Workshop Proceedings, September 2003

Physical Interaction (PI03)
Workshop on Real World User Interfaces

The workshop is held at the MobileHCI conference 2003 in Udine (Italy)

Workshop Organizers:

Albrecht Schmidt, University of Munich, Germany, Albrecht.Schmidt@acm.org
Steve Benford, Nottingham University, UK, sdb@cs.nott.ac.uk
Andreas Butz, Universität des Saarlandes, Germany, butz@cs.uni-sb.de

Further Information is available at:
http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

Table of Contents

Tejp: Designing for Embodied Interaction with Personal Information Layers in Public Spaces

Margot Jacobs
Play Studio, Interactive Institute
Hugo Grauers gata 3, 412 96 Göteborg, Sweden
www.playresearch.com
+46-(0)734055867
margot.jacobs@tii.se

Lalya Gaye, Lars Erik Holmquist
Future Applications Lab, Viktoria Institute,
Box 620, 405 30 Göteborg, Sweden
www.viktoria.se/fal
+46-(0)31-7735562
{lalya, leh}@viktoria.se

ABSTRACT

The project Tejp explores various possibilities for overlaying personal traces on public spaces through physical interaction and parasitic methods. A series of low-tech prototypes drawing inspiration from existing urban practices, are tested in real settings with the general public. We stage the prototypes as props and people as performers to discover and uncover emerging behaviours, and to derive design implications for location-based information systems that make full use of the physicality of public spaces.

Keywords

Embodied physical interaction, personalisation of public space, physical environment, location-based information systems, interaction design, urban sub-cultural practices

1. INTRODUCTION

1.1 Motivation

Prompted by the excess of commercial media pervading the public arena, the research and design project *Tejp* aims at encouraging playful ways for indivduals to personalise territory and at providing a space and sounding board for existing social relationships between residents, passers-by and potential players in the public arena. The aim of the project is to provide technological tools and situations for layering personal traces on public physical spaces for others to discover. We promote technologically enabled personal expression that can happen any day, in any public space without the need for group organization or ownership of specific devices.

Builiding upon existing alternative channels of public expresion and communication for everyday people such as graffiti, stickers, posters and bulletin boards, we perform a series of experiments by deploying low-tech prototypes in the public realm. This allows us to explore various types of physical interaction for authoring, layering and accessing digital information on public space, and to uncover how this interaction influences meaning, content and user's behaviour.

This paper was presented at "Physical Interaction (PI03) – Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors. Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

Figure 1. Found example of existing parasitic means of personalising public space.

1.2 Related Work

Tejp is related to projects in the fields of ubiquitous computing and augmented reality [e.g. 5, 9 11 12], interaction design [e.g. 1, 6, 7] and in particular to location-based information systems [e.g. 4], which connect digital information to physical space. For example, the project Hear&There enables people to leave virtual audio imprints at particular places, with content that is created by the users themselves, located with GPS on a map through a PDA, and accessed through headphones. GeoNotes focuses on the "communicatory, social and navigational implications of the mass usage" [4] of text-based systems, addressing the issue of information overload and introducing the notion of social filtering.

In Tejp, we concentrate on the tangible and semantic aspects of interaction between the user and that information space, and take advantage of the physicality of the environment. The ultimate goal is to facilitate the personalisation of public spaces by everyday people, exploring parasitic communication as means of adding and revealing layers of content on physical space.

1.3 Approach

Each place has emotions, meanings, and content that are related or attributed to them. These are linked to the context in which they are created and are affected by the cultural and cognitive filters through which they are perceived. For example, the interpretation of traffic signs depends upon placement and knowledge of traffic laws. Reciprocally, their presence determines the use and meaning of crossroads at which they are placed. Inspired by parasitic media theory [10] and the work of the Situationists [2], we are interested in the creation of new such layers of content for places, without altering their intrinsic structure or function (Fig. 1). We aim at providing people with a means of superimposing personal layers of content on the physical environment (like communities do with stickers and posters, f. ex.) as well as revealing hidden layers of understanding (as if x-raying unperceivable facets of reality in public spaces), thereby supporting the insertion of new meanings into places and their personalisation.

In terms of content, we recognise the need for incorporating dimensions of subtlety, abstraction and even poetry or subversion into the palette of possibilities, in order for such systems to be meaningful for user. Technology, body, media and context of use play a major role in how the user can express themselves with location-based information systems, mostly due to the fact that the interaction they enable can limit, widen or at least influence content and engagement. Hence, there is a strong need to focus on and explore more physical, embodied means of layering information on physical space, instead of assuming the prevalence of screen-based interaction with PDAs or head-mounted displays.

Rather than focusing on the technology or prescribing interaction procedures, we design a space of possibilities for the interaction to take place within and for meaningful, embodied use to emerge. Through the exploration of this space, we aim to derive informed design implications for location-based information systems, enabling the communication of personal experience and expression.

2. METHOD

The methods we employ actively involve people as performers in an iterative prototyping process, while permitting and encouraging the public to expand their notions of what is permissible in the public realm.

2.1 Low-Tech Prototypes as Props

In order to explore new and provocative ideas about authoring, accessing and layering digital information on public physical space as well as the behaviours this incurs, we created a series of various low-tech prototypes to test on-site with users.

These prototypes are used as props, allowing us to gain insight into the user's motivations and experiences by observing how people interact with them in real settings. Thereby, we explore how the physical attributes of the props and the technological options we propose through them influence information content, users' behaviours and conveyed meaning. Performing these experiments in the public setting also reveals the relationship between these aspects of interaction and the actual context of use.

In each case, we focus on different design parameters, such as the medium that is used or the synchronicity of the communication. These parameters are not orthogonal in terms of their influence on the resulting interaction, preventing us from isolating them and testing each parameter one at a time. Therefore we incorporate redundancy in the design of the prototypes as a means of confirming relationships of cause and effect through reoccurring patterns.

2.2 Designing for Embodied Interaction

We are designing for natural, embodied interaction [3] to occur, both socially and cognitively, with potential for adoption into everyday life. Therefore, it is important for the prototypes to be open, tangible [8] and inspired by existing cultural practice.

2.2.1 Tangibility

Designing for this context of use implies embedding the physicality and use of the devices into the physical environment. Instead of being displayed on PDA screens or virtual acoustic spaces, the content they convey are mediated as parasites of physical media and existing urban structures. The prototypes build upon those structures (walls, urban furniture, existing web of mobile communications, electrical apparatus, etc), using them as support by incorporating them into their functionality.

2.2.2 Openness

The prototypes are voluntarily kept simple in order to focus upon the interaction rather than the technology they represent. Each of them only hints at the nature of their application without imposing a procedure of use. Thus, they are open to unexpected emerging uses and behaviours, leaving space for appropriation and adaptability, as well as honest engagement, both experientially and creatively.

2.2.3 Cultural Grounding

The physical attributes of the prototypes, appearance, possible modes of interaction, and placement are inspired by existing urban practices and street culture, in order to ground the experiments in real people's understanding of public places and the urban environment.

Although emerging sub-cultural issues already play critical roles in our society, they are usually given little design attention. We access current communication procedures taking place in the public space by performing observations of sites and behaviours, as well as interviews of traditional and alternative, amateur and established street and graffiti artists. This gives us insight into sub-cultural motivations and perception of current modes of public messaging, as well as a confirmation of the need for new alternative channels. Building upon everyday urban (mal)practices and aesthetics of use, we allow for new experiences and ways of communicating, sharing, or revealing to occur.

2.3 Testing and Derivation of Design Implications

The prototypes are tested on site through specifically crafted tactics and placement. Testing procedures and experiments range from outdoors workshops, to stakeouts and video filming. A dozen of users are involved in order to ensure a certain critical mass, all of which are also asked to document their own experience with photographs, narratives, drawings, etc. Accidental protagonists are also observed.

We then derive design implications based upon reoccurring patterns of people's (mis)use of the prototypes and emerging narratives.

3. PROTOTYPES

Our series of prototypes vary in terms of technology, appearance, interaction and theme. They range in themes from intimacy and public interruption to give-and-take and hacking, exploring notions of appropriation, meaning of places and situations, and social rituals. The two following prototypes are described more in-depth as first examples of this series.

3.1 Tejp 1: Audio Tags

An *audio tag* is a small box containing an audio message that once recorded can be left at hidden places in public spaces. This personal message is whispered to by-passers as they lean towards the device. People then have the possibility to record over the existing messages with their own.

The prototypes, which are made from hacked low-cost gadgets, contain a sampler buffer, a small microphone, a small speaker, an IR proximity sensor and a recording button (Fig. 2). People can record their message by holding the button pressed, and fix the tags on walls or other structures in urban environment. The proximity sensor triggers the playback of the audio recording when someone is in its proximity. The size of the devices is about a few cm^3 only.

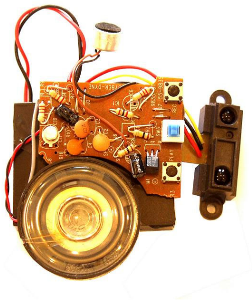


Figure 2: Electronics of an audio tag

3.2 Tejp 2: Glitch

In *Glitch*, interference caused when passers-by receive incoming messages and phone calls, are loudly broadcasted at a public place with high traffic potential, such as bus stops or busy street corners. If the speaker array is f. ex. linearly disposed along a usual pedestrian path, the glitches stalk the mobile user during the whole phase of mobile communication initiation.

The prototypes are arrays of powered-on loudspeakers picking up electromagnetic interferences from mobile phones. The experiment is split into two separate prototypes: one using a standard antenna and installed in a grid formation, and another one parasiting off existing metallic urban structures such as fences or rubbish bins in the city, re-using them as antennas (Fig. 3).

Figure 3: The two versions of Glitch: using a standard antenna and parasiting a metallic fence.

4. DESIGN SPACES

The aspects that these prototypes explore vary in terms of physical attributes, medium, and explicitness, affecting the type of content, emerging behaviours, and meaning that result from our experimentation. Mapping the design factors and test results to each other allows us to derive general design implications for embodied interaction with location based information systems.

4.1 Exploration Space

4.1.1 Audio Tags

Audio tags illustrate the notion of overlaying personal traces. Their small size yet identifiable design keeps their discovery serendipitous. By being placed on the physical environment and only making themselves heard within a certain radius, the tags open a space of intimacy inside the public realm.

The design of the audio-tags incorporates the following aspects:

- explicit vs. implicit interaction: leaning toward a wall to listen vs. triggering the audio by accident when just passing by in front of it
- small size factor, implying discreteness
- sound as media
- physical structures, such as walls, as a support to parasite
- asynchronous, distributed communication

4.1.2 Glitch

As opposed to overlaying information, Glitch is about revealing a hidden layer of personal communication in public space. Following the situationist tactic of détournement [2], it re-situates a familiar auditory phenomenon usually taking place at homes or offices, into the unexpected setting of outdoor urban environments. As the nature and origin of the noises are familiar to most people and easily identifiable, yet the speakers remain hidden, a situation of interruption is created, highlighting the virtual and pervasive layer of mobile phones communication.

Glitch is designed taking the following parameters into consideration:

- abstracted content, in the form of electroacoustic glitches
- sound as media
- electromagnetic bandwidth for mobile communication as a support to parasite
- synchronous, mobile communication
- recycling of urban structures as a part of the device

4.2 User Engagement Space

Observables of people's involvements inside the design parameters we have created are content, placement, modes of initiation, and interaction behaviours. People's perspective of the experiment can change depending on what role they play, from placers and seekers (insiders actively engaged in the experiments) to accidental finders (outsiders in the general public).

4.3 Analysis Space

The analysis of the observations is concerned with:

- types of emotion and meaning conveyed by the content
- relationships between content and placement, thus between meaning and context
- reactions to sound as media
- changed perception of place when interacting with device
- feelings of intimacy vs. disturbances as result of discreteness and initiation modes
- influence of the physicality of the device on the interaction

5. CONCLUSION

We introduced the project Tejp, which explores various possibilities for overlaying personal traces on public spaces through a series of on-site experiments. By focusing on tangible interactions between the user and the information space and grounding the experiments in cultural understanding of existing urban practices of personal expression, we design for embodied uses, meaningful content and interesting behaviours to emerge. Obvious issues of legality, acceptance, and fear of electronic objects deployed in the public realm are present and need to be addressed and solved. Ultimately, the results of the experiments will derive informed design implications for location-based information systems.

6. ACKNOWLEDGMENTS

We would like to thank all of the participants and interviewed individuals, specifically MABE, LADY and EMMI. We would also like to thank Tobias Skog, Ramia Mazé, Daniel Rehn, Daniel Skoglund, as well as PLAY, FAL, and the *Mobile Services* project members for their comments and support. This project is funded by the Swedish Foundation for Strategic Research through the *Mobile Services* project and the Interactive Institute.

7. REFERENCES

[1] BIT RADIO project, Bureau of Inverse Technology. http://www.bureauit.org/radio/

[2] Debord, G.-E. Methods of Détournement. *Les Lèvres Nues #8*, 1956.

[3] Dourish, P. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press Cambridge, USA, 2001.

[4] Espinoza, F., Persson, P., Sandin, A., Nyström, H., Cacciatore. E., and Bylund, M. GeoNotes: Social and Navigational Aspects of Location-Based Information Systems. *Proceedings of the Third Annual Conference on Ubiquitous Computing Ubicomp '01*, Atlanta, USA 2001.

[5] Feiner, S., MacIntyre, B., Höllerer, T., Webster, A. A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. *Proceedings of the First Annual International Symposium on Wearable Computing ISWC '97*, Cambridge, USA, 1997.

[6] Graffiti Writer project, Institute of Applied Autonomy. http://www.appliedautonomy.com/gw.html

[7] Hansson, R. and Skog, T. The LoveBomb: Encouraging the Communication of Emotions in Public Spaces. *Proceedings of Conference on Human Factors in Computing Systems CHI '01*, Seattle, USA, 2001.

[8] Ishii, H., and Ullmer, B. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. *Proceedings of Conference on Human Factors in Computing Systems CHI '97*. Atlanta, USA, 1997.

[9] Marmasse, N., Schmandt, C. Location-Aware Information Delivery with comMotion. *Proceedigns of the Second International Symposium on Hand-Held and Ubiquitous Computing HUC '2k*, Bristol, England, 2000.

[10] Martin, N. M. Parasitic Media: Creating Invisible Slicing Parasites and Other Forms of Tactical Augmentation. http://www.carbondefense.org/cdl_writing_7.html

[11] Rekimoto, J., Ayatsuka, Y. Augment-able Reality: Situated Communication through Physical and Digital Spaces. *Proceedings of the Second International Symposium on Wearable Computing ISWC '98*, Pittsburgh, USA, 1998.

[12] Rozier, J., Karahalios, K., and Donath, J. An Augmented Reality System of Linked Audio. *Proceedings of the International Conference for Auditory Display ICAD '00*, Atlanta, USA, 2000.

A Widget-Based Approach for Creating Voice Applications

Andreas Hartl
Telecooperation Group
Darmstadt University of Technology
Alexanderstrasse 6
Darmstadt, Germany
andreas@tk.informatik.tu-darmstadt.de

ABSTRACT

Voice based applications nowadays are difficult to author compared to conventional ones; yet there is an increasing need for such applications in mobile environment. A reason for this difficulty is that application developers must tackle with relatively low level programming interfaces for voice based applications. This paper introduces *"audio widgets"* – a notion of higher level building blocks for such applications that ease modularizing and re-use.

1. INTRODUCTION

Voice based applications are an important building block of future ubiquitous and mobile computing environments. The reason is that such applications may be used in situations where hands and eyes must be free to use – e.g. while driving a car – or that the device is too small to support display and/or keyboard. Our research group is currently developing such a device designed to be small enough to fit into a headset [8].

While the concepts of interfaces based on windows, icons, menus and pointers (*WIMP-interfaces*) are widely known and adopted by programmers, there is no such generally adopted higher-level paradigm for voice based user interfaces. Instead, developers have to deal with various elements. For handling input, developers must deal with the grammars supplying models to understand the users' utterances. In most cases, context-free grammars (CFGs) are used for this purpose which consist of hand-crafted rules derived from the developer's inherent knowledge of the language or from corpus-based linguistic knowledge. Especially n-gram based grammars may be a powerful alternative in cases when the word order of the input is irrelevant. The output of voice based applications usually consists of either pre-recorded sentences or of textual data converted into audio via a text-to-speech engine. For example whether there should be acoustical clues for the users or whether the output should be structured, developers must implement such solutions for themselves.

One of the downsides of these approaches for voice based applications is that re-usability of grammars is very limited. Developers are likely to re-implement most of the grammar if they encounter similar problems in different applications. Another pitfall is internationalization which may require large part of the user interface handling code to be rebuilt if other languages do not resemble the original language closely.

This paper presents an approach that implements the notion of widgets from GUIs to audio based user interfaces. Whereas the term widget originally is targeted towards window gadgets, we conceive it to be much more general, as a high level building block of user interfaces. The idea of "audio widgets" therefore is a generalization of the idea of graphical widgets; in both cases widgets are higher level building blocks that abstract the core functionality of the interface (drawing pixels or creating grammars). Furthermore, the idea may easily be extended to other modalities as well, e.g. gestures or hardware buttons.

2. EXISTING SOLUTIONS

2.1 Context Free Grammars

Most solutions commercially available support context free grammars. Rules for such grammars usually are authored in a BNF-like notation and may include references to other rules, text literals and variable names as well as designators for optional and repeatable elements.

When authoring a rule one can use the following techniques: sequences for elements that occur after another, alternatives, optional elements and iterations of elements. Often also some form of recursion is allowed.

The Java Speech API Grammar Format [10] (JSGF) is a platform and vendor independent textual representation of CFGs. It targets only voice input as the application may use an arbitrary Java API for generating the output. The Java Speech API provides means for loading and deleting grammars into a speech recognizer and the ability to create a grammar at runtime. JSGF grammars provide a set of rules which may be activated and deactivated once loaded. Each rule must have a unique name in order to reference it. In order to prevent naming conflicts, rules may be put into packages which provide separate namespaces. A rule may contain literal text, or reference to other rules. When using references right recursions are allowed. Programmers of the JSGF define the grammars in their application and apply them as necessary. They receive events once a particular grammar matches; programmers are free to interpret and handle these events as needed.

While JSGF extends the Java platform by additional means of input and output, VoiceXML [12] has a broader view towards voice based

This paper was presented at "Physical Interaction (PI03) Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors. Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

applications, it also targets output and navigation. VoiceXML applications are similar to web based applications: they are created by a server-side scripting technology and then interpreted by a *Voice Browser* which combines voice recognition and text-to-speech engines. VoiceXML applications consist of several forms that contain audio output and may define fields for entering data. VoiceXML provides several pre-defined fields for often used input elements (e.g. free form text, numbers, selections), developers can also use grammars to create their own fields. For navigation, VoiceXML forms can use URLs to link to others; also there is a special construct for menus that automatically creates the grammar necessary to select an entry. Currently VoiceXML only supports CFGs to an extent similar to JSGF and a grammar format for tone dial phones, although there are attempts to allow other grammars as well [11].

2.2 N-Grams

There are several types of statistics based grammars, the most widely used are N-Grams. N-Grams use a matrix that describes the probability of any word being entered based on an arbitrarily sized vector of previously recognized words. Compared to CFGs, n-gram based grammars are rarely used in commercial products.

N-grams are difficult to author as they usually are created using a sample text that is analyzed for its probabilities. This means that the text must be representative of future inputs. If an n-gram does not prove to be useful, one must alter the sample text and re-analyze it which leads to long development times when compared to CFGs. Yet there are some methods that ease creating n-grams either by using intermediate representations of the sample text, e.g. CFGs [4] or by using local n-grams which are merged together [9].

3. AUDIO WIDGETS

The idea of audio widgets was developed as part of our group's MUNDO project [5]. One of the goals of this project is to deliver a basic device that can operate hands- and eyes-free so that its user can focus on the task she is doing. Additionally, the device should work both in a disconnected and an online environment, with the ability to enhance its built in speech recognition by a server-based one if it is online.

We found that this was difficult to achieve using an approach based only on grammar descriptions for several reasons:

- Application designers must make a tradeoff between responsiveness and robustness of the system (the less rules the better) and ease of use (the more rules the better). In case of a dynamic environment which may change from online to disconnected mode and back again, the rule sets for both modes may differ strongly because of the limited capabilities of the built in recognition engine.
- If a recognition engine has special features (e.g. n-gram grammars), developers must explicitly write code using these features. Therefore, an upgraded engine may have little to no impact on existing programs.
- MUNDO will allow users to associate other devices to the basic device. Such other devices may feature additional output capabilities such as displays. If developers wanted to use such multimodal environments with grammar based applications, they would have to write an additional graphical user interface by hand.

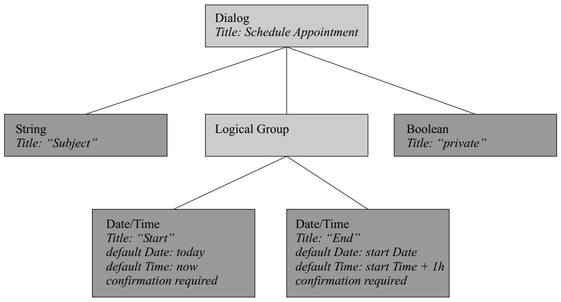


Figure 1: The logical tree of a dialog for a time scheduling application

In order to circumvent these issues, we use audio widgets as an higher level representation of the developer's intent. When using widgets, one must not change the application but only the widget instance to adapt to changes of the underlying recognition engine. Additionally application developers do not have learn how to author different grammars as this task is delegated to the widget developer.

3.1 Architecture

MUNDO defines an abstraction layer that differentiates between a *logical* user interface and its *realization*. Application developers specify the logical content using abstract widget items that form a tree. The tree contains the data types to be input and output as well as meta- data describing e.g. the title of the element, its priority and recommended presentations of the widget.

Figure 1 shows a tree for the logical user interface for scheduling appointments. The tree's leaves specify logical widgets that define the data type to be entered and the meta-data necessary to identify the widget. The title may be rendered into prompts, can be used to disambiguate input data (*"Set start date to tomorrow"*), etc. Other meta-data entries define default elements and whether the content should be entered explicitly or whether the user just needs to acknowledge precalculated data. The logical tree also serves as a means for structuring widgets into related elements. In the above figure, the date/time entries are more closely related to each other, so they are contained inside a logical group. The software that transforms the logical into the realization tree uses the grouping as a hint that it should place the widgets in the same hierarchy level.

At runtime we transform the tree for the logical user interface into a tree of concrete widgets based on the knowledge of the device. Such a transformation may map one or more logical widget to one realization widget and it may flatten or raise the hierarchy of the tree (c.f. [1, 2]).

Such a layer of indirection has been proven useful in the past [6] – a device is not required to support all logical widgets (as many of them may be mapped to one realization widget) thus saving space which is a crucial point for embedded devices. The varying hierarchy allows easy access to all information in cases when the environment permits it – e.g. if the server side speech recognition can resolve ambiguities – and fast access to crucial information in limited environments at the cost of users having to traverse additional hierarchy levels for less important information. Also, this architecture can be easily adapted to support additional modalities. For example, in order to render a logical widget onto a GUI, the GUI's own widgets can be used as realization widgets.

Finally, the runtime environment creates the concrete representation of the widgets specific for the environment out of the realization widget. For voice based applications, input widgets are converted either into CFGs of different grammar formats or into other grammar specifications available within the environment; output widgets typically are transformed into wave audio using text-to-speech (with parameters set by the widget) or prerecorded audio. Realization widgets are transformed into the appropriate representation as near to the device as possible – in most cases this will be done on the device itself.

While this paper explicitly targets voice based interaction, the architecture for mapping logical widgets is extensible. They may be mapped onto several other modalities apart from voice and graphics. We are currently extending the headset device by a limited number of buttons and a weel for entering continuous data. In a next step, we will adopt the widget paradigm to this additional input modality. One could also think of gesture based input widget or of an output widget that renders status information to a device glowing in different colors for different statuses.

3.2 Developing Widgets

Disconnecting the design of grammars from the design of audio widgets also means, that there is a need for a set of standard widgets. Also, it should be possible to create new widgets as special needs for inputting and outputting data arise. The design of realization widgets and logical widgets should be comparable, even it the latter ones deal with higher level abstractions. For logical widgets, developers must provide a mapping to the appropriate realization widgets, for realization ones a mapping to grammars and parameters of the speech synthesis.

Although developers can develop entirely new widgets just by creating new classes, it is advisable to subclass already existing ones. In most cases, the development of new widgets will result in adding additional constraints to already existing ones. By constraining widgets, one can easily model affordances into the user interface while utilizing the power of the general purpose widgets: they are fine tuned to the execution environment and abstract from the grammar used – in a mobile environment a widget can consist of a CFG minimizing the recognizer's complexity, in server based environments it may use a more powerful recognition engine without further modifications.

A likely candidate for subclassing is the *select one out of many list* which presents its users a set of mutually exclusive choices. In the general case, this list should support arbitrarily long lists which means that, when prompting the user, it only gives away its title. To create a logical widget for entering booleans, a developer can subclass the list and make the following modifications:

- Create a prompt that indicates that this is a question with the answers *"yes"* and *"no"*
- Fill the list with standard items that are possible answers. This may include items such as *"yeah"* and *"nope"*
- Provide a mapping from the list item selected by the user to the appropriate boolean value

By using a list instead of creating an entirely new widget, this new boolean widget benefits from various implementations of the existing widget.

3.3 Prototype

We have developed a prototype that implements a dialer functionality for a Voice over IP application using audio widgets. Currently, it uses only two types of widgets: a descriptive text that is rendered by a text-to-speech synthesizer and the "1 out of many list" mentioned above. The descriptive text is used in the prototype whenever the application wants to create a prompt for the user, the list contains the names of all members in our research group.

The prototype is written in Java and uses IBM's JSGF implementation based on ViaVoice. As a text-to-speech engine, it uses the default ViaVoice implementation without any modifications. For input recognition, it converts widgets into context free grammars. Currently there exists a 1:1 mapping from logical to realization widgets, although we have a graphical representation of logical widgets as well for debugging purposes. We are now working on a second widget-set for restricted environments such as the hands free device mentioned above.

The grammar of the list widget is created in such a way that users may say only a part of any list entry to select it, e.g. the first name. This may lead to ambiguities which the widget can deal with in a limited way. If a user's utterance affects several items, the widget creates a new grammar that allows only selecting one of the ambiguous items and prompts the user to resolve the ambiguity.

For comparison we have designed a similar user interface using VoiceXML. We have experienced that VoiceXML is efficient if the full content of a list item is necessary to select it; in such a case a `<menu>` element is sufficient. Depending on the voice browser, the `<menu>` element may even support selecting an item by DTMF tones. Yet, if one wants to allow ambiguities such as the widget-based application, the developer must write grammars similar to those automatically created by the widget. In that case the VoiceXML solution is notably bulky compared to the widget based prototype.

3.4 Future Work

As a next step, we plan to research basic building blocks of voice based applications and develop widgets after them. The work already done in this area for graphical user interfaces [3] may ease this task. Currently we have identified the following often used types of input widgets: free form input that pass the recognized input to the application without further modification, an audio input widget that simply records data, a list that lets users select one entry out of many, a list that allows multiple selections, a boolean input field for yes/no questions, an input field for numbers and a widget for selecting the date and/or the time.

Especially the date/time widget must deal with various types of input data: Not only absolute input like *"December, 24th, 2004"* but also with relative input *"tomorrow after the meeting"*. In order to correctly recognize the utterances, audio widgets must therefore have better knowledge of the environment they are deployed in than graphical ones. An easy way for achieving this is to give widgets the possibility to let widgets access the application's logical tree at runtime.

For the basics of outputting data, we are going to use audio output widgets that play back audio files and labels which are rendered by text-to-speech. Additionally, we are planing to deploy a widget for structured audio output [7] which allows browsing through larger amounts of text. We have yet to investigate if developers can create applications featuring ambient audio using any output widget and simply use it as a background or if a specialized ambient audio widget will be necessary. We also are going to extend the idea of widget based user interfaces towards other modalities.

As usual for user interface based research, it also will be necessary to make extensive usability tests (c.f. [13]). We are currently building a small set of audio based devices which support basic personal digital assistant functionality; the dialer prototype being one of its functionalities. Other applications will be accessing the personal calendar and scheduling new appointments, creating personal notes and listening to streaming audio. We plan to equip students with these devices and let them try to accomplish predesigned tasks; with a field test following where students should use the devices to ease their daily life at the university.

4. CONCLUSION

We have introduced an approach that introduces the concept of widgets to audio based applications. This creates a layer of abstraction between the application and the type of grammar used by the recognition engine thus enhancing portability and easing the process of developing the application.

We introduced trees of logical widgets which represent the data to be input and output along with meta-data augmenting the logical widgets. We described how these logical trees are transformed to trees of realization widgets based on the user's devices' capabilities and the supported modalities; the realization widgets then are transformed into the lower level representation required by the device.

We showed an example of how widget developers may use already existing widgets and subclass them in order to introduce additional constraints to the interface. We also showed a prototype application based on basic audio widgets and compared it to an application created using VoiceXML.

Finally, we have identified the main challenges of further developing audio widgets which are finding out the core widgets and evaluating user interfaces based on audio widgets.

5. REFERENCES

[1] J. Eisenstein and A. R. Puerta. Adaptation in automated user-interface design. In *Intelligent User Interfaces*, pages 74–81, 2000.

[2] J. Eisenstein, J. Vanderdonckt, and A. Puerta. Applying Model-Based Techniques to the Development of UIs of Mobile Computers. In *Proc. of Intelligent User Interfaces*, pages 69–76, Jan 2001.

[3] H.-W. Gellersen. *Methodische Entwicklung flexibler interaktiver Software*. PhD thesis, Karlsruhe, 1995.

[4] J. Gillett and W. Ward. A language model combining trigrams and stochastic context-free grammars. In *5-th International Conference on Spoken Language Processing*, pages 2319–2322, 1998.

[5] A. Hartl, E. Aitenbichler, G. Austaller, A. Heinemann, T. Limberger, E. Braun, and M. Mühlhäuser. Engineering Multimedia-Aware Personalized Ubiquitous Services. In *IEEE Fourth International Symposium on Multimedia Software Engineering*, pages 344–351, 2002.

[6] A. Hartl, G. Austaller, G. Kappel, C. Lechleitner, M. Mühlhäuser, S. Reich, and R. Rudisch. Gulliver – A Development Environment for WAP Based Applications. In *The Ninth International World Wide Web Conference.Amsterdam, NL*, 2000.

[7] F. James. *Representing Structured Information in Audio Interfaces: A Framework for Selecting Audio Marking Techniques to Represent Document Structures*. PhD thesis, Stanford, 1998.

[8] M. Mühlhäuser and E. Aitenbichler. The Talking Assistant Headset: A Novel Terminal for Ubiquitous Computing. In *Microsoft Summer Research Workshop, Cambridge*, Sep 2002.

[9] A. Nasr, Y. Estève, F. Bechet, T. Spriet, and R. de Mori. A Language Model Combining N-grams and Stochastic Finite State Automata. In *Eurospeech'99*, volume 5, pages 2175–2178, 1999.

[10] Sun Microsystems. Java Speech Grammar Format Version 1.0. http://java.sun.com/products/java-media/speech/forDevelopers/JSGF.pdf, Oct 1998.

[11] W3C. Stochastic Language Models (N-Gram) Specification (W3C Working Draft). http://www.w3.org/TR/2001/WD-ngram-spec-20010103/, Jan 2001.

[12] W3C. Voice Extensible Markup Language (VoiceXML) Version 2.0. http://www.w3.org/TR/2003/CR-voicexml20-20030220/, Feb 2003.

[13] N. Yankelovich, G.-A. Levow, and M. Marx. Designing SpeechActs: Issues in Speech User Interfaces. In *CHI*, pages 369–376, 1995.

Designing for Physical Interaction and Contingent Encounters in a Mobile Gaming Situation

Liselott Brunnberg
Mobility Studio, The Interactive Institute
P.O. Box 24082
104 50 Stockholm, Sweden
+46 70 320 68 06
liselott@tii.se

Kristina Hulterström
Mobility Studio, The Interactive Institute
P.O. Box 24082
104 50 Stockholm, Sweden
+46 70 845 15 94
kristina.hulterstrom@tii.se

ABSTRACT

Backseat Gaming is a project set out to investigate how the highway experience can be used as resource in a mobile augmented reality game. At the same time as it opens new possibilities for novel and engaging mobile experiences it also introduce many design challenges. In this paper, we present challenges and implications on the design of two different games, when using the vivid and dynamic mobile context as resource. Essential issues concerns how to adapt the game to the temporality and unpredictability of different mobile situations, safety and the way the interaction is designed and implemented in order to benefit from the dynamic and vivid mobile context.

Keywords

Mobile game, augmented reality, social interaction, tangible interaction, physical user interfaces.

1. INTRODUCTION

Backseat Gaming is a project set out to investigate how the highway experience, created during car travel, can be used as resource in a mobile augmented reality game. Despite changed circumstances when being mobile there are only a few examples of games that exploit the benefits of incorporating different aspects of mobility within the experience [e.g. 3, 6, 9]. We believe that a mobile game can become compelling, in a new way, if it is aware of the vivid and dynamic mobile context. Car travelling is a good example, where changing scenes, sense of motion and contingent encounters provide for a special experience in a true mobile situation.

The first prototype developed within the Backseat Gaming project made use of the changing scenery and sense of motion created during car travel as a resource in the game. We were concerned with the fictitious connection between the game and the surrounding world and how this spatial relation was *interpreted, explored* and *manipulated* during the game play. User feedback showed positive reactions both towards the idea of using road objects and car travel as gaming resources as well as the idea of the roadside as a fascinating game world to explore. We concluded that the game concept was a plausible design approach worth investigating further. The second prototype, which is based on the preceding version, benefit from contingent encounters with other players by using them as a resource in the game. Contingent encounters are central in the highway experience [1]. By creating an ad hoc peer-to-peer multiplayer game we explore how contingent encounters and the motion of the accompanying traffic can be used in an engaging mobile gaming experience.

Using the highway experience as resource in a mobile game raises several design challenges both regarding game design and the interaction. A central design challenge concerning the first prototype was to understand the characteristics of the linkage between roadside objects and the game, in order to create a satisfactory user experience. It was essential that users were able to interpret the objects correctly, enjoyed the exploration of the game space, and could manipulate the relationship in an engaging manner. The temporality and unpredictability of contingent encounters on the road call for new design challenges. Due to high relative speed, people meet for very short period of time. Still, some other encounters persist. The nature of contingent encounters inspired us to explore an alternative interface and ability to interact in order to benefit from the highway experience in the game-play. Important design criteria for the interaction concern the fictitious connection between game and the surrounding physical world, contextual situations of the game event, social interaction, awareness and body constraints. An additional design challenge concern safety, it is essential that the game-play doesn't affect the driving of the vehicle.

This paper focuses on design challenges and implications when using real world context, i.e. the highway experience, as game resource in a true mobile situation. We will shortly present the implementation and findings from the first prototype. We will then discuss design challenges and implications when integrating contingent encounters as resource in the game. We believe that these findings also could apply to other mobile situations and be useful for the design of future context aware mobile experiences.

2. THE FIRST PROTOTYPE

The first prototype [3] realise a game consisting of a framing story and physical game locations where local stories are told and game manipulation is pursued. The framing story is told when the game starts to provide the player with an understanding of the rules and goals of the game. When the car approaches a game location an animated local story is triggered. The player has to attend the story in order to find virtual objects at the locations. A manipulative event is triggered when the player comes even

This paper was presented at "Physical Interaction (PI03) – Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors. Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

closer to the location. The device automatically changes to a small and virtual window. The player can now aim at objects in the physical environment, which have been described in the local story, to find virtual objects and to make them appear on the screen. By pressing a button on the device, the player can now attack or pick up the object. The game is implemented on a Pocket PC. The device is aware of its geographical position by means of a GPS-receiver and its aiming direction by means of a digital compass mounted on the backside of the Pocket PC.

Figure 1: Hardware

Figure 2: Gaming device

2.1 Summary of user feedback

A test was carried out in order to acquire feedback on how players enjoyed, understood and handled the game. The study of the test provided knowledge about individual gaming situations although the number of test situations was limited and an indication of user experience of mobile context-dependent gaming in a road setting. The test showed positive reactions both towards the idea of using road objects and car travel as gaming resources as well as the idea of the roadside as a fascinating game world to explore. It was possible for the children to understand the game concept, interpret the objects as well as technically manipulate the fictitious relation successfully.

Figure 3: Kids playing Backseat Gaming

The players adopted different gaming strategies depending on the nature of the objects involved. There was a noticeable difference between the ways they moved the device, and how they fixed their gaze, during different types of manipulative events. When a singular virtual object was placed in close proximity of a specific physical object, e.g. a virtual document dropped at an old oak tree, the players gaze moved back and forth between the screen and the physical object to make sure that they aimed in the right direction. If the game event consisted of several virtual objects spread over a larger physical space such as an allotment area inhabited by several virtual creatures, the exploration was cumbersome. They focused either on the screen waiting for objects to show up or out through the window, peppering the environment, without checking whether there were any virtual objects on the screen. The complex and vivid mobile context bring about a need for different gaming strategies. We conclude that the intention of establishing an engaging fictitious connection between the game and the surrounding physical roadside was successful even with a light-version of augmented reality technology, but the test also indicate that the game could benefit from more non-visual feedback in order to even further augment the fictitious connection to the real world and to facilitate for the player to cope with different contextual situations.

3. THE SECOND PROTOTYPE

3.1 Design Challenges

The second prototype makes in addition to the roadside also use of contingent encounters as a resource in the game. The intention is to investigate how contingent encounter can add to the gaming experience in a true mobile situation. Contingent encounters such as rapid meetings, protracted overtaking or gatherings i.e. traffic jams or red light accumulations constitute an essential part of the travel experience. The design challenge when using traffic encounters as a resource in the game lye in their temporal and unpredictable nature. Encounters can occur anywhere and anytime during the journey, still encounters essential for the game i.e. other players, might not occur during long periods of time or not at all. Due to high relative speed, people meet for very short period of time. Still, some other encounters persists, two vehicles might for example end up in a caravan driving towards the same direction for a longer period of time [5]. It is difficult to predict when an encounter will occur and end. Integrating encounters as a resource in a mobile game involves a game design that take in to account sudden appearance of potential players, momentary respective continuous encounters as well as sudden and unexpected interruptions between players. It is also essential that the game-play won't affect the driving of the vehicle. Safety is an important issue when designing applications for use in the car. It is easy to imagine a situation where the player tries to affect the driver to change the driving of the vehicle in order to profit the game play.

Crucial for the success of the game is the design of the interface and the ability to interact during game-play. Firstly, it needs to be designed to support the concept of using the travel experience as resource in the game. Secondly, it needs to be adapted to the context of traveling in a car. To design an interface that supports the concept of using the travel experience as resource in the game motivates several design criteria. As indicated in the tests of the first prototype, the interaction during the game-play need to support different game situations and strategies adapted to the physical context. It also inspired experiments with more non-visual feedback in order to further augment the fictional connection between the game and the surrounding road context. To encourage and facilitate for the user to focus on what is happening outside the car rather than on a screen during game events is even more essential when making use of contingent encounters in the game. This is due to our motivation to spur social interaction and awareness of other players during encounters, which can happen suddenly, during swift periods of time. At the same time it is also important to cultivate the fantasy and imagination of the game and to provide the player with proper feedback and understanding of the game-play. Additionally, the ability to interact needs to be designed with the context of traveling in a car in mind. This concern the players constrained position in the car and different safety issues. However, this will not be discussed here as this paper focus is on the use of the real world as resource within the game.

3.2 The Game

The game consists as the first prototype of a framing story and physical game locations where local stories are told and game manipulation is pursued. The game also consists of multiplayer events automatically taking place when the players are in the proximity of each other. Physical game locations involve the player in the game play when no other players are in the proximity. The framing story is told when the game starts to provide the player with an understanding of the rules and goals of the game. The player's goal is to gain as high power, counted in power-points, as possible before getting to the big yearly meeting for witches and warlocks. High power can be gained both by achieving knowledge, such as new spells, gather powerful objects or by being the most powerful in battles. High power will gain the witch or warlock high status at the meeting. In the beginning of the game the player takes on the role of a witch or a warlock possessing different magical specialities. The character always carries a sack to collect objects in. Different objects can be used to help the character to gain power. The objects can be picked up from the roadside respective stolen or exchanged with other players during multiplayer events.

The game is implemented on the same technical platform as the first prototype but with a few changes. Gaming activity between players during multiplayer events is accomplished through peer-to-peer wireless ad hoc networking. The application uses a rapid mutual peer discovery protocol in order to quickly detect and connect the players when they meet [5]. An external button is also integrated in order to accomplish a squeezable interface and an intuitive interaction during brief encounters.

3.3 Design Implications for Contingent Encounters

When players come within proximity of each other, approximately within 150 meter, a multiplayer event will be triggered. The game is currently limited to take place only between two players. It is designed in such way that a multiplayer session can be played regardless of the duration of the encounter. An encounter will result in a battle between the players in the purpose of enchanting the others character and thereby capture some of the other characters skills or power. If a battle is ended earlier, because of disconnection, the character with most hits will simply be rewarded with power-points. When one character have become enchanted the battle will end, the players can now involve themselves in exchanging objects with each other or fight for objects found on the roadside. By gaining the right objects a character can break its spell. The rules of the game is designed in such way that a player will not lose or win anything by the actual disconnection from the other player, this with the intention of not affecting the driving of the vehicle with the game-play. A player should for example not have any advantage for intentionally breaking the connection with another player trying to escape from an enchantment neither should there be any disadvantage if one vehicles happen to drive in different directions in a crossing.

3.4 Design Implications for Interaction

Our main purpose when designing the interface and the ability to interact within the game was the intention to support the concept of using the travel experience as resource. With this in mind, we set out the following prerequisites for the design:

- The user interface should be designed to support the fictitious connection between the game and the physical world and on the same time cultivate the player's fantasy and imagination.
- It should take different contextual situations in to account, it should support both interaction with the roadside as well as interaction with other player during momentary respective continuous encounters.
- It should support awareness and social interaction between players.
- It should relate to the theme of the game.

The intention to preserve the connection with the physical world was of specific relevance when designing the interaction for the multiplayer events, as contingent meetings can be very brief and we wanted to encourage social interaction between the players. We believe that seeing the other player during interaction will increase the gaming experience and spur social interaction. We believe that split attention between the screen and the outside world, which was the result of the graphical interface used in the previous prototype, would limit the social interaction possibilities of the game, especially during brief encounters. Thus, to encourage the user to focus directly on what is happening outside the car rather than on the screen during interaction, the current prototype has a tangible interface. The tangible interface is realized by detaching the digital compass from its previous mounted location on the back of the Pocket PC. The compass is instead used as a separate item connected through a longer cable to the Pocket PC. The digital compass is in this way turned in to a sort of magic tool. The design of the detached interface is intended to spur the users to interact socially by gestures during the longer meetings. In swift meetings, when the period of time for interaction with other players is limited, the player could concentrate on spotting the other player and act instantly without looking at the display.

To further encourage the player to interact directly with the physical world, we have used sound rather than graphics as feedback on the interaction. Sound indicates the direction to game related objects. We have also used sound as a two-sided feedback, meaning that both players taking part in a multiplayer event will hear a sound as a result of an action. The feedback was designed with the purpose of increasing the awareness and feeling of presence of the other player and to encourage social interaction. Sound is also used to make the players aware of an approaching player. As two cars come within proximity of each other, the players will hear sound as an indication of the other player's presence. To further support awareness it is possible to imagine a small light on the roof of the cars etc. to help the players to immediately spot their competitor. To instantly be aware of the other players presence is of special importance during brief meetings that are too short to be spent searching for the right car.

We have implemented the interaction as a choice of different weapons. To follow up on the theme of the game, the tangible interface can be used as a magical wand, a magical hoover and a sqeezer. Three basic features intended for use in different kind of situations distinguish the weapons. The wand can be used to cast magic spells on other players. To cast a spell, the magic wand should be swung to follow a particular pattern. This is a rather slow procedure suitable for encounters that lasts for a longer period of time, such as when two vehicles end up in a caravan driving towards the same direction. Casting a spell is not easy, but those who learn to master the magic wand will be very powerful. The wand is also intended to incite social interaction by gestures. To pick up objects along the road, the best choice of weapon will be the magical hoover. The hoover can also be used to exchange things with other players and to place things along the road. It is easy to use and can be used in almost any kind of traffic encounter. The squeezer is preferable for very brief meetings when the interaction time is limited. To fire the squeezer, the interface should be squeezed. To squeeze the interface is easier and much less time consuming than moving the interface to follow some predefined pattern. Consequently, the squeezer is suitable for encounters that last for a very short period of time, possibly less than a second. In order to cultivate the fantasy and imagination of the game and at the same time preserve the connection with the real world we have chosen to use the screen as interface in between different gaming events. Additionally, there must be a clear connection between the screen interface and the use of the tangible interface. We have used screen based graphics to show the local stories before physical game locations and to reveal the other character in multiplayer events. Graphical feedback showing the result of the interaction and other information, such as objects in possession of the player, is visible on the screen after the interactive events.

4. RELATED WORK

A number of research projects explore aspects of integrating tangible, social and human to physical world interaction into digital and ubiquitous games. Examples include Touch-space [4] and Pirates! [2] Touch-space is a system which constitutes a game space where physical and social aspect of traditional game play is integrated with fantasy features of traditional computer entertainment. Pirates! is a wireless multi-player game also exploring novel ways to maintain social aspects of traditional game play in a computer game. Pirates! take place within physical space and uses proximity to locations or other players to activate events in the game. However, common among these projects are that they explore interaction between human to human and human to physical world within a very enclosed space as that of a room relying on pre-set infrastructures. Exploring the possibilities of using travel experience as a resource in a gaming situation constitutes a different design challenge than the ones in a pre-set room.

Games exploiting issues of incorporating different aspects of mobility and the physicality within the experience in an outdoor setting include Can you see me now? and Bystander [6], both part of the Citywide project [8]. These two games explore collaboration between online participants and mobile participants on the street. Commercially available Botfighters [9] from It's Alive use location and proximity of players as a resource in the game. The location is determined with GSM mobile phone positioning, which is too inaccurate for the purpose of our research. Additionally, our design is inspired by research on tangible and graspable interfaces, as for example work made by Ishii and Ullmer [7].

5. CONCLUSION

We have presented design challenges and implications considered when using the highway experience as resource in a game. Using the real world context as resource in a mobile game includes a vide variety of design challenges concerning how to adapt the game design to the temporality and unpredictability of different mobile situations, safety and the way the interaction is designed and implemented in order to benefit from the dynamic and vivid mobile context. When designing the interaction we have carefully considered issues such as how to support the fictitious connection between the game and the real world and simultaneously cultivate the player's fantasy and imagination, and how to support social interaction and awareness between players. We have used a tangible interface that directly links the digital world and the physical world and provides a seamless method of allowing natural physical and social interaction between people [4]. Traveling in a car constitute a true mobile situation, by studying the highway experience as resource highlight several design issues regarding the benefits and challenges of incorporating different aspects of mobility within a digital experience. We believe that the conclusions made within this setting could apply to other mobile situations and be useful for the design of future context aware mobile experiences.

6. ACKNOWLEDGMENTS

We thank SITI for partially funding the research in the project. We would also like to thank members of the Mobility studio for ideas, aid and valuable comments.

7. REFERENCES

[1] Appleyard, D., Lynch, K. et al (1964), The View from the Road, *MIT press.*

[2] Björk, S. et al. Pirates! Using the Physical World as a Game Board. Interact, 2001,

[3] Brunnberg, L., and Juhlin, O. Motion and Spatiality in a Gaming Situation – Enhancing Mobile Computer Games with the Highway Experience. Interact, 2003.

[4] Cheok, A.D. et al. Touch-Space: Mixed Reality Game Space Based on Ubiquitous, Tangible, and Social Computing. Personal and Ubiquitous Computing, 2002.

[5] Esbjörnsson, M. et al. Adding Value to Traffic Encounters: A Design Rationale for Mobile Ad Hoc Computing Services. In Proceedings of IRIS'26, Helsinki, 2003.

[6] Flintham, M. et al. Where On-line meets on-the-streets: experiences with mobile mixed reality games. Human Factors in Computing Systems, CHI'03, 2003.

[7] Ishii H, Ullmer B. Tangible bits: towards seamless interfaces between people, bits and atoms. Human Factors in Computing Systems, CHI'97, 1997.

[8] Izadi, S. et al. Citywide: supporting interactive digital experiences across physical space. Personal and Ubiquitous Computing, 2002.

[9] www.itsalive.com – verified 8th June 2003.

Pin&Mix: When Pins Become Interaction Components…

Kristof Van Laerhoven, Nicolas Villar and Hans-Werner Gellersen
Computing Department
Lancaster University
LA1 5SQ Lancaster
United Kingdom

{kristof, villar, hwg}@comp.lancs.ac.uk

ABSTRACT

This paper introduces a study into the realization of physical interaction components, based on a technology for providing network connectivity and power to small objects via a layered surface. Small pin-like components can be activated and networked by attaching them to the same, augmented surface, and can be used to dynamically create an interlinked set of atomic interaction components. The physical connection becomes thus also a digital link between components. To demonstrate our proposed platform, we have built atomic interface components in the form of dials and multicolour lights that are activated and integrated in a network by simply pushing their pin connectors in an augmented surface. They allow to pick and mix colours using the red, green, and blue primaries, as a physical alternative for the traditional WIMP colour mixer tools.

Keywords

physical interaction, pin-based connectors, tangible interfaces.

1. INTRODUCTION

This paper is based on a networking technology that exploits the familiarity and ease of attaching pin-like objects to a surface, called Pin&Play [5, 6]. This technology is applied to create a modular platform to network and organise a set of physical interaction components.

Prior examples include an augmented notice board with drawing pins, where 'pinning' a document to the wall also introduces the pin's digital self to the network [6], or wall-switches that can be powered, replaced and networked via augmented wallpaper [5].

Likely advantages of using this surface-and-pin metaphor to create and modify an interface with physical widget-like objects [3], are similar to the advantages of creating a graphical user interface in visual software development. It reduces the required expertise to build an interface to a minimum, and increases the time to develop or alter the interface.

After briefly discussing the core ingredients of Pin&Play, a more detailed description of a specific case study will illustrate how such a platform could be realized. This extended focus on an actual implementation aims at deepening the understanding of both practical and conceptual issues that may be involved.

2. PIN&PLAY

The technology used in this paper is based on Pin&Play [5,6], which uses a surface where layers, instead of routes, carry data and power to devices that get attached to the surface. The devices can access these layers by insulated pins that punch into the surface to the appropriate layer. This method literally expands the network to a two-dimensional plane rather than wires or circuits, while staying in the wired networking category: Devices will connect physically and digitally anywhere on the surface. Another interesting property of Pin&Play is that orientation of the object's placement does not matter either.

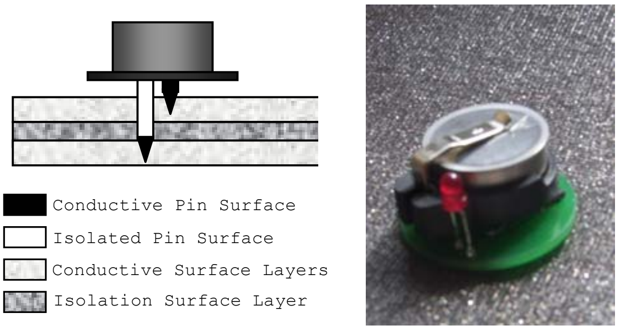


Figure 1. Left: diagram of a Pin&Play device accessing the network layers. Right: An actual Pin&Play pin attached to the surface. It holds a coin-sized unit (iButton) for memory and application specific processes, and an LED for basic user interaction.

The protocol that is used in Pin&Play is responsible for providing both a communication network and power to any device that gets connected. The Dallas MicroLAN [1] is a protocol standard that needs two connections to network and power devices: one for ground, and one for the communication signal. This signal is furthermore pulled to a high state most of the time, so that the attached devices are able to 'steal' power from that same communications bus.

The conductive layers of the surface are connected to a network master that controls all traffic over the surface. All other devices are slaves that do not have the authority to send messages autonomously, unless the master asked them to. The complexity

This paper was presented at "Physical Interaction (PI03) – Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors.
Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

of this master node depends on the specific network application, but can be implemented on embedded devices.

The actual surface material that was used in this paper consists of woven threads of nylon that have been coated with silver (AG) to give it its conductive properties [4]. To insulate two layers of this fabric, a common corkboard was used to attach the conductive layers to. Other insulators could be used as well, to make the surface flexible for instance [7].

3. PIN&PLAY AS A COLOUR PICKER

In this section, Pin&Play will be used to build a physical colour picker tool, named Pin&Mix. Pin&Mix playfully demonstrates the concept of additive colour mixing, where the three additive primary colours - Red, Green and Blue - can be overlapped in varying intensities to create a wider spectrum of colours.

The surface (referred to as 'Canvas') is used to display points of light (Mixers) that can be manipulated to display any of 256 colours by adding varying intensities of primary light using 'Primary' Pins. Up to three Primary Pins can be associated with a Mixer Pin to change its colour in real-time. The Mixer Pins reflect these changes, and remember their colour. They can be removed from the canvas, and when placed on the surface again they revert back to their last set colour which can once again be modified by using Primary Pins.

3.1 The Canvas

The canvas in this example is made of a traditional corkboard, coated on both sides with a conductive fabric. These two conductive layers are attached to a computer running the Pin&Play master software that polls for devices, reads theirs states, and executes device-specific scripts. In the case of Pin&Mix, the master reads the values form the Primary Pins (i.e., the position of their dial) and writes them directly to the Mixer Pin (i.e., to one of the three colour components of the LED).

3.2 The Mixer Pin

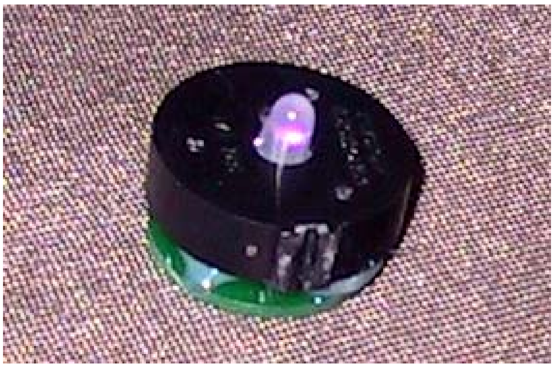


Figure 2. The Mixer Pin contains one multi-colour RGB LED, and three digital potentiometers that drive the red, green and blue values for the LED.

The Mixer Pin embodies a 'physical pixel' that can be set (or mixed) to any colour by specifying its red, green, and blue components. When attached to the Pin&Play surface, it acts as a receiver for commands from the surface's master. Figure 2 shows the Mixer Pin, while Figure 3 depicts its schematics.

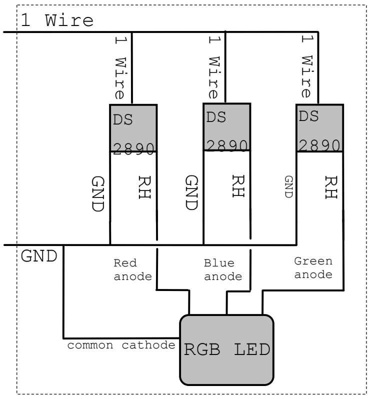


Figure 3. The schematics for the Mixer Pin, using three primary colours (Red Green and Blue).

At the heart of the Mixer Pin are three digital potentiometers, each with 256 wiper positions that can be set over the MicroLAN network to determine the colour of a Multicolour RGB LED (see the schematic in Figure 3). Three DS2890 DigiPots were used as potentiometers, which take their power from the MicroLAN network. The LED has a single common cathode, and four anodes (two for Blue, one for Green, and one for Red). The voltage supplied to each anode will determine the Red, Green or Blue intensity of the LED, and thus its perceived colour.

3.3 The Primary Pins

The form factor of the Primaries is that of a dial that can be attached anywhere on the board. Although all Primary Pins are identical in implementation, each one represents one of the three primary colours: Red, Green and Blue. Figure 6 shows a close-up of such a Primary Pin.

Figure 4. The Primary Pin contains a dial that is connected via a potentiometer to an analog-to-digital converter. It controls the intensity of one of three primary colours for the Mixer Pin.

As soon as a Primary Pin is attached to the Canvas, it becomes 'alive' (i.e. powered, and detected by the network master). Turning the dial will result in increasing and decreasing the mixing of the assigned primary colour in the assigned Mixer Pin. Currently, the assignment of the primary colour (red, green or blue) and Mixer Pin for every Primary Pin is done when the

Primary Pins are created (i.e. their target colour and Mixer Pins are fixed).

Every Primary Pin consists of a rotary potentiometer capable of varying its resistance between 0 Ohms when turned fully to the left and 100 Ohms when turned fully to the right, with an analogue range of values in between, resulting in voltages from 0 to 5.12 V.

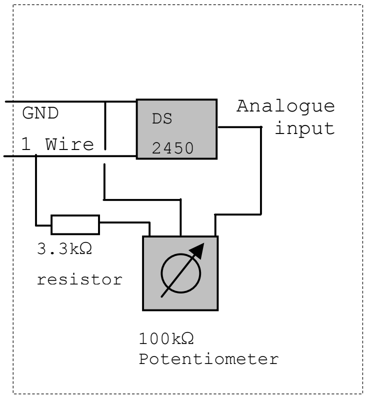


Figure 5. The Primary Pin schematics.

A DS2450, Dallas Semiconductor MicroLAN Analogue to Digital converter is used to measure these changes in voltage and translate them into a digital value that can be transmitted over the MicroLAN network. The DS2450 provides four channels, each capable of measuring voltages in the range of 0 to 5.12 volts. For our purposes, only one of these channels is used.

A 3.3kOhm resistor was placed in a series between the point of connection with the MicroLAN network that supplies the power the potentiometer to prevent shortcuts. These components are connected as shown on Figure 5. For aesthetic purposes, a metallic numbered dial is used as casing over all of these components. Turning the dial causes the potentiometer to rotate.

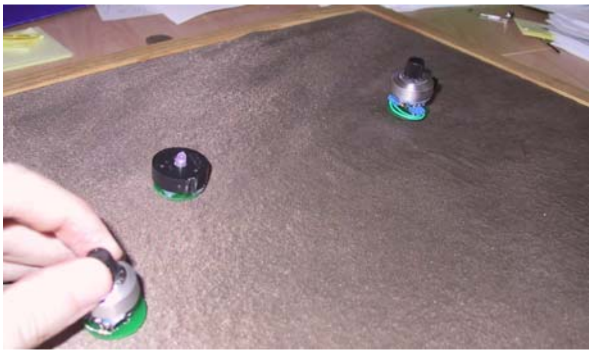


Figure 6. The surface (Canvas) with two types of interaction components: the Primary Pins (of which two are visible in the picture) are constantly being polled by the network master, and their values are mapped onto the digital potentiometers of the Mixer Pin.

The latency of the setup as depicted in Figure 8 (corkboard-sized surface with three Primary Pins and one Mixer Pin) allows three cycles per second where one cycle involves:

- Reading the dial positions of all available Primary Pins. If a Primary Pin is removed from the surface, the previous value remains in the network master's memory.
- Setting the colour of the Mixer Pin using the previously read values (if the right Mixer Pin is connected to the surface).

4. SUMMARY

We proposed a new approach to integrate a physical network of interface components, using a surface with layers of conductive textile to bring power and communication capabilities to every element of the interface. The surface has little restrictions on size, shape, or other properties, which makes it applicable in many environments. The choice of protocol and the fact that the interconnection of components is wired, means that the devices can be small and robust at an early prototyping stage.

To offer a better understanding of how Pin&Play technology could support physical interaction, a physical, modular alternative to a GUI colour picking tool was implemented, using one pin with an RGB LED as an output component and three others with dials to change the LED's colour.

5. ACKNOWLEDGMENTS

This research was in part funded by the European Commission's IST FET framework, under contract code IST-2001-37007 (under project name: Pin&Play), and by Equator, an EPSRC funded IRC project (GR/N15986/01 - "Technological Innovation in Physical and Digital Life").

6. REFERENCES

[1] Dallas MicroLAN: Dallas Reference Manual MicroLAN Starter Kit, also known as DS9091K http://pdfserv.maxim-ic.com/arpdf/DS9091K.pdf

[2] Dallas Thermochron® iButtons®: DS1921H Datasheet: http://pdfserv.maxim-ic.com/arpdf/DS1921H-DS1921Z.pdf

[3] Greenberg, S. and Fitchett, C. (2001) Phidgets: Easy Development of Physical Interfaces through Physical Widgets. Proceedings of the UIST 2001 14th Annual ACM Symposium on User Interface Software and Technology, November 11-14, Orlando, Florida, p209-218, ACM Press.

[4] Shieldex Coated Fabrics. http://www.shieldex.com, product name: Tüll.

[5] Van Laerhoven, K., Villar, N., Schmidt, A. Gellersen, H.-W., Håkansson, M, and Holmquist, L. E. (2003) Pin&Play: The Surface as Network Medium. IEEE Communications Magazine Vol.41 No.4. IEEE Press, pp. 90-96.

[6] Van Laerhoven, K., Schmidt, A. Gellersen, H.-W. (2002) Pin&Play: Networking Objects through Pins. In Proc. of Ubicomp 2002, G. Boriello and L.E. Holmquist (Eds). Lecture Notes in Computer Science; Vol. 2498, Göteborg, Sweden. Springer Verlag, September 2002, pp.219 - 229.

[7] Van Laerhoven, K., Villar, N., Gellersen, H.-W. (2003) A Layered Approach to Wearable Textile Networks. In Proceedings of the IEE Eurowearable 2003; Birmingham, UK, to appear.

TUISTER: a Tangible UI for Hierarchical Structures *

Andreas Butz, Markus Groß and Antonio Krüger,
Saarland University, Computer Science, 66123 Saarbrücken, Germany
email:(butz,gross,krueger)@cs.uni-sb.de

ABSTRACT

Tangible user interfaces provide access to virtual information through intuitive physical manipulation. However, feedback is mostly provided by displays in the environment instead of the TUI itself. In this paper we describe the design of Tuister, a tangible user interface with multiple embedded displays and sensors. We explain how Tuister can be used to browse and access hierarchical structures and briefly describe the current state of a prototype we're building.

1. INTRODUCTION

Ubiquitous computing environments pose new challenges to human-computer interaction. The increasing amount of technology embedded in the environment necessitates new interaction metaphors going beyond traditional GUI paradigms. Graspable or tangible user interfaces (TUI) are physical objects, equipped with or tracked by sensing and computing resources. They serve as dedicated physical interface widgets, allowing direct physical manipulation and spatial arrangements. It is most likely that in the next decade TUI will play an important role when interacting with computational processes in the environment. Several multi Purpose TUI designs have been proposed, such as the classical Bricks [2] or the Toolstone [8]. Often these designs focus on haptic input capabilities. Visual feedback is mostly conveyed by other devices in the environment, such as regular screens or projection surfaces. Other approaches use sensor equipped PDAs to construct a TUI providing visual feedback on the device itself [5]. The improvements made towards inexpensive and reliable organic displays will allow for direct visual feedback on a broader range of TUI in the near future. This will lead to TUI which can be used differently in different situations. They will also belong to and remain with the user instead of the environment and will reconfigure themselves according to their situational context [7].

This paper proposes the design of such a TUI named Tuister. We will first describe its basic design (section 2) and the basic modes of operation and then show how these can be mapped to the action of browsing a hierarchical structure (section 3). Finally the current state of a prototype is presented in section 4. Although we think that Tuister is especially good at browsing hierarchical structures, we will propose further applications at the end of the paper in section 5.

2. DESIGN OF THE TUISTER

Originally we were inspired by [1] to design a TUI in the form of a cube with multiple square displays. A discussion with cognitive psychologists made us aware of the fact that a cube could provide too many interdependent degrees of freedom, making it difficult to locate and remember a certain display position. Although from any given position there are only 6 possible operations with a cube (turning clockwise and counterclockwise or tilting north, south, east or west), experiments with dice have shown that people tend to quickly loose overview over the history of movements, i.e. where they came from, and which side moved where in a complex series of motions. This prompted us to choose a design with only one main axis providing just one degree of freedom.

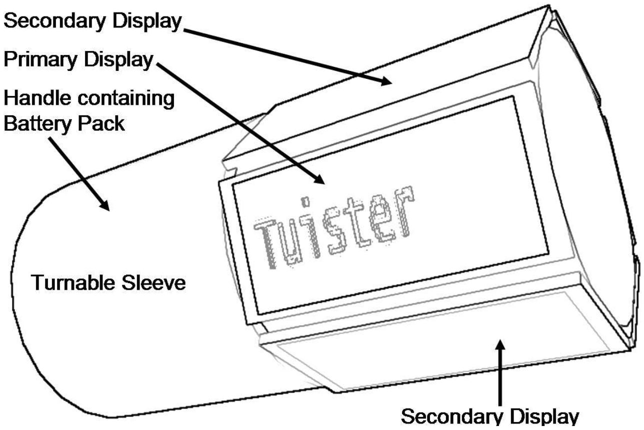


Figure 1: The basic design of Tuister

---

[*] This paper was presented at "Physical Interaction (PI03) - Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors. Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

As shown in figure 1 the Tuister is a cylindrical device of a size, such that it can be comfortably held in both hands. It consists of two parts, the *display part* and the *handle*. The display part is surrounded by discrete displays or enveloped by a continuous display. It can be twisted against the handle to an unlimited number of turns. The display part has sensors to determine its current absolute orientation in space. Technically, these can be gravitation/acceleration and magnetic sensors, but also a magnetic tracker, ultrasonic tracker, or gyroscope would do. The handle is just a plain cylinder and has no further marks or displays.

The display part can also determine its relative rotation against the handle. Together with the absolute orientation, this allows to track, which part was turned in space, and which was held stable. The picture in figure 1 as well as some of the description in the following section assume a right handed user. The main operation of Tuister is to twist the display part and the handle against each other and to rotate the whole device.

The display part has one primary display, which is the one facing the user in a comfortable reading position. In the case of a continuous display, there will still be a primary display region. If the Tuister is held at a comfortable reading distance, it will mostly be looked at from above at an angle of about 45 degrees from horizontal. By analyzing the input from the orientation sensors, the primary display can be determined, assuming that a right handed user will hold the display part in her dominant right hand.

The two displays above and below the primary display provide a visual context, since they are still partially readable. The effect of this is comparable to the perspective wall, firstly described in [6]. The form factor of a display head with six discrete displays (six-sided polyhedron with caps at both ends) has already been shown to be useful in [10]. There, a 3D-widget is described which can be used for interaction with text blocks related to anatomical 3D-models presented in the same scene.

Left handed users might prefer to hold the display part in their left hand, and operating the handle with their non-dominant right hand. This can be achieved by electronically switching the display direction by 180 degrees and by physically turning the device around accordingly. When text is displayed, the display direction also ensures that the whole device is held in the right direction, since otherwise all text would appear upside down. The direction is important for determining the primary display.

3. BROWSING A HIERARCHICAL MENU

One application we had in mind when designing the Tuister, is the browsing of hierarchical structures, such as nested menus. For the following description, we assume a right handed user. It is also helpful here, to think of the representation of the nested menu as a horizontal cone tree [9], with its root on the left and its leaves on the right side. The following interaction scheme is somewhat related to the manipulation of a cone tree. Another example of a 3D widget with a similar working principle can be found at [3] and is shown in figure 3.

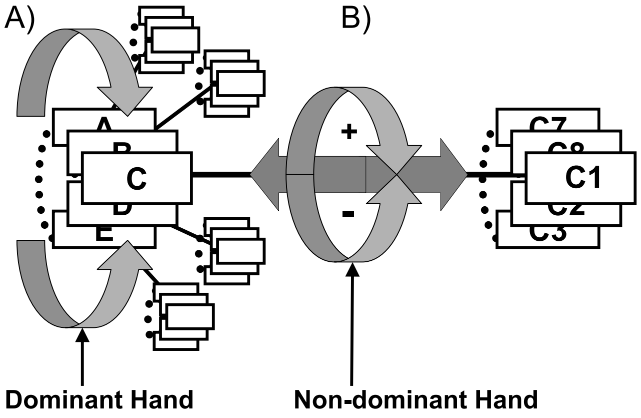


Figure 2: Browsing a hierarchical menu with two hands

The Tuister is held in both hands and the primary display initially displays the first entry of the first level of the nested menu. By holding the handle fixed in the left hand and turning the display part with the right hand, the user scrolls the first level of the menu (figure 2A). The direct metaphor for this is a selection dial. When the desired menu entry appears on the primary display, the display part is held fixed in the right hand, and the handle is turned clockwise. This selects the entry and moves to the corresponding submenu. Turning the handle counterclockwise will move the user one step up in the menu hierarchy (figure 2B). The metaphor for this motion is turning a screw.

The interaction thus uses two metaphors related to the act of turning something. Turning the display part corresponds to turning a knob that chooses between things, such as the tuning knob on an analog radio receiver or the program dial on an old washing machine. Turning the handle part uses the metaphor of a screw which is fastened clockwise and unfastened counterclockwise. Digging deeper into the hierarchy thus corresponds to fastening the handle, while going back up in the hierarchy corresponds to unfastening. We expect these metaphors to be intuitive and and understandable to a general audience after short explanation, although the second might not be obvious entirely without explanation. It will be interesting to study learning times for novice users, given no explanation at all, just a functional explanation, or the analogy of fastening a screw.

One side effect of the physical construction of the Tuister is that one part can be set in rotation and then let run freely. For the display part, this means that a menu level with very many entries, such as an alphabetical list of names, can be comfortably scrolled. For the handle, this means that we can unwind very quickly from very deep levels of the menu.

Another property we hope to verify in user studies is the fact that specific menu selections can be remembered as their corresponding sequence of turns and can eventually migrate from the cognitive memory to the much faster motor memory, just as the complex operations for solving Rubik's cube have entered the motor memory of some of us eventually.

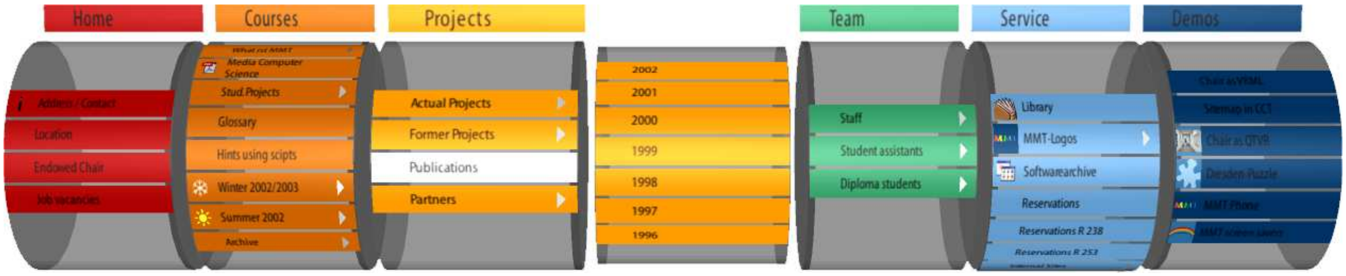


Figure 3: Browsing a hierarchical 3D menu with the mouse

4. PROTOTYPE

In this section we will describe the physical components of Tuister at the time of writing. Some of the electrical and physical design issues are still open, but most of them have already been solved and only need further refinement.

The current prototype will have a length of about $12cm$ and a diameter of about $6cm$, defined by the size of the displays in use. The display part consists of six organic displays with low power consumption and high brightness (see also figure 5). Each of those displays has a resolution of 64 by 16 pixels and is capable of displaying a short line of text, a few symbols or a small graphics. The handle part is firmly mounted to the display part and contains a battery pack with four AAA standard batteries. It is covered by a turnable sleeve that allows the battery pack and the attached displays to be rotated freely and without limitation with one hand, while the sleeve itself is held by the other hand (and of course vice-versa).

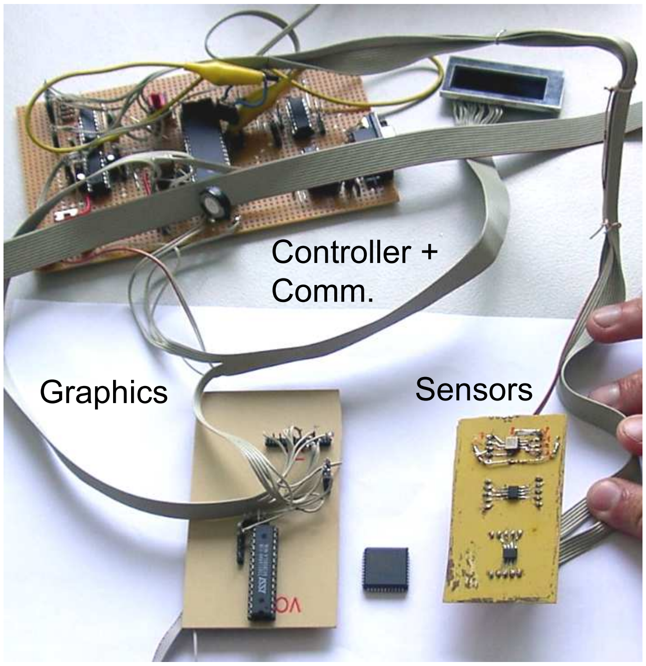


Figure 4: The hardware components of Tuister

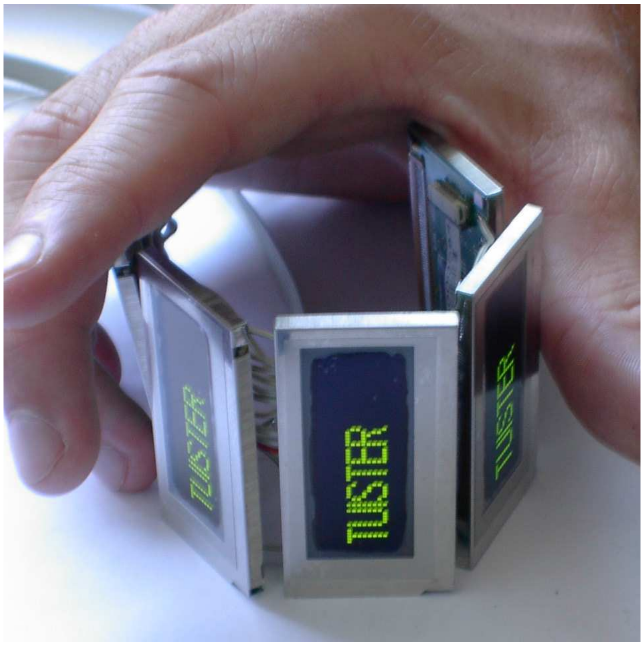


Figure 5: The display part consists of six organic displays in a hexagonal arrangement

Several sensors help to detect the absolute motions of the Tuister. 3D-acceleration and 3D-magnetic sensors will be embedded in the display part. At the moment these are working outside of the display part on an experimental circuit board (see figure 4), but embedding them should be straight forward. Both sensor types track the absolute orientation in space and movements of the whole device, such as twisting, turning and shaking of the Tuister. An additional optical sensor will track the movement of the sleeve relative to the rest of the device. Together with the absolute orientation, it is possible to distinguish, whether the sleeve or the inner part is rotated. As we will see in the next section this will allow the use of different metaphors for rotating both parts of Tuister. The functional diagram in figure 6 presents the main electrical components of Tuister. A custom-made graphics card, consisting of RAM and a graphics chip, provides the six OLED displays with the bitmaps stored in the RAM. All communication is managed by a Microcontroller, which also stays in contact with a host computer over a serial line. At the moment this connection is cable-based, but we plan to replace the cable with a wire-

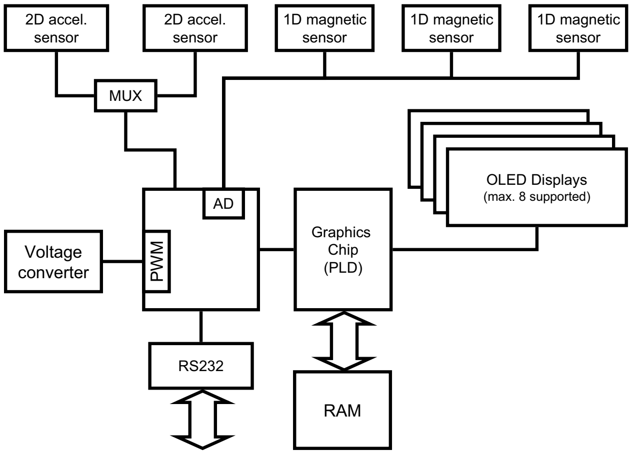


Figure 6: Functional diagram of the electric components

less link (probably with Bluetooth). The Microcontroller preprocesses the raw sensor data and hands it over to the host PC. In turn it receives bitmaps addressed to any of the six displays. Similarly to conventional input facilities, this implies that the interface logic runs in the environment (e.g. on a PC) and not on the device itself. Since Tuister is designed to remain with the user, it will reconfigure itself and eventually also integrate its interface into all instrumented environments the user visits.

5. OUTLOOK

At the moment the Tuister has materialized to the extent shown in figure 4 and as a virtual design study, but we hope to finish the engineering task in a few months. Some issues have still to be solved, e.g. how to design the sleeve in order to have enough haptical feedback and to be able to easily twist the Tuister at the same time. Further steps require to establish a wireless link between the device and the host computer. This will enable us to implement the menu browsing example and to start user studies.

Other interaction scenarios await further investigation. One idea is to use the Tuister to browse graphical material (e.g. maps) or tables (e.g. bus schedules). The device could also be used as pointing device in multi-modal system dialogues. This will enable usage pattern similar to those of the bluewand [4] with the advantage that visual feedback is available directly on the device.

The current physical dimensions are only determined by the size of electronic parts. With time, Tuisters might assume the form factors of pens or jewellery. They only need to remain big enough to be comfortably read.

6. REFERENCES

[1] K. Camarata, E. Y.-L. Do, B. R. Johnson, and M. D. Gross. Navigational blocks: navigating information space with tangible media. In *Proceedings of the 7th international conference on Intelligent user interfaces*, pages 31–38. ACM Press, 2002.

[2] G. W. Fitzmaurice, H. Ishii, and W. Buxton. Bricks: Laying the foundations for graspable user interfaces. In *Proc. of CHI-95*. ACM Press, 1995.

[3] L. für Multimedia-Technik. Sitemap. URL: http://www-mmt.inf.tu-dresden.de/.

[4] T. Fuhrmann, M. Klein, and M. Odendahl. The bluewand as interface for ubiquitous and wearable computing environments. In *Proceedings of the 5th European Personal Mobile Communications Conference (EPMCC'03)*, 2003.

[5] L. E. Holmquist, J. Sanneblad, and L. Gaye. Total recall: In-place viewing of captured whiteboard annotations. In *Extended Abstracts of CHI'2003*, 2003.

[6] J. D. Mackinlay, G. G. Robertson, and S. K. Card. The perspective wall: Detail and context smoothly integrated. In *Proc. of CHI-91*, pages 173–179, New Orleans, LA, 1991.

[7] H. Nakashima. Cyber assist for our future. In *Proceedings of the Cyber Assist International Symposium*, pages 7 – 11, Tokyo, 2001.

[8] J. Rekimoto and E. Sciammarella. Effective use of the physical manipulation vocabularies of input devices. In *Proc. of UIST 2000*, 2000.

[9] G. Robertson, J. D. Mackinlay, and S. K. Card. Conetrees: Animated 3d visualizations of hierarchical information. In *Proc. of CHI-91*, pages 189–194, New Orleans, LA, 1991. ACM Press.

[10] T. Strothotte. *Abstraction in Interactive Computational Visualization: Exploring Complex Information Spaces.* Springer, 1998.

VisualPen: A Physical Interface for natural human-computer interaction

Francesco La Rosa
Department of Mathematics,
University of Messina
C.da Papardo, Salita Sperone
98166, Messina - Italy.
flarosa@ingegneria.unime.it

Carlo Costanzo
Faculty of Engineering,
University of Messina
C.da Papardo, Salita Sperone
98166, Messina - Italy.
ccostanzo@ingegneria.unime.it

Giancarlo Iannizzotto
Department of Mathematics,
University of Messina
C.da Papardo, Salita Sperone
98166, Messina - Italy.
ianni@ingegneria.unime.it

ABSTRACT

In this paper we describe a physical user interface system for easy and natural user-computer interaction. VisualPen is a vision-based system for real-time detection and tracking of a stylus that completely replaces mouse and keyboard, thus providing a valid input device for mobile computers, and its low computational complexity renders it suitable also for PDAs. The system can be operated from a wide range of distances (either from a desk or from a wall-mounted projection panel) and is able to work with all lightning conditions. The architecture of the system is here described, and experimental results in several tests are presented and commented.

Keywords

Perceptual User Interfaces, Vision-based User Interfaces, stylus, IR

1. INTRODUCTION

Human-computer interaction has not changed its basic paradigm for nearly two decades: mouse, keyboard and icons are still the foundations of almost any computer interface. However in the last years an increasing number of researchers in various areas of computer science is developing new technologies to add perceptual capabilities such speech and vision to human-computer interfaces: such *perceptual user interfaces* are likely to be the next major paradigm in human-computer interaction. In particular, computer vision and other direct sensing technologies have progressed to the point where it is possible to detect several aspects of a user's activity, reliably and in real time, thus producing an increasing interest for vision based human-computer interaction: a technology which exploits a camera to sense the user's intentional actions and responds in real time.

Several classes of such reactive systems can be found in literature [10, 4, 9, 14, 13], including among others, those exploiting facial pointing and other head gestures [8], facial expressions, finger pointing and selection [15, 12, 5], full-body gestures [2, 1] and even more complex interactions such as overall user behaviour (mainly used for surveillance and elder/impaired people care). Most of these approaches appear promising and quite simple to implement with off-the-shelf devices such as webcams; yet, unfortunately, most algorithms heavily depend on lightning constancy, so, very often, when illumination cannot be controlled they became unreliable. Moreover, when CPU power consumption is a major issue (mobile applications running on wearable computers, PDAs, and similar devices) the high computational complexity of most adopted image processing algorithms makes them mostly inapplicable.

Another issue, largely neglected by several researchers, is related to the real naturalness of the tracked gestures and to ergonomics. As regards hand gesture - based visual interaction, for example, most of researchers initially concentrated on bare [11, 7, 6, 3] (or even gloved) hand gesture recognition, regardless of what kind of gestures was more natural for what applications. After the initial enthusiasm, which led to extremely interesting, accurate and complex solutions, some researchers realised that in several cases the main issue is *how to make easier and painless the interaction for the user*, instead of *how to astonish him with special effects*. Our Group did not escape this destiny : after having developed a system for visual human-computer interaction based on finger pointing and bare hand gesture recognition [5], we realized that most "gesture units" associated to human-computer interaction are better performed by the user when holding in hand a physical device, such as a stylus or a pen. This is probably related to the way we learn to write and draw, some kind of "legacy" very hard to change: pen, pencils, pieces of chalk, remain undoubtely the more "human" approach to writing and drawing. Experience with PDAs confirms that millions of everyday-users do not require any input device but a stylus and a LCD touch screen.

Following this intuition, we developed the device described in this paper: an optical stylus - based system that completely replaces mouse and keyboard. The system, that uses a camera to track in real time an IR emitting stylus, is able to work with all lightning conditions and can be used on whatever (if any) surface (e.g. walls, writing desks, pro-

This paper was presented at "Physical Interaction (PI03) - Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors. Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/ .

jection screens, a notepad...). The necessary feedback to the user can be provided by a video projector, a traditional CRT or LCD screen, or by more innovative devices such as head-mounted displays. The computational complexity of the exploited algorithms is so low that the system can be easily ported to a PDA without heavily affecting its power consumption.

The following two sections give a detailed description of our approach and present some experimental results. The final section draws our conclusions and outlines further research developments.

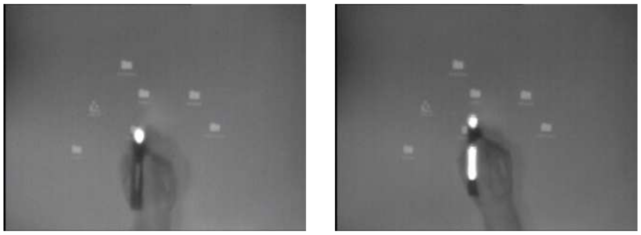
2. VISUALPEN

VisualPen is a vision-based system for real-time detection and tracking of a pen that allows to a user to interact with a kind of "virtual screen" projected on a flat surface without mouse or other pointing or keying devices (e.g. mouse, keyboard, etc.). The user can use the pen as a complete substitute for the mouse: it's possible to control the position of the cursor by moving the pen over the screen, to generate the events click and double-click and therefore to select and drag an icon, to open any folder, to draw and write. VisualPen is a system as simple to use as the mouse, but at the same time it is much more natural than the devices normally used to write or to draw. We all have experienced the difficulty to draw using the mouse and the trouble to use keyboard and mouse in order to write a text. Visual Pen puts togheter the naturalness of use of an everyday-life object, a pen, with the versatility of a personal computer and the the possibility of a distance interaction and collaborative work (it is possible to have more than a VirtualPen working at the same time). The system comprises a multimedia video projector, a gray-level video camera to acquire the scene and a pen with two IR emitting led. The scheme in Fig. 3 describes the main operational phases of VisualPen: the first phase consists of acquiring the image to be processed; the decision to implement a low-cost system and thus to use entry-level hardware means limiting the acquisition resolution to 320x240 so as to reduce the computational cost while meeting the real-time constraints. The decision to acquire in gray levels is due to the poor lighting of the environment in which VisualPen is likely to be used. A direct consequence of the poor lighting is the impossibility of distinguishing between colors. In addition, the projected images alter the scene being filmed even further. These considerations led us to exclude the use of color for the segmentation of the images acquired in these conditions.

Poor lighting and the need to make the system robust to abrupt background changes due to variations in the image being projected onto the screen make it necessary to have an additional lighting system for the projection surface or, as in our work, to add to the pointing device (the stylus) a visually detectable beacon to facilitate detection and tracking.

Two IR LEDs are mounted on the device: the first (Fig. 1a), of circular shape, is used to track the pen and is switched on for the whole period of activity; the second (of rectangular shape) is switched on by the user to generate a click (Fig. 1b).

We adopted IR leds because the IR radiation leaves the scene unaltered to the human eye therefore does not affect the projection itself. By filtering out the visible component of light while capturing the image, we then obtain an image from which it is very simple to detect and to track the led and the pen. We therefore placed a low-cost infrared filter in front of the videocamera lenses: the effect is to eliminate most of the visible light component, which is mainly represented by the projected images. Segmentation of the scene is performed by means of thresholding(Fig. 2), search of connected components and edge extraction.



(a) Pointing (b) Click event

Figure 1: acquired IR Images

(a) Pointing (b) Click event

Figure 2: Thresholded images

The edges resulting from the segmentation are then processed by the Classification algorithm which returns the position of the pen and the type of event.

The Classification phase discriminates the click event using the number of active leds in the same frame: pointing is characterized by only one LED active, click by two LEDs active. To simplify the detection of the second led we use informations on the shape of leds because the two leds have different shape (circular shape and rectangular shape). Shape analisys of the retrivied contours in the image of segmentation is based on the equations 1, 2, the leds are discriminated by the different ratio (*factor*) between the principal axis of the leds. The measure of principal axis of the leds is calculated using some statistical moments (M_{ij}) (zero, first and second order).

$$\begin{aligned}
M_{pq} &= \left(\sum_x \sum_y x^p y^q I(x,y)\right)\\
x_c &= \left(\frac{M_{10}}{M_{00}}\right)\\
y_c &= \left(\frac{M_{01}}{M_{00}}\right)\\
invm_{00} &= \left(\frac{1}{M_{00}}\right) \qquad (1)\\
a &= M_{20} * invm_{00}\\
b &= M_{11} * invm_{00}\\
c &= M_{02} * invm_{00}
\end{aligned}$$

$$\begin{aligned} square &= \sqrt{4*b^2+(a-c)^2}\\ \theta &= \arctan\frac{2*b}{a-c+square}\\ cs &= \cos\theta\\ sn &= \sin\theta\\ rotatea &= cs^2*M_{20}+2*cs*sn*M_{11}+sn^2*M_{02}\\ rotatec &= sn^2*M_{20}-2*cs*sn*M_{11}+cs^2*M_{02}\\ length &= 4*\sqrt{rotatea*invm_{00}}\\ width &= 4*\sqrt{rotatec*invm_{00}}\\ factor &= \frac{length}{width} \end{aligned} \tag{2}$$

Before passing the recognized command to the operating system, the coordinates supplied by VisualPen need to be corrected because the multimedia video projector and the gray-level videocamera are not orthogonal to the projected surface and generate a trapezoidal distortion (see Figs. 6, 7). To do so, we must determine the correction parameters. A test image is projected during the initialization phase and the user is asked to touch four highlighted points with his pen in the sequence indicated.

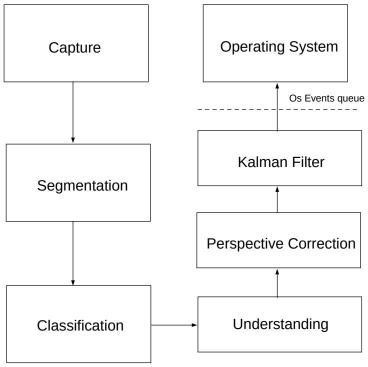


Figure 3: The main operational phases of VisualPen

In this way, once the positions of the pen in the image acquired by the camera and the four points in the test image are known, it is possible to determine the parameters of geometric transformation between the acquired image and the reference system.

The proportionality factors and the offset are given in equation 3, in which d_{12}, d_{23}, d_{34} and d_{14} represent the reciprocal distances between the four points in the test image, and D_{12}, D_{23}, D_{34} and D_{14} indicate the reciprocal distances between the four points in the acquired image.

$$\begin{aligned} w &= (D_{12}-(D_{12}-D_{34})*\frac{Y-Y_1}{Y_4-Y_1})\\ h &= (\frac{D_{14}+D_{23}}{2})\\ factorx &= (\frac{d_{12}}{w})\\ factory &= (\frac{d_{14}}{h})\\ offsetx &= X_1*factor-x_1\\ offsety &= Y_1*factor-y_1 \end{aligned} \tag{3}$$

Once *offsetx, offsety, factorx, factory* and the co-ordinates of a point (X,Y) in the acquired image are known, it is possible to determine the co-ordinates of the corresponding point on the screen (x,y) by means of the equation 4.

$$\begin{aligned} x &= X*factorx-offsetx\\ y &= Y*factory-offsety \end{aligned} \tag{4}$$

When a led has been detected, the information (after correction) about its position and velocity (interframe displacement) is passed to the tracker module. As shown in Fig. 5, this module comprises an estimator, a controller and a measure module connected in the conventional closed-loop fashion commonly adopted for visual object tracking. At each frame the Kalman Tracker, on the basis of the previous observations (measures), produces an estimate of the new status of the pen, the accuracy of which tends to improve at each iteration (in the ideal case, the error tends to zero) thanks to the information provided by each new measurement. Let us now define the status vector representing the status of the system to be tracked.

The status vector have a total of 4 elements, as expressed by the equation 5, it comprises 4 variables considered in the time instants i.

$$X_i = \begin{pmatrix} x_i & y_i & \delta x_i & \delta y_i \end{pmatrix} \tag{5}$$

In eqn.(5), (x_i, y_i) and $(\delta x_i, \delta y_i)$ are respectively the position and the velocity of the led (in screen co-ordinates).

The Prediction-Assimilation algorithm is outlined in Figure 4: Z is the vector of our measures, so it has the same composition as X_i in equation (5). The matrix G_i represents the linear relation between the measure and the status: in our case, $G_i = I$ (I is the Identity matrix). w_i and v_i represent the noise associated with the status and the observation process. We assume that they both have a Gaussian probability distribution, zero mean and variances, respectively: B_i and R_i. The variance of X_i is P_i. The model adopted for prediction is a linear and its parameters were determined experimentally.

Prediction-Assimilation paradigm

$$\begin{aligned} \widetilde{X}_i &= A_{i-1}X_{i-1}+w_{i-1} && \textit{Prediction}\\ Z_{i-1} &= G_{i-1}X_{i-1}+v_{i-1} && \textit{Observation model}\\ & w_i \text{ and } v_i \text{ have zero mean} &&\\ & \text{and variances: } B_i \text{ and } R_i &&\\ & X_i \text{ has a variance of: } P_i &&\\ \widetilde{P}_i &= AP_{i-1}A^T+BB^T && \textit{Riccati eqn.}\\ K_i &= \widetilde{P}_{i-1}G_i^T\left(G_i\widetilde{P}_iG^T+R_i\right)^{-1} && \textit{Kalman Gain}\\ \widehat{X}_i &= \widetilde{X}_i+K_i\left(Z_i-G_i\widetilde{X}_i\right) && \textit{Assimilation}\\ P_i &= (I-K_iG_i)\,\widetilde{P}_i && \end{aligned}$$

Figure 4: Prediction - Assimilation algorithm

The performance of the Kalman tracker described above is closely related to the hypothesis that both the noise vectors and the status vector have a Gaussian distribution. At this stage we will not address this issue, since the performance of the Kalman tracker is reasonable for our purposes; several different solutions do, however, exist for this problem.

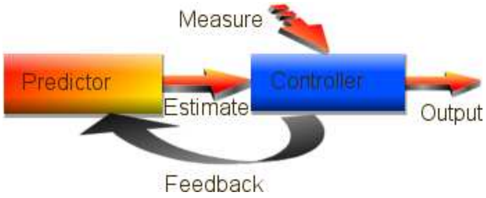


Figure 5: The prediction-measure-assimilation scheme.

3. EXPERIMENTAL RESULTS

The system was tested during and after development by several users for a considerable number of hours in numerous environments with different external lighting conditions. As VisualPen replaces the input devices in almost all their functions it was used to interact with the graphic interface of the operating system and most commonly used applications. For example, the system was used to open, select and drag icons, windows and other graphic objects on the desktop. The use of VisualPen is of particular interest in applications of free-hand interaction such as drawing in graphic processing applications (see Fig.6) and hand-writing in sign recognition software (e.g. PenReader) (see Fig. 7).

Figure 6: Drawing with Paint

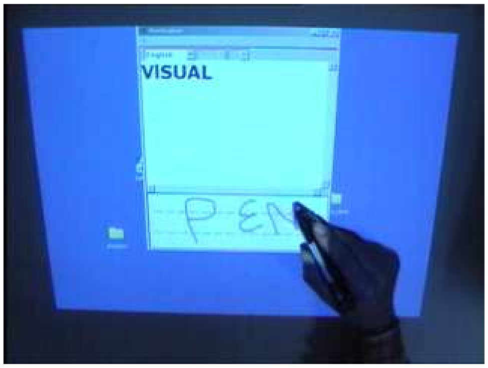


Figure 7: Use of a hand-writing

Tests were carried out on projections onto a desk, a wall and a projection screen to show the possibility of using VisualPen in different environments and situations. To evaluate the performance of the system in terms of accuracy and repeatability a considerable number of tests were carried out. To produce a quantitative evaluation we compared the output of VisualPen with a ground-truth reference. So we predisposed three classes of tests that can meaningfully characterize our system. At first we considered a segment of horizontal straight line that must be followed tracing it for its entire length with the pen. The measures have been realized asking 10 users to test 5 times the system following free hand the prefixed trajectories, that have been shown on the projection surface, and the system has stored during the tests the coordinates of the output points.

An estimation of the whole error (due both to the system and to the accuracy of the user) can be evaluated from the comparison between the acquired coordinates and those of the reference curve; carrying out then a statistical analysis on a considerable number of measures we obtained informations about the precision of the system calculating the standard deviations of the errors for each point along the reference trajectory; such errors are expressed in pixel or fractions of pixel.

For the second class of test we considered an arc of ellipse to be followed - again - free hand. Both these classes show, particularly in the second half of the abscissas, a defect of accuracy due to the uncertainty of the user. Nevertheless, the extreme naturalness of VisualPen allows to maintain the error under 3 pixels. We considered a third class of tests to try to render negligible eventual systematic errors unconsciously introduced from the users in order to estimate the intrinsic error of the system. This time the users must follow the same segment of horizontal straight line of the first class of tests, but with the pen constrained to slide on a fixed guide.

The analysis of 50 measures carried out for each class of tests shows that the standard deviation of the error is maintained always inferior to 3 pixels and that the total medium value on the three class of measures is approximately 1.5 pixels.

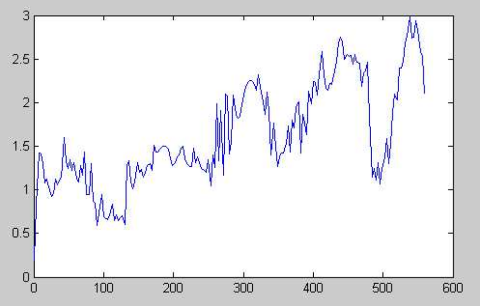


Figure 8: Standard deviation of the error made tracing free hand a segment of a straight line

Fig.8 shows results, along the 561 points of abscissa, of the standard deviation of the error made tracing free hand a segment of a straight line. The increment of the error in the second half of the segment is probably generated from a decay of the attention of users. Fig. 9 instead shows results, along the 466 points of the arc of ellipse, showing also in this time an increment of the error in the second half of the curve. Fig. 10 finally shows the obtained results along the previous segment constrained this time to a sliding guide. It is interesting to note that removing the error due to the users,

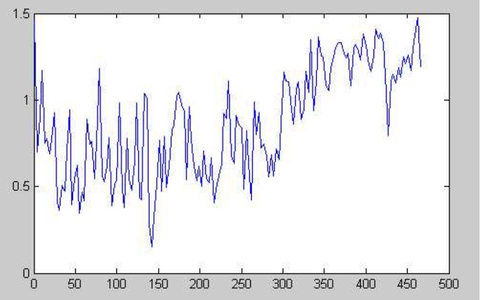


Figure 9: Standard deviation of the error made tracing free hand an arc of ellipse

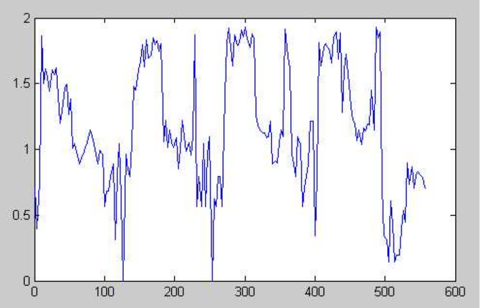


Figure 10: Standard deviation of the error made tracing constrained to a sliding guide a segment of a straight line

the system shows an intrinsic error that oscillates around 1 pixel. Such error is due to the different resolutions of the acquired image and of the projected image. To solve this problem we would need to use an algorithm that allows to obtain a sub-pixel accuracy. This kind of algoithm is usually very computationally intensive thus revealing unsuitable for our purpouses. We therefore decided to keep this error.

4. CONCLUSIONS

In this paper we presented a system for human-computer interaction that provides a more easy and suitable input device, we explained the insensitivity to lighting and the low computational complexity that permits a large number of application scenarios in several environments and with different types of devices like PDAs or other mobile device. We supplied measures of the accuracy in three classes of tests obtaining always good results that suggest the use of this system also in applications traditionally linked to mouse or keyboard. We are currently investigate the application of VisualPen to collaborative work sessions and to interact with Virtual and Augmented Reality Environments.

5. REFERENCES

[1] A. Bobick and J. Davis. The representation and recognition of action using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):257–267, March 2001.

[2] G. Bradski and J. Davis. Motion segmentation and pose recognition with motion history gradients. In *Proc. 5th IEEE Workshop on Applications of Computer Vision (WACV 2000)*, Palm Springs, CA, USA, December 2000.

[3] L. Bretzner, I. Laptev, T. Lindeberg, S. Lenman, and Y. Sundblad. A prototype system for computer vision based human computer interaction. Technical report, ISRN KTH/NA/P-01/09-SE, Stockholm, Sweden, 2001.

[4] R. Cipolla and A. Pentland. *Computer Vision for Human-Machine Interaction.* Cambridge University Press, 1998.

[5] C. Costanzo, G. Iannizzotto, and F. LaRosa. Virtualboard: Real-time visual gesture recognition for natural human-computer interaction. In *Proc. of the IEEE IPDPS'03*, Nice, France, 2003.

[6] C. Hardenberg and F. Bérard. Bare-hand human computer interaction. In *Proc. of PUI01*, Orlando, Florida, USA, November 2001.

[7] G. Iannizzotto, M. Villari, and L. Vita. Hand tracking for human-computer interaction with graylevel visualglove: Turning back to the simple way. In *Proc. PUI01*, Orlando, Florida, USA, November 2001.

[8] R. Kjeldsen. Head gestures for computer control. In *Proc. of the RATFG-RTS*, Vancouver, BC, Canada, 2001.

[9] M. Kohler and S. Schroter. A survey of video-based gesture recognition - stereo and mono systems. Technical report, Fachbereich Informatik, Dortmund University, 44221 Dortmund, Germany, 1998.

[10] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: a review. *IEEE Trans. on PAMI*, 19(7):677–695, July 1997.

[11] M. C. R. Grzeszczuk, G. Bradski and J. Bouguet. Stereo based gesture recognition invariant to 3d pose and lighting. In *Proc. of the IEEE CVPR2000*, 2000.

[12] C. P. Rick Kjeldsen, Anthony Levas. Dynamically reconfigurable vision-based user interfaces. In *Proc. of the 3rd International Conference on Vision Systems (ICVS'03)*, Graz, Austria, April 2003.

[13] J. H. Rick Kjeldsen. Design issues for vision-based computer interaction systems. In *Proc. PUI01*, Orlando, Florida, USA, November 2001.

[14] Y. Wu and T. Huang. Vision-based gesture recognition: A review. In *Int. Gesture Workshop (GW99)*, Gif-sur-Yvette, France, 1999.

[15] Z. Zhang, Y. Wu, Y. Shan, and S. Shafer. Visual panel: Virtual mouse, keyboard and 3d controller with an ordinary piece of paper. In *Proc. PUI01*, Orlando, Florida, USA, November 2001.

Enlargement of a Movement Area and Processing Multi Persons in a Vision-Based Face Tracking System

Yasuto Nakanishi Atsushi Nagasawa Hideki Koike
Graduate School of Info. Sys., Univ. of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo Japan
+81424-43-5653
{naka, atsushi, koike}@vogue.is.uec.ac.jp

Yoichi Sato
Inst. of Industrial Science, Univ. of Tokyo
4-6-1 Komaba, Meguro, Tokyo Japan
+813-5452-6461
ysato@iis.u-tokyo.ac.jp

ABSTRACT

We have developed a stereo-based face tracking system which can track the 3D position and orientation of a user in real-time, and have improved it for a large display. Our original system could track only one person, and the area in which he/she could move around was small. In this paper, we describe the enlargement of the movement area and the ability to track multiple persons in our face tracking system. Our tracking system incorporates the dynamic update of template images for tracking facial features so that the system can successfully track a user who moves around in front of the camera units; it also utilizes some image-processing boards for tracking multiple persons. These features would be necessary for ubiquitous computing environments using large-sided displays.

Keywords

Perceptual user interface, face tracking, image-processing.

1. INTRODUCTION

Recently, large-sized displays such as plasma displays or LCD projectors that can project images to a large area have become popular. They are often used in public places, e.g., train stations or shopping malls for displaying information. However, a large-sized display with equipped with a touch sensor has become popular gradually, it needs the positive action of a user to do so. Large-sized displays in public spaces often show advertisements or information in the form of movies, and they would draw attention than general information kiosks. A face-tracking system would be able to give information to users who do not approach to the display actively or to those who happened to pass the display. Using the eyes or the face as a source of input in advanced user interfaces has long been a topic of interest to the human computer interaction field. With using a face-tracking system, we can also record the number of occasions when each advertisement or information has been watched. It would be possible to offer the information about which advertisement is popular among people and the information about marketing to the person who offered advertisements or information [1].

We have also developed a face-tracking system that utilizes incorporates dynamic update of template images for tracking facial features so that it can successfully track a user's face for a large angle of rotation, and implemented several prototype applications [8]. However, our system tracked the face of a single user and did not work on multiple people simultaneously. We utilized some fixed parameters for finding facial features and for processing template-matching method, and the area in which our system tracked a user's face was not so large that he/she could move around in front of the system. A large display in a public space would be watched with multi persons, and they will watch it from various distance. In this paper, we describe enlargement of a movement area and tracking multi persons in our face tracking system.

There are several kinds of commercial products that detect a human's head position and orientation using magnetic sensors and link mechanisms, and there are much research based on computer vision techniques [1-4, 8, 9]. Haro presented a real-time pupil detector and tracker that utilized a probabilistic framework [3]. They used an infrared lighting camera to capture the physiological properties of eyes, Karman trackers to model eye/head dynamics, and a probabilistic-based appearance model to represent eye appearance. Kawato proposed an approach that tracks a point between the eyes and then locates the eyes [4]. It utilizes an image filter, i.e., the circle-frequency filter to detect "between-eyes," and stores the small area around it as a template for template matching. Stiefelhagen presented an eye tracker without special lights that employs neural networks to estimate a user's eye gaze using the images of both of the user's eyes as input [9]. They trained several neural networks to estimate a user's eye gaze on a computer screen using the eye images obtained with their eye tracker. However, most of these systems utilize a monocular image and it is very difficult to compute the full 3D locations and orientation of a face or to detect the eye gaze direction accurately and robustly. The most relevant work to us is by [6]; that work employs the template matching method for detecting the edges of eyes and a mouth by using a stereo camera pair. Their system tracks the 3D coordinates of the facial features and aims to utilize them as a visual human interface for a cooperative task with a robot. These studies, however, assume that a user sits down in front of a computer monitor. Our purpose in this research is to develop a face-

This paper was presented at "Physical Interaction (PI03) - Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors.
Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

tracking system not for a personal display, but rather for a large-sized display. Darrel et al. explored three different interface paradigms using a face-tracking system: direct manipulation, gaze-mediated agent dialog, and perceptually-driven remote presence, and showed a face-tracking system is an important module in designing perceptual interfaces for intelligent environments [2]. However, most of the previously developed face-tracking systems were designed to be used by a user sitting in front of a monitor; therefore, they might not be suitable for applications with a large-sized display as an ubiquitous computing environment.

2. Our Face Tracking System

2.1 Overview

Our vision-based face tracking system can track the position and orientation of a user in real-time (30 frames/sec) [4]. The configuration of our tracking system is similar to the one proposed by Matsumoto et al. [5], but our tracking system is capable of tracking a user's face for wider angles of rotation by introducing dynamic update of template images [8]. Our system runs on a PC (Pentium4-2.8GHz, Linux OS) equipped with a HITACHI IP5010 image-processing board, which is used while being connected to two NTSC cameras. It is equipped with 40 frame memories of 512 x 512 pixels. In order to reduce the processing time of face tracking, we use the lower resolution image whose size is 256 x 220 pixels. We use a camera unit that consists of two 3CCD black-and-white cameras and two near-infrared lights; the disparity of the two cameras is 16 cm (Figure 1). The cameras are equipped with infrared filters. These filters transmit only the light whose wavelength is close to infrared rays. By using this filter, the camera takes only the infrared light that reflects in the face of the user, thereby enabling us to eliminate the background images.

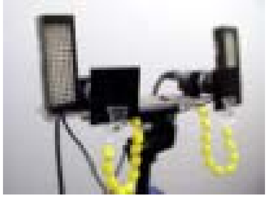


Figure 1. The stereo camera unit in our face-tracking system.

2.2 Stereo Tracking Algorithm

In order to search facial features from the camera images, we first select the region of the face. This is done by binarizing an input image from each camera while changing the threshold of binarization iteratively. Then, within this extracted facial region, we identify the location of pupils with the algorithm proposed by Stiefelhagen [9]. We search for the pupils by looking for two dark regions that satisfy the creation of anthropometric constraints and lie within a certain area of the face. After the pupils are located in the camera image, we identify the location of the mouth based on histogram projection in two orthogonal directions.

After storing the template images, we perform the template matching with four template images of eye edges and with two template images of mouth edges for each camera image. This search process using template matching is computationally expensive. Therefore, search areas are defined in our method and the eye edges and the mouth edges are searched for only within these areas instead of over an entire region of the user's face. In this process, each feature is assumed to have a small motion between the current frame and the previous one. We perform the template matching only in the areas around the eye and mouth locations that were found in the previous frame. The areas of a fixed size, e.g., 48 x 38 pixels in our current implementation, are set so that they include the locations of the edges of the eyes and the mouth obtained at the previous frame. We utilize a function of normalized correlation equipped in the image-processing board in template matching, and six 2D locations are found for each camera image. Then the 3D coordinate of each feature is determined based on triangulation.

The locations of the eye and mouth edges found in template matching are obtained independently, and the provided 3D coordinates do not always correspond to the model of the face registered at the initialization. There might be the case that multiple candidates exist for matching and that inappropriate points are detected, and it would not be appropriate that we utilize those locations. We utilize the 3D model of the face stored at the initialization to cope with this problem. We revise the coordinates provided in template matching so that they retain the nature of the rigid body model. We use the algorithm that lets a rigid body model fit the last state in the previous frame using the virtual springs proposed in [5].

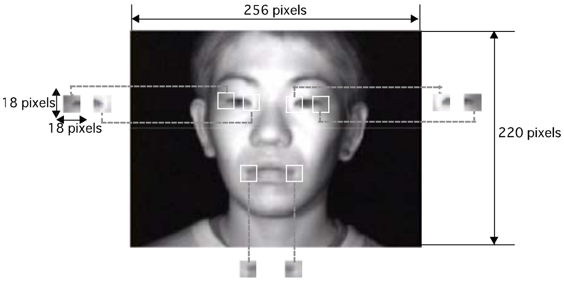


Figure 2. Samples of obtained initial templates.

2.3 Enlargement of a Movement Area

Our previous system has fixed parameters for searching facial features and for template matching. When a user move forward and back in front of our camera unit, the sizes of template images don't match with the user's face size, and our previous system fails to keep tracking.

In order to enlarge an area where a user can move, we dynamically change those parameters and store new template images according to the distance between the user and our camera unit. We set five areas, and change the parameters according to the area in which a user is (Figure 3).

Figure 4 shows the new algorithm for searching facial features. After selecting the region of the face, we estimate one area in which a user is using the face region size.

The threshold values for this estimation were decided with averaging face region sizes of five persons at each area.

According to the selected area, we select parameters for searching eye edges and mouth edges and template image size. With the selected parameters, we detect facial features and store the initial template images. This makes it possible to detect facial features even if a user enters into any area, however, we should also update template images dynamically in order to keep tracking a user when he/she moves. Figure 4 also shows the new algorithm for tracking eye edges and mouth edges.

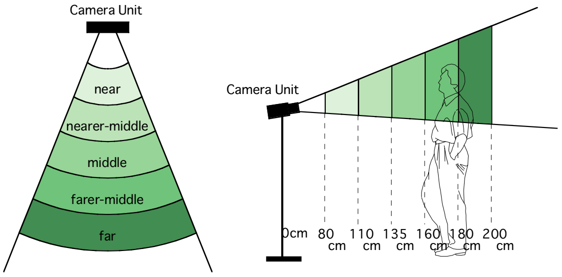


Figure 3 The five areas in our face-tracking system.

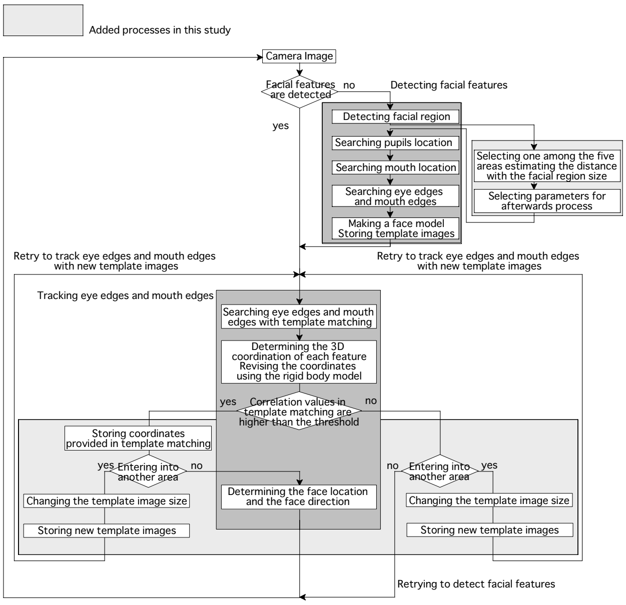


Figure 4 The improved algorithms.

After searching eye edges and mouth edges and obtaining the 3D coordinate of each feature, we check the obtained correlation values in template matching.

When the correlation values are higher than the predefined threshold, we store the provided 3D face location once and check whether the user enters into another area with the location. When the user remains one area, we determine where the user looks at using the face location and the face direction. When the user enters into another area, we change the template image size, store new template image at the obtained 3D coordinates and retry template matching with the new template images.

When the correlation values are lower than the predefined threshold, we check whether the user enters into another area with the obtained face location. When the user remains one area, we retry to detect facial features and restart tracking. When the user enters into another area, we change the template image size, store new template images at the obtained 3D coordinates and retry template matching with the new template images.

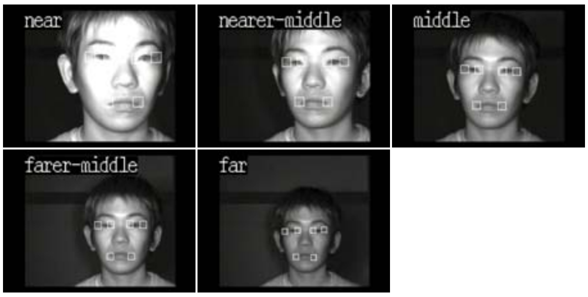


Figure 5 Examples of the tracking result in each area.

In our previous system, the area size that a user could move forward and back was 30 cm, which is just the “near” area among in Figure 3. In our current system, the current area size has become 120 cm, which is four times of the previous system.

2.4 Processing Multi Persons

Our previous system tracked only one person, and its one reason was that one user's face occupied most of the camera image as shown Figure 2. With the enlargement of the movement area, it has become possible to get an image of two person’s faces with our camera unit.

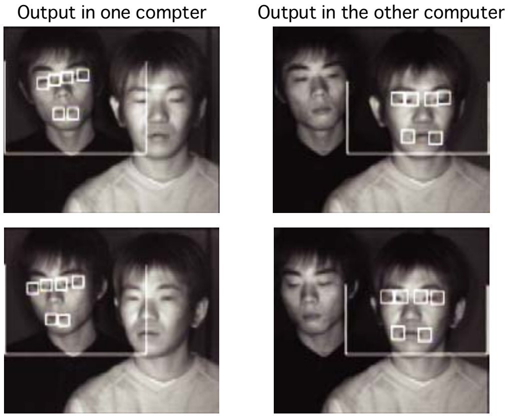


Figure 6 Output image in each PC

The image processing board that we utilize doesn’t correspond to a multi-thread program, and the processing speed has become slow when tracking two persons with processing in one loop. For these reasons, we utilize two PCs equipped with an image-processing board in the current implementation. The camera image is distributed to the two image-processing board, and each board processes different area among the same image.

When each board processes half of an image, there was a case that one person's face was in the half of the image from one camera but was not in the half of the image from the other camera. In the current implementation, each board processes an area of two-thirds of an input image (Figure 6).

3. Discussions and Conclusions

In this paper, we presented a stereo-based face tracking system which can track the 3D position and orientation of a user in real-time, and the expansion for enlargement of a movement area and for processing multi persons. Our tracking system incorporates dynamic update of template images for tracking facial features so that the system can successfully track moving users' faces. Another advantage of our tracking system is that it does not require a user in the area to manually initialize tracking process, which would be critical for natural and intuitive interaction in ubiquitous computing environments. Some researches can track multi persons, however, they tracks only 2D locations and assume that users sit down in front of a computer monitor. On the other hand, our system track 3D locations and we assume that users are in front of a large display move around.

In the current implementation, we utilize the face region size for estimating one area in which a user is, and the threshold values for this estimation process are heuristic. With using other systems such that can detect pupil easily [3, 4] or such that know a position of a user with other stereo camera units jointly, our tracking system will become more robust. For tracking two persons, we utilize two PCs equipped with an image-processing board. Because this system constitution is not suited for tracking more persons' faces, we are porting our system by using a software image-processing library instead of the image-processing board [10]. With this porting, it will become easy to add more camera units and develop a multi-thread program for tracking multi persons. We will enlarge the area in which a user can move from side to side by adding some camera units and will track more persons' faces

4. REFERENCES

[1] Crowley C, J., Coutaz, J. and Berard, F., Things that See: Machine Perception for Human Computer Interaction, Communications of ACM, 43, 3, 54-64,2000.

[2] Darrell, T., Tollmar, K., Bentley, F., Checkac, N., Morency, L., Rahimi, A. and Oh, A., Face-responsive interfaces: from direct manipulation to perceptive presence, Proceedings of 4th International Conference on Ubiquitous Computing (UBICOMP2002), pp. 137-151, 2002.

[3] Haro, A., Flickner, F, M., and Essa, I., Detecting and Tracking Eyes By Using Their Physiological Properties, Dynamics, and Appearance, Proceedings IEEE CS Conference on Computer Vision and Pattern Recognition (CVPR 2000), pp. 163-168, 2000.

[4] Kawato, S. and Ohya, J., Two-step Approach for Real-time Eye Tracking with a New Filtering Technique, Proceedings of IEEE International Conference on System, Man & Cybernetics, Nashville, pp. 1366-1371, 2000.

[5] Author1., Author2. and Author3, Vision-based face tracking system for window interface: prototype application and empirical studies, Extended Abstracts of 2001 ACM Symposium on Human Factors in Computing Systems (SIGCHI2001), pp. 359-360, 2001.

[6] Matsumoto Y. and Zelinsky, A., An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement, Proceedings of IEEE International Conference on Face and Gesture Recognition (FG'2000), pp. 499-505, 2000.

[7] Morimoto, C. H., Koons, D., Amir, A. and Flickner, M., Pupil detection and tracking using multiple light sources, Technical Report RJ-10117, IBM Almaden Research Center, 1998.

[8] Author1, Author2, Author3, Author4 and Author5, Vision-Based Face Tracking System for Large Displays, Proceedings of 4th International Conference on Ubiquitous Computing (UBICOMP2002), pp. 152-159, 2002.

[9] Stiefelhagen, R., Yang, J. and Waibel, A., Tracking Eyes and Monitoring Eye Gaze, In Proceedings of Workshop on Perceptive User Interfaces, pp. 98-100, 1997.

[10] IntelOpenCV, http://www.intel.com/research/mrl/research/opencv/

A user interaction paradigm for physical browsing and near-object control based on tags

Pasi Välkkynen, Ilkka Korhonen, Johan Plomp*, Timo Tuomisto,
Luc Cluitmans, Heikki Ailisto*, and Heikki Seppä[?]

VTT Information Technology
P.O.Box 1206
FIN-33101Tampere, Finland
+358 3 316 3111

?) VTT Information Technology
P.O.Box 1200
FIN-02044 VTT, Finland
+358 9 4561

*) VTT Electronics
P.O.Box 1100
FIN-90571 Oulu, Finland
+358 8 551 2111

Firstname.Surname@vtt.fi

ABSTRACT

In this paper, we present a user interaction paradigm for physical browsing and universal remote control. The paradigm is based on three simple actions for selecting objects: pointing, scanning and touching. We also analyse how RFID technology can be used to implement this paradigm. In a few scenarios, we show the potential of augmenting physical objects and environment with digital information.

Categories and Subject Descriptors

H.5.2. [**Information Systems**]: User Interfaces – *Interaction styles.*

General Terms

Human Factors.

Keywords

Physical browsing, pointing, tangible user interface, mobile phone, PDA, natural UI.

1. INTRODUCTION

Want et al. summarise the goal of *augmented reality* and *physically-based user interfaces*:

"The goal of these projects is to seamlessly blend the affordances and strengths of physically manipulatable objects with virtual environments or artifacts, thereby leveraging the particular strengths of each." [5]

Physical browsing can be defined as getting hyperlink information from physical objects. This can happen if the object has a way to communicate a URL to a user, which requests it. This URL can be transmitted for example with *an information tag* and it can be read with a mobile device like a cell phone. We define an information tag (hereafter: a tag) as a small and inexpensive unique identifier, which 1) is attached to a physical object but has limited or no interaction with the object itself, 2) contains some information, which is typically related to the object, and 3) can be read from near vicinity.

A tag may be for example a barcode, RFID (radio frequency identifier) tag or an IR (infrared) beacon. Based on the tag information, the user can then for example load the page corresponding to the URL to his device and get electronic information from a physical object. This is a powerful paradigm, which adds the power of World Wide Web to the interaction with physical objects – information signs, consumer goods, etc.

Another aspect of physically based user interfaces is controlling or interacting with physical artefacts using a user interaction device such as a PDA. An example of this is using a PDA as a user interface to a household appliance. This approach can be seen as a *universal remote control*. In this scenario, a universal remote control is a device, which may control or interact with all kinds of objects by using suitable communication mechanisms. A major challenge in this paradigm is the establishment of the communication between the object and the UI device.

In the world of millions of objects to be augmented with digital presence, tags represent a key enabling technology for physically based user-interfaces. Traditionally, RFID tags have been used to track objects and cargo in industry and commerce. In research projects they have also been used for physical browsing and providing services related to for example conference rooms [5]. RFID tag readers are not yet very common in consumer products but as the tags become more widespread, PDAs and cell phones may have readers and there will be a common way to access the tags.

Previously, Want et al. [5] developed *Xerox tags*, a system, which the creators describe as "bridging physical and virtual worlds". The system combines RFID tags and readers, RF networking, infrared beacons and portable computing. They have created several example applications to demonstrate the possibilities of the system. In the *Cooltown* project [3], a method called eSquirt was developed. It allows the users to collect links (URLs) from infrared beacons attached to physical objects like walls, printers, radios, pictures and others. Cooltown's user interaction theme is based on adding hyperlinks on physical locations. In addition, *barcodes* can be used to transfer information between physical objects and mobile devices. The user reads the barcodes with a

This paper was presented at "Physical Interaction (PI03) - Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors.
Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

wireless reader and the code is sent to a server. The server then transmits the information about tagged object to the user's cell phone, email, or some other information application or device.

Bowman and Hodges [1] have studied similar interactions in virtual environments whereas for example Mazalek et al. [2] have created tangible interfaces. Our paradigm lies somewhere between these two approaches, combining physical and virtual.

In this paper we represent and analyse a paradigm for physical user interaction based on using tags for augmenting physical objects with digital presence. Especially, we will present three paradigms for choosing the object of interest. We also discuss RFID tags as one possibility for implementing this paradigm.

2. INTERACTION METHODS

There are two approaches to using tags in physically based user interfaces: information related approach and control related approach. Essential for both uses is the requirement for choosing the object (tag) of interest. In our concept, there are three methods for choosing tags with readers: 1) scanning, 2) pointing and 3) touching. We suggest that for any tagging technology these paradigms should be supported to provide optimal support for natural interaction with physical objects.

2.1 ScanMe

Scanning is one way to choose the tag of interest. When a user enters an environment, he can use his reader to scan the environment for tags. The services provided by the tags will then be presented on the user's UI device. Thus the presence of the tags is communicated to the user and he can then choose the tag (object) of interest by using his UI device. Effectively, this means choosing a physical object in the digital world. This method can be called ScanMe paradigm (see Figure 1).

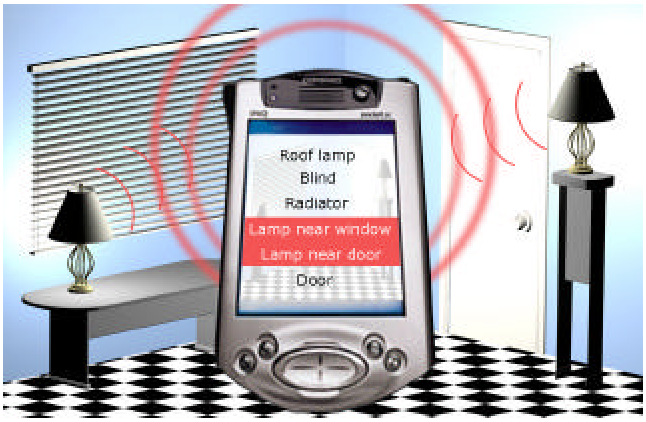


Figure 1: ScanMe

Technically ScanMe is supported by methods supporting omnidirectional or at least wide search beam communications, which is true especially for RF based methods. In ScanMe, all tags within reading range would respond to the scan, even if they were behind small objects like packaging[1]. A major issue with ScanMe is, however, the universal naming problem — association between virtual and physical objects. The tags must be named somehow so that the user can understand what physical object is associated with the information on the menu.

2.2 PointMe

If the tag is visible, pointing is a natural way to access it. In PointMe paradigm, the user can point and hence choose a tag with a UI device, which has an optical beam, e.g. infra red or laser, for pointing (see Figure 2). Pointing requires direct line of sight to the tag, but it works through transparent surfaces. Like in scanning, the tag can be accessed within the range of the reader. The PointMe paradigm may be typically implemented with IR alone, or by combinations of IR, laser beam, and RF technologies. In the latter case, the optical mechanism is used for choosing the tag while the RF communication is used from tag to UI device communication.

PointMe tags can be accessed directly by pointing and selecting. Depending on the width of the beam there is also a selection problem if there are more than one tag in the place the user points at; in this case, a scanning-like menu of the tags could be presented. In any case, there is an application dependent need for a compromise between the beam width (larger beam leading to more inaccurate selection) and the usability issues (requirement for very exact pointing may lower the usability). Typically, in the PointMe paradigm the tag of interest is chosen without ambiguity and hence the related service may be launched immediately to the UI device if required. For example if the tag responds by sending a URL pointing to product information, it could be loaded into the browser of the device immediately. In more complex situations, a user interface to the tag's services could be presented.

Figure 2: PointMe

2.3 TouchMe

In TouchMe paradigm, the tag (object) of interest is chosen by (virtually) touching it with a UI device. Like pointing, touching requires that the user identify the location of the tag. However, the tag itself does not necessarily have to be visible. RFID tags may be made into TouchMe tags by limiting the reading range. This

---

[1] Potentially, with some technologies and in the presence of a multitude of tags, there may be occasions that not all the tags successfully reply to the scan, e.g. due to communication channel overload. This would represent a problem to the UI paradigm unless there is some way of warning the UI device of the unread tags, in which case the scan could be repeated until all tags are successfully read.

can be done either by limiting the power used or by tag antenna design.

Touching is an unambiguous way to select the right tag and object. It eliminates the possibility of multiple tags responding, but the touching range limits its use. Typically, it is the most powerful paradigm in the case where a multitude of objects is close to each other, e.g. in a supermarket for downloading product information.

2.4 Universal remote control concept

The ScanMe, PointMe and TouchMe paradigms may easily be applied in the concept of physical browsing, i.e. in information related applications. However, tags and the above UI paradigms are also powerful in the concept of a *universal remote control*.

In this scenario a generic remote control is a device, which may dynamically control or interact with previously unknown objects by using suitable communication mechanisms. The basic challenges for such universal remote control are:

1. *Discovery*: how to choose the object of interest (in the physical space) by using the UI device (which is functional in the virtual space), or how to provide mapping between the physical and virtual space objects.
2. *Connectivity*: how to establish the communication channel between the object and the UI device in case the communication protocol is not know a priori, or if many communication mechanisms are supported (e.g. IrDA, Bluetooth).
3. *Communication protocol*: how to make the UI device and the object to communicate with the same vocabulary.
4. *User interface*: how to present the information to and allow control by the user on the UI device in an intuitive way.

We suggest that tags can be used as a simple mechanism to address these challenges. A tag attached to the device can hold or provide a pointer to the necessary communication parameters to be used in the control, such as communication mechanism, address, protocol and its parameters. If the tag contains a pointer to these parameters (for example in the Internet), it is possible to take into account the UI device characteristics and to download a proper UI to the device. The usage is as follows:

1. Our UI device (e.g. a PDA) includes a tag reader. In addition, it has some other communication mechanisms.
2. When the user chooses the object of interest, he scans the tag with his UI device by using ScanMe, PointMe or TouchMe paradigm. The most essential feature to the user in this procedure is that the selection is simplified as much as possible and the selection is done primarily in the physical space.
3. The tag replies to the tag reader with information about the necessary communication parameters for further control or communication needs. These may include actual communication parameters, or a URL to download these parameters and/or the device UI.
4. The UI device interprets the communications parameters, downloads (if needed) the drivers and UIs, and starts the communication with the object by using the defined method.

The main advantage from the users perspective is that the only action required from the user is to choose the object in the step 2 – all the rest may be implemented to happen automatically. There are two main advantages from the technological perspective. The first is a simple and standard[2] mechanism for device discovery supporting custom methods for communication. The second advantage is flexibility for supporting multiple devices, languages, etc. (especially in case the returned parameter is the URL of the method).

3. IMPLEMENTATION OF TAGS

The primary feature of tags is their extreme locality: they are only accessible within near vicinity, and hence they are closely related to a certain place or object. Indoor positioning and user identification can be used in similar manner as we suggest tags to be used. However, tags have some advantages over other technologies that can be used to identify a user and her indoor positioning. Some advantages of tags are their efficiency, simplicity and low cost both in computing power and monetary terms.

The most important tagging technologies currently are RFID tags and optically readable tags (barcodes or other kinds of glyphs). Both kinds of tags can be used to easily augment physical objects and the environment on a small scale. The RFID technology is becoming a challenger for barcodes in many applications, and its features allow its usage beyond possibilities of the barcodes.

RFID tags are typically passive components; i.e. they do not have their own power source; they get all the power they need from the device that is reading them. At present, the information content of a tag is typically static but the technology allows dynamic updates to the contents, e.g. updating information or adding some sensor readings from attached sensors. It naturally supports ScanMe and TouchMe concepts (the latter is achieved either by decreasing the reading power to the minimum or by modifying the antenna of the tag to be less sensitive). Support for tag selection by optical methods allowing the PointMe paradigm is being researched.

The central features of RFID tags may be summarised as follows:

1. *Visibility*. RFID tags don't need to be visible so they may be attached below the surface of the object. However, they are not readable through thick materials or metal.
2. *Range*. The maximum range of RFID tags is about four meters with 500 mW reading power [7]. It is possible to use tags that respond to touching or RF requests at very short distances. This kind of tag can be used as a TouchMe tag.
3. *Data storage capacity*. RFID tags usually have greater data storage capacity than barcodes or glyphs. The capacity may beat the range of a few kilobits [5].
4. *Sensors*. RFID tags can be connected to sensors. These sensors can be used as a condition for triggering the tag, or for reading and transmitting sensor data.
5. *Antenna*. The antenna is by far the largest element of the RFID tag, typically about one square inch. It can be made flexible and it may be attached to almost any surfaces.
6. *Price*. The prices of tags are in the order of tens of cents. In large mass production the price may be cut to a few cents.

Different RFID tags respond to different triggers. Still, their basic technology can be the same. This is a major advantage while keeping the price of the tags and their readers low.

---

[2] Here it is assumed that an industry standard for a suitable tagging technology becomes accepted and agreed.

4. SCENARIOS

The scenarios in this chapter provide use cases to illustrate the use of tags for physical user interfaces and to emphasise the need for different object selection paradigms.

4.1 Physical browsing

The user notices an interesting advertisement of a new movie (see Figure 2). She points her PDA at the advertisement and presses a button. The tag responds with an URL to a web page of the movie. The PDA immediately launches a web browser and loads and displays the page. The page contains links to the movie's web page, to a local theatre and to a sample video clip. The advertisement could also have direct physical links to aforementioned things. For example, it could have a tag, which would respond directly with the URL of the video clip, whereas the tag at the movie's name would open its web page. Physical objects could act this way like user interfaces to different kinds of information.

4.2 Shopping

The user goes to a shop in which the items are augmented with RFID tags. She sees a new chocolate brand but the trade description of the chocolate bar is not in any language she knows. However, she is very allergic to nuts and must know whether the product contains nuts. So, she touches the TouchMe tag in the chocolate bar with her PDA and gets a link to the page in which all ingredients are described. This page is provided by the shop chain, but it could also be provided by the manufacturer.

4.3 Universal remote control

The user walks into a room and wants to turn on some of the lamps of the room. He notices standard RFID stickers attached to the lamps, points the first lamp with his phone and presses a button. The tag attached to the lamp transmits an URL to the controlling method; i.e. the tag itself does not control anything. As toggle between on/off are the only options for controlling the lamp no specific UI display on the phone is needed.

To identify what controllable devices there are in the room, the user first uses his mobile phone's scan function. The RF reader of the phone sends a scan request to all tags in vicinity. The ScanMe tags respond, in this case with an URL, which is a link to their control and user interface. The mobile device constructs a menu of these responses and displays it to the user. The user then selects the desired item from the menu and his cell phone loads the user interface for that device. It should be noted that the user should not get a list of URLs for choosing. Instead, the mobile device should use these URLs to get a description of the item (i.e. a "link text"). This description would be displayed in the menu and with it a new URL, which points to the user interface of the device, for example the lighting of the room.

5. DISCUSSION

Digital augmentation of everyday objects represents a new powerful paradigm. However, there are some central usability issues involved in making digital augmentation natural. In this paper we have discussed use of tags in physical user interfaces and presented three paradigms for choosing the object. Still, some generic design issues should be kept in mind. First, the users should be able to find out if there are tags in their environment or recognise tagged objects from those not tagged. The users should also understand what the tag would do if it were addressed. This is not always clear from the tag's context. These are the basic issues of *visibility, affordances and mappings.* Visibility means that a user can see what can be done with an object. The term affordances refers to the perceived and actual properties of an object, primarily those fundamental properties that determine how the object could possibly be used. Mapping refers to mapping between control and action, i.e. relationship between doing something and getting a result from it. [4] The question with physical browsing is how do we communicate these issues to the user. Clearly, some standardisation for example in representing different kind of tags would help to solve these issues.

Currently RFID tag readers are not available as embedded in the mobile gadgets. However, especially when the RFID tags extend their range into higher radio frequencies (especially to 2.4GHz) it becomes feasible to integrate the reader with the handsets. This is required for the scenarios presented above to become reality in large term. However, despite the great number of mobile handsets sold so far, the number of potential objects to be tagged and hence augmented outnumbers them by far. Hence, it is especially the price of the tags and only secondarily the price of the reader which will decide which tagging technology is the winning technology in large-scale applications.

To conclude, we have presented a tag-based user interaction paradigm for physical browsing and near-object control. We suggest that a concept of physical user interaction should optimally support object selection by scanning, pointing and touching to fully utilise richness of natural interaction. Finally, we believe that new RFID technology developments are making it a potent technology for implementing physical browsing and digital augmentation.

6. REFERENCES

[1] Bowman, D. & Hodges, L. F. *An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments*. Proceedings of 1997 Symp. on Interactive 3D Graphics, Providence, RI, 1997. pp 35–38.

[2] Mazalek, A., Davenport, G., Ishii, H. *Tangible Viewpoints: a Physical Approach to Multimedia Stories.* Proceedings of the Tenth ACM International Conference on Multimedia, Juan-les-Pins, France, 2002. pp. 153-160.

[3] Kindberg, T., Barton, J., Morgan, J., Becker, G., Caswell, D., Debaty, P., Gopal, G., Frid, M., Krishnan, V., Morris, H., Schettino, J., Serra, B., Spasojevic, M. *People, Places, Things: Web Presence for the Real World.* Proceedings of third Annual Wireless and Mobile Computer Systems and Applications. Monterey CA, USA, Dec. 2000. p. 19.

[4] Norman, D. *The Psychology of Everyday Things.* 1988. Basic Books, 257 p.

[5] Want, R., Fishkin, K., Gujar, A., Harrison, B.L. *Bridging Physical and Virtual Worlds with Electronic Tags.* Proceedings of 1999 Conference on Human Factors in Computing Systems. pp. 370-377.

[6] Wicol Ltd Oy. 2003. [online]. http://www.wicol.net/

[7] Extending RFID's Reach in Europe. RFID Journal March 10, 2002.http://www.rfidjournal.com/article/articleview/328/1/1/

Body Mnemonics

Jussi Ängeslevä, Sile O'Modhrain, Ian Oakley, Stephen Hughes
Media Lab Europe
Sugar House Lane
Dublin 8
+353 (1) 474 2890
{jussi, sile,ian,steveh}@medialabeurope.org

ABSTRACT

Body Mnemonics is an interface design concept for portable devices that uses the body space of the user as an interface. In this system information can be stored and subsequently accessed by moving a device to different locations around one's body. The system is designed to ease cognitive load by relying on our proprioceptive sense, vibrotactile feedback, and the use of the body image of the user as a mnemonic frame of reference. Feedback from interviews conducted while developing this system suggests that the body view is a very personal artefact and that it is rich in meaning. It therefore has the potential to serve as a powerful memory aid. We are currently integrating an inertial measurement system into a portable device to enable us to conduct studies to validate our approach.

General Terms

Design, Human Computer Interaction.

Keywords

Portable device, proprioception, spatial interaction, ubiquitous computing

1. INTRODUCTION

Body mnemonics began as a interaction design proposal at the Royal College of Art and a working prototype is currently under development at Media Lab Europe. The project explores a novel interaction paradigm for portable devices that has wide-ranging applications. It is intended to improve the usability and reduce the attentional load of mobile interfaces. The mobile market is rapidly expanding and novel interface designs addressing the specific problems of the field are likely to find immediate real world applications.

Initially, the project examined the potential for a spatial interface from a psychological perspective; whether or not it makes sense to a user. The second stage is currently under way and is concerned with a technological implementation and validation of the system. The final stage will involve a number of evaluations of the system in a real world context to establish the appropriateness of our design.

This paper was accepted at "Physical Interaction (PI03) - Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors. Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

2. DESIGN CONCEPT

Body Mnemonics is a project that continues the work conducted by the primary author exploring embodied interfaces; using the real space and the body of the user. Interface design for portable devices is an ideal challenge for this approach, as the perceptual bandwidth [1] provided by the physical design of the devices is very limited. It typically comes in the form of a small touch screen, a few buttons and low quality audio output. Exacerbating this problem is the fact that portable devices are often used in situations where the user is simultaneously engaged in other cognitive tasks. Consequently, it is desirable to reduce the attentional requirements of the interface.

The fundamental concept in Body Mnemonics is that information can be accessed and stored in the space defined by the user's arm's reach - known as the reach envelope[1] - by moving a hand held device as illustrated in figure 1. The concept is very similar to a traditional desktop metaphor, where objects can be moved and activated with a cursor.

Using body space, however, has several advantages over on-screen interface in mobile scenarios. Firstly, the dimensions of the portable device limit its screen size. Expanding the working space to the reach envelope has the potential to enable direct access to significantly larger amount of data [2].

Secondly, the movement of one's hands in the body space can be perceived through the proprioceptive sense, our innate awareness of the position of our body and limbs [3]. This activity can take place in the background of our awareness and frees our vision for other tasks.

Thirdly, body image, the cultural construct through which we view our bodies, can be used as an aid to the storage and subsequent recall of information. This rich wealth of symbolism is influenced by factors such as up-bringing, education, constitution, body decoration and different hobbies and skills.

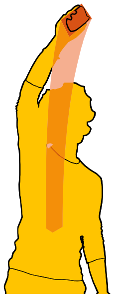


Fig. 1 Moving the portable device in one's hand to activate and store different information.

For example, the associations that might be meaningful to a skydiver, used to finding critical controls on their chest could differ radically from those of somebody with a tattoo in this area. A doctor with detailed anatomical knowledge might have yet another association.

Finally, body image can be related to the mnemonic device known as the method of loci [4]. This millennia old technique uses space to organise information and distribute memory. The memorised material is associated with different places and objects in an imaginary or real space. Journeying through the space is a key to the recall of the material. We propose that the body space can be harnessed in a similar manner, using different body parts as mnemonic cues to help access information.

3. INTERVIEWS

In order to gain insight into the validity and feasibility of this approach, preliminary interviews and questionnaires were conducted in the Royal College of Art to investigate the different mapping strategies that people might develop. The interviews were relatively unstructured, representing our desire to explore the conceptual space of the project. 10 subjects were interviewed face to face in either one or two sessions and questionnaires were sent via email to 35 individuals. 15 replies were received. All the participants were experienced computer users and were between 21 and 38 years old. In all cases participants were asked what applications, urls or other data they would store on their body and where and how they would position them. They were also given a more specific task; to distribute their personal phone book entries within their body space. The replies received were very varied, concentrating on many different forms of content and revealing four basic mapping strategies: *emotional, associative, functional and logical.*

3.1 Body mapping strategies

Emotional mappings tapped into the culturally shared symbolic perception of the body and resulted in such structures as "*husband and children in the heart area*" and "*my dad by my head cause he always knows best"*, reflecting the personal meaning of the stored information.

Associative organisations were based on the same kind of emotional history but were connected to specific past experiences, and hence made sense only to the individual. For example: *"my sister and my close friend [I would store on my neck], because they gave me necklace and pendant separately but I always wear them together*".

Functional mappings were connected to specific tools or to ergonomic or behavioural characteristics of the body. For example: *"MP3 archive to my left ear", "to-do-list to the back of my head, because I scratch my head when I try to remember".*

Logical mappings treated the space as having some associational starting point, and then built complex information in relation to it. For example, *"Right side is generally the more logical side, analysis, work etc. and left the emotional and fun side. My dad would go to the right side of the head, mom on the left, and sister somewhere around the head too, as they together form my family."*

During the face to face interviews whole body maps were composed around a specific content arrangement task. These described the associative maps of a variety of domains including music genres, application shortcuts, phonebook entries and bookmarks. Typical arrangements are illustrated in Figure 2.

Even though the association tasks in the study were conducted as mental exercises only, and hence lacked the depth and environmental influence that real life usage of the system would provide, the results were strongly suggestive that the body could serve as a versatile canvas on which to store information.

An underlying trend was that certain body parts possessed a strong association with a specific person or function, and this would set the frame of reference around which related information would be stored in a logical or spatial framework, using top-down or left-to-right symmetries, circular envelopes or the body shape.

Two participants in the study felt that using the body as an interface was not appropriate. One saw the method as cumbersome, and the body lacking relevant associations to the kind of content available on a portable device. The other was concerned that the body would form an overly emotive frame of reference. Arranging the contact details of one's acquaintances on the body might convey too open and strong a statement about his or her interpersonal relationships.

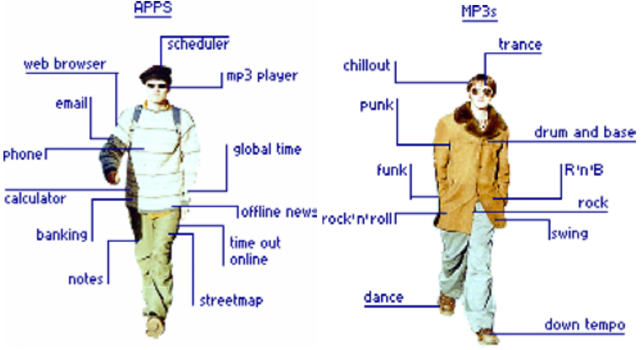


Fig. 2 Example body maps organising music archive according to genres and applications by function.

4. RESULTS AND FUTURE WORK

The interviews and questionnaires conducted suggest that the human body can provide a versatile associational space for meaningful organisation of a variety of forms of data. The range of different strategies reported in the study also supports our hypothesis that the body space is a very individual culturally defined construct, and thus can provide a highly personalised and meaningful interface. Further studies are needed to establish to what extent the associational process works in relation to personal experience and to what extent there are universal guidelines

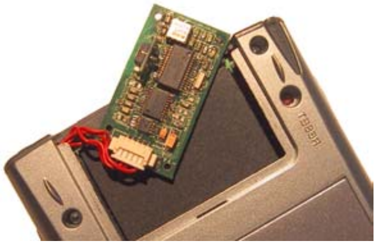


Fig. 3 Three axis inertial measurement unit developed to detect the motions of a HP Jordana PocketPC.

according to which data can be stored.

More importantly, however, we are currently working on building a functional prototype of the system using inertial sensing [5]. This is shown in Figure 3 and will enable us to begin empirical evaluation of our design. Only through real life testing can we establish whether or not this novel application of body space perception provides real world usability benefits

5. ACKNOWLEDGMENTS

We would like to thank the Interaction Design department at the Royal College of Art for supporting the project during its early stages of development.

6. REFERENCES

[1] Biocca, F., A. Tang, and D. Lamas, *Evolution of the Mobile Infosphere: Iterative design of a high infomation-bandwidth, mobile augmented reality interface.* 2001.

[2] Yee, K.-P. *Peephole Displays: Pen interaction on spatially aware handheld computers*. in *CHI 2003*. 2003. Ft. Lauderdale, Florida: ACM.

[3] Goldstein, E.B., *Blackwell handbook of perception*. 2001, Malden, Mass.: Blackwell. xii, 788.

[4] Yates, F.A., *The art of memory*. 1966, London,: Routledge and K. Paul. xv, 400.

[5] Benbasat, A.Y., *An Inertial Measurement Unit for User Interfaces*. 2000, MIT.

Realising Physical Selection for Mobile Devices

Heikki Ailisto, Ilkka Korhonen*, Johan Plomp, Lauri Pohjanheimo, Esko Strömmer

VTT Electronics
P.O.Box 1100
FIN-90571 Oulu, Finland
+358 8 551 2111
Firstname.Surname@vtt.fi

* VTT Information Technology
P.O.Box 1206
FIN-33101 TAMPERE, Finland
+358 3 316 3352
ilkka.korhonen@vtt.fi

ABSTRACT

Physical selection offers a promising method for using mobile devices, such as smart phones and personal digital assistants, as tools for communication between a human and the digitally augmented objects and services in the environment. In this paper, we analyse the concept from the technological perspective, and focus on different technologies, which may be used to implement the physical selection paradigm: visual patterns, electromagnetic methods or infrared.

Categories and Subject Descriptors

H.5.2. [**Information Systems**]: User Interfaces – *Interaction styles.*

General Terms

Algorithms, Design, Experimentation, Human Factors.

Keywords

Physical selection, tangible user interface, mobile phone, natural interaction, RFID, IR, barcodes.

1. INTRODUCTION

Ubiquitous computing inherently includes natural interaction between humans and digital devices embedded in their environment. The desktop metaphor [8] works well in the office, but it is not so well suited to ubiquitous and mobile computing [11]. The limited size of the mobile devices restricts the display area and handheld devices do not support the use of mouse or other common ways of pointing. Also the use of QWERTY keyboards is limited by the size of the mobile devices.

The mobile devices should be able to communicate with the devices and services available locally. Since the location[1] varies, the environment is inherently dynamic. In this respect the situation is very different from an office computer, where the tools (services), e.g. word processing, spreadsheet calculation, and email, are fairly stable. For example using multi-level menus for selection in stable environment is not difficult after the user familiarises her/himself with the tools. In a dynamically changing i.e. mobile environment this is not the case. Therefore, all the means to facilitate usage should be employed. One of these means is tying the available services to their physical counterparts.

This paper was presented at "Physical Interaction (PI03) -Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors. Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

Ideas close to physical selection have been suggested [10,11,4]. Ulmer and Ishii [10] developed the idea of Phicons, which serve as physical icons for the containment, transport and manipulation of online media in an office environment. Their paper does not discuss the role of mobile personal devices, such as smart phones or Personal Digital Assistants (PDAs), but instead rely on fixed devices, such as digital whiteboards, projectors, and printers. Kinderberg and co-workers study infrastructures to support "web presence" for the real world [4], their main idea being connecting physical objects with corresponding web sites. Infrared (IR) beacons, electronic tags or barcodes are suggested for creating the connection. We estimate the application potential to be much larger than accessing web pages associated with physical places or objects.

In this paper the employment of widely used and increasingly popular mobile devices, such as smart phones and PDAs, as a tool for physical selection is suggested. In the physical selection paradigm the interaction between the personal mobile device and a target object or device in the real world is initiated by a physical operation, such as pointing or touching. The function is analogous to selection in the virtual world of a desktop, hence the name.

We briefly describe three examples of using physical selection, derive requirements based on use cases, and then focus on analysing the potential of different technologies which may be used to implement the concept. Finally, the potential of physical browsing as well as future direction of the research is discussed.

2. EXAMPLES OF USE CASES

Three potential use-cases of physical selection are presented. The cases are

Use Case #1: updating the context profile of a mobile phone. The context profile of a mobile phone should relate to the current situation defined largely by the location and the task at hand. The location specific context could be e.g. an office, meeting room, car or home. Changing or updating the context profile of a mobile phone could be done by pointing it at a Context Tag and accepting the new profile, which is downloaded from the tag or from a location specified by the tag. A natural place for Context Tags would be near beside door posts of rooms. In a similar way, task or situation context could be chosen by pointing at physical symbols of each named context with the mobile device.

Use Case #2: Activating a phone call to a person by pointing at her/his picture or a tag in a business card. This would ease the dialling process, which is also error prone especially when the user is moving or preoccupied by some other task. A similar case would be launching any application or function on a

---

[1] Instead of location, also situation or task can be the variable.

mobile device while pointing at a tag, e.g. starting a web-browser and downloading the web-page related to the current object.

Use Case #3: Using a mobile device as a universal remote control for objects, which do not have a complete UI of their own (e.g. home appliances). The control of many everyday devices – thermostats, videos, ovens, washing machines – may in the future be partly delegated to mobile devices. The controlled device could have a tag[2], and by selecting the tag the user would launch a control UI on his/her activating device. This UI can provide significantly more freedom in personalisation and adaptation than any built-in UI can realistically do.

3. REQUIREMENTS FOR PHYSICAL SELECTION

There are many issues related to the implementation of the physical selection paradigm. These include:

1. Physical selection may be based on proximity or pointing. In the case of proximity, the selection is activated by bringing the activating device, e.g. a smart phone, close to the target device. Respectively, in the case of pointing, the activating device is aimed at the target device. In both cases, the maximum distance of activation may vary, but for proximity type of selection, it would be natural to assume "almost touching" as the prevalent case where as in the case of pointing, a maximum distance of up to a few metres seems natural. In the case of pointing, sensitivity to aiming errors and feedback of the aiming direction, e.g., with a visible laser beam, may be important for usability.
2. The key information transfer characteristics between the activating device and the target object include unidirectional or bidirectional data transfer, maximum data rate, maximum communication distance, which may be different from the maximum distance of activation, and latency in awakening the communication. It should be noted, that the means used for activating the communication channel may be different from the means of communication.
3. The information storage and processing capacity defines to a great extent the capabilities of the target device and thus the potential use of it. The target device may have fixed or dynamic information content, and the amount of information, measured in bits or characters, may vary from one bit to large text files, maps or even program files. The target object may be just an information storage, or it may have processing capability or even "smartness". One further characteristic is the stand-alone or front-end-of-a-system nature of the target object. Typically, a tag of the business card in use case #2 could be a stand-alone target device, while the use case #3 would require a target device with an application interface to the system to be controlled by the UI.
4. The manufacturing cost of the tags is an essential factor as the potential objects to be digitally augmented are numerous i.e. not only traditional digital devices but also other devices, printed commercials, consumer goods, places, things, etc. If the paradigm is aimed to cover the whole range of possibilities, the production cost needs to be rather in the order of cents than tens of cents.
5. The power economy of the tags is another essential feature related to the issues mentioned above at point 4: in the scenario of the world equipped with millions of tags, the maintenance and installation costs easily become a bottleneck. Hence, attention should be paid to minimise the need for battery recharge or change, and preferably other (ambient) power sources should be used.

Other important factors relevant especially for applications in the near future include compliance with standards such as those for RFID (Radio-Frequency Identification) and IrDA, compatibility with existing or future infrastructure, and prevalence and universality of pointing devices.

In the following, we aim to analyse potential implementation alternatives of physical selection in terms of the issues mentioned above and in the light of the three use cases.

4. IMPLEMENTATION ALTERNATIVES

The three main alternatives for implementing physical selection are visual codes, infrared communication and electro-magnetic methods. Wired communication methods are left out, since they require clearly more actions from the user than the physical selection paradigm implies.

4.1 Visual codes

The common barcode is the best known visual code. Barcode is a one-dimensional code consisting of vertical stripes and gaps, which can be read by optical laser scanners or digital cameras. Another type of visual code is a two-dimensional matrix code, typically square shaped and containing a matrix of pixels [7]. Optical Character Recognition (OCR) code consists basically of characters, which can be read by humans and machines.

The introduction of mobile devices with embedded digital cameras has made visual codes a feasible solution for physical selection. A code can be read with the camera and analysed by image recognition software.

Visual tags are naturally suitable for unidirectional communication only, as they are usually printed on a paper or other surface and the data in them can not be changed afterwards [5]. When printed on paper or adhesive tape, the tag is very thin, and it can be attached almost anywhere. The most significant differences between barcode, matrix code and OCR lay in the information density of the tag and the processing power needed to perform the image recognition. Barcodes have typically less than 20 digits or characters, while matrix tags can contain a few hundred characters. The data content of an OCR is limited by the resolution of the reading device (camera) and the available processing power needed for analysing the code. Visual codes do not have any processing capability and they do not contain active components, thus their lifetime is very long and they are inexpensive. The reading distance ranges from contact to around 20 centimetres with hand held readers and it can be up to several meters in the case of a digital camera, depending on the size of

---

[2] A tag should in this control application support bidirectional communications and also allow control of the device which it is attached to. This may be reached either by use of some advanced technology for tagging (e.g. IrDA) or by a combination of a tag (e.g. RFID) and some other communications mechanism (e.g. Bluetooth). In the latter case the tag would contain the necessary communication parameters to launch the communications in the actual communication channel (BT).

the code and resolution of the camera. By nature, visual codes are closer to the pointing class than the proximity type of selection.

Barcodes are widely used for labelling physical objects everywhere. There are already a myriad of barcode readers, even toys, on the market. Commercial image recognition software is also available.

4.2 Electromagnetic technologies

RFID systems incorporate small modules called *tags* that communicate with a compatible module called a *reader* [3]. The communication is usually based on a magnetic field generated by the reader (inductive coupling), but with very short operating ranges it is also possible to apply capacitive coupling. Operating ranges up to several meters can be achieved by long range RFID tags based on UHF (ultra high frequency) technologies [2]. The tags are typically *passive*, which means that they receive the energy needed for the operation from the electromagnetic field generated by the reader module, eliminating the need for a separate power supply. In addition, there are *active* RFID tags that incorporate a separate power supply for increasing the operating range or data processing capability. RFID technology can be applied for physical selection by integrating a tag in the ambient device and a reader in the mobile device or vice versa.

Typical tags based on inductive coupling incorporate an antenna and one IC (Integrated Circuit) chip providing data transfer, storage and possibly also processing capability. Usually the data transfer is unidirectional from the tag to the reader, but also bidirectional tags exist. The operating range is typically from a few millimetres to several tens of centimetres depending on the antenna, operating frequency, modulation method, operating power and bit rate. Examples of operating frequencies typically used are 125 kHz and 13.56 MHz. Originally the RFID tags were aimed at the electrical labelling of physical objects, replacing visual barcodes. Currently, the RFID technology has established itself in a wide range of applications, e.g. automated vehicle identification, smart cards, access systems and toys. There are several manufacturers providing RFID ICs, tags and systems. The basic advantages of the inductive RFID technology compared to other electromagnetic technologies are low price, small size, operation without a power supply and good commercial availability. These advantages make the inductive RFID technology very attractive from the viewpoint of physical selection applications based on the proximity concept.

In addition to the RFID technologies, there are some technologies based on magnetic induction and particularly aimed for short-range communication. In general, compared to RF (Radio Frequency) based technologies, magnetic induction has some advantages in short-range (below 3 m) wireless communication such as power consumption, interference and security [1]. There are also some commercial components available which are applicable in physical selection applications.

Longer operating ranges than by magnetic induction can be achieved by UHF-based technologies such as Bluetooth, other wireless personal area network (WPAN) technologies and long-range RFID technologies. The operating range of these technologies is typically several meters, which is too long for most of the physical selection applications. However, it is possible e.g. to reduce the operating range by external shielding or to use the received signal strength indication (RSSI) if available. Examples of the operating frequencies of WPANs and long-range RFID tags are 868 MHz, 915 MHz or 2.45 GHz. One possible disadvantage of Bluetooth, concerning especially ambient devices, is the high power consumption. However, the backscattering technology used in the long-range RFID tags enable an operating range up to several meters even without any external power source. Components and modules are available from several manufacturers.

4.3 Infrared technologies

Infrared (IR) is widely used in local data transfer applications such as remote control of home appliances and communication between more sophisticated devices, such as laptops and mobile phones. In the latter case, the IrDA standard is widely accepted and it has a high penetration in PC, mobile phone and PDA environments. Due to the spatial resolution inherent to the IR technology, IR is a potential technology for implementing physical selection applications based on the pointing concept.

An IR tag capable of communicating with a compatible reader module in the mobile device would consist of a power source, an IR transceiver and a microcontroller. The size of the tag depends on the implementation and intended use, but the smallest tags could easily be attached practically anywhere. The data transfer can be unidirectional or bidirectional. The operation range can be several meters, but a free line-of-sight (LOS) is required between the mobile device and the ambient device. In the IrDA standard, the specified maximum data rate is 16 Mbit/s and the guaranteed operating range varies from 0.2 to 5 meters, depending on the used version. One possible problem of IrDA, concerning especially the ambient device, is its high power consumption. For reducing the mean power consumption and thus extending the lifetime of the battery, if used, the IR tags can be woken up by the signal from the reader module [6,9]. It is also possible that the tag wakes up periodically for sending its identification signal to the mobile device in its operating range.

In general, IR technologies are very commonplace. Many home appliances can be controlled by their IR remote controller. Several mobile phones and laptops incorporate an IrDA port, and with suitable software they could act as tag readers. Components and modules are also available from several manufacturers.

4.4 Comparison of the technologies

The three most potential commercial technologies for implementing physical selection are compared in Table 1.

Bluetooth is included for reference since it is the best known local wireless communication technology. Obviously, exact and unambiguous values are impossible to give for many characteristics and this is why qualitative descriptions are used instead of numbers. When a cell in the table has two entries, the more typical, standard or existing one is without parenthesis, and the less typical, non-standard or emerging one is in parenthesis.

In the use case #1 *Updating the context profile of a mobile* device tags are used in a variety of places, usually without easy access to a power supply. To create sufficient infrastructure, a large amount of tags is needed. This suggests that the optimal technical solutions are based on visual codes or RFID tags although the use of infrared tags is also possible.

All suggested technologies apply to the *use case #2*. Several sub-cases of this use-case seem to be easier to use from a distance and that makes visual codes or infrared as a pointing based technology more suitable than electro-magnetic methods. When the premium is on the cost, barcodes seem to be the optimal solution.

The *UI for devices and services without display and keys* use-case #3 is the most demanding of the three cases presented. Bidirectional communication, and a demand for data processing capabilities on the tag side rule out the visual code option. Of the two remaining alternatives, infrared seems to be more compelling because of the standardised bidirectional communication and the ability of the tag to act as a front-end for the device in question.

Table 1. Comparison of potential commercial technologies for physical selection (Bluetooth included as a reference).

	Visual code	IrDA	RFID, inductive	Bluetooth
Selection concept	Proximity/ pointing	pointing	proximity	none
Data transfer type	unidirectional	Bidirectional	unidirectional (bidirect.)	bidirect.
Data rate	medium	high	medium	high
Latency	very short	medium	short	long
Operating range	short-long	medium (long)	short (medium)	medium (long)
Data storage type	fixed	dynamic	fixed (dynamic)	dynamic
Data storage capacity	limited	not limited	limited (not limited)	not limited
Data processing	none	yes	yes, limited	yes
Unit costs	very small	medium	low	medium-high
Power consumption	no	medium	no (low)	medium-high
Interference hazard	no	medium	low-medium	medium-high
Support in PDAs or m-phones	some (camera phones)	yes	no (future phones may have)	some (high-end m-phones)

5. DISCUSSION

Physical selection is a potential paradigm for human computer interaction in the ubiquitous computing domain. After analysing three potential use cases, some important issues related to the requirements of implementing physical selection could be identified. These are the principal way of selection - proximity or pointing; information transfer characteristics - unidirectional vs. bidirectional, data rate and latency; information storage and processing capacity; manufacturing costs and power economy. Furthermore, conformity with standards and existing infra structure are of importance.

Three implementation methods, namely visual codes, electro-magnetic means and infrared technology offer suitable characteristics for different applications. For example, visual codes are best suited for cases where cost critical unidirectional pointing type selection is needed, whereas RFID tags are best suited for unidirectional proximity based use cases. Infrared lends itself naturally for pointing based bidirectional control applications.

The physical selection paradigm seems well suited for cases where the user is on the move and uses a mobile device, such as a PDA or a smart mobile phone, for interacting with the digitally augmented environment. The vast and ever growing number of smart mobile devices with local communication capabilities, such as IrDA, Bluetooth, cameras for visual code reading, and in the future also RFID based techniques, offers a technical basis for this new paradigm. The simultaneous proliferation of low-cost tags makes the paradigm even more tempting.

We will continue our research on issues like the implementation of physical selection (IrDA and RFID based), usability, and identification of applications benefiting from this paradigm.

6. REFERENCES

[1] Bunszel, C. Magnetic induction: a low-power wireless alternative, R.F. Design 24, 11 (Nov 2001), 78-80.

[2] Extending RFID's Reach in Europe. RFID Journal March 10, 2002.http://www.rfidjournal.com/article/articleview/328/1/1/

[3] Finkenzeller, K. RFID Handbook, Radio-Frequency Identification Fundamentals and Applications. John Wiley & Son Ltd, England, 1999.

[4] Kinderberg, T., et. al. People, Places, Things: Web Presence for Real World, in IEEE Workshop on Mobile Computing Systems and Applications WMCSA'00 (Monterey CA, December 2000), IEEE Press, 19-28.

[5] Ljungstrand, P. and Holmquist, L.E. WebStickers: Using Physical Objects as WWW Bookmarks. Extended Abstracts of ACM Computer-Human Interaction (CHI) '99, ACM Press, 1999.

[6] Ma, H. and Paradiso, J. A. The FindIT Flashlight: Responsive Tagging Based on Optically Triggered Microprocessor Wakeup, in UbiComp 2002, LNCS 2498, 160-167.

[7] Plain-Jones, C. Data Matrix Identification, Sensor Review 15, 1 (1995), 12-15.

[8] Smith, D. Designing the Star User interface. Byte April 1982, 242-282.

[9] Strömmer, E., and Suojanen, M. Micropower IR Tag - A New Technology for Ad-Hoc Interconnections between Hand-Held Terminals and Smart Objects, in Smart Objects Conference sOc'2003 (Grenoble, France, May 20 03).

[10] Ullmer, B., Ishii, H. and Glas, D. mediaBlocks: Physical Containers, Transports, and Controls for Online Media, in Proceedings of SIGGRAPH '98 (Orlando, FL, July 1998), ACM Press, 379-386.

[11] Want, R., Weiser, M. and Mynatt, E. Activating everyday objects in Darpa/NIST Smart Spaces Workshop (Gaithersburg, MA, July 1998), USC Information Sciences Institute, 140-143.

Configurability in and Integration with the Environment: Diverse Physical Interfaces for Architecture Design

Giulio Iacucci
Department of Information Processing Science
University of Oulu
PO Box 3000 90014 Oulu, Finland
Tel: + 358 8 553 1900
giulio.iacucci@oulu.fi

Kresimir Matkovic, Thomas Psik, Ina Wagner
Institute of Design and Assessment of Technology
Vienna University of Technology
Argentinierstraße 8/187
1040 Vienna, Austria
Tel: +43 1 5880118702
{matkovic, tpsik, iwagner}@pop.tuwien.ac.at

ABSTRACT

We present several physical interfaces and how we assembled them to develop an environment for learning architecture design. Among others we are developing applications with barcode scanners, touch sensors, RFID tags, infrared remote control, video tracking, GPS receivers, and sensor boxes with electronic compass and acceleration sensors. In the environment input and output components are connected through an infrastructure. In a first round of experiments we have co-developed the components with the students for their practical design projects. We reflect on which features of the diverse prototype we developed contribute to understand configurability and integration with the environment.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems – *Human factors.*

General Terms

Design, Human Factors.

Keywords

Physical interfaces, configurability, integration, field study.

1. INTRODUCTION

Despite the advances and the large interest in physical interaction technologies, field studies of everyday use are rare. On one hand technology is presented and there are approaches to implement it, on the other hand serious ethnographic studies of current work stress the importance of the physical environment and the limits of desktops centric development. The contribution of our paper is to be placed in the scarcely populated area in between where field studies inform the development of prototypes that are experimented in everyday settings. In our research we are prototyping systems that have the ambition of freeing people from the desktop computer (or avoiding constraining them in front of it). However our primary interest is not the technology itself but to design for a particular community of practice and learn from field studies how new technology can be integrated seamlessly in everyday practices.

The case we present is part of a design project to develop a ubiquitous and mixed media environment for inspirational learning. One of the application sites is the architecture department at the Academy of fine Arts in Vienna. After careful observations of student projects at the Academy, we have prototyped various applications to be experimented by the students. We have experimented with several technologies to support physical interaction: among others barcode scanners, touch sensors, RFID tags, infrared, remote control, video tracking, GPS receivers, and a sensor box with electronic compass and acceleration sensors. In particular we describe: physical interfaces to digital media, the texture painter, the tangible image query, a mobile application to record path and organize media from visits.

Which features of these prototypes contribute to configurability? Which features contribute to explain their different integration in the user's environment and activities?

2. THE SETTING

We have observed seven student projects in the first half of 2002. After this period we introduced and observed the use of several tangible computing prototypes (see [2] for a detailed report on the field study). We will now describe the setting and the student projects. In the projects, groups of students have to work out designs of interventions for remote physical locations. During the project they concretise solutions experimenting with several representation techniques. Visit at sites are frequent and there are weekly feedback meetings with staff and external reviewers. The goal of the students is to be creative in getting ideas and develop them into a convincing solution.

The diversity of material and media is an important characteristic that is exploited in the handicrafts they produce. Students work with and produce texts, diagrams, comics, videos, sketches, models, screenshots, virtual models, and prototypes – material of different degrees of abstraction, different scale and materiality.

We used participant observation to study current practices and the use of prototypes. We set out to observe the students not

This paper was presented at "Physical Interaction (PI03) - Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors.
Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

only in topical events as presentations or meetings but also in everyday work. Inspired by interaction analysis we used a digital still camera and a video camera to record audiovisual material to analyse selected episodes.

3. TECHNOLOGICAL COMPONENTS OF THE ENVIRONMENT

3.1 Infrastructure, Configuration and Database

The environment we are developing is composed by a variety of interaction components (physical inputs, media playing and projecting applications, described in paragraphs 3.2 and 3.3), a configuration component, a hypermedia database, and an Infrastructure to connect all components. The infrastructure provides registration for all components and messaging between them. The configuration component is answering the need to have a consistent and transparent management of associations between input and output components. As the same input might be used by different people (or the same person) to trigger different actions, there is a need to store the configurations in a central place from where they can be loaded at anytime from any computer. The hypermedia database (HMDB) provides a shared database of digital media and meta-information. The multimedia objects can be grouped and organized hierarchically and can be linked using region and time based anchors.

3.2 Physical Interfaces to Digital Media

3.2.1 Animating bAR code

This interaction technology provides a way to link physical objects and digital material. Media files can be associated to physical barcodes in the environment. A barcode scanner is attached to a PC through the keyboard wedge. The input component running on the PC is able to capture one or multiple barcodes and send them in a message to other components in the environment. This technology has been used to animate physical models and diagrams with digital media. This is achieved by sticking barcodes on the models or diagrams to which users previously have associated media files as videos, sounds or pictures. Users can then scan the barcodes on the models during presentations or discussions and trigger the playing of media files. Users can associate to a media file a defined series of barcodes, so that the scanning of two physical objects in a different sequence or the scanning of the two objects different times can play different media.

3.2.2 Infrared Remote Control

This component is operated from a mobile devices giving the possibility to control application from anywhere in a room. The infrared remote control is used to send to other components messages as “Play”, “Stop”, “Forward”, “Up”, “Down”, “Right" or numbers. An infrared receiver is attached to a PC where an input component is running that receives the signals and sends them to components in a message. This component has been used to navigate through multimedia material. Users are able to navigate through linked media objects. With the forward button users will skip to next anchors and by pressing play the link of the current anchor will be used to open the linked media object.

3.2.3 RFID Tags

The RFID Tag/Tag reader combination is a component, which allows students to tag objects in the studio. This component consists of numerous tags, which can be attached to the objects, and several tag readers placed around the working environment. Whenever a tag is placed on one of the tag readers a specific action is triggered. This technology seems very suitable for selecting among various choices, e.g. students having small tagged objects with different video files associated, and can select a file by simple placing the appropriate object on a tag reader.

3.2.4 Touch Sensors

Touch sensors we are using are small sensors based on “qprox” [4] technology. The sensors are actually small copper plates which have to e covered with an insulating material. This makes it possible to integrate the sensors in students’ models. Sensors remain invisible, but they react if someone touches the model at the certain place. In this way touch sensors offer an alternative to barcodes or tags. They support physical – embodied - interactions with the artefact into which they have been integrated for retrieving and displaying media files. They proved effective in presentations since they introduce an element of surprise. Their invisibility, which is the main advantage, is a disadvantage at the same time. Namely, student has to now where the sensors are in order to activate them. It is possible to mark the sensor positions on the surface, but it is not always convenient from the students’ point of view. On the other hand, an unknown model (which is supposed to be equipped with sensors) invites the users to play with it, and explore it. There is a wireless communication between touch sensors and central computer. In this way the models equipped with sensors do not differ from conventional models from outside at all.

3.2.5 The control cube

The control cube emerged from the touch sensors idea. We put the sensors in a cube, added 6 tilt sensors, and made it possible to recognize which side of the cube is facing up. Such a device can be used for selection between six choices. The student simple turns the cube, and the side facing up determines which action will be triggered. Touch sensors integrated in the cube can be used to browse through a set associated with a cube side. E. g. if there are six collections of images, the user can select collection by turning the cube, and then, once a collection is chosen, the user can navigate through the collection using two touch sensors, one for “next” and one for “previous” image.

3.2.6 Different physical inputs

Barcodes are particularly suitable for animating physical models and diagrams with digital media, as barcodes can be attached in the environment. However barcode scanners have a limited range because they are attached through cable to the PC. Similarly RFID tags can be easily attached to objects, whereas users generally move the tags on the reader (instead users move the scanner to the barcode). Note that this is actually inverse process, if we are using barcodes we are walking around holding the barcode scanner in the hand and shooting the barcodes. In the RFID approach, tag readers (equivalent to barcode scanner) are more or les fixed, and we have to move objects. The third possibility, touch sensors, are similar to a bar code in a way that we do not have to move the object, but we do not have to carry the bar code reader around neither. Actually although all those technologies offer similar functionality they have significantly different qualities and each of them is used in different situations.

The infrared remote control is completely different device, it is not directly bound to the models, although it can trigger actions which will enhance the environment.

3.3 The Texture Painter

Using a brush, which is tracked, this application allows 'painting' on objects such as models or parts of the physical space, applying textures, images or video, scaling and rotating them (Figure 2). Students started animating their models with the help of the Texture Painter. One student studied soccer games to identify the most exciting camera views and to understand which kind of atmosphere the players need. He used the camera views to find out where to place few spectators so that the stadium looks jammed. He built a simple model of a stadium and used model and images together with the Texture Painter for projecting different atmospheres into this 'fragmented stadium'. Another student painted images of his interventions into projected images of two residential buildings, projecting detailed plans into the space between them.

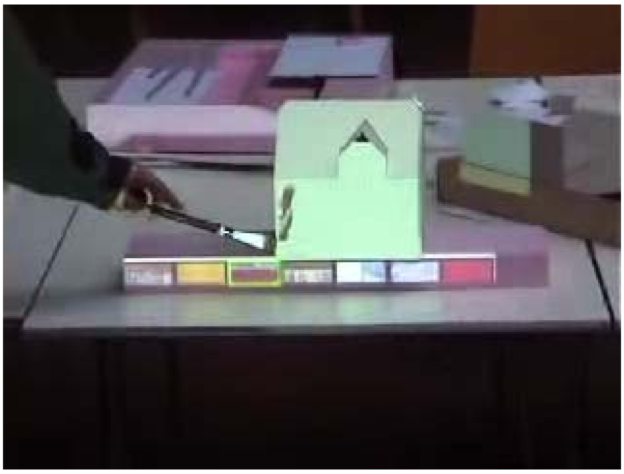


Figure 2. Painting objects electronically using a physical brush

3.4 Tangible image query

This is a physical interface for browsing the HMDB in an interactive way. It consists of a web camera integrated into a small table. A user may use small coloured cubes to specify a colour layout that is used to search in the HMDB for similar images. The search is based on the Visual Image Query method described in [5]. It is also possible to use colour on a simple sheet of paper or any set of reasonably sized coloured objects for creating colour patterns. This resonates with observed practice – architects often using material that is at hand for illustrating ideas and qualities, such as density, fragility, opaqueness, etc. Since the definition of the colour layout is done in a rather rough way, the results of the search are a source of surprise and inspiration for the user.

3.5 Sensors recording walking path and directions

The e-Path for iPAQ Pocket PC is an application to support visitors in organizing media material created during visits. With the e-Path application users can log the path of a visit through GPS, log position and direction for recorded media like photographs and sounds, and create a "hyper document" of the visit that can be used to store the media files and information of the visit in the HMDB.

Figure 3. Mobile unit to record paths and directions

The e-Path can currently make use of external sensors (optional) like a CompactFlashGPS reader card and the VTT SoapBox (Figure 3). VTT Electronics has developed a general-purpose SoapBox module (Sensing, Operating and Activating Peripheral Box, [3]) that is a light, matchbox-size device with a processor, a set of sensors, and wireless and wired data communications.

4. DISCUSSION

The diversity of tangible computing application that we developed brought us to consider how they can relate differently to the environment and to user practices. To further detail these relations we analyse how the prototypes can be differently integrated in the environment and how configurability is supported. This analysis provided us with four characterizing features of the diversity of the prototypes.

Integration with artefacts and environment. The physical interfaces vary in the way they are integrated in the environment and on artefacts. *Invisible* – as touch sensors were inside the models. *Virtually connected* – as in the Texture Painter application that is projecting images on objects. *Aesthetically or functionally integrated.* In some cases the barcode was part of the diagram that student created in an aesthetical and functional role. *Detached from the environment*. Even though the Infrared Remote Control is a physical interface it is detached from the environment.

Distributed Intelligence. What kind of intelligence is behind the application and where it resides? The answer to this question contribute to explain how far computing has been distributed away from the desktop computer, and to which extend it is embedded in the environment. *The PC backstage*. Although hidden in the interaction the PC is present in different ways in almost all of the applications (texture painter, tangible image query). *Embedded in objects*. The Control Cube contains sensors and a radio sender to communicate signals to another computing unit. *Mobile devices*. The e-Path runs on a Pocket PC, other examples of devices are the Infrared Remote Control, the RFID tag reader.

Component Structure. **T**he applications can be composed by more or less independent components or be more like an organic stand alone application. Analysing this aspect is important when developing configurable and tailorable solutions. *Component Structure:* As in our case barcode scanner, tag reader, touch sensors infrared remote control can be configured to control the same applications. *Stand Alone:* As in the Texture Painter application.

Bodily Actions. The bodily interaction with the applications was differently characterized. *Reader to Object – Object to Reader* we experimented with two types of readers. With the RF-ID reader the actor moves a tagged object on a fixed "reader", with the barcode scanner the actor moves the reader to the tag. While both technologies can be made worked either way, the two approaches are conceptually different. *Sensed – Performing Body*. In the case of the e-Path the actor performs a path and points to directions all of which is recorded. In the Texture Painter the actor performs movements with the brush that are sensed by a video camera or infrared tracking. *Remote Control* – the Infrared Remote Control and the Control Box are examples of a mobile control device that can be operated moving in the environment.

We have observed in the field study how the physical environment is in constant reconfigurations. Configurability is not only supported by producing software in component structure and providing platform for configuration (e.g. the infrastructure and configuration component). As the applications have physical interfaces the physical realm needs to be considered as well. While we have not precise guidelines for digital-physical configurability with our analysis we suggest at considering as features to be considered how computing is distributed in the environment (embeddedness) and to which extend the physical interfaces are integrated in the environment and physical artefacts. Finally the diverse bodily interaction with the prototype is a characterizing feature to understand how physical interfaces become part of a a social environment and of people activities.

5. ACKNOWLEDGMENTS

The Atelier project is funded by the EU IST Disappearing Computer program. We are grateful to our partners Malmö Högskola Creative Environments K3, Akademie der Bildenden Künste Wien Institut für Kunst und Architektur, Imagination Computer Services GesmbH, Consorzio Milano Ricerche, Dipartimento Di Informatica Sistemistica e Communicazione, Interactive Institute, Space & Virtuality Studio. Finally we would like to acknowledge Infotech Oulu for supporting this research at the University of Oulu.

6. REFERENCES

[1] Dourish, P. Where the action is: the foundations of embodied interaction. MIT Press, 2001.

[2] Iacucci, G., Wagner, I., Supporting Collaboration Ubiquitously: An augmented learning environment for architecture students, In: the Proceedings of the 8th European Conference of Computer-supported Cooperative Work, 14.-18. September 2003, Helsinki, Finland, in press.

[3] Tuulari, Esa; Ylisaukko-oja, Arto, SoapBox: A Platform for Ubiquitous Computing Research and Applications. Lecture Notes in Computer Sc. 2414: Pervasive Computing. Zürich, CH, August 26-28, 2002. Mattern, F. Naghshineh,M. (eds.). Springer (2002), pp. 125 – 138.

[4] K. Matković, L. Neumann, J. Siegler, M. Kompast, W. Purgathofer, „Visual image Query", In the Proceedings of Smart Graphics 2002, Hawthorn, NY, USA

[5] Quantum Research Group, http://www.qprox.com

A Framework for Tangible User Interfaces

Boriana Koleva, Steve Benford, Kher Hui Ng, Tom Rodden
The University of Nottingham
School of Computer Science & IT
Nottingham, NG8 1BB, U.K.
+44 115 8466530

{bnk, sdb, khn, tar}@cs.nott.ac.uk

ABSTRACT

This paper extends our understanding of tangible user interfaces (TUIs) by considering the different ways in which physical and digital objects can be computationally coupled. It proposes a framework based around the degree of coherence between physical and digital objects. Links between physical and digital objects are described in terms of a set of underlying properties (transformation, sensing, configurability, lifetime, autonomy, cardinality and link source). We use our framework to classify a representative selection of existing TUI systems. This classification raises key implications for the field of tangible computing. In particular our focus on enriching physical-digital links highlights the need to consider the asymmetry of these links, issues surrounding their configuration and the need to represent their nature to developers and users.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *theory and methods.*

General Terms

Design, Theory

Keywords

Tangible user interfaces, design framework, interaction models

1. INTRODUCTION

Tangible computing [11] allows users to interact directly with computational artifacts by manipulating everyday physical objects rather than using traditional graphical interfaces and dedicated physical interface devices such as mice and keyboards. A variety of systems have been developed to date that illustrate the tangible interface principle. Some notable examples include:

- The TangibleGeospace application of the metaDesk [22], where physical representations of geographical features are used to manipulate a digital map;
- Illuminating Light [24], where physical models of optical elements are used to create a simulated optical layout;
- The Passage system [16], which provides a mechanism for transporting digital information by linking it to physical objects;
- WebStickers [10], where everyday physical object act as bookmarks for web pages;
- The tangible tools (tongs, eraser and magnet) provided by the Surface Drawing system [21];
- Illuminating Clay [20] where users interact directly with a clay model of a landscape and observe the effects of geometric changes.
- Storytent [8], where physical balloons are used as identifiers for virtual balloon objects and flashlights are used as pointing devices for manipulating them.

Just as concrete examples of tangible user interfaces (TUIs) are proliferating, so conceptual frameworks are also emerging to help researchers, designers and developers classify what constitutes a TUI and to understand the various ways in which physical objects can be combined with digital information. Like the MVC interaction model for GUIs [4] and the PAC interaction model for dialog design [5], these frameworks seek to highlight the main components of TUIs.

Ullmer and Ishii have proposed the model-control-representation (physical and digital) (MCRpd) interaction model for tangible interfaces [23], which highlights the integration of physical representation and control with this type of interface. Holmquist et al have suggested a broader taxonomy of how physical and digital objects can be coupled [10]. They propose the categories of *containers* as generic objects for moving information between devices or platforms, *tokens* as objects for accessing stored information (the nature of which is physically reflected in the token) and *tools* as object for manipulating digital information.

This paper aims to further extend our understanding of the different ways in which physical and digital objects can be computationally coupled. It introduces a framework that is based around the idea of the "degree of coherence" between physical and digital objects. This is further broken down into the concept of *links* between physical and digital objects that are described in terms of a set of underlying *properties*. We use our framework to classify a representative selection of TUI systems. In turn, this classification raises key implications for the field of tangible computing.

This paper was presented at "Physical Interaction (PI03) - Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors.
Further information and online proceedings are available at
http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

2. DEGREE OF COHERENCE

The first concept that we introduce as a means of distinguishing between different types of tangible UIs is “degree of coherence”. It is proposed that relationships between physical and digital objects can be rated along a coherence continuum, where the level of coherence represents the extent to which linked physical and digital objects might be perceived as being the same thing. That is whether the physical and the digital artifact are seen as one common object that exists in both the physical and the digital domain or whether they are seen as separate but temporarily interlinked objects. Figure 1 shows the coherence continuum along with some proposed categories of TUI types.

Interface objects that establish the weakest level of coherence with the computational artefacts they operate are termed “general purpose tools”. Using such a tool a user can select to manipulate any one of many digital objects and perform different transformations (depending on the application). Examples include traditional physical interface devices such as mice and joysticks.

The next category along the coherence continuum, named “specialised tools”, encompasses interface objects that have a more specialised function but still temporarily connect to potentially many different digital objects. Examples include the tongs, eraser and magnet from the Surface Drawing system [21] and the optical instruments, such as mirrors, beam-splitters, lenses etc., from the Illuminating Light system [24].

The “identifier” category represents interface objects which act as bookmarks for retrieving computational artefacts. The passenger objects in the Passage system [16] and the bar-coded everyday objects in the WebSticker system [10] are examples of this category. Here the physical object is a token representing a digital artefact and the two are often more permanently coupled.

Interface objects that belong to the “proxy” category are even more coherent with the digital objects they are coupled to. This is because proxies are more permanently associated with, and allow a more extensive manipulation of, their digital counterpart (more than identification for subsequent retrieval). Examples include the physical building models in the Geospace application of the metaDesk [22] and the pucks on the Senseboard as used to represent conference papers [14].

The “projection” category encompasses relationships where the digital artefact is seen as a direct representation of some properties of the physical object and so its existence is dependent on the physical object. An example is the representation of human activity in a physical foyer as a digital pattern projected on the wall of the ambientRoom [12].

Finally we can create the illusion that two coupled objects are one and the same if they are visible only one at a time, making smooth transitions between the physical and the digital space. For example a physical object may pass though a traversable interface [15] and appear as a virtual object on the other side of the display (in the virtual space) or the spaces can be superimposed in such a way that all actions appear coherent (e.g. highly registered augmented reality).

Unpacking this idea of coherence a bit further, we propose that what distinguishes the different categories are underlying differences in what interactions are sensed, the type of effects mediated between the coupled objects, the duration of the coupling, autonomy of the digital artefact and configurability of the coupling. The following section develops this idea further, proposing a detailed set of properties that we can use to describe and design different types of links.

3. COHERENCE AND LINK PROPERTIES

3.1 Transformation

This property describes whether the effect mediated between linked objects is literal or transformed. If actions are mediated literally, movement of the physical object for example, will result in the same movement of the digital object. This is the case with the phicons and virtual building models in Tangible Geospace [22] and the manipulation of the CUBIK interface [17]. Here the shape of a physical cube is altered by pushing and pulling its sides and these manipulations are directly mediated to a virtual cube whose shape changes in a corresponding manner.

On the other hand, the effect between linked objects can be transformed. For example positioning a physical object on a predefined place may trigger an animation of a digital object and/or for the digital object to emit sounds. Another example is the magnet tool in the Surface Drawing system [21], which changes the meaning of the drawing action to that of altering existing geometry. Waving the magnet near the region of a drawing pulls that region closer to the magnet.

3.2 Sensing of Interaction

This property describes what interactions with the interface object and its surrounding environment are sensed and transmitted to the destination object. This can range from detecting and responding to the presence of the source object in a specified area [16] to mediating manipulations in the full 6 degrees of freedom. E.g. translations and rotations in a plane are sensed for the metaDesk phicons [22] and for the CUBIC interface [17] scaling in the X, Y and Z axes are transmitted. An example of sensing actions in the surrounding environment is using video processing to detect gestures such as pointing at the physical objects, e.g. the DigitalDesk [26].

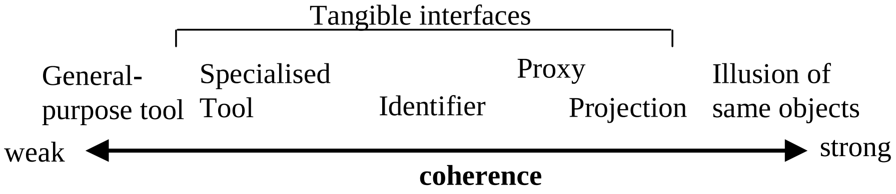


Figure 1: TUI categories along the coherence continuum

3.3 Configurability of Transformation

This property describes whether the transformation mediated between two linked objects remains fixed for the lifetime of the link or whether it is configurable over time. For example in the Illuminating Light system [24] each tangible object has a fixed transformation associated with it. E.g. a representation of an optical-grade mirror always has the effect of reflecting the virtual beams of light. On the other hand the pen for interacting with the Toshiba tablet PC changes its effect from a left to a right click on a digital objects when its button is pushed down.

3.4 Lifetime of link

This property describes for how long a physical and a digital object remain linked. A physical object may be consecutively linked to different digital artefacts in the lifetime of the application. For example a flashlight in the Storytent [8] can be used to select different balloon objects. Alternatively, the physical and digital object may remain linked for the lifetime of the application. This is exemplified in the Tangible Geospace [22] where the phicons are permanently bound to their digital counterparts. Finally, a link may retain its nature across many applications, potentially even permanently.

3.5 Autonomy

This property describes to what extent the existence of the destination object is reliant upon the existence of the link and the source object. For example, a digital object may be created only as a result of its link to a physical object. This is the case with the balloon objects in the Storytent – a virtual balloon is created whenever a physical balloon is brought into the tent and then it is deleted when that physical balloon leaves the tent. The destination object may also be a representation/projection of some of the characteristics of the source object, e.g. the digital pattern projected on the wall of the ambientRoom [12] reflects human activity in a physical space. In such cases if the source object ceases to exist, the destination object would also disappear or become meaningless.

3.6 Cardinality of Link

This describes whether an object is linked to one or more objects. One-to-one relationships seem to be most common. For example in the Tangible Geospace application on the metaDesk [22] each phicon, a small physical model of a particular building, was bound to the digital representation of that building. However it is also possible to link a physical object to multiple digital objects, e.g. Passage objects [16] could have been implemented so that a single physical "passenger" can identify a selection of digital documents (i.e. play the role of a folder). We can describe such a configuration by saying that a link has multiple destinations.

3.7 Link Source

So far we have only discussed cases where there is a physical interface object that mediates transformations to a digital object. However, it is also possible for digital objects to affect the state of the physical world. For example, haptic interfaces such as the PHANToM [18] provide a tangible feedback to the person manipulating digital objects and ambient displays such as Natalie Jeremijenko's Live Wire [25] provide tangible feedback of activities in digital space (Ethernet traffic in this case) through physical motion, sound and touch.

The link source property describes whether the source of the effect is the physical or the digital object.

4. REVIEWING CURRENT SYSTEMS

We now use our proposed link properties to classify current TUIs. First, we use the link source property to broadly divide systems into those where the source is a physical object and those where the source is a digital object. Thus Table 1 summarises the properties of the example TUI systems that have been discussed for far in which physical objects control digital ones. Table 2, on the other hand, introduces systems where the source is a digital object that has an effect in physical space. The examples in both tables are broadly listed in order of increasing coherence.

As object relationships with a cardinality of one to many are rare in current systems, this property has been omitted from the tables and all examples illustrate links where a single physical object is coupled to a single digital artefact.

5. IMPLICATIONS FOR TUIs

Our framework is based on unpacking the nature of the links between tangible and digital objects and using this to classify TUIs. This represents a shift in focus from many of the current perspectives. Not unsurprising much of the existing work on tangible interfaces has tended to focus on realising the physical artefacts associated with tangible interfaces. Current frameworks such as those proposed by Ullmer and Ishii [23] and Holmquist [11] while based on a connection between physical and digital tend to leave the nature of this connection as implicit with little reflection on the different ways in which this connection may be manifest. Understanding TUIs based on the richness of potential links between the digital and the physical provides us with a slightly different perspective on their nature and outlines a number of significant future research directions. In this section we briefly reflect on three initial examples by considering the asymmetry of the links, the configuration of the links involved and the need for users and developers to understand the nature of the link between the physical and the digital.

5.1 Tangibles that Push Back

Comparing tables 1 and 2 reveals a significant asymmetry in how the physical and digital are linked. While there are many examples of using physical objects to control digital objects, tangible interfaces that react to changes in digital information are relatively rare – there are few examples of tangible interfaces that "push back". It is therefore a challenge to develop techniques that will allow us to create digital artefacts that will push back on the physical space. These will be useful for providing tactile information, maintaining synchronisation between digital and physical objects and showing or monitoring digital activity through physical movement.

Examples of push back technologies from related fields include haptic devices for virtual reality such as the PHANToM [18], tangible interfaces for remote collaboration such as the PsyBench [3], and also ambient displays such as Pinwheels [6] and Table Fountain [7]. However, it remains a challenging problem as to how to push back through everyday objects such as blocks on a table or post-it notes on a board. Promising approaches include the use of airflow, waterflow, electromagnets and actuator arrays [13].

5.2 Mobility, Reconfiguration and TUIs

Many of the examples of tangible interfaces have tended to be based on stable arrangements between the physical and digital. TUIs such as the MetaDesk [22] have tended to be constructed as installations to be experienced by users as stand alone applications. However, TUIs have also become closely associated with Ubiquitous computing and examples such as the ambientROOM [12] outline the potential of TUIs and demonstrate how the digital may be physically manifest. However, less consideration has been given to the ubiquity of information and what happens when the physical element of TUI moves from one context into another. The main use of mobility of physical object has been to act as a token to access digital data.

How do the physical devices within the Ambient Room act when they are placed within a second room? Do their existing connections with the digital material in one ambient room persist into the second room offering remote availability or are new links established reflecting different digital effects? Considering the lifetime, autonomy and configurability within the potential links allow us to chart and understand this design space and consider how we may support the mobility of TUIs.

5.3 Understanding the effects of TUIs

Our turn to reasoning about the links between the digital and the physical within TUIs also seeks to develop a richer understanding of the interactive nature of TUI. As Bellotti et al argue existing work on sensed environments have tended to not provide mechanisms to allow users to make sense of the interaction [2]. Essentially, as we establish richer forms of links between the physical and the digital we need to carefully consider how the variability inherent in these different links are conveyed to users and how they might make sense of their interactions with TUIs.

This issue is manifest both within toolkits to realise tangible user interfaces and how TUIs present themselves to users. Toolkits such as Phidgets [9] provide a rich set of physical objects linked to digital objects. The connection between the real and physical is manifest through only one mechanism. Few structures are given to manage a variety of forms of link. The iStuff toolkit [1] exploits different types of events within an event heap to allow a richer set of connections to be established. However, it is unclear which of these connections are desirable and how these should be structured. We would suggest our framework offers a way of exploring this design space.

In order for the link between the physical and the digital to be intelligible to the user we must carefully consider how these effects are conveyed to users. What feedback is provided? How might users understand the extent of physical manipulation? How might the properties of the link be presented to users and how these properties might be explored? Previous work has considered how the properties associated with boundaries between real and virtual environments might be presented to users [15] and we would suggest similar explorations are needed for TUIs.

Category	Example	Transf.	Scope of Interaction	Config.	Lifetime	Autonomy
General purpose tool	mouse	Transformed	Translations in X-Z plane	Configurable	Temporary	Autonomous
	Tablet pen	Transformed	Drag, tap, tap with button pressed	Configurable	Temporary	Autonomous
Specialised Tool	tongs, eraser, magnet	Transformed	Translations in X-Z plane	Fixed	Temporary	Autonomous
	Storytent torches	Literal (ish)	Translations in X-Y plane	Fixed	Temporary	Autonomous
	Illuminating Light	Literal	Translations in X-Z plane and rotations Y	Fixed	Temporary	Autonomous
Identifier	Passage	Literal	Presence	Fixed	Semi perm.	Autonomous
	WebStickers	Literal	Presence	Fixed	Permanent	Autonomous
	Storytent ballons	Literal	Presence	Fixed	Permanent	Dependent
Proxy	metaDesk phicons	Literal	Translations in XZ plane and rotations around Y	Fixed	Permanent	Autonomous
	CUBIC	Literal	Scaling in X, Y and Z axis	Fixed	Permanent	Autonomous
Projection	Display in ambientRoom	Transformed	Human movement	Fixed	Permanent	Dependent
Illusion of same object	Traversable interface	Literal	Crossing boundary	Fixed	Permanent	Dependent

Table 1: Classification of systems with links with a physical source

Category	Example	Transf.	Scope of Interaction	Config.	Lifetime	Autonomy
Proxy	PSyBench objects	Literal	Translations in X-Z plane	Fixed	Permanent	Autonomous
Projection	Pinwheels LiveWire	Transformed	LAN traffic	Fixed	Permanent	Dependent
Illusion of same object	Traversable interface	Literal	Crossing boundary between virtual and physical space	Fixed	Permanent	Dependent

Table 2: Classification of systems with links with a digital source

6. CONCLUSIONS

We have proposed a framework for TUIs based around the idea of the *degree of coherence* between physical and digital objects. This was further broken down into the concept of *links* between physical and digital objects that are described in terms of a set of underlying *properties*. We used this proposed framework to classify current TUIs, which raised a number of broad implications for the field of tangible interaction. The focus on enriching the link between physical and digital highlighted the need to consider the asymmetry of these links, issues surrounding the configuration of these links and the need to represent the nature of these links to developers and users. We suggest that these areas represent potentially fruitful directions for future research.

7. REFERENCES

[1] Anderson Ballagas, R., Ringel, M, Stone, M., Borchers, J (2003), "iStuff: A Physical User Interface Toolkit for Ubiquitous Computing Environments", CHI 2003

[2] Bellotti V, Back M, W. Edwards K, Grinter R.E., Henderson A. (2002), Lopes C "Making sense of sensing systems: five questions for designers and researchers" Proceedings of the SIGCHI conference on Human factors in computing systems: 2002, ACM Press, pp 415-422.

[3] Brave, S., Ishii, H., and Dahley, A. (1998), Tangible Interfaces for Remote Collaboration and Communication, Proc. CSCW '98, Seattle, Washington, USA, ACM Press, pp. 169 – 178

[4] Burbeck, S. (1987), Applications Programming in Smalltalk-80: How to use Model-View-Controller, http://st-www.cs.uiuc.edu/users/smarch/st-docs/mvc.html

[5] Coutaz, J. (1987), PAC, an Object Oriented Model for Dialog Design, Proc. Interact'87, pp. 431-436

[6] Dahley, A., Wisneski, C. and Ishii, H. (1998). Water Lamp and Pinwheels: Ambient Projection of Digital Information into Architectural Space, Proc. CHI'98, ACM Press, May 1998, pp. 464-471

[7] Gellersen, H.W., and Schmidt, A. (2002), Look who's visiting: supporting visitor awareness in the web, IJHCS, 56(1), January 2002, pp. 25 - 46

[8] Green, J., Schnädelbach, H., Koleva, B., Benford, S., Pridmore, T. & Medina, K. (2002), Camping in the Digital wilderness: tents and flashlights as interfaces to virtual worlds, Proc. CHI'02, Minneapolis, Minnesota, ACM Press, pp. 780-781

[9] Greenberg S, Fitchett C (2001) "Phidgets: easy development of physical interfaces through physical widgets", Proceedings of ACM UIST 2001, pp 209-218

[10] Holmquist, L., Redström, J. & Ljungstrand, P. (1999), Token-Based Access to Digital Information, Proc. HUC'99, Karlsruhe

[11] Ishii, H. and Ullmer, B. (1997): "Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms", Proc. CHI'97, ACM Press, pp. 234-241

[12] Ishii, H., Wisneski, C., Brave, S., Dahley, A., Gorbert, M., Ullmer, B. & Yarin, P. (1998), ambientROOM: Integrating Ambient Media with Architectural Space, Proc. CHI'98, ACM Press, pp. 173-174

[13] Iwata, H., Yano, H., Nakaizumi, F. and Kawamura R. (2001), Project FEELEX: Adding Haptic Surface to Graphics, ACM SIGGRAPH 2001, ACM press, pp. 469 – 475

[14] Jacob, R., Ishii, H., Pangaro, G. & Patten, J. (2002), A Tangible Interface for Organising Information Using a Grid", Proc. CHI'02, ACM Press, pp. 339-346

[15] Koleva, B. N., Schnädelbach, H. M., Benford, S. D. & Greenhalgh, C. (2000), Traversable interfaces between real and virtual worlds, Proc CHI 2000, Hague, Netherlands, ACM Press, pp.233-240

[16] Konomi, S., Müller-Tomfelde, C. & Streitz, N. (1999), Passage: Physical Transportation of Digital Information in Cooperative Buildings, Proc. CoBuild'99, pp. 45-54

[17] Lertsithichai & Seegmiller, (2002), Cubic interface, Proc.. CHI'02, 20 – 25 April 2002, Minneapolis, Minnesota, ACM Press, pp.

[18] Massie, T. & Salisbury, J. (1994), The PHANToM haptic interface: a device for probing virtual objects, Dynamic Systems and Control, pp. 295-301

[19] Patten, J., Ishii, H., Hines, J. & Pangaro, G. (2001), Sensetable: A Wireless Object Tracking Platform for Tangible User Interfaces, Proc. CHI'01, Seattle, WA, ACM Press, pp.253-260

[20] Piper, B., Ratti, C. and Ishii, H. (2002), Illuminating Clay: a 3-D Tangible Interface for landscape analysis, Proc. CHI'02, 20 – 25 April 2002, Minneapolis, Minnesota, ACM Press, pp. 355 - 362

[21] Schkolne, S., Pruett, M. & Schröder, P. (2001), Surface Drawing: Creating Organic 3D Shapes with the Hand and Tangible Tools, Proc. CHI'01, Seattle, WA, ACM Press, pp. 261-267

[22] Ullmer, B. & Ishii, H. (1997), The metaDESK: Models and Prototypes for Tangible User Interfaces, Proc. UIST '97, ACM press, pp. 223-232

[23] Ullmer, B. & Ishii, H. (2001), Emerging Frameworks for Tangible User Interfaces", in J. Carrol (ed) Human Computer Interaction in the New Millennium, Addison-Wesley, pp. 579-601

[24] Underkoffler, J. & Ishii, H. (1998), Illuminating Light: An Optical Design Tool with a Luminous-Tangible Interface, Proc. CHI'98, ACM Press, pp. 464-471

[25] Weiser, M & Brown, J.S. (1995), Designing Calm Technology

[26] Wellner, P. (1993), Interacting with Paper on the DigitalDesk, Communications of the ACM, Vol. 36, No 7, July 1993, pp. 87-96

Sensing Opportunities for Physical Interaction

Florian Michahelles and Bernt Schiele
Perceptual Computing & Computer Vision Group,
ETH Zurich, Switzerland
{michahelles, schiele}@inf.ethz.ch

ABSTRACT

Today's technology offers a wide variety of sensors. Although many sensing applications have been produced, there is no support for the design of applications offering physical interaction. In order to make a step towards such a design framework this paper analyzes different means of sensing of humans and human activity. In particular we identify six sensing goals, referred to dimensions of sensing: ID (1), Object Use (2), Location (3), Bio Signs/Emotions (4), Activity (5) and Interaction among humans (6). Those dimensions together with different sensor placements are used to review and analyze ubiquitous computing research related to physical interaction and sensing. The final discussion draws conclusions from this analysis with respect to appropriateness of sensors and sensor placement for different sensing dimensions.

1. INTRODUCTION 1PAGE

Mobile devices give access to computing services without the constraint of sitting in front of a desktop computer. This poses new challenges for human-computer interaction: mobile users can be busy with real-world activities at any time while using mobile devices, such as crossing a busy street, discussing in a meeting or riding a bicycle. Previous work [15, 29, 24, 13] proposes physical interaction, e.g. tilting a device for configuring a device's functionality, as new and convenient forms of interaction for mobile user scenarios. The notion of implicit interaction takes this a step further and suggests to sense "an action, performed by the user that is not primarily aimed to interact with a computerized system but which such a system understands as input." [28]. That means, the user interacts with physical objects in a natural way, but a computer system also can extract inputs from these actions for the use in meaningful applications. System inputs generated from interaction with physical objects already have been used for coupling physical objects with computer applications as tangible user interfaces [18], computer-assisted furniture assembly [1], future restaurant scenarios [17], tracking a patient's medicine cabinet [31] or work-flow monitoring in a chemical lab [3]. Empowering a computer system to process physical user inputs requires augmentation of today's computer nerve-endings, such as mouse and keyboard, by sensors: perception and interpretation of real world phenomena enables a computer system to participate in the user's physical environment and serve the user in an appropriate way. Today's technology offers an astounding variety of sensors more or less suited for different applications: accelerometer, oximeter, microphone, gyroscope, temperature, skin resistance, etc. However, this variety of sensors makes it difficult for an application designer to choose the most appropriate subset which depends on the application. Although quite a variety of applications have been produced, there is no support for the design of applications offering physical interaction, such as toolkits or style-guides available for GUI development.

This paper is a first step to develop a conceptual framework, that allows to categorize existing sensors and evaluates their utility in various applications. Eventually, this framework shall guide application designers to choose meaningful sensor subsets, inspire new systems and evaluate existing applications. The paper is structured as follows: Section 2 briefly summarizes related work, section 3 describes a conceptual categorization framework of sensors and reviews existing ubicomp applications by means of the framework. Section 4 presents an evaluation of sensing technology in respect to the framework. Section 5 concludes the paper.

2. RELATED WORK

Related research attempts have been made to develop frameworks and infrastructure for reusable sensing mechanisms. The context toolkit [26] supports the development of context-aware applications with useful abstractions from the actual sensors. However, it mostly deals with the context recognition on an abstract level decoupled from the variety of sensor technology. Furthermore, it limits applications to single sensor usage as only one context abstraction can be mapped to one physical sensor. In contrast to that, the TEA architecture [32] focuses on low-level abstractions for simple sensors, which depends to much on the used sensors and, as such, does not provide reusable perception mechanisms either. The sensor classification scheme [37] facilitates the comparison and classification of sensors.

This paper was presented at "Physical Interaction (PI03) Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors. Further information and online proceedings are available at `http://www.medien.informatik.uni-muenchen.de/en/events/pi03/`

3. FINDING THE APPROPRIATE SENSORS

Advances in sensor technology such as form-factor, power consumption, processing requirements and cost-effective fabrication offer a wide variety of integration into devices and appliances. An application that enables implicit interaction uses sensors as nerve-endings to perceive the environment. But what are the appropriate sensors? Instead of an technology-oriented view we take the perspectives of a designer and an engineer: sensing goals, referred to as dimensions of sensing, and actual placement of sensors.

3.1 Logical View: Dimensions of Sensing

Typically, application designers are more interested in the opportunities sensors can offer, than in the actual technology itself. As physical interaction shall happen between human and machine, all characteristics that describe the user's situation are of interest to the application. For that, in the last years a very general definition has been established [9]: "Context is any information [...] to characterize the situation of an entity". Unfortunately this definition is too general and does not really help for application design. Thus, we identify six sensing goals, referred to as dimensions of sensing, that give a more precise description of user context.

The first dimension is a user's *ID* - this has been widely used already, e.g. for customizing and personalizing services without requiring explicit user inputs [25, 4]. In fact, we use a more general definition of *ID* ranging from differentiating people to actual identification. The second dimension is *Location*; it has been the dominant implicit input used in ubiquitous computing applications [35, 8]. It does include 3D coordinates but also semantic location descriptions. The third dimension, *Activity*, describes the task the user is performing which ranges from simple moving patterns [33] to precise job descriptions. The fourth dimension, *Object Use*, comprises collocation of a user to an object [25], carrying an object [20] and the actual use [1]. The fifth dimension, *Bio Signs/Emotions*, describes the internal state of the user. Research in this area is still in its infancy. First results could be obtained with heart-rate and skin-resistance, for reasoning about a user's affects [23]. The sixth dimension, *Human Interaction*, characterizes the relationship between humans including simple collocation, listening to a speaker, gaze, and actual interaction as discussion. In section 3.3 we will use these six dimensions of sensing together with choices of sensor placement to categorize current sensing technology.

3.2 Physical View: Placement of Sensors

An application engineer has to consider the possibilities of sensor placement in the physical world: e.g. a traffic jam can be remotely detected by a camera or locally at each car through mutually exchanged distance and speed information. Both choices are appropriate for the intended purpose but have different side-effects: the camera has to be mounted once and works for all cars, but only at one location. As a side-effect its use could be extended to other applications, e.g. criminal search. The local set-up instead requires individual effort at each car but users have the choice to participate in the system or not and it works everywhere.

We identify four different categories of sensor placement. *In Environment* refers to stationary installed sensors, e.g. in the floor, walls, where placement can only be changed with effort. Whereas *In Environment* installations work with all users at the stationary location *on Human* has the opposite characteristics: only users wearing the sensors can participate, but therefore they are not bound to a location. *On Object* is in between the two previous categories, as objects can be personal and can be carried with a human (e.g. key), but also stay at a certain location (e.g. chair). This distinction depends on the object. Additionally, *mutual collaboration* defines sensing system that always require more than one unit in order to operate properly, e.g. triangulation of signal strength for localization.

3.3 Review: Sensors in Ubicomp Research

Based on own experience with sensors and literature review we compiled a table (Fig. 1) characterizing sensor technology in respect to the six sensing dimensions and the four sensor placement possibilities. This table should be used as reference for application developers during the process of finding the appropriate sensor for their application.

In each table cell sensors are aligned due to bandwidth consumption and quality of perception in respect to the dimension. The alignment due to precision and bandwidth should be seen as rough estimation for relative comparison between sensors occurring in the same table field. For recognizing a person's *ID* the table shows four choices of sensors for installation *in environment* in the upper left cell. Obviously, the best results can be achieved with biometric sensors [36], such as finger print or iris scan, as represented by vertical alignment in the cell. Methods based on vision [10], audio or load-cells embedded into the floor [6] deliver less quality. Horizontal alignment in the cell shows, that data generated by load-cells and finger print sensors consumes lower bandwidth than methods based on vision or audio. Inertial sensors placed *on object* and *on human* can be used to sense typical movements, e.g. perceiving the signature at a pen, for identification. [27] reports about using vision, [19] about using audio worn *on human* for people identification. Location systems [16] can also be used for identifying people at different locations. As these systems require both sensors worn by human and installed base stations those system appear in the *mutual collaboration* column. For detecting *object use* load-cells [30] have been proven useful installed both *in environment* and *on object*. Object classification with vision is well established in static settings, occlusion during dynamic use can be challenging. Audio is another option, if the object use generates characteristic sounds. The use of inertial force sensors placed *on object* has been successfully used [15, 29, 24, 13] for *object use*. Obviously, motion during *object use* can be also sensed *on human* but with less quality. Audio *on human* is also possible [21] but is an indirect measurement compared to *on object* placement. Location systems can give hints as well for *object use*, e.g. teleporting X Windows to user's current location [25]. *Location* is the most explored sensing dimension in ubiquitous computing. Load-cells [30], vision [5] and audio [7] have been explored in different project. Coarse *location* can be also gained through passive-infra-red sensors, mechanical switches or IR-barriers.

Placement	Quality of Sensing	Installed in Environment	On Object	On Human	Mutual Collaboration
ID	high low	biometric sensing vision audio load-cell	inertial sensors	audio inertial sensors	location systems
Obj. Use	high low	load-cell audio vision switch/lightbarriers	inertial sensors load-cell force/distance/capacity light	audio inertial sensors	location systems
Location	high low	load-cell radar, laser Vision PIR audio switches, IR-barrier	GPS pressure, humidity	GPS pressure, humidity	location systems diff. GPS
Bio/Emots	high low	vision audio	force/load touch	GSR oximetry inertial sensors temperature	-----
Activity	high low	Smart Board Load-cell vision PIR, pressure, capacity	-----	inertial sensors strain strips GPS	location systems
Interaction (humans)	high low	Load-cells vision audio	-----	inertial sensors vision micro GPS	-----
		Bandwidth low high	Bandwidth low high	Bandwidth low high	Bandwidth low high

Table 1: Placement vs. Dimension

On object and *on human* the primary outdoors is GPS[1], more low-level information deliver humidity, inertial [34] or pressure sensors. The variety of location systems based on *mutual collaboration* is huge: differential GPS, ultra-sound, radio etc. Sensing *bio signs/emotions* with *in environment* sensor-settings is difficult: [10] and [12] report vision and audio for reasoning on user's *bio signs/emotions.* Augmented objects measuring force and touch [2] can give some hints about *bio signs/emotions.* However, most promising are *on human* measurements such as [14, 22]. *Activity* can be well sensed with special purpose system, such as commercially available smart white boards. Load-cells, passive infra-red, pressure and capacity sensors can be used for low-level detection only. *On human* sensing has been well explored for motion activity [11]. Location system can give hints reasoned from semantical location descriptions. *Interaction* among humans has not been explored very well. *In environment* sensing systems based on vision, load-cells and audio could help to perceive characteristics of interaction, such as collocation, gestures or speech. The *on object* field is blank as objects are not involved here. *On human* the same sensors can be used as for *activity* if measurements are correlated among interactors. Location system do not really help here, as collocation is not significant for interaction.

4. DISCUSSION

As a result of the review presented in the previous section, this section discusses the appropriateness of different sensor placement for the six sensing dimensions. Table 2 evaluates our classification of sensing technologies due to the dimensions of sensing and sensor placement. In particular we differentiate between *not applicable*, if a combination does not make sense, *possible* for instances with very low quality of perception, and *good* and *very good* for more promising solutions.

It points out, that *in environment* placement is the primary choice for *ID* sensing. As our focus is on human sensing it is not surprising that *on object* is well suited for *object use*. *On human* is suited for direct measurements of human-centric sensing aspects, such as *bio signs/emotions* and *activity*. *Mutual collaboration* sensors, such as the location systems perform best *location* sensing, but also can give hints for other dimension. Quite interestingly, each sensor placement is meaningful for it least one sensing dimension.

Placement	In Environment	On Object	On Human	Mutual
ID	++	---	+	+
Obj. Use	o	++	+	o
Location	+	o / +	o / +	++
Bio/Emots	+	o	++	---
Activity	+	---	++	o
Interaction	+ / ++	---	+	o

--- not applicable, **o** possible, **+** good, **++** very good

Table 2: Evaluation

Looking placements more globally, table 2 depicts that *in environment* and *on human* offer the best sensing results. Analyzing the dominant factors for each placement, it points out that video and audio are most prominent for *in environment* sensing. However, the perception quality depends

---

[1] Actually GPS is a mutual technology requiring a receiver in collaboration with satellites in space. However, as satellites are so ubiquitous and invisible anyway we consider them as a "natural" resource and view the receivers only.

on computational expensive recognition, as video and audio per se can only provide indirect measurements which are less power than e.g. direct *activity* with inertial sensors. Nevertheless, once an environment has been augmented with sensors, e.g. Smart Rooms, applications work without additional instrumentation of users or objects. It also can give hints for human-human interaction from an outer view.

As physical interaction with everyday object mostly involves movements, such as grasping, moving or turning, the dominant sensor technology for *object use* are inertial sensors. *On human* placement is suited for various sensors such as inertial sensors, audio, bio sensors and also video to a certain extend. In regards of human sensing *on human* also represents the closest to phenomena placement. Due to the high relevance of location in the real world *mutual collaboration* sensors can provide coarse information about *object use*, *activity* and *in environment*. This also explains why in the first years of context-aware computing mostly location was regarded as context. It can do a lot but in direct comparison with *on object* and *on human* sensing location system are in an inferior position.

5. CONCLUSION

This paper is a first step to systematize the use of sensor technology. Therefore, six dimension of sensing have been identified representing the sensing goals for physical interaction. We reviewed existing ubiquitous computing research for an evaluation of sensing technology with respect to the dimensions of sensing and physical sensor placement opportunities. This enables to support application designers finding appropriate sensors during system design.

Acknowledgments

The Smart-Its project is funded in part by the Commission of the European Union under contract IST-2000-25428, and by the Swiss Federal Office for Education and Science (BBW 00.0281).

6. REFERENCES

[1] S. Antifakos, F. Michahelles, and B. Schiele. Proactive Instructions for Furniture Assembly. In *UbiComp*, Gothenburg, Sweden, 2002.

[2] W. Ark, DC. Dryer, and DJ. Lu. The Emotion Mouse. In *Proc. of HCI International*, 1999.

[3] L. Arnstein, G. Borriello, S. Consolvo, C. Hung, and J. Su. Labscape: A Smart Environment for the Cell Biology Laboratory. *IEEE Perv. Comp.*, 1(3), 2002.

[4] T. Bohnenberger, A. Jameson, A. Krüger, and A. Butz. User acceptance of a decision-theoretic, location-aware shopping guide. In *IUI 2002*. New York, 2002.

[5] B. Brumitt, B. Meyers, J. Krumm, J. Kern, and S. A. Shafer. Easyliving: Technologies for intelligent environments. In *HUC*, pages 12–29, 2000.

[6] Philippe Cattin. *Biometric Authentication System Using Human Gait.* PhD thesis, ETH Zurich, 2001.

[7] T. Darrell, D. Demirdjian, N. Checka, and P. Felzenswalb. Plan-view Trajectory Estimation with Dense Stereo Background Models. In *ICCV*, 2001.

[8] N. Davies, K. Mitchell, K. Cheverst, and G. Blair. Developing a Context Sensitive Tourist Guide. In *MobileHCI*, 1998.

[9] A. Dey and GD. Abowd. Towards a Better Understanding of Context and Contex-Awareness. Technical Report 22, GeorgiaTech, 1999.

[10] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE T. Patt. Analy. and Mach. Intell.*, 21:974 – 989, 1999.

[11] J. Farringdon, AJ. Moore, N. Tilbury, J. Church, and PD. Biemond. Wearable sensor badge and sensor jacket for context awareness. In *ISWC*, 1999.

[12] R. Fernandez and R.W. Picard. Modeling Drivers' Speech under Stress. *Speech Comm.*, 40:145–159, 2003.

[13] BL. Harrison, KP. Fishkin, A. Gujar, C. Mochon, and R. Want. Squeeze me, hold me, tilt me. In *CHI*, 1998.

[14] J. Healey and RW. Picard. Startlecam: A cybernetic wearable camera. In *ISWC*, pages 42–49, 1998.

[15] K. Hickley, J. Pierce, M. Sinclair, and E. Horvitz. Sensing Techniques for Mobile Interaction. In *UIST*, pages 91–100, 2000.

[16] J. Hightower and G. Borriello. Location systems for ubiquitous computing. *Computer*, 34(8):57–66, 2001.

[17] LE. Holmquist, R. Maz, and S. Ljungblad. Designing Tomorrow's Smart Products - Experience with the Smart-Its Platform. In *DUX*, San Francisco, 2003.

[18] H. Ishii and B. Ullmer. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *CHI*, pages 234–241, 1997.

[19] N. Kern, B. Schiele, H. Junker, P. Lukowicz, and G. Troster. Wearable Sensing to Annotate Meeting Recording. In *ISWC*, Seattle, 2002.

[20] M. Langheinrich, F. Mattern, K. Romer, and H. Vogt. First steps towards an event-based infrastructure for smart things. In *Ubiquitous Computing Workshop at PACT*, 2000.

[21] P. Lukowicz, H. Junker, M. Staeger, T. v. Bueren, and G. Troester. WearNET: A Distributed Multi-sensor System for Context Aware Wearables. In *UbiComp 2002*, 2002.

[22] F. Michahelles, P. Matter, A. Schmidt, and B. Schiele. Applying Wearable Sensors to Avalanche Rescue: First Experiences with a Novel Wearable Avalanche Beacon. In *IMC Workshop*, Rostock, Germany, 2003.

[23] RW. Picard and J. Klein. Computers that Recognise and Respond to User Emotion: Theoretical and Practical Implications. *Interacting with Computers*, 14(2):141–169, 2002.

[24] J. Rekimoto. Tilting operations for small screen interfaces. In *UIST*, pages 167–168, 1996.

[25] T. Richardson, F. Bennett, G. Mapp, and A. Hopper. Teleporting in an X Window System Environment. *IEEE Personal Comm.*, 1994.

[26] D. Salber, AK. Dey, and GD. Abowd. The context toolkit: Aiding the development of context-enabled applications. In *CHI*, pages 434–441, 1999.

[27] J. Scheirer, R. Fernandez, and RW. Picard. Expression Glasses: A Wearable Device for Facial Expression Recognition. In *CHI Short Papers*, Pittsburgh, 1999.

[28] A. Schmidt. Implicit human-computer interaction through context. In *MobileHCI*, Edinburgh, 1999.

[29] A. Schmidt, K. A. Aidoo, A. Takaluoma, U. Tuomela, K. Van Laerhoven, and W. Van de Velde. Advanced interaction in context. *LNCS*, 1999.

[30] A. Schmidt, M. Strohbach, K. Van Laerhoven, A. Friday, and HW. Gellersen. Context acquisition based on load sensing. In *Ubicomp 2002*, 2002.

[31] F. Siegemund and C. Floerkemeier. Interaction in Pervasive Computing Settings using Bluetooth enabled Active Tags and Passive RFID Technology together with Mobile Phones. In *PerCom*, Fort Worth, USA, 2003.

[32] TEA - Technology for Enabling Awareness. `http://www.teco.edu/tea/`.

[33] K. Van Laerhoven, K. Aidoo, and S. Lowette. Real-time analysis of Data from Many Sensors with Neural Networks. In *ISWC*, Zurich, October 2001.

[34] E. Vildjiounaite, E. Malm, J. Kaartinen, and P. Alahuhta. Location Estimation Indoors by Means of Small Computing Power Devices, Accelerometers, Magnetic Sensors and Map Knowledge. In *Pervasive*, pages 211–224, Zurich, 2002.

[35] R. Want, A. Hopper, V. Falcao, and J. Gibbons. The active badge location system. Technical report, 1992.

[36] J. Wayman, A. K. Jain, D. Maltoni, and D. Maio. *Biometric Systems: Technology, Design and Performance Evaluation*. Springer, 2003.

[37] R.M. White. A Sensor Classification Scheme. *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, pages 124–126, 1987.

Designing Physical Interaction with Sensor Drawbacks in Mind

Stavros Antifakos[1], Jan Borchers[2] and Bernt Schiele[1]
Perceptual Computing and Computer Vision Group[1]
Human Computer Interaction Laboratory[2]
ETH Zurich, Switzerland
{antifakos, borchers, schiele}@inf.ethz.ch

ABSTRACT

Physical interaction often relies on information stemming from sensors perceiving the real world. Sensors however, have imperfections resulting in drawbacks such as uncertainty and latency. Consequently, the improvement of sensors and perception methods is important. In this paper we argue however that imperfections of sensing will remain and that the key to better physical interaction lies in taking into account those sensor drawbacks explicitly during the design of the interaction. In order to take a first step in this direction we analyze sensor drawbacks and their effects on physical interaction. Based on this discussion we propose example solutions to the arising problems.

1. INTRODUCTION

As computer systems gradually find their place in everyday life, interaction with real world user interfaces becomes more and more important. The term *physical interaction* embodies the paradigm, where real world artifacts become part of the user interface. Researchers from the fields of ubiquitous computing, interaction design, augmented reality and human-computer-interaction all are working on new interaction metaphors based on physical interaction [10].

Acquiring the users input from the real world is one of the challenges that most of the projects in physical interaction have to face. Most often sensors are used to capture real world actions. For this, research can draw on work from the fields of sensor technology and pattern recognition. Appropriate sensor selection can often simplify the recognition of real world actions. Pattern recognition techniques offer methods for acquiring more complex actions.

Although it is often taken for granted many modern household and office appliances already have physical interfaces beyond switches and dials. “Intelligent” air-condition systems switch off as soon as someone opens the window in the room. Refrigerators sound an alarm tone when the door is left open too long. The simplicity with witch these two examples acquire information about the users actions, is why they work so well. Another type of applications are automatic doors, which open when someone is nearby or water dispensers, which start as soon as someone’s hand is under the faucet. Often though, problems occur. Automatic doors don’t open before you get really close to them, or they open when you are only standing close by. Water dispensers don’t react until you have found the exact position. Frustration or even changes in people’s behavior can be the result. A count on a Swiss train showed, that approx. 70% of the passengers wave their hand in front of the infrared sensor to make the door open.

These applications rely on sensed information to react. This information is often uncertain or ambiguous. As physical interaction is mostly based on sensor systems to recognize actions in the world, the problem of uncertainty will have to be addressed in many of these systems. In this paper we propose how incorporating sensor drawbacks can lead to better physical interaction.

Sensors drawbacks are only one part of the problem that HCI researchers face. Many other technical problems related to bandwidth, connectivity, latency, and power will have to be dealt with. Here we focus on the drawbacks of sensing systems

In particular we analyze the effects that sensing systems have on human-computer interaction. We give an overview over sensing tasks of interest for HCI, and discuss main sensor drawbacks. We then propose a set of design ideas to make physical interaction clearer and more useable.

2. SENSOR CHARACTERISTICS

By analyzing the effects sensor systems have on physical interaction we hope to bridge the gap between sensor research and physical interaction design. In this section we analyze sensing systems with respect to important sensing tasks for physical interfaces.

In Table 1 a few representative perception tasks for physical interaction interfaces are listed. The perception tasks are sorted by increasing task complexity. The sensing system is decomposed in the actual sensors and the recognition system. The column “recognition system” describes the actual sensor data processing algorithm used. This can either be a complex statistical pattern recognition system or a simple thresholding algorithm amongst others. The examples are either applications or projects that make use of the perception tasks for their interfaces.

This paper was presented at "Physical Interaction (PI03) - Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors. Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

Table 1: Perception tasks ordered by increasing complexity

Perception Task	Sensors	Recognition System	Main Drawbacks	Examples
Wireless switches/sliders	Switches/sliders		Latency	iStuff [7], Phydgets **Error! Reference source not found.**
Detecting persons presence	IR-sensor	Threshold	Latency, ambiguity	Automatic door, water faucet, auto. urinal
	Floor weight switch	Threshold	Latency	
Object location	IR-based location	Triangulation, time of flight	Precision	MediaCup [5]
	Ultrasonic location		Precision	
Object identification	Rf-id tags	Database	Too many items	Smart shopping cart
	Weight sensors	Weight matching	Ambiguity	Weight surfaces [6], Tangible bricks [2]
Object movements	Inertial (accel. & gyros)	Dead reckoning	Sensor precision, ambiguities	MediaCup [5]
People's location	GPS	Map lookup	Robustness, doesn't work indoors	Tourist guide [11]
	Ultrasonic, …	Triangulation	Precision	Active bat, …
Gestures (depends on number of commands)	Inertial	Various HMM, in general pattern recognition	Recognition rate, latency, ambiguity	Sign language recognition, interactive narrative systems …
	Vision			
	Combined			
Handwriting recognition	Scanner, touch sensitive screen	Optical character recognition (OCR)	Learning time, latency, recognition rate	PDA input systems, recognizing handwritten notes
Speech recognition (commands)	Audio	Signal matching HMM, …	Robustness	Voice dialing
Speech recognition in general	Audio	HMM's, other pattern recg.	Recognition rate, latency	Taking down a text
Situation detection	Audio & video	Statistical pattern recognition	Recognition rate	Context sensitive notification
Activity recognition	Video, inertial	Statistical pattern recognition	Recognition rate	Context sensitive notification, surveillance

Most problems with the use of sensors for physical interaction arise because of the uncertainty which is inherent in sensor systems. Where uncertainty comes from can be analyzed by regarding the different aspects of sensor uncertainty. Here we differentiate between four aspects, namely robustness, recognition rate, precision, and ambiguity.

Robustness of a system describes how well a system performs over all external conditions that make sense for a specific task. This includes changes of lighting conditions, changes in the noise level, and changes of the number of people in the environment. The *recognition rate* quantifies the performance of a classification system for a recognition task under fixed external conditions. Often the rate is gained during experiments with a set of example test data. In contrast the *precision* of a sensing system describes how well the output of the sensor system represents the real world phenomenon. Finally, *ambiguity* describes how well different physical phenomena can be held apart using a certain sensor configuration. When two real world actions have similar effects on the sensor system they become hard to distinguish, i.e. ambiguous.

Table 2 gives an overview of the different aspects of uncertainty. It relates the different aspects to the level from which they emerge in the sensing system.

If a system is *not robust* towards environmental changes this can be very annoying and surprising to the user. For example mobile phone voice dialing systems are expected to work wherever you are. However close to a noisy street, it would be surprising if such a system would work. Similarly the Global Positioning System (GPS) will not always have satellite contact in a city with tall skyscrapers.

Similarly, a sensor system with a *low recognition rate* can become interruptive. When using a gesture based interface the user doesn't want to have to repeat every third gesture just because the system didn't understand. In the example of a retina scanner for access control purposes a low recognition rate would definitely be cumbersome.

The *precision* of a sensor influences the previously mentioned recognition rate, but also has its own influences on physical interaction systems. Sensors with inherently low precision will only be used for tasks with minor importance, if at all. Variations in precision during use of the sensor will result in effects of annoyance and surprise to the user. Most indoor positioning systems still are not very precise. This may be one reason for them not being successful in many commercial applications.

Sensing *ambiguities* in systems can also have disturbing effects on physical interaction. The automatic opening of a train door

Table 2: Aspects of uncertainty in sensor systems

Sensing system level	Uncertainty aspect
Task	*Robustness*
Classification System	*Recognition Rate*
Sensor	*Precision*
Physical Signal	*Ambiguity*

when a passenger is turning a page of his newspaper is at least surprising, sometimes even bothering. Ambiguities during speech input can have serious implications. If you are talking to your office neighbor about "deleting files", you definitely don't want your system to take such a drastic action.

All sensor drawbacks mentioned until now were directly related to the uncertainty of sensing systems. Beyond uncertainty, the *latency* of a sensor system is important for physical interaction. Latency directly influences the interactivity of a system. Slow systems become cumbersome to use, as the flow of action is always interrupted by pauses. In the worst case systems may even loose causality. This happens when delays become so long that the user is unable to make the connection between his input and the systems reaction.

Although sensor system researchers are making vast amounts of progress, it is clear that uncertainties will always remain. We believe that these remaining uncertainties and sensor latency needs to be incorporated into interaction design. By integrating these sensor drawbacks systems will become less interruptive and less cumbersome.

The effects that sensor uncertainty and latency have on the interaction experience range from being surprising over disturbing to totally interrupting. These effects are well known from the general field of HCI and have been discussed widely under the names of *principle of least surprise* [14] and *principal of fluid interaction* [15]. Systems designed with these principals in mind are easier to use and find greater acceptance. Designing physical interaction systems with these principles in mind is only the next step. In the next section we propose simple design guidelines to comply with these principles.

3. DESIGNING INTERACTION WITH SENSOR DRAWBACKS IN MIND

To enable more usable physical interaction, sensor and interaction researchers need to work together. On the sensor side researchers need to become aware of the effects of their systems on interaction. For example, robustness of systems could be increased if the exact task profile for the usage of the system is clearly defined. Defining sensor system characteristics should always be done with respect to a given task.

A paramount goal of interaction design is to keep the user's mental model [14] of the system as simple as possible. The system should appear causal to the user. This is why the *principle of least surprise* has such great importance. Keeping sensor drawbacks in mind during the design of physical interaction results in a more precise anticipation of the mental model the user will have.

There are several ways of dealing with sensor drawbacks in interactive systems. MacColl et al. [1] describe the basic idea of explicitly presenting uncertainty. They describe four ways to present uncertain information: pessimistic, optimistic, cautious, and opportunistic. Horvitz [13] proposes systems that vary their self-initiative depending on uncertainty and the expected utility of an action. Mankoff et al. [17] present the technique of mediation, where the user can chooses from several possible recognition results.

While studying the effects of uncertainty as a whole is important, we believe that the different aspects of uncertainty identified in the previous section need to be considered individually and in more detail. In the remainder of this section we do this by presenting examples.

When insufficient *robustness* is the cause of error, the user should be informed why the system is not working. It may simply be that the system doesn't work in a loud environment, as in the example of using voice dialing close to a noisy street. Telling the user the reason for the lack of robustness, gives him the possibility to change the setting. Many GPS Systems for example, let the user know how well the system is working by showing the number of satellites available.

When systems have low *recognition rates* they become interrupting whenever they make wrong decisions. Presenting the results non-destructively is one way around constantly interrupting the user. For example, handwriting recognition in the Interactive Workspaces Project [9] presents the recognized text beside the handwritten text. In this way the user is not interrupted by wrong recognitions but can still be aware of the systems recognition.

Applying sensor systems with low *precision* needs to be done with grate care with respect to the effects wrong results may have. Automatic system actions have to be designed with the precision of the sensor in mind. Drastic actions should only be taken when precision is high.

For both systems with low recognition rates and low precision it may be useful to display a confidence level of the system. For a recognition system this may be the recognition probability or an external evaluation. For a sensing system with imprecision on the sensor level the momentary precision depending on the external conditions could be shown to the user.

In systems with inherent *ambiguity* interaction should be designed to minimize surprising the user by unexpected actions. This can be done by informing the user about what is sensed. For example, a train passenger will be less surprised about an automatically opening door when he turns a page in his newspaper, if he knows how far the sensor reaches. A method for actually reducing ambiguity has already been used in speech recognition systems. Here *quasimodes* [12] let the user activate the system by pressing a key on the keyboard. Using this no more mistakes happen when the system was listening and you thought it wasn't.

Beyond uncertainty, dealing with *latency* of sensing systems is a highly important task. Offering the user immediate feedback is often invaluable. Giving the user a notion of how long an action will take can also be encoded in feedback. In [16] a tactile display is presented which gives feedback of how far a task on a mobile device has advanced. The stronger the device shakes the further the task has progressed.

Informing the user on how far his request to the system has advanced, could be accomplished using a *feedback chain.* The idea of a feedback chain is to inform the user about how far his command to the system has been processed. This could be done on each relevant level. For example, pressing a wireless button would give an immediate local feedback by lighting an LED on the device. When the system recognizes the event, further feedback could be given on an ambient display. When finally the command is processed it could be reported back to the user by letting the wireless button shake or flash an LED. Such a feedback chain would help the user build a mental model of the system. Beyond that it could be a useful tool for anticipating system errors.

4. DISCUSSION

Although sensor drawbacks will always exist it is still necessary to further investigate innovative sensing approaches using various sensors and algorithms. Most importantly sensor systems need to be evaluated on real interaction tasks. For this, sensor researchers need to work together with interaction designers.

In the HCI community, the task conditions in which interaction will take place need to be clearly stated. Beyond this, the drawbacks inherent to sensor systems need to be taken into account. Thus an informed design can take place. Overall, the loop between interaction designers and sensor researchers needs to be lived.

In the examples presented in this paper we didn’t distinguish between explicit and implicit interaction [8]. Sensor drawbacks equally effect both explicit and implicit interaction. In many cases implicit interaction is effected more than explicit interaction as it often heavily relies on sensor information. Presenting feedback explicitly as proposed in this paper could be a solution to make implicit interaction more useable. In applications where sensor drawbacks cause too many problems implicit interaction will have to be extended by disambiguating explicit mechanisms.

5. CONCLUSIONS

In this paper we take first steps towards dealing with sensor drawbacks in physical interaction. We analyzed sensor drawbacks to discover the effects they have on HCI. By decomposing the notion of uncertainty into its aspects it was possible to clarify where the different effects on interaction come from. To deal with the problems we present design techniques that mostly have already been used in different applications.

Most importantly, we believe that the quite separate research communities for sensing and HCI need to work together more closely. On one side the sensing community needs to take the tasks settings into account while evaluating their work. On the other side the HCI community needs to be aware of potential sensor drawbacks and take them into account appropriately during their design. We believe that this contribution is a first step in answering the five questions in designing physical interaction posed by Bellotti et al. [4].

6. REFERENCES

[1] MacColl I, Chalmers M., Rogers Y., and Smith H. *Seamful ubiquity: Beyond seamless integration*. Workshop at UbiComp 02, Gothenburg, Sweden, 2002.

[2] Fitzmaurice G. W., Ishii H., and Buxton W. *Bricks: Laying the Foundations for Graspable User Interfaces.* In CHI 1995, Denver, Colorado, USA.

[3] Greenberg S., and Boyle M. *Customizable Physical Interfaces for Interacting with conventional Applications.* In UIST 2002, Paris, France.

[4] Bellotti V., Back M., Edwards W. K., Grinter R. E., Henderson A., and Lopes C. *Making Sense of Sensing Systems: Five Questions for Designers and Researchers.* In CHI 2002, Minnesota, USA.

[5] Beigl M., Gellersen H-W., and Schmidt A. *MediaCups: Experience with Design and Use of Computer-Augmented Everyday Objects.* Computer Networks, Vol 35(4), 2001.

[6] Schmidt A., Strohbach M., Van Laerhoven K., and Gellersen H-W. *Ubiquitous Interaction – Using Surfaces in Everyday Environments as Pointing Devices.* 7th ERCIM Workshop, User Interfaces For All, 23 - 25 October, 2002.

[7] Ballagas R., Ringel M., Stone M., and Borchers J. *iStuff: A Physical User Interface Toolkit for Ubiquitous Computing Environments*. In CHI 2003, Ft. Lauderdale, Florida, USA.

[8] Schmidt A. *Implicit Human Computer Interaction Through Context*. Personal Technologies, Volume 4(2), June 2000.

[9] Johanson B., Fox A., and Winograd T. *The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms.* In IEEE Pervasive Computing Magazine 1(2), April-June, 2002.

[10] Svanaes D., and Verplank W. *In Search of Metaphors for Tangible User Interfaces.* In DARE 2000, Copenhagen.

[11] Davies N., Cheverest K., and Blair G. *Developing a Context Sensitive Tourist Guide.* In Workshop on HCI and Mobile Devices, Glasgow, U.K. 1998.

[12] Raskin J. *The Humane Interface: New Directions for Designing Interactive Systems.* Addison-Wesley, April 2000.

[13] Horvitz E. *Uncertainty, Action, and Interaction: In Pursuit of Mixed-Initiative Computing.* In IEEE Intelligent Systems, pp.17-20, Sept./Oct. 1999.

[14] Norman D. A. *The Design of Everyday Things*. Doubleday: New York, 1990.

[15] Guimbretière F., Stone M., and Winograd T. *Fluid Interaction with High-Resolution Wall-Size Displays.* Proc. of the ACM Symposium, New York, pp 21-30, 2001.

[16] Poupyrev I., Maruyama S., and Rekimoto J. *Ambient Touch: Designing Tactile Interfaces for Handheld Devices.* In UIST 2002, Paris, France.

[17] Mankoff J., Hudson S. E., Abowd G. D. *Interaction Techniques for ambiguity resolution in recognition-based interfaces.* In UIST 2000. CHI Letters 2(2).

Controlling multiple devices

Rainer Wasinger
DFKI GmbH
Stuhlsatzenhausweg 3
66123 Saarbrücken, Germany
+49 (681) 302-3393
rainer.wasinger@dfki.de

Christian Kray
Saarland University
P.O.Box 15 11 50
66041 Saarbrücken, Germany
+49 (681) 302-2865
kray@cs.uni-sb.de

Christoph Endres
Saarland University
P.O.Box 15 11 50
66041 Saarbrücken, Germany
+49 (681) 302-4134
endres@cs.uni-sb.de

ABSTRACT

Two major differences between ubiquitous computing and a traditional desktop scenario consist of the number of users interacting simultaneously with a system, and the number of devices that they use. This paper focuses on the physical user interfaces problem of how device control is allocated, shared, and released by services and users. Based on a classification of different types of devices, we analyze in which ways a device can be controlled. We then identify several influencing factors in allocating devices, and conclude by sketching out a high-level strategy for the (semi) automatic handling of device allocation.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: Input devices and strategies, interaction styles.

General Terms

Management, design, human factors.

Keywords

Human computer interaction, ubiquitous computing, device control.

1. INTRODUCTION

Until recently, interfaces for single users in a stationary setting were the focus for much of the research in human-computer interaction. This usually took the form of a single person using a single desktop computer. Although research is now being conducted on interaction with multiple devices [3], interfaces for mobile applications [4], and computer-supported collaborative work (support for multiple users in a distributed setting) [1,2], the research community is only slowly advancing on the combination of these features in an intelligent environments setting that supports multiple collocated users.

A common feature of intelligent environments today is that they consist of multiple devices, and that multiple users can simultaneously request services from the environment. In comparison to typical desktop scenarios, devices are now distributed throughout the environment, and like the user, some of them are now highly mobile. Devices and users may enter or leave an environment without warning, and individual users may bring along their own devices. In comparison to interfaces for typical desktop scenarios, the interfaces for ubiquitous computing are increasingly transparent, and as dynamic as the devices they contain. In the right light, the effect of such environment characteristics can result in a much more natural and flexible means of interaction for users with the environment around them, and an enriched range of available services. However, for this to be true, there are different concerns that first need to be addressed regarding multiple users and multiple devices. This is independent of the type of environment that exists, be it at work, at home, in a museum, while shopping, or even outdoors.

To illustrate such concerns, let us assume that all members of a household wish to retire to the living room. One person wishes to watch TV, while another wishes to have a book read out to them. Yet another would like to play chess, and a final person wants to surf the Internet. One concern with this scenario is who controls what device at which time, and what this control might look like? Can devices be shared, and if so which ones, and by how many users? Will different services offered by the environment require the same devices? Will the delivery of multiple services affect the quality of other services currently being requested? This paper outlines the underlying concepts and relationships that will help to address these issues in the future.

In section 2, we define the interacting components of a physical user interface. In section 3, we discuss how an intelligent environment may incorporate device allocation, device sharing and device release for multiple users. Sections 4 and 5 discuss factors concerning the allocation of devices to users and services, and how such an allocation strategy might conceptually look like.

2. INTERACTING ELEMENTS OF A PHYSICAL USER INTERFACE

An intelligent environment can be seen to encompass three essential interacting elements – devices, services, and users. In comparison to traditional desktop environments, these components are largely decoupled from traditional graphical interfaces, and interactions primarily take place through (partially) transparent interfaces. We can distinguish between several classes of *interface devices*, depending on their type and their individual profile properties. When a device is primarily concerned with the handling of input and output, as in the case of cameras, microphones and displays, their *type* can be classified as

This paper was presented at “Physical Interaction (PI03) – Workshop on Real World User Interfaces”, a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the authors.
Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

dedicated. However, when the primary role of a device is to fulfill other functions in everyday life, they may be classified as *non-dedicated*. Non-dedicated devices can be further classified based on whether they have been augmented or *enhanced*. Enhanced devices can be grouped as either *active* when they pursue interaction with their environment (e.g. a smart bookshelf, or touch-sensitive table), or *passive* when the environment must pursue interaction with them (e.g. an RFID tagged book). A *non-enhanced* device in comparison would simply be a non-tagged ordinary coffee mug. This device taxonomy is shown in Figure 1 below.

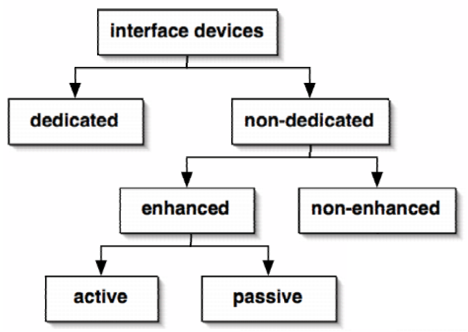


Figure 1. Device taxonomy

An extended categorization of devices would see the incorporation of *device profiling*. This would cover individual properties of the device such as whether the device is suitable for *private* or *public* use, or whether or not it is *shareable*. Other properties in a device profile may include the devices ability to *cater for different human senses* (e.g. sight, sound, touch, smell, taste), and information regarding device *ownership*. Device ownership is particularly important for devices such as PDAs and bluetooth headsets, which may be owned by single users, and only be brought into an environment as additional infrastructure. This is in comparison to devices such as large displays and speakers that belong to an environment's own infrastructure.

Services represent the functionality of an intelligent environment, and rely on the availability of underlying devices to form a channel of communication with its users. For example, watching television requires both audio and visual output devices to function. In contrast to typical single-user desktop scenarios, the spatial location of both the user and the required devices is more dynamic, and this also gives rise to a range of external factors that need to be considered when allocating services to users, such as whether one service will interfere with another service, or how many services a device can physically support.

In comparison to single-user desktop scenarios where a sole user controls all devices, scenarios catering for the simultaneous support of several users must share devices among *multiple users*, all of whom may be moving around in the environment. Users may be *collaborating* with one another, or interacting *independently*. They may be *distributed* (in different environments), or *collocated* (in the same environment). Users may have a preference for using particular devices (e.g. a large screen over a small screen). Some users will have a preference for the set of input modalities they wish to use (e.g. a disabled person with poor vision), and other users will place a different emphasis on the type of services available (e.g. users that prefer reading books compared to watching TV).

An important aspect that arises in intelligent environments is which user or service controls what device, and how devices are shared among users and services. *Control* refers to the allocation of a device to a particular user and/or service so that the user or service can use it for interaction. Some devices may support multiple users, so the notion of device sharing is also of relevance. *Sharing* may be either user independent, cooperative, or parallel. These concepts are discussed in more detail below.

3. CONTROL AND SHARING

The entities and concepts we defined in the previous sections (e.g. users, devices, and control) form a complex and interacting system that is strongly influenced by situational factors. In this section, we propose a preliminary analysis of these interactions.

In an intelligent environment, both users and services (the system) may request control of a device, and in different ways. One such form of request is *user-initiated*, in which a user asks for a specific service, and directly specifies which device(s) should be used. However, there are several ways to specify a device, ranging from spoken commands ("Show my email on the big plasma display.") to multi-modal references ("Show my email on that [pointing gesture] screen.") and physical acts such as picking up a pointing device. The set of possible (physical or non-physical) actions for obtaining device control depend on the type of device (see Figure 1) and its profile properties. For example, while a user can pick up small devices such as remote controls, larger devices like touch-sensitive tables cannot be picked up.

Another form of request is *system-initiated*, in that the system (or a service) automatically allocates a set of devices for a given task. The resulting assignment may however displease the user – even if multiple situational factors are taken into account – and the user may feel controlled by the system. In addition, a combined *user-system initiated* approach is possible, where the user directly specifies some devices while others are selected by the system. While this may combine the problems inherent to both approaches, it may also remedy some. For example, if a user can specify at least some devices, they may less likely feel that they are not in control. In addition, the mixed allocation of devices would free the user from specifying *all* devices that s/he wants to use for a task, which could be tedious (e.g. "I want to browse the web using this screen, this loudspeaker, this keyboard…").

As shown in Figure 2, the control of a device can be either *exclusive* or *shared*. In the first case, a single person uses the device, while in the later case several users may access the device either *cooperatively* (e.g. playing a game together) or in *parallel* (e.g. two users browsing the web in two separate windows on a single large screen). In principle, the methods for allocating device control also apply to device sharing, with the exception that not all devices are shareable (e.g. a headphone), and purely system-driven decisions on device sharing would most likely alienate users. As an example, consider a user reading their email on a desktop monitor, as the layout of the screen is suddenly changed so that only part of it still displays the email while the rest is used for a video game that two other users want to play.

The final step in the handling of device control consists of releasing the control of a device. Again, the considerations we presented for allocating control also apply to the release process in that either the user or the system may explicitly or implicitly release control of a device. In addition, there may be a strong spatial-temporal component in the process, such as when a user simply walks away from a set of devices or does not use the device(s) for a longer period of time. In this case, control of the devices should also be implicitly released.

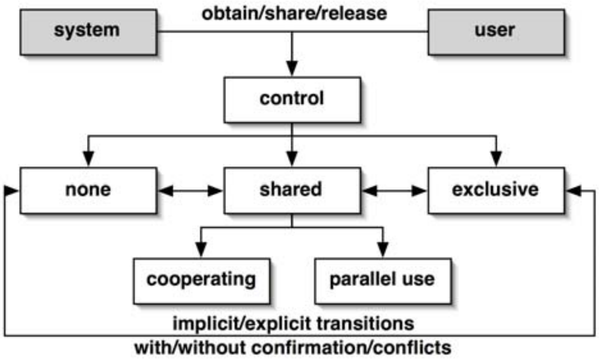


Figure 2. Assignment of device control

An orthogonal dimension to the processes of obtaining, sharing, and releasing control is the way in which changes in control are *confirmed* and *authorized*, as well as how conflicting requests are handled. Furthermore, we can distinguish between two different ways in which users are informed about a change in control. The change is either communicated *explicitly* (e.g. the system generates speech output such as "the plasma screen is now in use by Brian."), or *implicitly*, i.e. the device is simply allocated to another user without notification. The most appropriate way to communicate change in control also depends on who initiated the change, for example if the change was initiated by the system, an explicit explanation may be beneficial to avoid alienating users.

4. FACTORS INFLUENCING THE CONTROL OF MULTIPLE DEVICES

The control of multiple devices by users and services in an intelligent environment can (as shown above) be divided into the areas of device allocation, sharing and release. These forms of control are however influenced by a multitude of factors characteristic of a dynamically changing environment. There is for example a need for constant re-evaluation and adaptation of the allocation of resources, due to fluctuations in users and devices as they move in and out of an intelligent environment. In this section, we describe service and user implications on device control as required for physical user interfaces, and also consider social issues and spatial/temporal constraints relating to multiple users.

Both services and users may have preferences for different types of devices. One way to accommodate for this is through *device modeling*, by listing the properties of each device (e.g. shareable/non-shareable, private/public, modalities being catered for), and making the model accessible by each service and user. An added level of complexity arises when a device is only partly shareable. Some devices such as touchscreen displays although being shareable on the presentation side (screen can be split in two halves), are currently still difficult to share in parallel on the input side, due to the touch sensors only identifying a single user's interactions at a time. Another factor is that of *resource limitation*. If the devices that a service requires are no longer available, the system will have to either consider redistributing the already allocated devices, or inform the user of an expected waiting time. Such a redistribution of devices may be classified as *resource adaptation*.

Similar to devices, users must also be modeled if the system is to best understand their needs, and this information must be merged with any prerequisites the user may currently have. An important issue is that users need to be provided with system resources in a fair manner, and must also ”feel” that this is the case, especially in times of device conflict. The system must be able to make distinctions between the desired needs of a user, i.e. *soft prerequisites*, and the required needs of a user, i.e. *hard prerequisites*. For example, a distinction may be made between a user who desires a large screen to read their email simply because the screen is large, compared to a visually impaired user who requires a large screen in order to see anything at all. Distinctions may also be required to classify the *value of a user's work* (e.g. an intern playing solitaire, compared to the CEO's secretary updating business spreadsheets), and the *access rights* a user or service may have to devices (e.g. should a service be allowed to assign the personal PDA of a user to another user's use?)

In contrast to single-user scenarios, multiple users also require certain *social aspects* to be considered when allocating the control of devices, such as *privacy*, *background noise* to other users, and *urgency*. Social implications can affect either the user themself / *1st party* (e.g. introverted users, and users desiring privacy while reading emails), or cooperating users / *2nd parties* (e.g. does one input device such as a microphone dominate over another input device such as a keyboard), or other users / *3rd parties* (e.g. one user watching television while another is trying to read). Social aspects may also apply to the type of service such as bank transfers or the editing of finance spreadsheets, and to the type of task within a service such as entering a PIN number or password.

Spatial influences can also have a large effect on allocating device control to multiple users. While a system must try and distribute users to areas that best support the service, it must also consider any desires of the user, and try not to force a user to move ”too” far away from their current position. Spatial concerns become more complex when devices are already in use by other users, as the system must then try and predict for optimal allocation of resources for the present time, and also for the future. Decisions must also be made as to when a person wishes to move their service to another part of the environment, or has stopped using a set of services altogether (e.g. a user going to the toilet compared to a user who no longer wants to watch television). It must also weigh up the need for some users to relocate to other areas in order to accommodate for additional users in the environment.

Temporal influences include for example the urgency in which a user requires a service or set of devices. Temporal conflicts may arise when there are too few devices for a required service, and may require decisions to be made by the system as to how long a user must wait before either an alternative user's service is disrupted, or other users are relocated. The importance of the new user's task is also relevant in such a situation, as user disruptions are only rarely appropriate. For example, a conflict may arise if one user wants to watch the news (which is only broadcast at specific times) while another is already playing a computer game. Providing user feedback on expected waiting times and feedback regarding the information that the system is grounding its decisions on, along with the ability for the user to schedule events in the future, will all help a user feel more in control in such a situation.

5. CONCEPTUALIZING AN ALLOCATION STRATEGY

Strategies allowing for multiple users and services to obtain, share and release control of multiple devices must be flexible and fair. This section illustrates a basic conceptual strategy to help the understanding on how important factors such as those described in section 4 may fit together in a practical implementation. As shown in Figure 3, the strategy is flexible in that the user can select either a service (e.g. “I want to watch TV”), a service and a set of devices (e.g. ”I want to watch TV on that display and those speakers”), or just a set of devices (e.g. ”I want to use that display and those speakers”). This is achieved through the notion of a service/device request, in which the system tries to fill in the “UNKNOWN” fields, based on implicit and explicit user input. In this strategy, devices are generally associated to a service. This means that if a user only selects a set of devices and the system cannot implicitly or explicitly determine what the user wants the device(s) for, the device(s) will be reallocated when required by another service. As described in section 4, the prerequisites for devices and users need to be considered, as too the social implications that may arise to any 1st, 2nd or 3rd parties involved. Spatial and temporal constraints are also considered, and only then are the devices allocated to a user. Conflicts will undoubtedly also exist in a system that allows for multiple users interacting with multiple devices, and solutions to these (if at all adequately resolvable) may take the form of removing soft prerequisites in the search for appropriate devices, calculating new optimal device allocations, or simply informing the user of expected waiting times. Transforming this conceptual strategy into a concrete solution will form a major part of our future work.

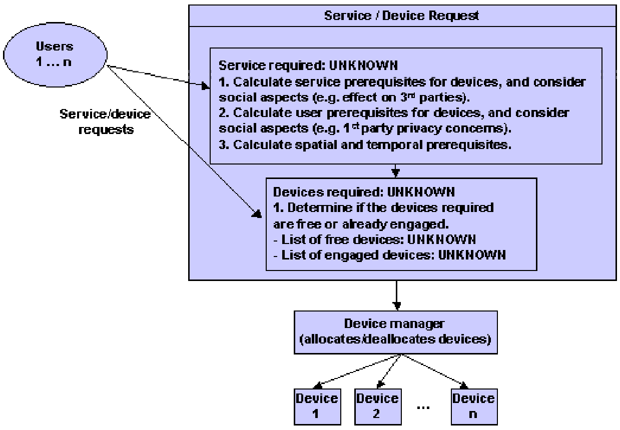


Figure 3. Outline of a device allocation strategy

6. CONCLUSIONS

Physical user interfaces are an important aspect to modern intelligent environments, and have been shown to be very dynamic and difficult to model. The main components of such interfaces are the devices themselves. This paper addresses the concerns on allocating, sharing and releasing multiple devices to multiple users and services in such a physical user interface setting. We illustrate the factors affecting this process, and also sketch out how they may conceptually fit together in a practical solution.

7. REFERENCES

[1] Bentley, R., Rodden, T., Sawyer, P., Sommerville, I., “An architecture for tailoring cooperative multi-user displays”, *Proc. of CSCW*, 1992, pp. 187-194.

[2] Bowers, J., Rodden, T., “Exploding the interface: Experiences of a CSCW network”, *Proc. of SIGCHI*, 1993, pp. 255-262.

[3] Brumitt, B., Meyers, B., Krumm, J., Kern, A., and Shafer, S., “EasyLiving: Technologies for Intelligent Environments”, *Handheld and Ubiquitous Computing*, 2000, pp. 12-29.

[4] Dey, A., Ljundstrand, P., Schmidt, A., “Distributed and Disappearing User Interfaces in Ubiquitous Computing”, *CHI Workshop*, 2001.

Study of Tangible User Interface for handling tridimensionnal Objects.

Ludovic Garreau
LIPSI - ESTIA
Technopole Izarbel
64210 Bidart, France
l.garreau@estia.fr

Nadine Couture
LIPSI - ESTIA
Technopole Izarbel
64210 Bidart, France
n.couture@estia.fr

ABSTRACT

The field of interaction system is very wide from the mouse to the data glove while in passing by the cubic mouse [7], metaphors ("WIM" [17], voodoo Dolls[13]) and tangible user interface. In these systems of interaction, the designer of these interfaces seek to simplify the interaction by using human knowledge in order to create behavioral interface. In this paper we propose a new kind of tangible interface based on the handling of physical object that we call "interacteurs". In that goal, after a short introduction on the context of the study, we introduce the concept of Tangible User Interface with samples. Then we present our reflexion on TUI for Handling : from the hardware to the software. For that, we define a typology of interacteurs based on concepts proposed in Design For assembly (DFA) methods, and we explain our idea of a tangible user interface: ESKUA.

Keywords

Computer Human Interaction, Tangible User Interface, Interacteur, Handling, Visualization.

1. INTRODUCTION

The scientific community generally agrees that traditional input devices, mouse and keyboard, are limited and that it is necessary to create new devices, in particular when three-dimensional (called 3D in following) scenes are to be visualized. New Human Computer Interaction systems have appeared, proposing new concepts for hardware (input and output devices) well as for software (graphical interfaces). With those systems, authors don't search only to create new hardware or software, but want to propose new manners to interact with virtual objects.

Historically, the first systems of interaction between human and the computers are at the beginning of the years 1960. One of the first was the Sketchpad (Sutherland, 1963) the purpose of which was to make it possible to the user to interact in a direct way with the software interface using an optical pen. This system, like the current systems "mouse and keyboard" are clearly limited by the interface of visualization, the screen, and by a space of interaction in two dimensions. The traditional devices evolved logically towards the systems of 3D mouse and joystick which pains to be essential because their uses creates a considerable cognitive difference between the action carried out on the mouse by the user and the result in the 3D numerical scene. To palliate these disadvantages, two currents emerged: virtual Reality, Mixed Reality (Head Mount Data, panoramic screen, Workbench) and the tangible user interface (cf. 2) (Aish, 1967) quoted in [4].

As Fuchs [8], we think that the systems directed towards visualization, often gathered by the term of "Virtual Reality", require complex haptic interfaces in their realization for an interaction of quality with the digital model. Ware and Rose [23] showed that the use of real objects, included in tangible user interface, clearly improves the performances of the users at the time of the phases of handling of virtual objects.

We propose a tangible user interface which enable to have a physical perception of the data constraints during the "virtual" phase of combination and handling by the use of special physical icons that we call *interacteurs* (cf 3.1).

2. THE TANGIBLE USER INTERFACE

Ullmer and Ishii in [21] define system based on the real object as a **T**angible **U**ser **I**nterface (called **TUI** in following) by analogy with a **G**raphical **U**ser **I**nterface (usually called **GUI**) (cf. figure 1). Ullmer and Ishii describe the TUI like a physical realization of the graphical user interface (GUI).

2.1 Description

TUIs (Latin tangere: the capability to be touched) seek to make intuitive interfaces whose finality is to couple the physical reality and the numerical one in order to simplify interaction. TUIs are based on the use of real objects which allow a representation of the data and a physical control of numerical information [19].

So by joining the idea developed in [19], we think that the TUIs can be one more in the combination and visualization of several virtual objects. With the real objects, handling is simple to realize but nevertheless brings to identify the difficulties of combination concerning the questions of sym-

This paper was presented at "Physical Interaction (PI03) - Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003.The copyright remains with the authors.

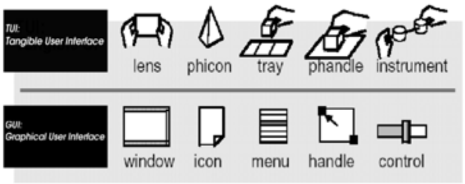


Figure 1: Analogy between GUI and TUI
(reproduce from [18])

metry, occlusions, and setting in position of the parts (cf 3.1). Moreover, it is now admitted [23], [11] that the use of real objects for displacements and the control of the virtual objects is more powerful than the traditional systems (Virtual Reality, 3D mouse).

2.2 TUI for handling

A TUI is made up mainly of two parts: the tangible part (to allow the interaction) and the visualization part (the feedback of handling). The realization of TUIs consists in using tangible objects (cubic, plane, etc.) who have a true and figurative meaning from their forms and their colors in the real world. The tangible object positions and orientations are captured by the way of sensors or by the use of camera(s) and then transmitted to the computer. To name the tangible part, Ernst, Schfer and Bruns [6] use the term *graspable* and Ullmer et al. [20] reference the words *"Artifact, Container, Tangible, Object, Phicons, Objects, props"*. We choose as Ullmer and Ishii in [18] to use the term "artifact" to designate physical objects in TUI. The artifacts are at the same time an input device and an output device. They symbolize the concept, and provide control of the virtual data. So, there are as much artifact as different data and control, for example the luminous room [22], the metadesk [18] and Ernst's TUI [6]. The handling of the artifacts modifies the virtual objects (input device), moreover only by casted one's eye over the artifacts, contrary to a 3D mouse, provide information to the user (position, orientation, etc.) on the virtual objects (output device). In [22], authors says "The students and professionals who have experimented with it (authors' note : The luminous Room) affirm that its direct manipulation style -"like working with real thing"- both fosters and takes advantage of the spatial understanding inherent to work with real optics".

The user is not "attracted" by a feedback different than output device (for example : screen). The consequence of this was that the user gathers his thought on the action and doesn't search the result on an another device. Anderson et al. (figure 2, [1]) illustrate the idea that the application and the computer are only one successive stage with the combination of artifact.

To be closer to our domain, in the next part we study TUI created and usable for handling.

2.3 Samples of TUI

We describe components of TUI and discuss about specific application. The main difficulty is to find the good artifact which allows convinient handling and symbolizes a right representation of the manipulated data. The closest related works are the system developed at MERL[1] Laboratory [1] and the "Active Cubes" [10].

2.3.1 Application of Merl

An application developed within the laboratory of Mitsubishi is based on the principle of combination of part (figure 2). In [1], they present *physical modules which describe, interpret and decorate the structures in which they are assembled* . Here, the *containers* are blocks of style LEGO^{TM} which determine and communicate their own *structure* with a computer once the finished combination. A software based on rules interprets these *structures* like a construction (building), analyzes their architectural devices, then adds the geometrical details and the decorative elements (for example: texture). The recovery of the geometry 3D is then reduced to the problem to determine the identity and the connection of the blocks and to communicate this information with a principal computer. The three principal problems are : the connection, communication and the duration of the estimation of the geometry.

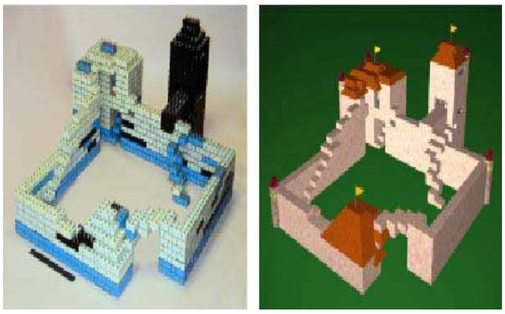


Figure 2: Merl project
(reproduce from [1])

2.3.2 ActiveCube

The "Cognitive Cubes" [16] is describe by this author like an application of ActiveCube, a Lego^{TM}-like TUI for the description of 3D form. The TUI consists of a set of plastics cubes (all 5 cm/edge) that can be connected to one another using simple male-female connectors on each face, forming both a physical form and a network topology. Each cube and cube face has a unique ID. A host PC is connected to a special base cube and communicate with the small CPUs in each cube. Since all cubes have the same size. Some ActiveCube have environment sensors (light, detector of obstacle, etc.) which increase the possibility of interaction. For instance, to approach one hand close to a sensor of some ActiveCube decreases the lighting of the structure. The intuitive use becomes less significant, just as the cognitive gap between the perception of the artifact and the perception of the data. We think the artifact lost the concept of " intuitive interface", because action of sensor can't be think ahead with the artifact form.

2.3.3 Other applications

There are also applications using of the specific and different tangible artifacts which are convert by a system of camera into a digital information and an action for the numerical

---

[1] MERL: Mitsubishi Electric Research Laboratory

data. Thus, this 3D model allows of "data life" to the artifact. Indeed, the 3D model has the physical properties of the artifact and thus modifies the numerical data such as for example in "Luminous Room" of MIT ([22]). In this application the artifact are blocks which represent mirrors with various optical properties (prism, mirror, semi-reflective mirror, etc). Once the combination created, the computer traces, actually increased, the advance of a luminous ray according to the artifact (mirror, prism). Another application within the framework of European project BREVIE (bridging real and virtual with graspable to use interface) presented in [15], [6] and [2] based on the principle of the realization of a digital 3D model with real object. The goal of this project is to allow an automatic transition between two distinct worlds for the students: the physical pneumatic circuit and the draftsmanship (3D representation).

The two first applications make it possible to combine artifacts by connected electronic and manipulate the data in intuitive way. The combination by means of connectors does not seem to us relevant, because it induces too many restrictions. Indeed, users can assemble two artifacts only face to face. The two last are TUI with artifacts closer than real object that the artifacts of ActiveCube are. Thus, in our sense, all these TUIs are not conceived and are not directed for the combination of parts. We think there is a potential to create a new interface based on TUI if we find suitable artifact.

3. ESKUA: PROPOSAL

ESKUA (figure 3) is a TUI which increase the ease of assembling and handling a combination of parts. ESKUA means "the hand" in Euskara, the Basque Language, and is the french acronym of "Exprimentation d'un Système Kinésique Utilisable pour l'Assemblage".

Figure 3: Our idea of ESKUA.

3.1 System Description

We use the term "interacteur" to define our artifacts. We choose this term because an interacteur is an object to do interaction (inter + to act on). Afterward, to describe our artifact we use the term *interacteur*. ESKUA is composed of interacteurs, a video capture system and work area. The actions which the user carries out on interacteurs (displacement, combination and rotation) are reproduced in 3D on the display screen. The capture of the position and the orientation of the set of artifact is based on a system of video capture. Its low cost and its upgrading capabilities (a number of cameras, the liberty for the choice of interacteur forms, identification with colors) seem to us very interesting assets.

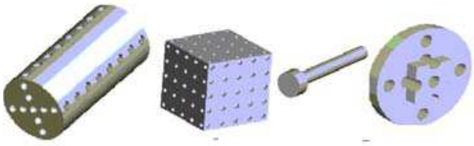


Figure 4: Samples of interacteurs : cylinder, parallelepiped, screw, graft.

3.1.1 Interacteurs

ESKUA associates each interacteur with one or a set of virtual objects. Our interacteur can be defined as "figurative", because their forms are primitive like cylinder or parallelepipeds (figure 4). They symbolize for the user a more complex virtual objects. The use of interacteur raises one question : "Why we associate an object to a concept and vice versa? ". Here, the concept is defined as [14] *"any unit of thought; a mental image formed by generalization"*. Indeed, we think that the concept makes it possible to associate a virtual object and real object. Thus the concept is the link between an interacteur and a 3D model. It seems to us [9] that during the fast identification of an interacteur we analyze the outline which is consisted of its form and its size. That's why we chose two criteria to design are interacteur : form and size.

We show in [12], that even in a complex application area (see part 3.2) two forms are rich enough for combinations as far as interacteurs of different sizes are provided. The user can add to an interacteur a surface provided with guideway or alignment, called *graft*, in order to modify the way he gathers (figure 4). The addition of graft makes it possible to limit once the number of interacteurs and secondly to offer several solutions for the assembly of parts.

A study leads us to define our interacteurs with the following characteristic. In order to be combine, there are many piercings in each interacteurs. Actually, they are proposed in three sizes : Small, Medium and Large (see table in figure 5). Our interacteurs are designed with Catia V5, CAD software from Dassault System [3]. They are machined by four axes machining centres and are made in isotropic material with a 0,65 g/cm^3 volume mass, a polyurethane resin (KUVO-15040). It's a good deal between cost and two mechanical properties important for us : small weight and wear resistant.
Here is an accurate description of our interacteurs. The size Small, Medium and Large are respectively 42, 70 et 98 millimeters in diameter for cylinders and edges of 42, 70 et 98 millimeters for the front of parallelepiped. The different depths are 14mm, 28mm, 42mm (only for the small size of parallelepiped), 56mm, 70mm (for the medium size of parallelepiped) and 98mm. The piercing is 6mm in diameter. Piercings are space out 14mm.

3.1.2 Video Capture

The user is allowed to associate any of interacteur to a virtual 3D model. Thus, all handling on interacteurs (rotation, translation) is reproduced on the 3D models. For motion capture, we intend to use model-based systems. In [5], the authors use a hand model in order to capture the hand

Size	Depth	**Cylinder**	**Parallelepiped**
		Name	Name
Small (42mm)	14mm	$C_{s,1}$	$P_{s,1}$
	28mm	$C_{s,2}$	$P_{s,2}$
	42mm		$P_{s,3}$
	56mm	$C_{s,4}$	$P_{s,4}$
	98mm	$C_{s,7}$	$P_{s,7}$
Medium (70mm)	14mm	$C_{m,1}$	$P_{m,1}$
	28mm	$C_{m,2}$	$P_{m,2}$
	56mm	$C_{m,4}$	$P_{m,4}$
	70mm		$P_{m,5}$
	98mm	$C_{m,7}$	$P_{m,7}$
Large (98mm)	14mm	$C_{l,1}$	$P_{l,1}$
	28mm	$C_{l,2}$	$P_{l,2}$
	56mm	$C_{l,4}$	$P_{l,4}$
	98mm	$C_{l,7}$	$P_{l,7}$

Figure 5: Table of interacteurs.

movements. Given a hand model in a starting pose and an input image, a model-based algorithm will make the model gradually converge to a final hand pose. We want to adapt this approach in our system. The interacteurs don't lose their forms contrary to the hand, but they are move in the space. Thus, the difference between two captured images are the translatory motion (left/right, front/back) and rotations. However, it doesn't provide enough informations to get the orientation of this interacteur. For example : a rotation of 90 degrees between two captured image is not visible. To adapt this technique for interacteurs, we will use piercing as a texture to capture more informations. Finally, we will use marks, by drawing symbols on each face, in order to recognize easily faces and their orientations.

3.1.3 Work area

We have to study and design this material part of our system. Today, we imagine it as shown figure 3.

3.2 Application area

In previous work [12], we have shown that it is interesting to tackle the general problem, to first concentrate on a particular field, in our case, mechanical design. In this specific application field we have verified that our platform concept makes it possible for a CAD designer to carry out the combination of CAD parts of a product. The proposed working environment directly confronts the designer with constraints of combination/assembly which are usually occulted by the functionalities of existing CAD software. For example, the difficulties of setting in relative position of two parts before fixing or the difficulties of insertion of a part compared to the others such as the inaccessibility or the collisions will be potentially identifiable by the designer during his handling. We think that the handling of physical objects makes it possible to bring back the combination run of the product in the real world and leads the designer to raise questions in a "natural" way by carrying out the gestures related to the assembly. To bring closer the user of ESKUA of the real activity practise by the fitter, we propose clamping system between interacteurs who are representative of the various existing technical solutions. For that, we propose (figure 4, third drawing) various sizes and types of fastenings like nut, screw, stud, spring retaining ring, etc.

The interacteurs symbolizing the parts are bored in several places (figure 4) in order to allow their combination by the preceding fastenings. With ESKUA, designer can carry out his assembly by allotting a type of interacteur to one or more parts CAD, and by handling these physical objects to carry out the assembly of the product. So the user is confronted with the real constraints of the operations of assemblies such as, for example, the difficulties of setting in relative position of the parts, maintains it in a joint way of certain elements. To simulate a such difficult task, people think ESKUA requires plenty different forms of interacteur. But, we show in [12] that two forms can be enough, parallelepiped and cylinder, for assembling. This proof is based on the analysis and description of DFA methods. In addition, certain complex parts can be represented by several interacteurs assembled between them.

4. CONCLUSION AND FUTURE WORK

In this paper, we have presenting our idea of a TUI for handling and fit together 3D models. Thus, we propose ESKUA. ESKUA means "the hand" in Euskarian, the Basque language, and is the french acronym of "Expérimentation d'un Système Kinésique Utilisable pour l'Assemblage". For now, we are working on the engineering achievements (hardware and software) of ESKUA in order to confront our ideas to the reality. We will propose a prototype version (few interacteur and a beta version software) dedicated to the CAE Domain.

5. REFERENCES

[1] D. Anderson, J. Frankel, J. Marks, A. Agarwala, P. Beardsley, J. Hodgins, D. Leigh, K. Ryall, E. Sullivan, and J. Yedida. Tangible interaction graphical interpretation: A new approach to 3d modeling. In *siggraph*, 2000.

[2] F. W. Bruns. Integrated real and virtual prototyping. *Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (IECON 98), Aachen, Germany.*, 1998.

[3] D. S. Catia. Available on World Wild Web [may the 28th 2002]. http://www.catia.com/.

[4] A. Cho, R. Lederer, E. Sharlin, B. Watson, L. Liu, and S. Sutphen. Designing virtual worlds ? the cognitive map probe. In *Proc. 13th Western Canadian Computer Graphics Symposium (Silver Star, Canada, March)*, 2002.

[5] J.-B. de la Rivière and P. Guitton. Hand posture recognition in large display VR environments. In *Gesture Workshop*, April 2003.

[6] H. Ernst, K. Schfer, and W. Bruns. Creating virtual worlds with a graspable user interface. *Interactions in Virtual Worlds, Proceeding of the Twente Workshop on Language Technology, TWLT 15, Enscheide: Universiteit Twente, S*, pages 45–57, 1999.

[7] B. Frölich, J. Plate, J. Wind, and al. Cubic-mouse-based interactaction in virtual

environments. IEEE *Computer Graphics and Applications*, pages 12–15, juillet aout 2000.

[8] P. Fuchs, G. Moreau, and J. Papin. *Le traité de la réalité virtuelle*. Les Presse de l'École des Mines, novembre 2001.

[9] L. Garreau and N. Couture. Quel est le bon compromis entre la maquette et l'interacteur unique et symbolique ?. Note interne, LIPSI-ESTIA, juillet 2003.

[10] Y. Kitamura, Y. Itoh, and F. Kishino. Real time interaction with activecube. *CHI 2001 Extended abstracts*, 2001.

[11] K. Kiyokawa, H. Takemura, and N. Yokoya. Seamlessdesign for 3d object creation. *IEEE MultiMedia*, 7(1):22 – 33, January 2000.

[12] J. Legardeur, L. Garreau, and N. Couture. Des interacteurs pour l'assemblage mécanique en cao. *3rd International Conference: Integrated Design and Production, CPI'2003, Mekns, Maroc*, 2003.

[13] J. Pierce, R. Pausch, and B. Stearns. Voodoo dolls: Seamless interaction at multiple scales in virtual environments. In *Symposium on Interactive 3D Graphics*, 1999.

[14] O. québécois de la langue franaise. Grand dictionnaire terminologique. Available on World Wild Web [may the 16th 2002], *http://www.granddictionnaire.com/_fs_global_01.htm*.

[15] K. Schmudlach, E. Hornecker, H. Ernst, and F. Bruns. Bridging reality and virtuality in vocational training. *CHI 2000 Extended Abstracts (Interactive Poster).*, pages 137–138, 2000.

[16] E. Sharlin, Y. Itoh, B. Watson, Y. Kitamura, L. Liu, and S. Sutphen. Cognitive cubes: A tangible user interface for cognitive assessment. In *ACM CHI 2002 Conference Proceedings*, pages 347–354. ACM Press New York, NY, USA, April 20-25 2002.

[17] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a WIM: Interactive worlds in miniature. In *Proceedings CHI'95*, 1995.

[18] B. Ullmer and H. Ishii. The metadesk: Models and prototypse for tangibles user interfaces. In *Proceedings of UIST'97*, October 1997.

[19] B. Ullmer and H. Ishii. Emerging frameworks for tangible user interfaces. *IBM SYSTEMS JOURNAL*, 39(3& 4):915–931, 2000.

[20] B. Ullmer and H. Ishii. Emerging frameworks for tangible user interfaces. *Human-Computer Interaction in the New Millennium*, pages pp. 579–601, 2001.

[21] B. A. Ullmer. Models and mechanisms for tangible user interfaces. Master's thesis, University of Illinois, 1995.

[22] J. Underkoffler, B. Ullmer, and H. Ishii. Emancipated pixels : Real-word graphics in the luminous room. In ACM, editor, *Proceedings of SIGGRAPH'99*, August 1999.

[23] C. Ware and J. Rose. Rotating virtual objects with real handles. *ACM Transactions on CHI*, 6(2):162 – 180, 1999.

Exploiting Physicality: linking action and effect

Sara Price and Paul Marshall
Interact Lab, School of Cognitive and Computing Sciences
University of Sussex
Brighton, BN1 9QH, UK
+44 (0) 1273 873179
sarap, paulma@cogs.susx.ac.uk

If the real world becomes the user interface, in what ways can we use it to create new forms of interaction? We are interested in how novel technologies might be used to support learning, specifically in the cognitive benefits of using one kind of interface or another. Two approaches that have been taken are the tangibles approach, where objects in the real world can be digitally augmented and manipulated to produce digital effects, and sensor-based interaction, where people in the real world can be identified and tracked, and serve as the trigger themselves for digital events to take place. From a technical perspective these two approaches have much in common: sensors are used to track the location or state of an object or person, and this information is used to send information relevant to that state or location. However, the benefit in terms of users' understanding of these systems is as yet unclear.

A number of theorists have argued that productive learning can result from a cycle between engaged situated action [2] and more objective reflection [1, 4]. Intuitively, it might be expected that real world interfaces could support this type of learning well: the learner's primary focus can be on the real world, providing the facility to really contextualise learning, but at the same time to augment and enhance it, providing reflection prompts or further information in a timely fashion.

An interesting issue that we perceive is in how readily the learner understands the link between physical action and digital effect. Learners may less readily understand the link when it is triggered by a change in location in a large-scale environment, than by, for example, manipulating an object on a table-top tangible interface. As part of the Ambient Wood project, which aimed to explore the potential of technology to augment the information available on an ecology field trip [3], we investigated some of the issues related to this question.

Initially we had envisaged delivering information about local flora and fauna to children engaged in exploring a large-scale woodland environment via handheld computers [3]. However, it became apparent in a trial run of the experience that children did not readily make the link between information presented to them and the real objects in the world: simply because the children were standing next to a tree in the wood didn't mean that they were currently interested in that tree, and the information delivered to the handheld relating to it was frequently ignored.

In contrast, a probe tool designed to allow the children to measure light and moisture levels in the wood allowed a more direct and active link between physical action and digital effect (an abstract representation of light or moisture level). By recording the location where each reading was taken, we were also able to present children with a representation of the readings taken in different habitats in the wood, allowing comparison and facilitating reflection about the about the relationships between light and moisture levels and the local plant life. Thus, the relationship between physical action and digital effect occurred at two distinct levels with this tool. Immediate feedback while using the probe seemed to encourage greater levels of exploration, while the delayed feedback showing all readings together promoted reflective thought and discussion between the kids.

We pose the following questions for discussion:

- From a cognitive perspective, in what circumstances might it be beneficial for a learner to knowingly (deliberately) trigger an effect, and in what circumstances might an unknown trigger be beneficial?
- How might information more relevant to a learner's focus of attention be delivered in an exploratory activity?

REFERENCES

[1] Ackermann, E. Perspective-taking and object construction: two keys to learning. In Kafai, Y. and Resnick, M. eds. *Constructionism in Practice: Designing, Thinking and Learning in a Digital World*, Lawrence Erlbaum, Mahwah, NJ, 1996, 25-35

[2] Brown, J.S., Collins, A. and Duguid, P. Situated Cognition and the Culture of Learning. *Educational Researcher,* 18 (1) 32-42

[3] Rogers, Y., Price, S., Harris, E., Phelps, T., Underwood, M., Wilde, D., Smith, H., Muller, H., Randell, C., Stanton, D., Neale, H., Thompson, M., Weal, M.J. & Michaelides, D. Learning through digitally-augmented physical experiences: reflections on the ambient wood project, Equator IRC tech. report 2002

[4] Schön, D. *The reflective Practitioner: How Professionals Think in Action.* Basic Books, New York, 1983.

This paper was presented at "Physical Interaction (PI03) – Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003.
Further information and online proceedings are available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/

Interacting with the Disappeared Computer

Bert Bongers
Academy for Digital Communication
Padualaan 99 Utrecht, The Netherlands
bertbon@xs4all.nl

ABSTRACT

This paper describes an approach based on the decoupling of devices, functionality and the user interface, proposing a generic interface and mapping of real-world elements with the virtual world. The research project looks at interfaces for our technological environment in general, which can be called the electronic ecology or *e*-cology

1. INTRODUCTION

Due to the tendencies of increased networking and the continuing miniaturisation, the desktop computer is disappearing. In some situations (for instance when mobile) the computer has disappeared already. The physical presence of the appliance has shrunk to a point where all that remains is the *human interface*, or even beyond that point......

1.1 The Computer has Disappeared, Now what?

The danger is, as we have seen with other technologies in the past that have been miniaturised away, that when the computer disappears also the interface will disappear. The field of HCI research can take this as an opportunity. After all, the interface technology is extremely malleable and interfaces can be shaped taking the human (in)capabilities (both physically as well as mentally) as a starting point rather than the technology. Form follows function, not the technology - more than ever because the computer technology has virtually disappeared. It is sometimes said that the ideal interface has to be invisible (or disappeared?), but this is mainly a sentiment that stems from the frustration caused by interfaces that are badly designed (if at all) and are seemingly getting in the way. Generally computers do not do what the user wants, but what the engineers and designers *think* the users wants, or what the engineers and designers *want* the users to want. When the computer becomes ubiquitous the danger is that this misunderstanding also becomes ubiquitous. The need for a solid and understandable interface for ubiquitous computing is bigger than ever.

1.2 The Ubiquitous Interface

A spatial interface, such as used in what I call Interactivated Spaces, is a way of searching for solutions for the problem of how to control an invisible, ubiquitous system. Such an interface can be a combination of speech recognition, gestural control, and tangible interaction elements that are placed in the space or worn by the user.

1.3 A shift in thinking: from devices to functions

For several years there has been a tendency in technological developments towards the disappearance of devices, the functions of which are then incorporated in the remaining appliances. These resulting multifunctional appliances are therefore harder to operate. The tendency of the increased networking of appliances results in functions disappearing into the network. An example of this is the "voice mail box", storing messages somewhere in the network instead of on a tape or chip in an answering machine in the home. This has certain advantages, but the problematic issue is that the *interface* of the old answering machine, which gave access to the functionality of voice mail, has disappeared. Now, the functionality needs to be operated with an interface that was never designed for this - "to delete this message, press 5". This results in cumbersome switching between modes and modalities, instead of just having a "delete" button at hand.

The research project is making an analysis of this technological environment based on functionalities rather than based on the devices (which change and often have disappeared). This research attempts to separate functionality and technology.

1.4 The Generic Interface

Approaching our technological environment as outlined above, it becomes possible to create an overview of functionalities (whether in devices or in virtual, networked environments) and interfaces. The interface usually affords a two-way interaction: it has an input channel (buttons, dials) and an output channel (displays). A number of 'medium-independent' functions can be identified which are applicable to all sorts of content: for instance a "play" button that activates a message in a voice mail functionality, or plays music track in a CD playing functionality. In the design of the generic interface some medium-specific controls need to be implementable too. The generic interface is a contrast with the present situation. Currently, one may be walking around with a mobile telephone (interface: a few buttons, small display), a PDA (very few buttons, pen input, larger display), a laptop computer (trackpad, keyboard, even larger screen), a walkman or CD player (buttons, dials, headphones), a watch (tiny buttons, small display), et cetera. It is clear that there is a lot of overlap in the interfaces, which is the tradition. The strong point of this is, as seen from the user, that there is a fixed *mapping* between interface elements and functionality. When devices disappear, and a generic interface remains, this mapping needs to be designed and built in a different way, without losing the clarity and transparency.

1.5 Mapping of functionalities and interfaces

The focus of the research is on the development and testing of interfaces and interaction styles that link the functionality, content and control (mapping). Part of the approach is based on using real world objects, their (virtual) affordances and beacons, gesturally by pointing and linking. Objects and processes that are outside the field of view can be represented by icons, metaphors and maps. The articulations in this gestural input will be supported by a rich, multimodal feedback.

This paper was presented at "Physical Interaction (PI03) Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003.

How the move to physical user interfaces can make human computer interaction a more enjoyable experience

Rowanne Fleck
Interact Lab, University of Sussex, Brighton UK
rmmf20@sussex.ac.uk

1. INTRODUCTION

As the interface moves away from the WIMP paradigm and becomes more physical or tangible, the field of HCI is expanding to include alongside usability goals, user experience goals [8]. There are many aspects of these new physical interfaces that, in addition to other benefits over WIMP interfaces, I argue will provide more pleasurable interactive user experiences. Below are just a few.

Physical interfaces can offer a direct relationship between information and control, and even allow direct control of virtual objects through physical objects. Laurel discusses how traditionally interaction technology interfaces are considered as intermediaries between the person and what they want to achieve [7]. In contrast, in order to promote engagement in interaction she argues the system should provide the user with interactive first-person-ness, allowing them to act more directly. A physical interface could achieve this. Having good control is rated by people as one of the important aspects of a pleasurable product [3]. Schneiderman [9] suggests that people like to have control of their interactions as it gives them a sense of power over the system.

Physio-pleasure, to do with the body and sense organs, is one of the four types of pleasure involved in pleasurable experience [6]. An important part of the experience of a product is in feeling or touching it, the satisfying clunk of the car door shutting, the smell of a new magazine. Another of the four pleasures, ideo-pleasure highlights the importance of how something looks, its aesthetics. In the world of physical interfaces the look and feel of interaction can be taken to new levels.

The theory of flow describes situations in which optimal experience can be achieved [2]. One important factor is the provision of immediate feedback to actions; this is of course important for usability in any type of system. With a physical interface it is possible to provide feedback in a variety of modalities. An important aspect of a tangible interface as described by Ishii and Ullmer [5] is the seamless integration of representation and control. The physical state of the system partially embodies the underlying digital state allowing the user to feel and to see in a 3D environment the state of the system. The games industry already utilises this introducing more physical controllers and even providing haptic feedback.

People's attention is automatically drawn to things which are novel in our environment. We form schemas or expectations about what might happen next given the context which allows us to prepare. We are surprised if this expectation is not met, or uncertain if more than one expectation is aroused at once. Berylne [1] suggests this raises our arousal (readiness to react) levels and that slight transitory jumps in arousal can be pleasant because of the relief felt afterwards. This can be exploited in physical interfaces by coupling familiar physical objects with unexpected or unusual digital capabilities. Ambiguity can be introduced into design by placing something out of context or not presenting complete information about what the system knows or is displaying [4]. This can produce "intriguing, mysterious and delightful" results and encourage close personal engagement with the system.

For any interface the goal of the user is the most important thing. In an extension to Maslow's hierarchy of needs, [6] suggests that once we have functionality, then we will want usability and once we have that we will want it to be a pleasurable experience. By enabling more direct control and feedback on a number of levels, providing pleasure through the look and feel of the physical aspects of the interface and with the possibility to intrigue, it seems clear to me that the physical user interface can help take us the step beyond usability.

2. REFERENCES

[1] Berlyne, D.E., *Conflict, Arousal and Curiosity*, McGraw-Hill, 1960.
[2] Csikszentmihalyi, M., *Finding flow: the pschology of engagement with everyday life*, Basic Books, New York, 1997.
[3] Flyte and Nilsen, Understanding attributes that contribute to pleasure in product use. In W.S. Green and P.W. Jordan (Eds.), *Pleasure with products: Beyond Unsability*, Taylor & Francis, London, 2002, pp. Chapter 19.
[4] Gaver, W.W., Beaver, J. and Benford, S., Ambiguity as a resource for design. *CHI (in submission)*, 2003.
[5] Ishii, H. and Ullmer, B., Tangible bits: Towards seamless interfaces between people, bits and atoms. *Proceedings of CHI '97*, ACM, Atlanta, GA USA, 1997.
[6] Jordan, P.W., *Designing Pleasurable Products*, Taylor & Francis, London, 2000.
[7] Laurel, B., Interface as Mimesis. In D.A. Norman and S.W. Draper (Eds.), *User centred system design: new perspectives on human-computer interaction*, Erlbaum, Hillsdale, 1986.
[8] Preece, J., Rogers, Y. and Sharp, H., *Interaction design: beyond human-computer interaction*, John Wiley, Chichester, New York, 2002.
[9] Scheiderman, B., *Designing the user interface: Strategies for effective human-computer interaction*, 3rd edn., Addison-Wesley, Reading MA, 1998.

This paper was presented at "Physical Interaction (PI03) –Workshop on Real World User Interfaces", a workshop at the Mobile HCI Conference 2003 in Udine (Italy). September 8, 2003. The copyright remains with the author. Further information is available at http://www.medien.informatik.uni-muenchen.de/en/events/pi03/