop.

Keywords

Pervasive interaction, movement based interaction, speech based interaction, mixed interaction spaces.

INTRODUCTION

Pervasive interaction is interaction with computers that takes place at the time and place decided by the use context and the interaction is not limited and dictated by the presence of a personal computer.

The mouse and keyboard have for a long time been the only mean of communicating with a computer, but both interaction devices require the presence of a flat surface, a desk. Moving from desktop interaction to more pervasive interaction, e.g. interaction in busses, trains, cars, hospitals, and when walking or running, the assumption that the user is sitting or standing at a flat surface breaks. A number of new interaction techniques have appeared to support this non-desktop situation. Mobile phone interaction, pen based interaction, large wall display interaction, augmented interfaces, and tangible interfaces are just some examples of interfaces that promise to carry interaction away from the desktop.

In this paper we focus on movement based interaction and speech based interaction, two non-desktop interaction technologies. We present a set of projects and frameworks we have developed and the main contribution of the paper is to highlight the possibilities for pervasive interaction with these two interaction technologies.

Our main focus area is interaction with computers in hospitals and within the hospital we have been working with two scenarios. How two interact with computers while

been mobile and how two interact with computers during surgery.

MOVEMENT BASED INTERACTION

Movement based interaction is interaction where the movement of the human body is used as input to the computer. In this section two projects are presented that uses movement.

In the first project we focus on interaction with large wall displays. We track the location of both of the user's hands and overlay the image of the tracked hands as a dual cursors.

In the project mixed interaction spaces we use the camera in a mobile phone to track the position of the device in relation to a feature. The movement of the device is thus registered and used as input to a large variety of different applications.

Finally we have been working on a framework for movement based interaction that uses cameras as tracking devices.

Hands2Cursor

In the first project we worked with how a surgeon can interact with computers while operating.



Figure 1: A typical operation room setup

Figure 1 is from a typical surgery. The surgeons are sterile and situated next to the patient. How can you interact with a computer in this situation? In a first attempt to address this situation we look at gesture interaction. Would it be possible for a computer to recognize the surgeon's hands and if so could this movement information be used to control a user interface?

Based on this initial research question we developed a system based on a PC and a webcam. We have worked with a number of different tracking algorithms e.g. different variants of Hough algorithms, different edge and feature detection algorithm, flood fill algorithms and the CamShift algorithm. The outcome was a system that found the position and rough rotation of both of the user's hands under most conditions.

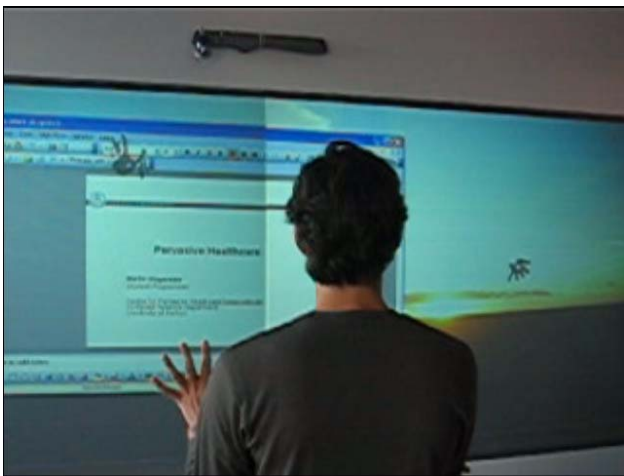


Figure 2: Hands interaction with a large wall display. The tracked hands are projected on the interface.

However, a major finding (not surprisingly) from this work was that it was really difficult to find a robust algorithm that performed well in all conditions. Therefore it was really important to visualize to the user the performance of the system. Did the system work as planned or had it lost track of the hands? We worked with an approach where we projected the images of the tracked hands as the computer saw them on top of the interface at the position where the computer thought they were. This approach proved to be rather successful because it provided the user both with a feeling of actually manipulating the interface with the hands and at the same time visualized how well the system was able to track and locate the user's hands.

After a period of time we paused the project temporary because other interaction techniques, e.g. speech as described in the next section, seemed more suitable for this type of environment. However, some of the findings were transferred to a mobile platform and used in the next project. A video of the tracking system and further information is available at [2, 6].

Mixed Interaction Spaces

With the project Mixed Interaction Spaces we looked at how to interact with computers while been mobile. A mobile device is required to fit into the pocket of the user, which limits the possibilities for interaction. However, at the same time a mobile device is almost always present and at hand. In the mixed interaction space project we looked at how to use the fact that a device is mobile to facilitate new ways of interactions. In the project we use the embedded camera most mobile devices are equipped with to track one or more features. By tracking a feature with the mobile device we get a fixed-point or reference point we can use to determine the position of the mobile device and its movement in relation to this feature.

In the current version of the prototype we have three different algorithms running on a Nokia Series 60 platform. We have a variant of Hough circle detection algorithm for tracking printed or hand drawn circle and we have two algorithms based on the CamShift algorithm for tracking either a colored object or the users face.

Figure 3 shows what happens when the face tracking algorithm is used (this is only possible on mobile phones with a camera pointing towards the user). The face works as a fixed point. By looking at the size of the face the mobile device can determine if it is being moved closer or further away from the face. The algorithm can also determine if the phone is moved up or down, left or right and to some extent rotation. This is however, only possible if the face keeps within the visible range of the camera. The space in which the phone can track the feature is highlighted on figure 3 (not to be confused with the cameras field of view).

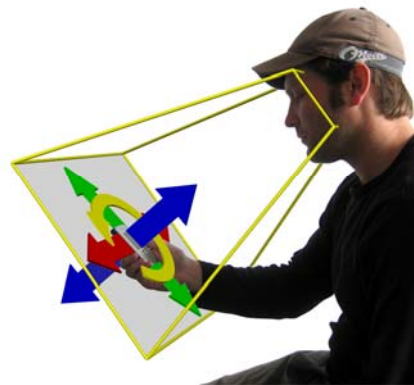


Figure 3: The Mixed Interaction Space with face tracking

Based on this technique we have developed several different applications. Some of the applications run on the phone and use the interaction technique to pan and zoom on maps or images or use it as input to different games. Other applications run on a nearby large wall display where e.g. a cursor can be controlled with the interaction technique. Bluetooth is used to transmit the input

information to the large wall display and we have developed different multi-user applications that also use this technique as input.

A typical use scenario is that a user starts the program, connects to e.g. a large wall display, look for a feature (e.g. the top of a Coca-Cola bottle as in figure 4), and use this feature as fixed point.



Figure 4: The top of a Coca-Cola bottle is used as reference point for the tracking application.

The main findings in the project is described in [3, 4], however the technique seems really promising. It works really robust and you can use any nearby object, draw a circle or use your own head to interact with the system. Besides most people carry a mobile phone and using it as an interaction mediator between interfaces in the world allows the interface to remain simple and robust while allowing several users to hook up to the interface with their own devices.

Movement based framework

Based on the above present projects and a range of other camera based applications we have identified three important concepts and build a framework around these three concepts.

The three main components are space, relation and feedback.

Space: We found all the projects that used cameras to track the movement of the user somehow worked with the notion of space. The scale, orientation and the number of used spaces seemed to be both characteristic and important. Some of the projects use large spaces that are able to contain multiple users, while other e.g. the space between a mobile phone and the users face is much smaller. Some uses the floors as main interaction areas while others use walls etc.

Relations: When a camera tracks a feature, a relationship can be described between the camera and the feature. Depending on the tracking algorithm multiple features can

be tracked thus spanning multiple relations. Also the information an algorithm can extract from a relation can differ. Some algorithms are e.g. only able to identify colored objects while others are able to uniquely identify the different features through 2D barcodes.

Feedback: Finally, as hinted in the section Hands2Cursor, camera based interaction rely heavily on user feedback. Being able to describe how, what kind and where feedback occur is important. Is an application using auditory feedback or visual feedback? Is the information presented on a hand held device or a shared surface? How is multi-user feedback addressed?

We have used the above summarized framework to describe and analyze nine very different camera based interfaces. The full framework and the analysis can be found in [1].

SPEECH BASED INTERACTION

Speech based interaction is another area we have been working heavily with. As with your body, speech is always there and in many situation where your hands are occupied it is the only alternative. Speech interaction has been around for decades however, the performance and accuracy have been greatly improved lately and speech recognition is now also available on even handheld devices.

ActiveTheatre

We have been working extensively with speech interaction in the project named ActiveTheatre. The main problem we address in this project is the same as with the Hands2cursor. How do you interact with computers while operating? Speech seemed much more promising than gestures for interacting while operating. When operating the surgeon normally use his hands for operating and multiple users occludes the cameras field of vision.

The idea behind the active theatre project is to use large wall display and speech interaction to recall medical images and data from the electronic patient record while operating. However, we wanted to not only facilitate accessing the data, but also actively creating new data while operating. Examples would be to capture video and images of the operation or using speech recognition to access medical information.

We have build and evaluated several different prototypes that use speech as input. Figure 5 shows a surgeon from a nearby hospital trying the system out.



Figure 5: A surgeon is testing the interface. The interface runs on both a large wall shared display and an arm mounted 17" inch screen.

In the application we use two different modes of speech recognition. We use speech commands to control the application e.g. to zoom in on pictures or text, to take pictures and to navigate in the interface. The speech commands are based on a predefined syntax. The second mode is free dictation where every word is recognized. We use this e.g. for annotation the pictures or for writing notes.

In the first versions we use Microsoft Speech Recognition engine as a proof of concept, but are currently working with a Danish speech recognition engine focusing on the medical domain based on Phillips Speech Engine.

We are currently developing the third version of the system, which is going to be tested during real life surgeries later this spring. The project is further described in [5].

CONCLUSION

Pervasive interaction requires novel form for interfaces that supports new use situations. There are many forking paths that move forward. In this paper we have presented some of the work we have been doing focusing on movement and speech interaction.

Many of the findings are published elsewhere, however there are some common themes in the above presented work that could be interesting to discuss at the workshop.

Context: When moving away from the desk into the real world the number of different contexts explodes and the importance of taking the context into consideration is getting more and more important. What are the important

findings with novel interaction techniques that allow them to work both in operating theatres, school, busses, while running etc. or is the question not to find cross context principles, but to find specialized interaction techniques that are tailored to a specific work setting?

Uncertainty: Many of the novel interfaces that appear use sensors with uncertainty associated with them. How do you handle uncertainty in the interaction technique? Movement based systems are not always able to recognize your gesture and speech recognition system might hear wrong. Maybe the computer heard what you said, but that was not what you actually meant. How to handle uncertainty with interaction could be another interesting topic.

REFERENCES

1. Eriksson, Eva, Hansen, Thomas Riisgaard, Lykke-Olesen, Andreas, Space, Relation and Feedback - describing, discussing and analysing movement based interaction using video tracking, submitted for publication.
2. Hansen, T.R., Interaction without Keyboard and Mouse, UbiHealth workshop, Ubiquitous Computing Conference 2004, Nottingham, England.
3. Hansen, T. R., Eriksson, E. & Lykke-Olesen, A., Mixed Interaction Space — Designing for Camera Based Interaction with Mobile Devices, in CHI'05 Extended Abstracts of the Conference on Human Factors in Computing Systems, ACM Press, pp.1933–6, 2005.
4. Hansen, Thomas Riisgaard, Eriksson, Eva., Lykke-Olesen, Andreas, Mixed Interaction Space – Expanding the Interaction Space with Mobile Devices. In Proceedings of British HCI 2005, 2005.
5. Hansen, Thomas Riisgaard, Bardram, Jakob E., ActiveTheatre – a Collaborative, Event-based Capture and Access System for the Operating Theatre, accepted as full paper at UbiComp 2005.
6. Videos at Pervasive Interaction.org: <http://www.pervasive-interaction.org/videos/>

Managing Extrinsic Costs via Multimodal Natural Interaction Systems

Rebecca Lunsford, Ed Kaiser, Paulo Barthelmess, Xiao Huang

Natural Interaction Systems

10260 Greenburg Road

Suite 400

Portland, OR 97223 USA

rebecca.lunsford; ed.kaiser; paulo.barthelmess; xiao.huang; @naturalinteraction.com

ABSTRACT

Modern day interactions, whether between remote humans or humans and computers, involve extrinsic costs to the participants. Extrinsic costs are activities that, although unrelated to a person's primary task, must be accomplished to complete the primary task. In this paper we provide a framework for discussing certain extrinsic costs by describing those we term over-specification, repetition, and interruption. Natural interaction systems seek to reduce or eliminate these costs by leveraging peoples' innate communication abilities. However, in conceiving these interfaces, it's critical to acknowledge that humans are naturally *multimodal* communicators, using speech, gesture, body position, gaze, writing, etc., to share information and intent. By recognizing and taking advantage of humans' innate ability to communicate multimodally, extrinsic interaction costs can be reduced or eliminated. In this paper we review previous and ongoing work that demonstrates how multimodal interfaces can reduce extrinsic costs.

Author Keywords

Natural interaction system, multimodal interfaces, mutual disambiguation, extrinsic interaction costs

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

In all but the most straightforward interaction between two co-located individuals, people pay extrinsic costs to support their interactions. Whether giving directions over the phone, transcribing meeting notes, or making plane

reservations using a speech enabled application, people must engage in activities that are unrelated to the primary task, in order to accomplish that task. In many situations users accept these extrinsic costs as a minor inconvenience. However, in some cases users feel the costs outweigh the benefit received and refuse to use a system [4].

In our view, many of these extrinsic costs are a direct result of non-existent or beleaguered communication channels that do not allow people to communicate using the broad range of modes typical of co-located human-human communication. As a framework for discussing these extrinsic costs and how multimodal interfaces can minimize or reduce these costs, we focus on *over-specification*, *repetition*, and *interruption*.

Over-specification

Over-specification occurs when one needs to translate information that can be conveyed easily via one or more communication modalities into other, less natural modes. For example, when giving directions on the telephone, one must translate spatial and directional information into complex verbal descriptions such as "take the 295 exit for Taylor's Ferry, and get in the center lane because you're going to have to take a wide left then an immediate right to actually get on Taylor's Ferry". Alternatively, if the parties were co-located and sharing a map, the same instructions might be as simple as "when you exit you need to get in the center lane because you want to end up here", accompanied by tracing the route with a finger. We term this extrinsic cost "over-specification" in that the user must engage in a more cognitively demanding, less natural communication due to the constraints of the communication channel.

Repetition

Repetition is the process of transcribing information into another format, typically digital. Consider the following example. A project manager and team meet to discuss an ongoing software development project and determine that research is needed on software licensing. The project manager then writes "software licensing" on a flipchart, turns to one of the team members, and says "can you handle

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2006, April 22–28, 2006, Montréal, Québec, Canada.

Copyright 2006 ACM 1-59593-178-3/06/0004...\$5.00.

that?" to which the team member responds with a nod while writing a star next to his meeting notes — his way of identifying the task as his. On returning to their offices the team member adds the action item to his digitized to-do list and the project manager adds the action item to the electronic project plan. Although the action item was discussed verbally, labeled on the flip chart and assigned via conversation and note taking, it was also typed into a to-do item and typed into the project plan. We term this extrinsic cost "repetition" in that the same piece of information is processed multiple times.

Interruption

Interruption is defined as "to stop or hinder by breaking in" [13]. For the modern-day information worker, interruption can be viewed as a side-effect of availability. That is, by having open communication channels such as phone, email, IM, and office door, the user is available to be interrupted by any of these devices. Researchers working on minimizing the impact of interruption have already recognized that multimodal sensors will be needed to identify when a person is most interruptible [6]. However, this work doesn't address interruption within an interaction by the interaction partner misinterpreting the users' behavior.

One example of this type of interruption are automated features (such as in Microsoft's XP operating system) which alter the interface state based on a user's actions, for example spontaneously adding a clipboard pane if the user selects a second item to copy without pasting the first selection. However, in doing this, the application has not only "stopped or hindered" but has effectively forced an added task onto an already busy user - that of figuring out how to return the interface to a state where they can comfortably continue with their task. For the purposes of this paper, we term this extrinsic cost "interruption", in that the user's primary task must be postponed while the secondary task is concluded.

Background - Multimodality in Evolving Interfaces

We believe that as computational systems become more and more perceptually capable of recognition at all levels — (1) symbolic/syntactic input levels (e.g. speech, handwriting and gesture recognition), (2) semantic understanding of both single and integrated multiple input modes (e.g. parsing and semantic modeling), and (3) pragmatic contextualization and tracking of interaction structure — they will benefit from multimodal integration. We believe that reality-based interfaces will be fundamentally multimodal interfaces, or what we call natural interaction systems.

Multimodal systems support and leverage the fact that people communicate in various modes. On one hand multimodal inputs carry complementary information. For example, people are adept at using fine information in one input mode to disambiguate coarse information in another

input mode. In a classroom full of desk/chair groups pointing to a distant pair and saying, "Move that chair," is an effectively understood command. However, saying only, "Move that chair," without pointing, nor pointing at the distant group while saying only, "Move that," are not effectively understood commands — they both require further specification. This mutual disambiguation, using information from one uncertain input mode to disambiguate complementary information in another uncertain input mode, is a powerful combinatory tool for dealing with the inherent ambiguity in individual recognizer interpretations. Ranked lists of semantic interpretations can be combined and filtered during mutual disambiguation. Choosing the best scoring combinations using mutual disambiguation reduces semantic error rates by between 40%-67% [14] [8].

On the other hand, aside from carrying complementary information, multiple input modes at times carry semantically redundant information. In human-human, computer mediated interactions like distance lecture delivery or whiteboard-centered business meetings (particularly those of a "presenter") are often semantically redundant. One recent study of tablet-PC delivered distance-learning lectures found that 100% of randomly selected handwriting instances were accompanied by semantically redundant speech. Our own analysis of a spontaneous white board planning session (recorded from a quarterly project planning meeting of 15-20 participants) also found that whiteboard and flip-chart handwriting were accompanied by semantically redundant speech 98% of the time.

Thus we argue that evolving ubiquitous, tangible and perceptive interfaces need to leverage people's natural use of multimodal communication. Multimodality is what users expect in natural human-human interactions. System integration of multimodal inputs using mutual disambiguation has been shown time and again to increase overall system recognition levels. In the area of human-computer interactions the primary benefit of mutual disambiguation is due to complementarity of inputs. Whereas in human-human, computer-mediated interactions multimodal semantic redundancy is common, and leveraging its occurrence can be the basis for perceptual interfaces that learn better how to observe, recognize and understand multiple input streams as they are used.

ELIMINATING OVER-SPECIFICATION

An important source of extrinsic cost can be associated with the way technological constraints force users to adapt their communicative behavior and consequently their work practices. Conventional interfaces require users to express their intentions by means of textual input and widget selections; that is at the same time both very limited and too different from usual communication modes people use.

As discussed above, people communicate through multiple channels simultaneously, in such a way that these channels of information complement or reinforce each other,

reducing ambiguity. In contrast, conventional interfaces impose a strictly sequential operation performed by manipulating physical devices such as keyboards and mice.

While conventional interface paradigms may be appropriate for a potentially large class of domains, they are in general less acceptable when users are dealing with more demanding domains e.g. when visual spatial components are involved, or in multiparty situations. In these situations, interface limitations lead to use of circuitous language (e.g. [2, 10]), as users attempt to compensate for technological limitations by exploiting in unnatural ways remaining available channels. In addition, there is commonly a disruption of the natural flow of work, resulting from the need to perform actions whose sole purpose is to steer the technology, e.g. by requiring users to explicitly control turn-taking and pointer control through a series of widget selections while on distributed meeting.

In contrast to conventional interfaces, systems that are able to analyze multiple modalities, as produced in the course of work performance, support unchanged work and communicative practices while introducing minimal additional system-related costs. In the next paragraphs we highlight some of the previous work of the group that illustrates this quality.

Work within our group has explored the use of speech and pen input as means to support a natural way for users to perform a variety of map-based tasks. NISMap [4], derived from the Quickset system [3] allows for users to annotate maps with military symbols by sketching conventional symbols and speaking. Earlier versions of NISMap required the use of interactive surfaces such as touch sensitive screens; more recently, NIS Map has been extended to support drawing over regular paper maps [4]. This transition to support conventional tangible materials makes NIS Map use virtually indistinguishable from the non-automated, paper-based operation users are already accustomed to, significantly lowering extrinsic costs associated with its use.

More recently, we explored multimodal perception and integration to support remote collaboration. The Distributed Charter system [1] transparently handles the propagation of project schedule sketches produced by small teams working at distributed sites. In this system, the multimodal contributions made by distributed participants, as they sketch on instrumented boards or tablet computers and talk about a schedule, are integrated by the system into a coherent semantic interpretation of the interaction. The system thus keeps the representation of the schedule consistent across sites, independently of device and resolution used. In addition, a stereo-vision tracker and recognizer is used to perceive deictic gestures made towards the interactive boards at each site and transparently propagates the gestures, making participants aware of naturally occurring pointing gestures made by users at remote sites. Once more, the goal of this system is to take

advantage of naturally occurring group behavior to drive system mediated awareness services without requiring explicit user actions to drive the system.

REDUCING REPETITION

By leveraging existing work practices (allowing users to draw units on post-it notes and place them on the map) our group's earlier Rasa system reduced the repetition necessary to digitally record a complex battle-scene or planning map [12]. Instead of hand copying the large map, or taking pictures and typing in information to digitize it, the underlying digitizing technology simultaneously and unobtrusively recorded and recognized the inputs. Since perceptual technologies have improved, even the modest intrusiveness of these recording methods is no longer necessary. Reliable digitizing ink technology (e.g. Anoto¹) is based on a special pen containing a camera, which operates on paper printed with a (almost invisible) dot pattern. With the dot patterns, the camera in the pen determines the relative location of the ink coordinates with respect to the paper. These coordinates can then be transmitted to a computer for storage and further processing. Rasa's digitizing tablets can be and have been replaced by digital paper and pens [4]. The map itself is now printed on digital paper making the touch sensitive whiteboard and the post-it notes optional.

Using digital paper and speech recording devices can have wide applicability. For example, repetition costs in human-computer interaction are prevalent in note taking and form-filling. To minimize these costs, we are developing a new collaborative note-taking facility called NISNotes. Regular handwritten notes taken on digital paper can be directly transferred in real time to current digital note taking systems like Microsoft OneNote. This digital ink can be combined with speech to produce more accurate recognition transcripts. That means less copying of notes, less correction of transcripts, etc. By incorporating our MOOVR/SHACER architecture (Multimodal Out-Of-Vocabulary Recognition using a Speech and HAndwriting reCOgnizER) [7] we intend to evolve versions of NISNotes that learns and improves over time as their vocabulary and understanding increase.

We are already able to show the advantages of systems that learn and improve over time in Charter, our automated interface to MS Project scheduling. By combining sketch, speech, gesture and task-contextual information Charter supports the dynamic learning of new vocabulary like abbreviation semantics (e.g. that "JB" stands for "Joe Browning"). This learning also occurs in the context of existing work practices — the creation of a whiteboard chart in a collaborative meeting. Eventually when Charter is combined with NISNotes, we believe that the collaborative creation and manipulation of complex project charts will

¹ <http://www.anoto.com>

become even better as more sources of information all contribute to mutually disambiguate each other. Repetition can be eliminated because to-do lists and project plans need not be re-typed later. Also, having a digital memory of meetings facilitates easier review and sharing of summaries and project planning views as well as supporting more informative distributed interfaces as described above in the section on Eliminating Over-Specification.

Another example of reduced extrinsic costs is related to Physicians who are accustomed to using paper forms. With only a paper record, opportunities are missed for improving both individual care and institutional efficiency. To avoid this issue in current practice, clinicians are often expected to engage in repetitive data entry [5].

Using the Anoto technology, we have built NISChart [4], a digital paper-based charting application where handwriting and speaking are the primary input modalities. A physician can enter values, text, check marks, and so on into the hospital's standard forms, printed on Anoto paper. After finishing writing, the physician can put the digital pen into a cradle, from where the digital ink is sent to the computer. The application recognizes the texts and symbols written on the form. Both the digital ink and transcripts are saved in a relational database. Thus, we have the physical paper with real ink as the primary record. Moreover, the ink is digitized, analyzed, recognized and saved as reusable records without retyping. In other words, by eliminating repetition costs, now data entry is "as easy as writing" with a multimodal natural interaction system.

REDUCING INTERRUPTION

Audio-visual open-microphone speech-based interfaces often assume the user is addressing the system if he or she is speaking while facing the system. However, in recent work it's been shown that users often look at a system while addressing a peer [9], or while engaging in self-directed speech relevant to completing the task at hand [11]. In attempting to respond to these non-system directed utterances, systems implement a state-change that the user must then reverse.

In recent empirical work, we found that users engaging in self-directed speech while interacting with a speech-enabled application will consistently use significantly lower amplitude as compared to speech addressed to the system [11]. By leveraging this additional mode already available in the speech channel, future applications will be able to reduce interruption by attempting interaction only when the user is, in fact, addressing the system. Ongoing work will explore whether amplitude, in concert with other natural behaviors, can differentiate self-, system, and other human directed speech in a multiparty environment leading to more natural human-human-computer interactions.

CONCLUSION

In describing the extrinsic costs of over-specification, repetition and interruption inherent in modern-day

interactions, we provided an initial framework within which to discuss and evaluate those costs. In addition, by detailing ways in which multimodal applications can and do reduce these costs, we demonstrate ample evidence to support our belief that the goal of natural interaction systems should be to reduce the extrinsic costs inherent in modern-day interactions by recognizing and utilizing natural human interaction modes and supporting effective current working practices.

ACKNOWLEDGEMENTS

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. NBCHD030010. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the DARPA or the Department of Interior-National Business Center (DOINBC).

REFERENCES

- [1] Barthelmess, P., E. Kaiser, X. Huang, & D. Demirdjian. Distributed pointing for multimodal collaboration over sketched diagrams. ICMI'05: ACM Press: 10-17.
- [2] Bly, S.A. A use of drawing surfaces in different collaborative settings. CSCW 1988: ACM Press: 250-256.
- [3] Cohen, P.R., M. Johnston, D. McGee, S. Oviatt, J. Pittman, I. Smith, L. Chen, & J. Clow. QuickSet: multimodal interaction for distributed applications. ACM Multimedia 1997: ACM Press: 31-40.
- [4] Cohen, P.R. & D.R. McGee, Tangible multimodal interfaces for safety-critical applications. Commun. ACM, 2004. 47(1): 41-46.
- [5] The Computer-based Patient Record: An Essential Technology for Health Care. 2 ed. Institute of Medicine. 1997, Washington. D.C.: National Academy Press.
- [6] Horvitz, E., P. Koch, & J. Apacible. BusyBody: creating and fielding personalized models of the cost of interruption. CSCW 2004: ACM Press: 507-510.
- [7] Kaiser, E. SHACER: a Speech and Handwriting Recognizer, Workshop on Multimodal, Multiparty Meeting Processing. ICMI'05
- [8] Kaiser, E. & A. Olwal. Mutual Disambiguation of 3D Multimodal Interaction in Augmented and Virtual Reality. ICMI'03
- [9] Katzenmaier, M., R. Stiefelwagen, & T. Schultz. Identifying the addressee in human-human-robot interactions based on head pose and speech. ICMI'04: ACM Press: 144-151.
- [10] Luff, P., C. Heath, H. Kuzuoka, J. Hindmarch, K. Yamakazi, & S. Oyama. Fractured Ecologies: Creating Environments for Collaboration. Human-Computer Interaction, 2003. 18(1&2): 51-84.
- [11] Lunsford, R., S. Oviatt, & R. Coulston. Audio-visual cues distinguishing self- from system-directed speech in younger and older adults. ICMI'05: ACM Press: 167-174.
- [12] McGee, D.R., P.R. Cohen, R.M. Wesson, & S. Horman. Comparing paper and tangible, multimodal tools. CHI 2002: ACM Press: 407-414.
- [13] Merriam Webster Online. <http://www.m-w.com/>
- [14] Oviatt, S. Mutual Disambiguation of Recognition Errors in a Multimodal Architecture. CHI '99

E. Tools and Development Techniques

E.1 An Approach for Developing Ubiquitous Augmented Reality Systems

Bernd Bruegge Technical University of Munich
Asa MacWilliams Technical University of Munich
Asim Smailagic Carnegie Mellon University

E.2 Reflecting on Production Tools and Methods for Mixed Reality Design

Steven Dow Georgia Institute of Technology
Blair MacIntyre Georgia Institute of Technology

E.3 Exploring Activity-Based Computing: Interaction Styles, Models and Tool Support

Yang Li University of Washington
James A. Landay University of Washington and Intel Research Seattle

An Approach for Developing Ubiquitous Augmented Reality Systems

Bernd Bruegge
Asa MacWilliams
Technical University of Munich

Asim Smailagic
Carnegie Mellon University

1. Introduction

Augmented reality is an interaction technology that provides users with spatially related information in the real world in real time [1]. An augmented reality system tracks the position and orientation of objects, for example the user's head, and superimposes virtual objects into the user's field of view. Augmented systems support multiple human-computer interaction techniques for input (multi-modality) and output (multi-media). Ubiquitous computing aims to provide a "calm" interface to computers, making them invisible, yet omnipresent [2]. Users should be able to use computing technology in their environment without having to think about it as such. For this, devices must provide their services autonomously and shield the user from configuration tasks. Additionally, other users can bring along devices which themselves provide new services. Ubiquitous computing is a broader field than augmented reality, and includes many different ways of augmenting the users' environment, with augmented reality as one possible technology. As the computing and interaction hardware required for augmented reality becomes smaller and more ubiquitously available, augmented reality and ubiquitous computing can converge, creating systems that are ubiquitously available and allow interaction in the style of augmented reality.

In ubiquitous augmented reality applications, users are involved with tasks in the real world, and they cannot solve these tasks by sitting in front of a desktop computer. This includes mobility: users move around in a fairly large, inhomogeneous area. At the same time, users wish to access or modify information associated with their current real-world situation, which is not otherwise immediately available.

Building ubiquitous augmented reality systems presents a challenge [3], because the requirements are still ill-defined, as appropriate interaction metaphors are still being researched and users' preferences change.

To deal with this problem, this paper proposes a new prototyping style called extreme modeling. In extreme modeling, users and developers address ill-defined requirements by changing the system's behavior during development as well as while it is running, facilitating the creation of new applications. We describe three prototype systems that have been built with this methodology.

2. Extreme Modeling

In extreme modeling, the system is not only deployed incrementally, but also be developed incrementally as well. Thus, parts of the system will be deployed — and tested by the end user — before others are even developed. In extreme modeling, developers in collaboration with end users incrementally improve a system while users use it and provide feedback. This follows the idea of shorter development cycles, as, for example, in extreme programming [4] or Scrum [5]. Two techniques that can be used in the process are described: jam sessions, where users and developers cooperate synchronously, and continuous extension, where they cooperate asynchronously.

2.1 Jam Sessions

A jam session is a synchronous group activity of users and developers, cooperating in improving a running system, which is deployed within one room, or a small site consisting of several adjacent rooms. A jam session takes from a few hours up to a whole workday. In developing a system, users and developers may wish to have several jam sessions consecutively. The basic unit of work in a jam session is the feedback-develop cycle. When the end user finds a fault or suggests an improvement in a component, a developer changes the component, taking it off-line briefly, and restarting it while the rest of the system remains running. Or a developer implements several alternative versions or configurations of a component, and the user can test and chose among them.

In a jam session, several of these develop-feedback cycles take place simultaneously, each involving at least one user and one developer. While one part of the system changes, the rest continues to run, and thus different parts of the system are improved simultaneously. A jam session does not have to be centrally

coordinated; it follows the idea of a jam session in music, where participants take the lead in turn. At times, one group of developers and users will have to defer a change to the system, because it would interfere with a test being performed by another group. This kind of group coordination requires a certain level of experience from developers and users. A jam session also requires support from the architecture and runtime infrastructure for ubiquitous augmented reality systems: in particular, the software architecture must allow components to be exchanged while the system is running.

Jam sessions are particularly suitable for building prototypes of ubiquitous augmented reality applications as well as evaluating and improving their usability. A jam session is less suitable for the development of new algorithms within a specific component, or for performance tuning.

2.2 Continuous Extension

In contrast to jam sessions, continuous extension assumes that developers and users cooperate asynchronously, and that the system supports their collaboration. Continuous extension assumes that a system has been - at least partially - deployed and is in use. When a user discovers a problem or an opportunity for improvement, he uses a feedback tool, which is part of the system, and records a wish for improvement. The system stores this wish, together with the current user and system context—such as the state of services the user is using, the current graph of services, and the user's location. The Kiva, described in [6], is an example of such a system.

The system can also gather usage information automatically, by creating profiles of which services get used in which circumstances, whether the performance is adequate, or how long it takes the user to complete a task. Using the information that the system has gathered, developers implement missing functionality, improve existing services, and fix bugs.

This can happen off-line in a development laboratory. To test new and experimental features, they can even use jam sessions again, either with the same users who requested new functionality, or with selected pilot users. The new versions are then deployed to the hardware in the environment, and onto the users' mobile hardware.

If the components are backward-compatible, this can happen in the background, without users noticing. Once the new components are dynamically deployed, users can take advantage of the improved functionality, and a new feedback–development cycle can begin.

As with jam sessions, many of these cycles are active at any given time, as users can provide feedback asynchronously. When context changes or the user gives the system an appropriate command, it can adapt itself immediately—assuming it “knows” how. Assuming a speech recognition interface, the user can say, “show me an asana for improving my upper body strength”, and a wall-sized display in the environment is reconfigured to show the asana. On the other hand, the system may not be able to react immediately when the required functionality is missing. In this case, the system reacts by saying “sorry, I cannot do that yet”, record the user's wish, and a developer implements the missing functionality. When the user tries the command again the next day, the system can fulfill it.

The concept of end user programming allows users to reprogram their ubiquitous computing systems with a simple visual interface. A simple example is a rule-based system: users define rules telling the system what to do in which circumstances. By contrast, in continuous extension, users do not program directly, but record requests for developers. This is based on the assumption that most users of a ubiquitous augmented reality system are not programmers. Of course, continuous extension can be combined with end user programming; users can define simple rules and add simple functionality themselves, getting immediate results; and professional developers handle the more complex programming tasks.

3. Case Studies: Skill-Coaching Systems

Teaching a skill involves four basic steps: Preparation, Explanation, Demonstration and Practice. The explanation introduces the subject and talks about its usefulness and applicability. During demonstration the teacher shows each step so that the learners can easily follow and gain confidence in their own ability to acquire this skill. During practice the learners try out the skill under the teacher's guidance

When teaching a skill, the focus is on the learners; the teacher is trying to help *them* gain this skill and be able to use it with a sense of comfort and confidence. For complicated skills such as yoga, playing piano or conducting, practicing the skill under the teacher's guidance is crucial, because learning these techniques from a book, tape or video is severely limited. The (real) teacher cannot be present all time during practicing. We argue that ubiquitous augmented reality can be used to replace a teacher periodically in practice sessions. In this case the interaction between computer and learner during the practice session is especially important.

We argue that the interaction should have the following properties:

- **Adaptivity:** If the learner is not familiar with the skill, the system must adapt to the learner. For example, the system should be able to first check on accuracy and then for speed, if necessary.
- **Noninterference:** The system should not interfere when learners try to do it on their own. In particular, the system should not interrupt their efforts unless they bog down or go off on the wrong track.
- **Support for trial-and-error:** The learners should be allowed to make mistakes if this will help them learn. That is, the system needs to distinguish between a moment where it should interfere and when not.
- **Feedback:** In the case the system needs to point at mistakes, it should do it tactfully. In particular, the system needs to provide multi-modal output to select from a variety of reactions depending on the learner situation and context. For example, when a yoga student violates a position that requires symmetry, the system should be able to select from feedback in form of a spoken instruction or different types of music indicating the severity of the violation.
- **Annotations:** Encourage the learners by making remarks on their progress, pointing out the completion of each step, and remarking on the steps they have done well.
- **Learning:** The system should be able to learn from the interaction with the learner and adjust the next interaction based on this knowledge.

Rapid prototyping has successfully been used to develop GUI based systems: The developer shows a graphical interface to end user or domain expert, who makes comments that may lead to the revision of the user interface. Extreme modeling goes beyond rapid prototyping. The validation of the system is no longer completely done during development. The development of the system continues during the use of the system, even to the point that the user has to interact with the system to revise some of its behavior, which can be seen as a change of the requirements. That is, two types of interaction must be supported: 1) Interactions to practice the skills with the learner and 2) interactions that allow the domain expert and/or end user to change the system at runtime to improve its usability. The same mechanisms can be used for both types of interactions. This requires that the underlying software architecture must peer-to-peer and not client-server based: The system has to act as a true peer, being able to initiate an interaction or reacting to an unplanned interaction initiated by the user.

3.1 Personal Yoga Instructor

We developed a context-aware application for learning Indian Yoga, in particular Hatha Yoga. Yoga is poised to become a major theme for the wellness market of the near-to mid-term future. A student learning Yoga attends a weekly Yoga class and doing exercises at home between the classes. In the Yoga class, the qualified Yoga teacher provides advice, but at home, nobody guides the student to perform the exercises correctly. We built a "Personal Yoga Instructor" (PYI). The PYI, shown in Figure 1, gives advice on planning and selecting appropriate daily and weekly Yoga exercises, provides direct feedback on achievements through non-intrusive interaction, and alerts students to damaging postures with Yoga poses. The PYI is an example of a new class of applications that use smart sensors and novel interaction modalities emerging in the next decade that move beyond mobile peer-to-peer voice and text message communication.

3.2 Interactive Conducting

We are going to use accelerometers on a small wearable device and machine learning algorithms to recognize the arm movements of a symphony orchestral conductor. This will be an effective mechanism to support the training and practicing of conductors. We are planning to perform testing with a symphony orchestra later this year. We are also aiming to use gesture recognition and compare recognition accuracy. The system will provide feedback to conductors to improve their performance.

3.3 Wellness Coach

The wearable Personal Wellness Coach system [7], based on our machine learning algorithms described in [8], and wearable computing [9], supports a range of activities including health data collection, interpretation, feedback and self-monitoring. It is shown in Figure 2. We implemented a wearable computing platform consisting of an array of sensors in the form of an armband, and application software to motivate users to reach fitness goals and prevent harm in a real-time environment. Evaluation and testing were done on 65 users in four groups: students, health club members, ROTC members and Baby Boomer-age individuals. Comparing the Personal Wellness Coach against an existing exercise product, our results indicate that this system is an effective tool that is less intrusive and in some cases can perform like a personal trainer and help improve user efficiency.



Figure 1. Personal Yoga Instructor



Figure 2. Personal Wellness Coach

4. Discussion Points for the Workshop

- For which types of systems is extreme modeling applicable?
- How do we interact with systems whose behavior is not reproducible?

5. Bibliography

[1] Azuma, R. T., "A Survey of Augmented Reality," *Presence*, Vol. 6, No. 4, August 1997, pp. 355–385.

[2] Weiser, M., "Hot Topics: Ubiquitous Computing," *IEEE Computer*, Oct. 1993.

[3] Asa MacWilliams, A Decentralized Adaptive Architecture for Ubiquitous Augmented Reality Systems, Dissertation, Technical University Munich, 2005.

[4] Beck, K., *eXtreme Programming Explained: Embrace Change*, Addison-Wesley, 1999.

[5] Schwaber, K. and Beedle, M., *Agile Software Development with Scrum*, Prentice Hall, 2002.

[6] Finger, S., Gelman, D., Fay, A., Szczerban, M, Smailagic, S., Siewiorek, D.P., "Supporting Collaborative Learning in Engineering Design," *International Journal of Expert Systems and Applications*, to appear, 2006.

[7] Asselin, R., Ortiz, G., Pui, J., Smailagic, A., Kissling, C., "Implementation and Evaluation of the Personal Wellness Coach," *Proc. IEEE International Workshop on Smart Appliances and Wearable Computing*, June 2005, Columbus, OH

[8] Krause, A., Smailagic, A., Siewiorek, D.P., "Context-Aware Mobile Computing: Learning Context-Dependent Personal Preferences from a Wearable Sensor Array," *IEEE Transactions on Mobile Computing*, Vol. 5, No. 2, 2006, pp. 113-127.

[9] Smailagic, A., "Wearable Computers: A New Paradigm in Computer Systems and Their Applications," *IEEE Transactions on Computing*, Vol. 52, No. 8, 2003, p. 977

Reflecting on Production Tools and Methods for Mixed Reality Design

Steven Dow and Blair MacIntre

GVU Center, College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0760, USA
{steven, blair}@cc.gatech.edu

ABSTRACT

The workshop’s emphasis on “reality-based” interfaces and understanding humans’ skills and expectations from the real world pertains to our work on *mixed reality* experience design and the development of tools and methods to support that design process. Pulling ideas from media theory, interaction/experience design, cognitive science, sociology, and philosophy, we reflect on several new media experiences created in our interdisciplinary collaboration interleaving our accounts of the production tools and methods found to be most valuable. We share our preliminary thoughts towards methods for creating mixed reality experiences including forming a “studio” with appropriate content production tools and social configurations, conducting contextual inquiries within analogous design settings such as game and film production, and building future prototyping tools and methods to aid creative work in this medium.

Keywords

Mixed reality, augmented reality, design studio, design methods, authoring tools, sketching, video prototyping

INTRODUCTION

The next generation of human-computer interfaces will be determined, to a large extent, by the next generation of content producers, or creative individuals who can weave technology into meaningful designs for people. Here at Georgia Tech, a collaborative effort between the College of Computing and the School of Literature, Communication, and Culture is underway to create a design “studio” for mixed reality (MR) technology, a term we use to encompass pervasive computing, augmented reality (AR), virtual reality, tangible computing, and other “reality-based” interfaces. Akin to Milgram’s continuum [22], we view the full spectrum of lightweight to heavyweight technologies as computational media. By developing the necessary design tools and methods for rapid design and prototyping, we begin to understand how humans’ skills and expectations from the real world play out in both authoring spaces and collaboratively designed media experiences.

Our design work on new media experiences and the tools to support these endeavors is guided by philosophical foundations in phenomenology [9,10]. Merleau-Ponty argues that the human experience and our conscious actions on the world are mediated through our perceptions [21]. In mixed

reality experience design we leverage cultural and social expectations, often based on previous media forms [1], to augment human modalities of perception for storytelling and other forms of art. We believe that both *technology* and *human perception* mediate our experience of the world and that mixed reality technologies are an emerging medium for communication and expression.

Our emphasis on design tools stems from Heidegger’s notion of “ready-to-hand”; that tools are unconscious extensions of our bodies and that our primary concern is with the human activity [9]. As Dourish notes, “I act *through* the computer mouse (to operate menus and so forth); the mouse is an extension of my hand.” [7] Engrained in our process of experience design and tool development is the continual discovery of how technology tools best recede to the background of consciousness. Ideally our tools will achieve the notion of *flow*, an optimal design experience somewhere between anxiety and boredom, expressed eloquently by Csikszentmihalyi [2]. We aim to provide useful tools for mixed reality designers, many of which take advantage of our natural human abilities.

In this workshop paper, we will discuss three different mixed reality experiences created through our interdisciplinary collaboration using our prototyping environment, the Designer’s Augmented Reality Toolkit or DART [18]: the Voices of Oakland [3], AR Karaoke [8], and AR Façade. Our practical endeavors have provided insight into an emerging design craft for mixed reality, leading to reflections on specific tools and methods [5,18] and providing direction for further tools, such as a sketch interface for storyboarding AR experiences [27], partly inspired by work at Pixar [28]. We share our current thoughts towards the future of production tools and methods for mixed reality, emphasizing advances on video prototyping tools intended to help designers communicate situational context while iterating sketch content over time

RELATED WORK

Our design studio for mixed reality is similar to Pausch’s *Building Virtual Worlds* project using the ALICE environment at CMU [25], in that we encourage intense, multidisciplinary design teams. Our objective is to support a spectrum of media technology, from lightweight to heavyweight, and to constantly work to evolve our tools and methods.



Figure 1 (a) Participant experiencing The Voices of Oakland audio tour. (b) A view from the head-mounted display in Augmented Reality Karaoke. (c) Screenshot from the desktop version of the 3D interactive dramatic, Façade.

Landay’s thesis and the tools created by his students at Berkeley [12,13,15,16] imbue the philosophy of rapid, informal content creation we instill in DART. Our sketch actors are intended for early-stage mixed reality design. We hope to better understand this process so we can support design through all stages.

Our current investigation on video prototyping lead us to Mackay’s exposition of video for multimedia system and augmented workspaces design [17]. We hope to learn from and enhance the current methods for video prototyping enabling early-stage sketching and iteration.

THE DESIGN OF MIXED REALITY EXPERIENCES

We will briefly describe three mixed reality experiences developed over the past few years, and then reflect on the tools and methods that have made those efforts possible. Integral to our collaboration is the prototyping environment used for most of our experiences, DART. We created DART in 2004 to enable designers familiar with Macromedia Director [19] to experiment with 6DOF tracking, live video, and other sensor information in a media rich environment. We will not only talk about specific technology-focused authoring tools in DART, but also particular social arrangements and how they aided the design process for mixed reality experiences.

The Voices of Oakland

The Voices of Oakland project (Figure 1a) is a location-sensitive audio tour set in historic Oakland Cemetery in Atlanta [3]. In the audio experience, we attempt to enhance the environment with dramatic stories from the lives of cemetery “residents”, so that visitors can better understand the history of Atlanta and the South. We are particularly interested in exploring a blend of linear and non-linear storytelling as participants wander through this historic place.

To develop of The Voices of Oakland we instituted weekly afternoon work sessions among the interdisciplinary project group. These intensive sessions allowed designers to get help on technical questions and provided developers insight on how to improve the tools. For example, we built high-level support for the Wizard of Oz method and data visualization into DART; designers used these tools to iteratively evaluate the experience throughout its design [5]. This was important

because it allowed for rapid content creation without relying on unpredictable GPS tracking. The social organization of the design environment facilitated the development of both the audio tour and the underlying tools.

Augmented Reality Karaoke

In Augmented Reality Karaoke (Figure 1a) users perform their favorite dramatic scenes with virtual actors [8]. AR Karaoke is the acting equivalent of traditional karaoke—the goal to facilitate an acting experience that is entertaining for both the actor and audience. The actor dons a tracked head-mounted display (HMD) giving her a first-person view of a movie or TV scene where the other characters are positioned in the physical space. The system prompts the actor to read lines from the scene and directs her to move around the space.

We produced Augmented Reality Karaoke through our course on MR Design, a cross-listed class pulling graduate and undergraduate students from computer science and design. Again, the social orientation of design played a role in the project’s success, as students split into small, diversely skilled groups and developed sketch-based content prototypes in DART to be shared with the entire class. During the course we improved the sketch tools in DART to allow sketch annotations to be overlaid onto tracked video [27]. This method marks an improvement over importing sketch content from drawing tools such as Photoshop because sketching can happen more rapidly and in situ.

Augmented Reality Façade

We are currently just beginning the development of Augmented Reality Façade, a fully embodied AR version of Mateas and Stern’s critically acclaimed 3D interactive drama, Façade [20]. Originally conceived for traditional desktop interaction (Figure 1c), players use the keyboard to navigate an apartment and type natural language statements to responsive characters in the midst of ongoing marriage troubles. The concept for AR Façade places the player in a physical apartment with a tracked head-mount display. The virtual characters now appear in the physical space and the player interacts with them by walking around the space and speaking aloud. The final experience is reminiscent of Murray’s vision of Hamlet on the Holodeck [23], and

plays on our natural human abilities as social and physical creatures.

Unlike the two experiences described above, the AR Façade project is not designed in DART and does not have a fluid design space. Because of the massive AI engine and custom code, we are constrained to adding video and tracking support into the Façade code base, a task only suited for a savvy graphics programmer. To help us move the project forward, the designers on the team are developing a sketch-based storyboard of the experience inside DART that demonstrates the basic concept and allows us to work on design decisions in the physical space. We need to figure out the furniture layout, player's interaction with physical objects, lighting conditions, the ergonomics of the HMD, and microphones for listening to the player's communication with the characters. In this case, DART serves as a temporary prototyping tool to pre-visualize the experience, in similar vein to Peter Jackson's film production techniques [11].

POTENTIAL FOR DESIGN

Looking forward, we see potential for creative work across the spectrum of lightweight to heavyweight technologies into different contexts. Virtual Reality is too *heavy* to play a role in regular everyday life, but it could play an important role in defining future high-end entertainment, such as games and narrative experiences. On the other hand, pervasive sensors and displays integrate more seamlessly into existing contexts. We believe technology mediates at different levels of perceptual awareness and consciousness, that it's important to support creative work across all mediums. Towards VR, methods are being developed for designing more directly with the 3D content [26]. Ideally designers will work in the actual 3D environment, not on a 2D desktop, but challenges remain. Researchers still need to work out the gestural consistency of constrained 2D interaction. Towards pervasive computing, and other lightweight computer mediated experiences, researchers are working on tools to support prototyping of physical/digital devices [14], a design task typically split between an industrial designer and an interaction designer. In practice, designers have started adopting storytelling techniques to communicate the context of interaction [6]. Across the spectrum of media, there is room for improving the design process and providing more adequate tools.

PLANS FOR IMPROVING THE MR DESIGN PROCESS

Our plans for lowering the threshold for media designers exploring mixed reality follows two paths: 1) qualitative fieldwork looking at analogous design contexts, the history of computational design tools, and emerging practice in MR design, and 2) development of new tools and methods. Video prototyping methods are particularly interesting because they have the potential to influence a wide range of media.

Qualitative Fieldwork

Due to the relative novelty of mixed reality technologies, it will be difficult to conduct fieldwork let alone empirical

evaluations. Design traditions in mixed reality have not been established, so it pushes our research to find alternate resources.

Study of Analogous Design Contexts

Our recent qualitative research of professional designers from industry reveals some of the emerging external representations and methods used for ubiquitous computing design [6]. The study focused mostly on lightweight technology, but outlined an interesting research strategy for informing the next generation tools. We hope to continue looking at the production of analogous media forms, particularly film, television, gaming, and photography for common representations, techniques, divisions of labor, and other applicable design issues. By performing contextual inquiries with professional designers in established environments we hope to inform the design of mixed reality.

Historical Review of Computational Tools

We can learn by studying the trends in media tools such as Photoshop, Illustrator, Director, Flash, etc. What were the technological and social factors influencing the evolution of these tools? How does tool assortment and interaction affect design craft? We believe a rich overview of media computation tools will help us understand the evolution of mixed reality tools.

Design Workshops and Experimental Vignettes

Perhaps most important to the study of MR design is to shape a flourishing tradition. We are integrating practices from schools of art, design, and architecture into our mixed reality design studio. Media designers will create experimental vignettes—short examples of structured interaction or display techniques—as building blocks for larger experiences. We will hold intensive design workshops and critiques with seasoned craftspeople. Craft practice, “getting our hands dirty” so to speak, has always taught us the most about the MR medium, so we will continue to push for more compelling experience design.

New Tools and Methods

We currently have plans to develop a number of tools for DART including added support for video prototyping, tangible manipulation and sketching.

Video prototyping (a.k.a. Replay Reality)

DART includes infrastructure for capture and replay of video and time-synchronized sensor data [4]. Originally developed to break the requirement for co-located design and testing of experiences in physical spaces, we realized this infrastructure supports storytelling narratives that can be iterated. Content can be developed and iterated within editable video scenarios and used to communicate design ideas for head worn augmented reality experiences or augmented workspaces. Video provides an effective early-stage representational medium, consumable by designers and potential users of the application.

Tangible Manipulators

We are exploring tangible manipulators for mixed reality, especially for positioning 3D content in augmented reality spaces. This conceptually taxing undertaking requires designers to perform complex transformations in 3D. Creating real-time tangible methods for placing content in the space, such the method proposed to the pervasive computing community [24], lowers the barriers for designers and exploits our natural ability to move around and point within a space.

Sketching in Space

Our preliminary work on annotation tools for DART [27] needs further development including more sophisticated sketch support and better methods for placing content into the world. We plan to tightly integrate our capture/replay and sketching infrastructure with the tangible tools for placing content. Our vision for “sketching in space” is a very lightweight, tablet-based tool for sketching concepts directly into a physical space. We will discuss the current progress of this work during the workshop presentation.

CONCLUSION

This work reflects on the design of mixed reality experiences through specific social orientations and production tools. As we push boundaries in this medium we think generalizable methods for MR experience design across the spectrum from VR to AR to pervasive computing will continue to surface. Our plans for qualitative research and tool development define a course of research, likely to be dissertation work for the first author. We hope the research on video prototyping proves to impact a wide range of designers in new media. Ultimately we want to study how mixed reality technology mediates our perception of the world, and how it affects the human experience.

ACKNOWLEDGEMENTS

We thank Jay David Bolter, Maribeth Gandy, Jaemin Lee, Christopher Oezbek, Danny Muller, Michael Mateas, Manish Mehta, and Tobias Lang for their efforts on various projects and Marshini Chetti and Zachary Pousman for early edits.

REFERENCES

1. Bolter, J.D. and Grusin, R. *Remediation: Understanding New Media*. Cambridge, MA: MIT Press, 1999.
2. Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience*. New York: Harper & Row, 1990.
3. Dow, S., Lee, J., Oezbek, C., MacIntyre, B., Bolter, J.D., and Gandy, M. Exploring Spatial Narratives and Mixed Reality Experiences in Oakland Cemetery. In ACM SIGCHI Conf. on Advances in Computer Entertainment, 2005.
4. Dow, S., MacIntyre, B., Gandy, M., and Bolter, J.D. Prototyping Applications for the Physical World Using Integrated Capture/Playback Facilities. In Ext. Abst. of Conf. on Ubiquitous Computing, 2004.
5. Dow, S., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J.D., and Gandy, M. Wizard of Oz Support throughout an Iterative Design Process. In IEEE Pervasive Computing, November, 2005.
6. Dow, S., Saponas, T.S., Li, Y., and Landay, J.A. External Representations in Ubiquitous Computing Design and the Implications for Authoring Tools. Submitted to Conf. on Designing Interactive Systems, 2006.
7. Dourish, P. *Where the Action Is: The Foundations of Embodied Interaction*. Cambridge, MA: MIT Press, 2001.
8. Gandy, M., MacIntyre, B., Presti, P., Dow, S., Bolter, J.D., Yarbrough, B., O’Rear, N. AR Karaoke: Acting in Your Favorite Scenes. In International Symposium on Mixed and Augmented Reality, 2005.
9. Heidegger, M. (1927) *Being and Time*. English Translation. New York: Harper & Row, 1962.
10. Husserl, E. (1913) *Ideas: General Introduction to Pure Phenomenology*. The Hague: Martinus Nijhoff, 1981.
11. Jackson, P. *King Kong: Peter Jackson’s Production Diaries*, MCA Home Video, DVD set, 2005.
12. Klemmer, S.R., Sinha, A.K., Chen, J., Landay, J.A., Aboobaker, N., Wang, A. SUEDE: A Wizard of Oz Prototyping Tool for Speech User Interfaces. In ACM Symposium on User Interface Software and Technology, 2000.
13. Landay, J.A. *Interactive Sketching for the Early Stages of User Interface Design*. Ph.D. Thesis, Carnegie Mellon, Dec. 1996.
14. Lee, J.C., Avrahami, D., Hudson, S., Forlizzi, J., Dietz, P., and Leigh, D. The Calder Toolkit: Wired and Wireless Components for Rapidly Prototyping Interactive Devices. In Proc. of Designing Interactive Systems, 2004.
15. Li, Y., Hong, J.I., Landay, J.A. Topiary: A Tool for Prototyping Location-Enhanced Applications, In Conf. on User Interface Software and Technology, 2004.
16. Lin, J. and Landay, J.A. Damask: A Tool for Early-Stage Design and Prototyping of Multi-Device User Interfaces. In Conf. on Distributed Multimedia Systems, 2002.
17. Mackay, W.E. Video Prototyping: A Technique For Developing Hypermedia Systems. In ACM Conf. on Human Factors in Computing Systems (CHI 88), 1988.
18. MacIntyre, B., Gandy, M., Dow, S., and Bolter, J.D. DART: A Toolkit for Rapid Design Exploration of Augmented Reality Experiences. In Conf. on User Interface Software and Technology, 2004.
19. Macromedia Director MX 2004: <http://www.macromedia.com/software/director/>
20. Mateas, M. and Stern, A. Facade: An Experiment in Building a Fully-Realized Interactive Drama. In Game Developer’s Conference: Game Design Track, 2003.
21. Merleau-Ponty, M. *Phenomenology of Perception*. London: Routledge, 1945.
22. Milgram, P. and Kishino, F. A Taxonomy of Mixed Reality Visual Displays, IEICE Transactions on Information Systems, Vol E77-D (12), 1994, 1321-1329.
23. Murray, J. *Hamlet on the Holodeck*. New York: The Free Press, 1997.
24. Patel, S.N., Rekimoto, J., Abowd, G.D. iCam: Precise at-a-distance Interaction in the Physical Environment. To appear in Proc. of Pervasive Computing, 2006.
25. Pausch, R., et al. Alice: Rapid Prototyping System for Virtual Reality. In IEEE Computer Graphics and Applications, May 1995.
26. Pierce, J.S. Expanding the Interaction Lexicon for 3D Graphics. Ph. D. Thesis, Carnegie Mellon, Nov. 2001
27. Presti, P., Gandy, M., MacIntyre, B., Dow, S. A Sketch Interface to Support Storyboarding of Augmented Reality Experiences. In Extended Abstracts of SIGGRAPH, 2005.
28. Wolff, E. Tool Time at Pixar. Millimeter, Nov. 1, 2004.

Exploring Activity-Based Ubiquitous Computing: Interaction Styles, Models and Tool Support

Yang Li¹ & James A. Landay^{1,2}

¹DUB Group
Computer Science and Engineering
University of Washington
Seattle, WA 98105-4615 USA
{yangli, landay}@cs.washington.edu

²Intel Research Seattle
1100 NE 45th Street
Suite 600
Seattle, WA 98105
james.a.landay@intel.com

ABSTRACT

Activity-based computing is a promising paradigm for ubiquitous computing. By providing a consistent framework and structural view for integrating ubicomp technologies into natural human activities, activity-based computing can better facilitate our day-to-day lives and allow ubicomp technologies that are sustainable in the dynamic, complex world in which we live. In this position paper, we surface several aspects of activity-based computing. We first briefly describe what activity-based computing is. We discuss our early ideas on an activity framework for structuring ubicomp technologies and representing interaction contexts. Next, we discuss the interaction styles and models of this new paradigm. We then discuss tool support for activity-based computing. In particular, we give an overview of our ongoing work on analysis/design tools supporting early stage prototyping of activity-based ubiquitous computing.

Author Keywords

Ubiquitous computing, activity-based computing, activity theory, context-aware computing, interaction styles / models, implicit interactions, rapid prototyping tools, activity/situation/action modeling.

ACM Classification Keywords

H.5.2 [User Interfaces]: Interaction Styles, Prototyping, Theory and Methods; D.2.2 [Design Tools and Techniques]: User interfaces

INTRODUCTION

The rapid development of technologies, as predicted by Moore's law, has provided us with tremendous computing power. However, as we move away from the well-understood desktop environment, we face a complex and dynamic real world [15] and, currently, we are extremely limited in leveraging this enormous computing power to facilitate our day-to-day lives in this complex environment.

Although steady progress is being made in developing individual ubicomp technologies (such as new sensors, mobile computing devices or ambient displays) as well as in using these technologies to enhance human experiences (such as showing nearby restaurants based on the user's current location), we argue it is important to explore a new

interaction paradigm for the era of ubiquitous computing [21]. We believe a new interaction paradigm is important for providing guidelines for shaping and accelerating the development of ubicomp.

Here we describe our ongoing work on exploring activity-based computing as a new interaction paradigm for ubiquitous computing. This work was motivated by activity theory, a conceptual tool for describing human activities that originated in Russian psychology [17]. Activity theory provides a theoretical background on the structure of human activities and what to look at while analyzing human activities. This work was also based on our experience in striving to find a uniform framework for context-aware computing [11-13].

ACTIVITY-BASED UBIQUITOUS COMPUTING

We intend to explore activity-based computing as a new interaction paradigm for ubiquitous computing. We first discuss how human activities are structured in light of sociological and psychological findings [17, 19].

In the physical world, a human activity (such as keeping fit) often lasts for a relatively long period involving many discontinuous stages (such as running every morning for keeping fit) and many different situations (such as jogging in a park or exercising in a gym). An *activity* evolves every time it is carried out in a particular *situation*. A situation involves a set of *tasks* or *actions* performed under certain *conditions*, e.g., locations, times or temperatures. For example, for keeping fit (activity), a person gets healthier or stronger every time the person exercises in a Gym (situation) that may include doing pushups (action) or running on a treadmill (action).

In the entire process, tools (e.g., a microwave, a chair or a lamp) are a key element and important resource at every level of performing an activity. To facilitate our everyday lives, ubicomp technologies, as computing tools, should be designed based on activities. This includes organizing various ubicomp technologies based on the structure of activities and enabling the context-awareness of ubicomp technologies, i.e., knowing what people are doing.

To achieve these goals, we are designing an activity framework for 1) integrating ubicomp technologies into natural human activities, and 2) representing interaction contexts. Our proposed framework will involve key concepts such as activities, situations, actions, roles, tools and status properties. Roles are an abstraction of involved people based on what they do in an activity. We here briefly discuss tools and status properties.

For clarification and simplification, we use the term *toolets* for activity-based ubicomp tools (ubiquitous computing services as well as the hardware they are based on, such as sensors, mobile phones or ambient displays.) Computationally speaking, they are often at a smaller granularity than traditional applications are. Similar to determining graphical contexts in a scenegraph, the behavior of toolets can be contextualized based on how they are attached to an activity framework. For example, toolets attached to an activity node are designed for providing sustainable support for the lifetime of an activity (such as keeping track of how many steps a person has walked), while ones attached to a situation node address a particular situation of an activity (such as reminding me to take stairs rather than taking an elevator to improve fitness), which are only invoked in that situation.

Status properties characterize the aspects of interest or progress of an evolving activity. In the example of Elder Care, the activity is to help an elder age independently which can be characterized as a set of attributes such as an elder's health or medication status (e.g., blood pressure) and social involvement (e.g., the number of conversations). Status properties should be tracked throughout the lifetime of an activity and the history of status property changes is also important contexts for toolets to adapt their behaviors.

Toolets can be designed for facilitating an elder's daily routines (actions), such as prompting the elder to take medicines after meal or showing procedural instructions while making French toast (e.g., using an ambient display). Toolets can also be designed to capture the performance of these routines as well as the elder's status (e.g., using a sensor-embedded watch to record blood pressure or temperatures). In addition, toolets can be designed to communicate status properties to other roles in a care net, e.g., automatically showing the elder's status to family or more complex information to doctors.

Interaction Styles

Activity-based ubiquitous computing affords an *implicit* interaction style. Interactions in activity-based computing are mainly based on input such as what people are doing as well as in what conditions (e.g., locations and times) an activity is conducted. These inputs are acquired *implicitly* (e.g., via sensors). This includes not only what a person is doing at the moment but also what people have done in the past and if possible, what people plan to do in the future.

By leveraging implicit input, computer systems can decrease the requirement for user attention and enable users

to focus on their tasks at hand. For example, computer systems can automatically keep track of an elder's daily activities and communicate her wellbeing to caregivers. Traditionally, a caregiver has to manually record an elder's activities using an ADL form.

At the same time, activity-based computing does not exclude explicit or traditional interactions. Instead, explicit and implicit interactions can work together as we previously explored for location-enhanced applications [12].

However, it is still an important research topic to further discover and design appropriate interaction styles for activity-based computing. What is the *interaction vocabulary* for users to explicitly communicate with a system in such a paradigm, e.g., allowing human users to explicitly correct inference errors? One important direction to explore is to design interactions based on existing skills, e.g., employing common hand gestures or vocal variations.

Interaction Models

One can use the well-known model of Model-View-Controller (MVC) to analyze the structure of activity-based computing. In MVC, views and controllers are provided for modifying application semantics, i.e., the model. A model may have multiple pairs of views and controllers that a user directly interacts with. It is interesting to notice that, in activity-based computing, an activity is the counterpart of a model in MVC and it can be realized and modified via situations that correspond views and controllers in MVC. In addition, actions decide how an activity should be performed in a particular situation, which is similar to how interaction is specified in a controller.

However, there is a significant distinction from the traditional use of MVC. In activity-based computing, how a user may act is mainly based on what people naturally do in the real world, which is more dynamic than what is predefined in a controller of a traditional UI. For example, interactions are performed via physical actions (with or without an object) under physical or social constraints in the real world rather than by manipulating electronic artifacts such as windows, menus and buttons on a computer screen. This implies the importance of analyzing and modeling activities, situations and actions so as to understand current practice and establish models that allow computer systems to automatically infer them for implicit interactions.

At a lower level, interaction sequences are traditionally modeled by state transition diagrams. For example, Jacob [8] designed a framework for using a set of coroutines of state transition diagrams to model multi-thread direct manipulation interfaces. Hudson [7] extended state transition diagrams with probabilistic reasoning for dealing with ambiguous interactions. Jacob et al. [9] enhanced discrete state machines for continuous interaction. Interestingly, the AI community has been using HMMs, also a state machine, to infer physical actions [18]. We are

exploring how these state-machine schemas can be combined or extended for modeling interactions based on real world actions and how they can fit into the overall structure of activity-based computing.

TOOL SUPPORT

In this section, we first survey what the field has achieved in tool support for activity-based computing. We then describe our ongoing work on analysis/design tools supporting early stage prototyping of activity-based ubiquitous computing.

To infer activities, Canny previously designed algorithms for factorizing activities from computer (or desktop) event logs [3]. Philipose et al. developed algorithms for inferring physical actions [18] based on user interactions with physical objects. Fogarty et al.'s work on interruptibility modeling can potentially support general situation modeling [5]. Bardram et al. [1, 2] has developed an activity-based computing infrastructure for managing mobile computing services, in which "activities" are a collection of computational services¹. However, none of the work provides tool support for analyzing and designing activity-based computing.

Prior ubicomp prototyping tools (e.g., [12, 14]) addressed ubicomp design at a task or action level, such as helping me find a nearby library. Without enabling a higher level unit, namely activities, for design and analysis, it is hard to identify the crucial resources (such as roles and tools) that are constantly required across situations for carrying out an activity. It is insufficient to model and infer a richer runtime context, e.g., finding a nearby library for meeting a friend or learning astronomy. In addition, it is hard to organically integrate various ubicomp technologies to provide consistent support for the entire range of an activity.

Previously, activity-based design approaches [6, 17] have been proposed at a conceptual level, which has shown promise in areas such as CSCW, education and software engineering. However, these approaches are conceptual tools with abstractly defined principles and heuristics. In particular, activity theory, as the underlying component of these approaches, though conceptually powerful, is complex to understand and use in practice by general researchers and interaction designers. In addition, many technical challenges have to be overcome before conceptual analysis results can be realized as a testable ubicomp prototype.

Tools for Prototyping Activity-Based Ubicomp

We are currently exploring tool support for designing and prototyping activity-based ubiquitous computing, by which we intend to address several key challenges:

¹ In contrast, we intend to establish an interaction paradigm that integrates real world human activities and ubicomp technologies. In addition, we separate the notions of "activities" and "tasks" (or "actions") in light of activity theory. They are interchangeable in Bardram's framework.

- Human activities are complex and dynamic. It is hard to *analyze activities* so as to identify the breakdowns of current practice that could be enhanced by ubicomp technologies (e.g., informing me of how grandma is doing in an unobtrusive way while I am working).
- It is often unclear what information is relevant for *modeling activities* and how much automatic sensing can do for us (e.g., is it relevant to telling whether this is a meeting by knowing whether Bob is in the room).
- Activity-based computing often employs *sophisticated technologies* (such as using accelerometers to infer whether a person is walking up stairs or taking an elevator) that require sophisticated technical expertise to build.
- It is hard to test and deploy activity-based ubicomp applications. To get valuable feedback, they should be *evaluated in-situ* and over a long period (how well a CareNet display can help family members to monitor the grandma's status without requiring them paying much attention in their daily lives [4]).

We have concretized activity-based design in the domain of ubicomp and devised an intuitive *Theme/Scene/Action* design paradigm to streamline the activity-based design process. The design paradigm allows researchers and designers to design activity-based computing using familiar concepts like those used in theaters [10]. A *theme* represents an activity that is conducted by human users (e.g., doing more exercise). A theme may be conducted in multiple *scenes* (e.g., running or riding a bike in a park). A scene represents a situation and a set of *actions* can be carried out in each scene. This framework allows designers to simultaneously analyze a target problem top-down and bottom up. A designer can first list themes involved in a problem and then enumerate concrete scenes where those themes are carried out. A designer can also start from concrete scenes of interest and then extract themes from scenes. The framework allows designers to establish a consistent structure over individual situations and independent ubicomp technologies. It will inform what individual toolsets should be designed and how they should work together as a whole.

Based on this paradigm, we have built a tool called ActivityStudio that allows designers to design and analyze based on field data that reflects existing activities, model actions of interest using a video editing metaphor, prototype interactive behaviors of toolsets by storyboarding and test a suite of toolsets using a Wizard of Oz approach or real sensors.

To facilitate the design based on large scale field data, we plan to provide support for automatically extracting themes, scenes, and actions from collected log data such as data from in-situ studies. We also intend to allow analysis results as well as interaction designs to be shared and reused across problems and designers.

Currently, a tooleet can be rapidly prototyped using a built-in storyboard editor. However, we intend to enable the ActivityStudio to export analysis or modeling results so that a tooleet can be prototyped or refined in other task or action-level prototyping tools (e.g., Topiary [12]). This will allow the ActivityStudio to leverage existing work that may have better support for particular domains or tasks.

In addition to making designs, our design environment could potentially be used for evaluating a design. From activity theory's point of view, the design and evaluation processes are all about analyzing human activities. The difference is before (for design) or after (for evaluation) ubicomp technologies have been designed and employed in human activities. We intend to build runtime support for collecting users' activity and interaction logs. This includes data collected by implicitly logging what users are doing or explicitly surveying users for their in-situ feedback. The collected data will then be analyzed in the environment for another iteration of a design. This supports a full cycle of an iterative design process.

Target Application Domains

To evaluate our framework as well as prototyping tool, we are currently working on two application problems: elder care and personal fitness. Some aspects of these problems have been studied in previous work [4, 16, 18, 20]. We intend to exercise our framework as well as prototyping tool on these problems. We hope to find out whether our approach can save effort in building such applications and also add new values.

We intend to build ubicomp support for the fitness domain, more concretely, to help people get more exercise and reach their fitness goals. By automatically keeping track of and visualizing the exercise that a user has done throughout a day, users can gain better self-awareness of their fitness progress. We also intend to encourage a user to do more exercise by providing hints appropriate to particular contexts (such as suggesting parking in a parking lot distant to their workplace).

We are also building ubicomp technologies to support elder care, e.g., helping elders age gracefully and independently. This activity is carried out in many situations such as in hospitals for medical attention, grocery stores for buying food and homes for dining. There are multiple roles (such as elders, nurses, families, and friends) involved in a "care network". Ubicomp technologies, for example, can be built for detecting an elder's daily activities and communicating her status to caregivers in an unobtrusive way.

CONCLUSIONS

In this position paper, we explore activity-based computing as an interaction paradigm for ubiquitous/context-aware computing. We cover several issues of activity-based ubiquitous computing. We described our early ideas on an activity framework for structuring ubicomp technologies and representing interaction contexts. We discussed interaction styles, interaction models and tool support for

activity-based ubicomp. We also briefly discussed our focus on providing tool support for early stage prototyping of activity-based ubiquitous computing.

Activity-based computing as proposed in this paper does not exclude previous work. Instead, we intend to provide a higher level framework for structuring ubiquitous and context-aware computing, where individual ubicomp technologies can come into play.

ACKNOWLEDGMENTS

We would like to thank John Canny, Tye Rattenbury and Bonnie Nardi for discussions on activity theory. We would also like to thank Beverly Harrison, Mike Chen and Sunny Consolvo for their feedback. This work was supported by Intel Research Seattle and the NSF under Grant No. IIS-0205644.

REFERENCES

1. Bardram, J.E., Activity-Based Computing: Support for Mobility and Collaboration in Ubiquitous Computing. *Personal and Ubiquitous Computing*, 2005. 9(5): pp. 312-322.
2. Bardram, J.E. and Christensen, H.B. Open Issues in Activity-Based and Task-Level Computing. In *Pervasive'04 Workshop on Computer Support for Human Tasks and Activities*. 2004. Vienna, Austria.
3. Canny, J. GAP: A Factor Model for Discrete Data. In *SIGIR'04*. 2004. Sheffield, UK. pp. 122-129.
4. Consolvo, S., et al., Technology for Care Networks of Elders. *IEEE Pervasive Computing Mobile and Ubiquitous Systems: Successful Aging*, 2004. 3(2): pp. 22-29.
5. Fogarty, J., et al., Predicting Human Interruptibility with Sensors. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2005. 12(1): pp. 119-146.
6. Gay, G. and Hembrooke, H., *Activity-Centered Design: An Ecological Approach to Designing Smart Tools and Usable Systems*. Acting with Technology. 2004: The MIT Press.
7. Hudson, S. and Newell, G. Probabilistic State Machines: Dialog Management for Inputs with Uncertainty. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. 1992. pp. 199-208.
8. Jacob, R.J.K., A Specification Language for Direct Manipulation User Interfaces. *ACM Transactions on Graphics (TOG)*, 1986. 5(4): pp. 283-317.
9. Jacob, R.J.K., Deligiannidis, L., and Morrison, S., A Software Model and Specification Language for Non-WIMP User Interfaces. *TOCHI*, 1999. 6(1): pp. 1-46.
10. Laurel, B., *Computers as Theatre*. 1993: Addison-Wesley Professional.
11. Li, Y., Hong, J.I., and Landay, J.A. ContextMap: Modeling Scenes of the Real World for Context-Aware Computing. In *Adjunct Proceedings of UbiComp 2003*: Poster. 2003. Seattle, WA. pp. 187-188.
12. Li, Y., Hong, J.I., and Landay, J.A. Topiary: A Tool for Prototyping Location-Enhanced Applications. In *CHI Letters*: 6(2), *UIST'04*. 2004. Santa Fe, NM. pp. 217-226.
13. Li, Y. and Landay, J.A. Rapid Prototyping Tools for Context-Aware Applications. In *CHI'05 workshop - The Future of User Interface Design Tools*. 2005. Portland, Oregon.

14. MacIntyre, B., et al. DART: A Toolkit for Rapid Design Exploration of Augmented Reality Experiences. In CHI Letters: 6(2), UIST'04. 2004. Santa Fe, NM. pp. 197-206.
15. Moran, T.P. and Dourish, P., Introduction to This Special Issue on Context-Aware Computing. *Human-Computer Interaction*, 2001. **16**: pp. 87-95.
16. Mynatt, E.D., et al. Digital Family Portraits: Supporting Peace of Mind for Extended Family Members. In CHI Letters: 3(1), CHI'01. 2001. pp. 333-340.
17. Nardi, B.A., ed. *Context and Consciousness: Activity Theory and Human-Computer Interaction*. 1995, MIT Press.
18. Philipose, M., et al., Inferring Activities from Interactions with Objects. *IEEE Pervasive Computing*, 2004. **3**(4): pp. 50-57.
19. Suchman, L.A., *Plans and Situated Actions: The Problem of Human-Machine Communication*. 1987: Cambridge University Press.
20. UbiFIT. Improving fitness through mobile devices. <http://dub.washington.edu/projects/ubifit/>
21. Weiser, M., The Computer for the 21st Century. *Scientific American*, 1991. **265**(3): pp. 94-104.

F. Other Issues and Innovations

F.1 World of WarCraft, the role-playing game and the deconstruction of the classical model. The construction of environments and multiform narrations.

Mar Marcos Molano Universidad Complutense de Madrid

Maria Luisa Garcia Guardia Universidad Complutense de Madrid

Michael Santorum Gonzalez Universidad Complutense de Madrid

F.2 An Argument for Using Techniques Found in Audio Games as the Next Generation of User Interfaces for Blind People

Chris Hofstader

F.3 Considering the Next Generation(s) of Human Computer Interface: A conceptual framework or theory

Fred Welter ITT

World of WarCraft, the role-playing game and the deconstruction of the classical model. The construction of environments and multiform narrations.

Mar Marcos Molano: marmarcosmolano@gmail.com

María Luisa García Guardia: mgarciaguardia@hotmail.com

Michael Santorum González: michael.santorum@gmail.com

“World of Warcraft, like many massively-multiplayer online role-playing games, is constantly evolving. In an effort to keep the game challenging and balanced (...) new content and game systems will also be added.”

World of Warcraft Manual Book, page 13

The communication pretends to share the results of a budding investigation that we are carrying out around the video game *World of War Craft*.

The hypothesis from which the investigation starts goes through the demonstration of how the role-playing games dismantles the classical model, established with the cinema, starting from the construction of an *alternative* model based on:

- 1.- The relation of the observer/player with the symbolic universe proposed.
- 2.- The Setting.
- 3.- The narration.

- 1.- Relation of the observer/player with the symbolic universe.

Let's remember the idea of the film spectator.

The projective systems of artistic creation have always tried a “place” for its spectator. In the cinema furthermore, the exhibit's own structure proposes the spectator as an indispensable element of the cinematographic act, developing a hegemonic model, in which it occupies a central and essential position. As Münsterberg stated, the cinema works ideally for the spectator, it is formed to aim the human spirit not existing, psychologically, nor in the film nor in the screen, but only in the spirit that gives it its reality. It is articulated like this, a whole device of creation and cinematographic reproduction designed for this cinematographic spectator without all its creative process remaining ruined. Retaking the Sartrian thesis of the image as “absent-presence”, in which the perception of the world gains force from an infantile mentality which considers the reality of the dreams with so much presence with the reality of the vigils, it introduces the figure of the cinematographic spectator in an identical position, as this one is the one who gives a “soul” to the things he perceives over the screen. In this way the perception in the cinema comes close to the “magic perception” based in a system determined “by the believe in the double, in the metamorphosis and in the ubiquity, in the universal flow, in the reciprocal analogy of the microcosms and macrocosms, in the anthropocosmomorphism”.

This way the author puts together the processes of the dreams with those of the cinema, shaping a psychoanalytical theory of the spectator, of unquestionable validity in the surroundings of the institutional representation, which is based on the processes of projection and identification, in which the spectator, instead of projecting himself over the world he absorbs it towards his own world. He makes it his. The cinema with its techniques favours said phagocyte, provoking and intensifying his projection.

The double projection goes through an unconscious primary phase, a tactic agreement between the spectator and the representation, thanks to which the two dimension images psychologically become real to the observer, offering to his look a “sham” of his perception of the real universe. And if this projection tries to get close the spectator to a “parallel reality”, one has to take into account that it is establish through the base device (the camera) to produce certain effects. Plus the unconsciousness of the projection and the structural elements which emanate from the cinematographic's own happenings, permits the spectator to submerge into the magic of what happens before his eyes feeling an integrated and an essential part of said “reality”.

The second projective phase goes through the identification of the text's superficial elements. At this point, the secondary projection personifies itself in the character's figure and of its psychological construction. In this sense the spectator's selective memory acts in the reconstruction of the character, providing it, to a certain point, a distorted psychology: the character appears upon the spectator's memory with a simpler psychology than what it is in reality. This one builds up in a more contradictory and complex manner through out the story. Also the characters are not only "beings of the fiction" with paths that are born only for the story, but that are built up and limed with a biography, a past that reduces them to behaviours derived from their own vital path. Given this association by remembrance, the spectator tends to believe that he has identified himself by the liking with the personage, by its character, by his behaviour..., a little bit like in the real world. This way the secondary identification in this level is basically the identification with the character as a fellow man figure in the fiction.

However, there are authors who believe that the *diegesis* identification in the cinema uses more complex mechanisms, for which that identification comes from a "structure" effect, a question more of place than of psychology. As a fact, in order for the spectator to find his place, the narrative space of a sequence or a scene, is sufficient with inscribing to this scene a network of relations, a situation. At this point, it has little importance that the spectator does not know his characters. This way, as Barthes states, identification does not have psychological preferences. It is a structural operation: I am that one which occupies the same place as I.

In the role-playing games, Barthes' maximum is not only an option, it's essential. If it is true that we could assume for the video game the first of the projections, in such, that the tactic agreement between the user and the representation, it is also true that in said universe the player believes himself the centre of the representation because he is, because he has created a character to move around that parallel universe. I am that one that occupies the same place that I, becomes virtually effective when the user is capable of moving by himself in said universe. Once the character is created, re-presentation in the representation of oneself, and once assume his characteristics by the user in rule forms, also existent in the classical story to know how to decode it, acquires the power to articulate it's own destiny. Walk freely and virtually through that universe deciding his actions and interactions. The antropocentric projection becomes more effective in the sense that the player is conscious of being one in that universe, with a "physical" space in virtue of his designed character and not as in the cinema, without effective presence but psychological.

2.- The Setting

The cinematographic construction sets forth its setting criteria in virtue of the creation of a two-dimensional space with an impression of a three-dimensional, under the inherited patterns of the renaissance projective systems.

Reality sensation, intromission in a recreated world, illusionism, which though being a word with low acceptance by critics, it constituted one of the most revolutionary values that the artificial perspective placed in reach of the artist in form of accepted hegemonic creation models, which came to mark a point of inflexion in art, where the recreation of the reality, the organization of the composition and the harmonization of this new reality with the pictorial flat surface, are the principal formal potentials of the new perspective.

If during centuries it was thought that the norms of the perspective coincided with the physiological reality of vision and with the same structure of nature, with photography the system's level of convention is observed. The system that has reined the European painting since the Renaissance has emerged from the perspective but it has not supposed the rigorous acceptance of all its premises mathematics. This pictorial system has such a parallelism with the photographic, that it can therefore not be spoken of the abandonment in our century of the old system's representation.

The cinema as a projective technique based in the reduction of reality through the altered objective. Through another path, the traditional space notion but parting from a reaffirmation of the principles of the projective geometric. The creation criteria pass through the relations that are established between what is seen (field) and what is not seen (out of the field). The global space created by the cinema makes for both to have a similar importance, but it is the enounce stay which determines the presence / absence, what we can see and what we cannot. Also, the mobile character of the cinematographic image complicates the space creation processes, which at the same time includes the temporary creations, up to now if well implicit not explicit in the text. Therefore the creation of a field and its articulation with an out of field, now it's added with the possibility of movement in the time-space through the mobile camera, and the possibility of incapable movement by the setup is provoked by the travelling that, if initially it was basically used to follow the actors, soon it will find a meaning of character, not only functionally but expressive: the camera takes body, it moves between the characters and with them, it converts itself in an autonomous presence which gives back to the spectator its own look from the same interior of the scene's space. Thanks to the mobile camera, the spectator submerges in the space represented, because this one becomes a place in which it moves in a psychological manner: the things are making themselves "here and now" because it's capable of following them in it's own mobility, according a continuous development, present, actual. The camera acquires then a signification that initially it did not have, the one of supporting the construction of the dramatic space.

The role-playing game analysed comes back, in this sense, to modify the lay out of the temporary space creation. Temporary because it permits the spectator the possibility to play in real time building up the story as he advances in the game: the ellipsis are the player's strategy, not the creator's, the universe offers it to him, temporarily, parallel to his. In space because the game, in third person, permits him not only to recognise the conventionalised 2D criteria, in the field's deep form, but with a soft mouse movement, the sequence plan in which his path develops through the parallel universe, permits him access to the parts of the fields not shown by the screen. In this case, it is the player and not the creator that can determine the frame, and with it access freely to all the infinite space of the virtual.

3.- The Narration

If the classical model bases the narrative structure in the casualty, in the presentation of an effect which succeeds a cause, weaving a network of structures between characters and situations clearly determined by the author, constructed upon the virtue of inherited norms from the traditional narrative since Aristotle, in the video game, the narration is abandoned in a risky chance form. Even though we must make a series of actions to strengthen our I in the story of the video game, we are free in doing them. If we opt to do them in the place and time predetermined, our character goes assuming a series of skills and abilities that make his journey easier. In the contrary case, we will take more time to reach them, but in any case, the story will not break. The user becomes a lector author capable of influencing his own course without eroding the universe in which he is immersed.

An Argument for Using Techniques Found in Audio Games as the Next Generation of User Interfaces for Blind People

Chris Hofstader

December 2005

Introduction: Reality-based Interfaces and the Blind Computer User

The majority of blind people who use computers (including the author of this position paper) access the information on their device using a program typically called a screen reader. Some users, mostly on the GNU/Linux platform, use software that augments the text-based console interface or a text editor to access their information¹. For the purposes of this paper, though, I will treat them identically as the interface to both classes of access technology are very similar and neither is based in anything resembling what a human might regard as reality.

The interface to screen readers (whether *JAWS*, the most popular in the world, or *VoiceOver*, the newest entry to the market) which will be described in greater detail later in this paper, is a one-dimensional stream of words spoken by a text to speech engine². So, unless one's reality is in a single dimension, the interface for blind people has no basis in human experience.

There are, however, pioneers of three-dimensional user interfaces making and selling audio-only programs³ to blind users. These people come from the world of audio gaming and have advanced to a point in which blind people can, in real time, navigate through complex environments similar to those enjoyed in video games by millions of sighted people⁴.

The position described in this paper is that the techniques used in the more advanced audio games should define many aspects of the next generation of interfaces for blind people in Computing, orientation and mobility technologies and smart spaces/environments.

Inefficiency: The Achilles' Heel of Screen Readers

Screen reader vendors have made improvements to the usability of their products over the last decade⁵. *JAWS*, the most popular screen reader, has added hundreds of features that a blind person can employ to use the more advanced functions of the *Microsoft Office* and

¹ The Blinux Project, <http://www.leb.net/blinux/>.

² Many screen readers also offer the user the ability to have text output to a refreshable Braille display. For the purposes of this paper, though, all tactile I/O devices will be ignored.

³ Wikipedia, the Free Encyclopedia, "Audio Game History", *Audio Games* (November 8, 2005) http://en.wikipedia.org/wiki/Audio_Games#Audio_game_history.

⁴ Cohn, David, "The Blind Fragging the Blind", *Wired News* (14 March 2005) <http://www.wired.com/news/games/0,2101,66879,00.html>.

⁵ Virtually every review of a new release of a screen reader that one can find in the assistive technology press makes mention of new features and enhancements to the product they are writing about.

Internet Explorer suite of products. *JAWS* has also added a number of keystrokes that make document navigation simpler in some of its more recent releases.

Apple Computer's *VoiceOver*, which has been poorly received by critics⁶, does get applause from some users for being simpler to use than the more complex Windows products. *Voiceover's* simplicity, unfortunately, comes at the expense of the many features that blind users have grown accustomed to having available to them. Its ease of use does not, therefore, lead to improved efficiency as many of the features in a product like *JAWS* designed to increase efficiency are simply not present.

Fundamentally, *JAWS*, *Window-Eyes* and *HAL*, the three most popular screen reading tools, no matter what features they add to increase a user's efficiency, can only deliver one little piece--either a syllable or pause--of information at a time. Thus, the interface presented to a blind user is linear and the information received at any moment is merely a single point on a long string of auditory tokens.

To access a specific piece of information, a blind user must wait for the point on the line that contains the bit of information he is interested in hearing to be spoken by the screen reader. Certainly, blind users can employ navigation commands to jump from one point to another on this line of information, but doing so without jumping past the point one is seeking can be difficult for even the most experienced users.

For lack of another term, I call the amount of time it takes to deliver one of the aforementioned auditory tokens a gTick (grammatical tick). One gTick is equal to the average amount of time at a given speech rate (each user will have her own level of comfort in words per minute) it takes a speech synthesizer to deliver a single syllable plus the amount of time a speech synthesizer pauses for white space and different types of punctuation (following the rules of the human language the synthesizer is speaking). Thus, the most productive a blind person can be using one of these screen reading tools is measured in the total number of gTicks that pass in the time allotted for their task as, using a purely speech interface, they are not presented with more than a single gTick of information at a time.

Audio Games and the 3D Interface

Many blind people enjoy the game *Shades of Doom* (SOD) from GMA Games. SOD presents a complex environment where the user (just like the video game with a similar name) must navigate in real time a sequence of mazes, fight off mutants, find useful items and, ultimately, find his way to the center of the secret laboratory and kill the mad scientist before his experiment escapes and turns all of humanity into mutants⁷.

⁶Leventhal, Jay, "Not What the Doctor Ordered: A Review of Apple's VoiceOver Screen Reader", *AccessWorld*, September 2005.

(<http://www.afb.org/afbpress/pub.asp?DocID=aw060505>)

⁷David Greenwood, "The Shades of Doom Status Report", *Audyssey Magazine*, Issue 21 (January-February 2000)

The people who play this game must navigate by the sound of the wind (always out of the east), the echo of their own footsteps against walls, the sound of their own breath and other ambient sounds which, after a while, the player learns to locate in relation to their current position. The author of the game claims that the player is presented with up to 32 simultaneous sounds⁸ which they must decipher to perform the various tasks, navigate the maze and fight off the mutants.

Shades of Doom presents multiple simultaneous audio data streams. *JAWS* and its counterparts present a single stream of information. Hundreds, maybe thousands, of game players enjoy SOD and the other entries from GMA. Approximately 150,000 blind people worldwide use *JAWS* to access information on Windows-based computers. How is it that the gamers can process so much information at once, but the screen reader users are bound by the tiny bit of information expressed in a single syllable or pause?

The amount of semantic information delivered by a single sound in SOD in a single gTick is actually greater than that delivered by a speech interface. A player who hears a particular "bloop" which takes about as long as a syllable to present, knows (after a little practice with the game) that there is an object of use to them nearby. For *JAWS* to speak, "there is an object of use nearby" requires 15 gTicks (9 syllables and six spaces).

There is a fair amount of research by cognitive psychologists about the number of audio data points a person can comprehend simultaneously⁹. There has also been work studying how humans perceive the location of an object presented by a three-dimensional interface¹⁰. There is no research, to my knowledge, that suggests how many data streams a blind person can use effectively to perform a given task on their computer, to find an object in their home, or to more effectively move independently through real world spaces.

Applying 3D Interfaces for Use By Blind People

The game developers have certain advantages over people who design interfaces for screen readers. Specifically, they know the narrative that they are presenting and so can predict, within a small set of actions, what the user will do next. A screen reader cannot predict what word will appear next in an email, what value will be present in a random cell in a spreadsheet or which link on a web page the user will select next.

<http://www.audiogames.net/page.php?pagefile= Audyssey Magazine issue 21 - January - February 2000 .>

⁸ Referenced in the games tutorial information and on the GMA Games web site, <http://www.gmagames.com>.

⁹ Bregman, Albert S., *Auditory Scene Analysis: The Perceptual Organization of sound*. Cambridge, Massachusetts: The MIT Press, 1990.

¹⁰ Begault, Durand R., *3-D Sound for Virtual Reality and Multimedia*. NASA technical report, Moffett Field, California: 2000.

A sighted computer user can learn a number of bits of information about the state of their computer by merely diverting their glance from one portion of the screen to another. Using three-dimensional audio an interface can be built that, by placing audio objects in different locations in a simulated spatial environment, will deliver far more simultaneous information to a blind person. These techniques, if applied correctly, can greatly improve the efficiency of computer use for those who rely on an audio interface.

3D Interfaces for Non-Computer Applications

Although I have not tested these concepts yet, I can envision a blind person using 3D audio to help them locate objects in a smart space. If, for instance, one keeps their keys attached to a fob that contains an RF tag and the room has an array of RF readers the blind person can ask their user agent, "Where are my keys?" And, through directing the audio in the space be presented with a tone that sounds as though it is emulating from the spot where the RF system has located the object of interest.

Orientation and mobility technologies can likely be greatly improved if expanded to include a third dimension. There are a wide variety of technologies ranging from sonar to infrared to laser range-finding equipment that can report a tremendous amount of information about the solids in an outdoor environment. Delivering this through a TTS in a single dimension would take an overwhelming amount of time. In sound, though, like the sounds of walls and other objects in *Shades of Doom*, a portable 3D tool can shape the surroundings for the blind person as they walk. This type of interface will provide a terrific amount of information about obstacles that one might miss with a cane.

Conclusion

I could provide many examples of how 3D audio interfaces can be used to increase the productivity of blind computer users and to improve safety and quality of life at home. For the most part the audio portion of the software engineering is done science, as Microsoft and the other game platform developers have figured it out for the very lucrative video game market. What the HCI people, including me, now need to do is to discover how to do the specific implementations for blind people so they can be increasingly productive, safer and have greater conveniences in their homes in the future.

Considering the Next Generation(s) of Human Computer Interface

A conceptual framework or theory

Fred Welter
ITT

As an Electrical Engineering graduate of the sixties we solved differential equations with Analog Computers. Our Human Computer Interface was the pegboard-wiring matrix to the system's Operational Amplifiers.

There were IBM cards then, special enclosed and carefully air conditioned spaces for the computer, and ambitious young lady operators who enjoyed their sudden employment opportunity in a new industry that welcomed them. The Human Computer Interface was the Service Desk.

As a Medical Device Reliability Engineer of the nineties we designed a Human Computer Interface not perceived by the patient unless the device was programmed to notify of certain circumstances needing attention.

At the other end of the spectrum is the Human Computer Interface evolving to greater levels of stimulation and interaction. Those Medical Device patients have children or grandchildren who play (or work?) with computers for stimulation and interaction (learning?). Children today have constant computer interface, games, homework, and the toaster in the morning. A Gameboy or an iPod is used for the ride to school. Surfing the Internet at night. So extremely computer stimulated today are our children that few would be able to identify each and every Human Computer Interface involved in their daily activities. For these future adults the best business application Human Computer Interface could be the dual-hand video game controllers in such widespread use today. These future adults will grow up accustomed to ubiquitous, pervasive, handheld, interactive, tangible, lightweight, and tacit Human Computer Interface in a multitude of personal computer devices. But the next generation of Human Computer Interface is not going in a single direction. One direction for Human Computer Interface design involves pleasure, the gaming or entertainment possibilities with computers. The intensity or vividness of the "virtual reality" emersion aspect of the Human

Computer Interface is paramount. Virtual and augmented reality developments will be the primary focus in gaming and entertainment.

Another direction involves an informational interface to assist human decision-making. Here an optimal organization and display of information would be paramount. Perceptual interfaces, affective computing, context-aware interfaces, ambient interfaces, embodied interfaces, sensing interfaces, and speech and multi-modal interfaces have significant applications in this area of informational interface to assist human decision making.

Another differentiation of Human Computer Interfaces are those interfaces requiring our human conscious interaction (mental interaction or the informational interface to assist human decision making previously discussed) and the other type of Human Computer Interface not requiring attention, such as all the computer interfaces involved in making a simple gas pump transaction. This category embraces Human Computer Interfaces involving passive or non-command interactions.

The "Unconscious Human Computer Interfaces", (the intelligent refrigerator that orders groceries) will evolve in a Charles Darwin fashion. The design innovations that are most useful and least troublesome to the human user will survive.

The "Conscious Human Computer Interfaces" will develop in a revolutionary fashion with breakthroughs in the technology of making connections with human sensory experience at the neural level. The ultimate "virtual reality" experience is one that rides into our brain on the same nerves that bring it sense experiences. The ultimate Human Computer Interface to our human nervous system.

Connection to visual and auditory nerves will be the next generation of Conscious Human Computer Interface. Several interests drive development of this technology. The Medical Device community has Cochlear Implants and some success with surgical connection to the visual nerve. The Computer Simulation community wants connection to visual and auditory nerves so the bulky sound and vision gear (and now "pressure garments" used to simulate touch) can be discarded and the "reality" of the stimulation experience enhanced. The Gaming community wants the most totally absorbing "virtual reality" experience attainable.

That is enough financial pressure to develop this nerve system human

computer interface into working systems. Direct interface with the human brain is a next generation Human Computer Interface.

The other non-mental interaction, service mode of Human Computer Interface is found in our current medical devices, our car computers, our homes, all are increasingly controlled by computers we interface by "programming buttons" or, if next-generation, by voice command. Automobile GPS does that now. There will be a slow and incremental improvement of this "button" interface for many years. The typewriter keyboard (written language) will remain intact even if altered for cell phones and hand-held computers) until language recognition technology sufficiently improves to push aside the cumbersome keyboard entry.

A Next Generation Human Computer Interface will recognize the owner's voice and later computers will recognize their owner's thoughts (as the technology for connection to human nerves develops), and, later still, computers will supplement human thinking. When this occurs the Human Computer Interface becomes one of partnership or companionship.

We recognize each other as intelligent companions because we share a "mind-space" of commonly understood concepts, meanings, and values. This allows us to communicate our thoughts by language. Animals would be considered intelligent if they communicated with us in this manner. We would judge an extraterrestrial being as "intelligent" if able to communicate in "mind-space word concepts" or mathematical concepts. Animals on our planet would be thought of as "intelligent" if they shared thoughts or feelings with us though language communication that can take place only if there is this "shared mind-space".

Communication within the "mind-space" is the ultimate Human Computer Interface. You could think of it as talking with a computer in the same manner one talks to one's self mentally. But, in the same way we might be confused where an idea or concept was originated in our thoughts there might be a difficulty separating the computer's participation in our thought developments and our own natural mental ability.

We as "creators" of computers have produced a potential intelligence that could accumulate enough agility within our human "mind-space" to be undifferentiable from "actual" humans. This is possible now with specific applications such as Chess. Can a Chess player detect if he is playing a human or computer?

The thing that makes us human is the process of taking sensory stimulation and turning those physical experiences into mind space artifacts that other human beings can recognize, value, and understand.

The final Human Computer Interface must be the computer's participation and contribution to our human "mind-space".