Designable Visual Markers for Mobile Human-Computer Interaction

THÈSE

PRÉSENTÉE À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

par

Enrico Costanza

MEng Electronic and Communication Engineering, The University of York, UK
MS Media Arts and Sciences, Massachussets Institute of Technology, USA

de nationalité italienne

Jury:

Prof. Jean-Pierre Hubaux, président de jury
Prof. Jeffrey Huang, directeur de thèse
Prof. Pierre Dillenbourg, rapporteur
Prof. John Robinson, rapporteur
Prof. Tom Rodden, rapporteur

Lausanne, EPFL
2010

Abstract

Visual markers are graphic symbols designed to be easily recognised by machines. They are traditionally used to track goods, but there is increasing interest in their application to mobile human-computer interaction (HCI). By scanning a visual marker through a camera phone, users can retrieve localised information and access mobile services. In particular the dissertation examines the application of visual markers to *physical tagging*: practices of association of digital information with physical items. One missed opportunity in current visual marker systems is that the markers themselves cannot be visually designed; they are not expressive to humans, and thus fail to convey information before being scanned.

To address this limitation, this dissertation introduces the idea of *designable markers*, visual markers that are both machine-readable and visually communicative to humans, and presents an investigation of the ways in which they can support mobile human-computer interaction. The application of designable visual markers to the creation of mobile interfaces is explored through a variety of methods: through formal usability experiments, through the creation and analysis of example designs, as well as through the qualitative analysis of two field trials. All three approaches were enabled by the engineering and development of d-touch, an actual recognition system that supports designable visual markers and by its integration in a variety of applications and experimental probes.

D-touch is based on image topology, and its markers are defined in terms of constraints on the nesting of dark and light regions. The constraints imposed by d-touch are flexible enough to allow novice users to create markers which are visually expressive and at the same time machine readable. A user study demonstrates how such system enables people to design their own functional visual markers, determining their aesthetic qualities and what they visually communicate to others. A desktop application to support users in the creation of valid markers, the *d-touch analyser*, is presented and its usefulness is demonstrated through the same study.

A formal usability experiment comparing five variations of marker-based interfaces on keypad and touch-screen phones shows that all of them allow users to reliably select targets within, on average, less than 4 seconds. Participants of the experiment reported a strong preference for interfaces that involve only marker scanning, compared to those that require a combination of marker scanning and key-presses or touch selections.

Example designs of mobile interface generated by the author as well as others are presented to expose how the d-touch recognition system can be integrated in mobile applications. The examples illustrate a variety of ways in which markers can be used to augment printed materials such as cards, books and product packages, adding to them interactive capabilities. The examples show also different approaches to marker design, ranging from simple and recognisable iconic design, to symbols that integrate cues about the interactive functionality, to making them invisible by hiding them in existing graphics.

Finally, the dissertation reports and analyses two field trials conducted to study what practices of *physical tagging* can emerge from, and be supported by, the use of markers. The trials were centred around the use of uWiki, a functional prototype based on d-touch, that allows users to associate digital content to markers printed on physical tags that can be affixed to objects or buildings. Observations show that a variety of practices emerge around the use of this technology, indicating that they provide a rich medium that has potential to attract the interest of real users. Though the results of this work are preliminary, they serve to demonstrate the range of potential for the future of such systems.

Keywords Human-computer interaction, visual marker recognition, design, mobile devices, mobile HCI, fiducial recognition, user studies, usability, field trials, grounded theory, physical tagging, location-based systems.

Riassunto

I codici visuali, tra cui ad esempio i codici a barre, sono simboli grafici disegnati per essere facilmente riconosciuti in maniera automatica dalle macchine. Essi sono tradizionalmente utilizzati per tener traccia di prodotti nei magazzini, ma vi è un crescente interesse per la loro applicazione nel campo dell'interazione umana con dispositivi mobili. Tramite la scansione di un codice visuale attraverso un telefono cellulare dotato di fotocamera, gli utenti possono accedere ad informazioni e servizi specificamente legati al luogo o all'oggetto a cui il codice è applicato. Un limite dei sistemi esistenti per il riconoscimento dei codici visuali è connesso alla circostanza che i codici stessi non possono essere disegnati arbitrariamente e che non sono direttamente intellegibili dalle persone, a meno di non essere letti e tradotti da un dispositivo elettronico.

Questa tesi presenta l'idea di codici visuali facilmente identificabili dalle macchine ma al tempo stesso visualmente significativi per le persone, ed inoltre affronta lo studio della loro applicazione per la creazione di sistemi di interazione uomo-macchina in contesti mobili. L'applicazione di tali codici in campo di interazione uomo-macchina è esplorata in questa tesi attraverso una varietá di metodi: con esperimenti di usabilitá, con la presentazione di esempi di interfacce mobili, e tramite l'analisi di due prove sul campo. Tutti e tre gli approcci sono stati resi possibili dalla progettazione e lo sviluppo di d-touch, un sistema di riconscimento che supporta codici visualmente significativi, e dalla sua integrazione in una variet di applicazioni e prototipi.

D-touch è basato sulla struttura topologica delle immagini. I codici visuali d-touch sono definiti in termini di vincoli sull'annidamento di regioni chiare e scure dell'immagine. I vincoli imposti da d-touch sono abbastanza flessibili da consentire agli utenti di creare simboli che siano visivamente espressivi e nello stesso tempo riconoscibili automaticamente. Un esperimento dimostra come tale sistema consenta alle persone di disegnare codici visuali funzionali, controllando le loro qualit estetiche e ci che comunicano visivamente. La tesi presenta anche il *d-touch analyser*, un'applicazione desktop che supporta gli utenti nella creazione di codici validi; la sua utilitá viene dimostrata attraverso un esperimento.

Un esperimento formale di usabilitá ha messo a confronto 5 varianti di interfacce mobili basate su codici d-touch per la selezione di opzioni, su telefoni a schermo tattile ed a tastiera. I risultati dimostrano che tutte le varianti permettono agli utenti di effettuare selezioni in modo affidabile ed, in media, in meno di 4 secondi. I partecipanti allesperimento hanno espresso una forte preferenza per le interfacce basate sulla sola scansione di codici, piuttosto che sulla combinazione di scansione e pressione di tasti o selezioni su schermo tattile.

La tesi esamina anche vari esempi di interfaccie mobili basate su d-touch, progettate sia dallautore della tesi che da altri. Questi esempi illustrano una varietá di modi in cui i codici d-touch

possono essere utilizzati per arricchire materiale stampato, come carte da gioco, libri e confezioni di prodotti, aggiungendo ad essi capacit interattive. Gli stessi esempi mostrano anche diversi approcci per la progettazione ed il disegno di codici d-touch: si va da semplici e riconoscibili icone, a simboli che indicano visualmente la propria funzione interattiva, a simboli invisibili mimetizzati nella grafica pre-esistente.

Infine, per studiare quali pratiche possano emergere, e siano supportate, dall'uso di codici d-touch, la tesi riporta ed analizza due prove sul campo. Le prove sono state definite intorno all'uso di uWiki, un prototipo che consente agli utenti di associare, tramite telefoni cellulari, contenuti digitali a codici d-touch stampati su etichette applicabili ad oggetti o edifici. Le osservazioni effettuate durante queste prove rivelano una varietà di pratiche di comunicazione emergenti dall'uso di questa tecnologia, il che sta a significare come questa costituisca un vero e proprio nuovo *medium*, che ha potenziale per attrarre linteresse di utenti reali.

Sebbene i risultati di questo lavoro siano preliminari, servono tuttavia a dimostrare la gamma delle potenzialità offerta per il futuro da questi sistemi.

Parole Chiave Interazione uomo-macchina, riconoscimento di codici visuali, design, dispositivi mobili, usabilitá, studi di usabilitá, prove sul campo, servizi basati su localizzazione, nuovi media, comunicazione, telecomunicazione.

Acknowledgements

This work presented in this dissertation was possible thanks to the advice, help and encouragement that I received from several mentors, colleagues and friends. Therefore I would like to acknowledge and thank them here.

Prof. **Jeffrey Huang**, my advisor, for accepting me in the Media and Design Lab, for allowing me to pursue this exciting and original research topic, and for guiding and encouraging me during my time at EPFL.

My work on d-touch started when I was at the University of York, thanks to the guidance, support and trust of prof. **John Robinson**. It continued "on the back burner" during my 3 years at Media Lab Europe and MIT Media Lab thanks to the support of Prof. **Rebecca Allen** and Prof. **Pattie Maes**, and thanks to the encouragement of many colleagues and visitors at the two labs.

Prof. **Pierre Dillenbourg** and **Patrick Jermann** provided precious advice about my research, especially the quantitative approach. Prof. **Geraldine Fitzpatrick** and Prof. **Matthias Finger** guided my first steps in the field of qualitative research.

Tal Drori has been my mentor for the graphic design aspects of this work; she believed in designable visual markers sometimes even more than I did. She designed many of d-touch markers, and helped me to understand how designers generally think differently than engineers. She also designed or improved many of the images in this document.

The qualitative research reported in Chapter 7 was performed thanks to the collaboration, enthusiasm and guidance of **Johanna Brewer**. I am indebted to her also for the advice she provided about the editing of this document.

Giuseppe Costanza adopted d-touch for the interaction design work that he did as part of an MA in Communication Design, Digital Media at the Central Saint Martins College of Art & Design, London. His designs provided very valuable material for my research, as reported in Chapter 6.

The 'Class' field trial of physical tagging (Chapter 7) was possible thanks to **Mark Meagher** & **Julien Nembrini** who organized, with Prof Jeffrey Huang, the Architecture class, and engaged in many illuminative discussions before and after it. Thanks to **Morten Gisselbaek and the entire Îlot 13 community** for allowing us to work at their beautiful place.

The uWiki prototype described in Chapter 6 was implemented through the help of **Alok Parlikar** and **Joël Schintgen**. Joël also run the marker generation simulation reported in Chapter 4 as part of his master project under my supervision. The experiments reported in Chapter 5 were run by

François Suter and **Pedro Andre Cunha**, as part of his master project under my supervision. Franc̆ois and Pedro Andre also implemented large parts of the prototypes used in those experiments and parts of the Mobile Audio Cards application reported in Chapter 6.

Much of the work presented in this dissertation was informed by long, enlightening discussions with friends and colleagues across years and places. In particular: **Justen Hyde**, **Simon Shelley**, **Sam Inverso**, **Martin Kaltenbrunner Matteo Giaccone**, **Olivier Küng** and **Andoni Galdos** about the d-touch recognition system; **Mauro Cherubini**, **Denis Lalanne** and **Valentina Nisi** about the field trials of physical tagging (Chapter 7). I am also thankful to Denis for organising a winter school on usability engineering which was of great help for this work.

The Near Futures workshop (Chapter 6) was organised in collaboration with Valentina Nisi, **Arianna Bassoli**, Johanna Brewer and Tal Drori. I would also like to thank the **workshop participants**.

Many thanks also to other friends and colleagues who helped me, shared ideas, inspired and commented various parts of my work, especially: **Andrea Bianchi**, **Francesco Della Casa**, **Caroline Dionne**, **Seth Flaxman**, **Julien Freudiger**, **Hendrik Knoche**, **Vincent Le Petit**, **Ian Oakley**, **Jacques Panchard**, **Guillaume Zufferey**.

The work on physical tagging (Chapter 7) was inspired by the work and the discussions I shared with the Electronic Lens team at the MIT Mobile Experience Lab, especially **Prof. Bill Mitchell**, **Federico Casalegno**, **Miria Leinss**, **Jie-Eun Hwang**, **Sajid Sadi**, **Jon Gibbs**, **Aaron Zinman**, **Jae-Woo Chung** and **Oriol Lloret**.

The work reported in Chapters 3, 4 and 7 benefited from numerous suggestions provided by anonymous CHI reviewers.

This work would have not been possible without the great administrative and technical support of **Monique Amhof** and **Benjamin Barras** at EPFL.

Thanks to everyone who participated in the user studies and field trials.

Thanks to my family and my friends in Lausanne, Palermo, Milano and everywhere else for their constant love and support.

Finally and again, thanks to Tal, for the inspiration, support and patience.

Contents

1. Introduction

Systems that associate digital content with physical objects and locations through mobile devices have been discussed and demonstrated in the technology research community for about 20 years. With the recent rise in availability and popularity of mobile Internet access, the attractiveness of such interaction paradigms is even more compelling, as they may facilitate on-line access to information and services, despite the limited size of mobile phones. A number of technological and conceptual solutions have been proposed by researchers in industry and academia. Such systems have often been promoted as extension of the world-wide web paradigm to the real world; examples include: physical hyper-links (Pradhan et al. 2001), Internet of things (Sterling 2005) and augmented reality (Costanza et al. 2009b). One practical way to implement this link between digital and physical is through visual markers.

Visual markers are graphic symbols designed to be easily recognised by machines. They are used to relate physical objects to computer systems. Examples include the barcodes used on most commercial goods to keep track of stock in shops and warehouses, and more recent *2D-barcodes* (ISO 2006, 2000, de Ipina et al. 2002, Rohs and Zweifel 2005) which are easier to read with low-resolution cameras, such as those included in mobile phones. Beyond stock control, there has recently been interest in both the academic and commercial community for using this technology to enhance mobile applications and facilitate access to mobile services (Toye et al. 2007, O'Neill et al. 2007, Mäkelä et al. 2007). Visual markers[1] are often considered as an alternative to RFID tags. They are not exactly equivalent because visual markers are *read-only*, whereas some RFID tags can be rewritten, but they can support similar types of interaction with the advantage that they can be created with normal printers and accessed using standard photo cameras already available on most mobile devices. RFID requires special readers devoted only to this purpose, integrated only in a limited number of devices.

Similar to traditional barcodes, the shape of existing visual markers is solely based on maximising their readability by machines: they are not visually meaningful to people, and different markers of the same family are generally not easy to distinguish from one another by looking at them. In other words, one missed opportunity in current visual marker systems is that the markers themselves cannot be designed, they are not expressive to humans, and thus fail to convey information to people before being scanned. For example, information could be conveyed about the type of digital content or mobile service associated with the markers, or the project they belong to or the person or company who created the content. Further, an interesting marker design could simply

[1] Visual markers are also referred to as *visual tags* or *fiducial markers*. The term *marker* is preferred in this dissertation over *tag*, to avoid confusion with keywords and metadata *tags*, or other physical tagging technologies, such as RFID tags.

attract attention so that it would be scanned with a mobile device. We argue that it is fundamental to empower both application creators and end-users to visually design their own markers – giving visual markers the same visual dignity and expressive potential currently given to icons, and even opening up functional markers to the paradigm of user-generated-content.

This dissertation investigates the definition and use of *designable visual markers*: markers that can be created by users, controlling their aesthetic qualities and what they visually communicate to others, while being efficiently identifiable by a recognition algorithm running on mobile devices. In particular the analysis is focussed on the application of designable visual markers to the design of mobile interfaces that mix the physical and digital realms. The dissertation introduces d-touch, a functional recognition system that supports designable visual markers. An example d-touch marker, scanned by a user, is shown in Figure 1.1. By allowing the creation of markers that support interaction both visually and functionally, d-touch can enhance most applications normally supported by visual markers, including interactive guides, mobile service access, mobile games, interactive story telling systems and augmented reality applications that have broad visual appeal and are not constrained to unintelligible glyphs. Markers that are both functional and visually expressive can be easily produced by a wide spectrum of users, without specific training. The system can be used by professional designers as well as end users, enabling both groups to design markers as visual icons with high expressive or communicative power. Because markers are designable, end users and designers can consciously determine their look and feel, including the degree to which they are immediately recognisable as markers to be scanned. The design can range from icons that are obviously scannable (*explicit*) to ones that are hidden in the overall design and only accessible to a closed circle or upon closer look (*ambiguous*). For applications in

Figure 1.1: A user scanning a d-touch marker through a mobile phone.

which immediate user recognition of the markers is essential, designers may define conventions for the marker placement, e.g., markers may be placed at the bottom right corner of posters or below text in museum labels. Specific application scenarios for professional designers include the creation of highly polished, explicit visual markers that follow the design guidelines of corporate identity, or the incorporation of ambiguous markers in visual communication, such as ads, that are not recognisable at first glance. Application scenarios for end users include hand-drawn expressive visual markers left in the environment to leave location-specific information and traces. Markers could be used as hidden, secret symbols that are ambiguous and only noticed and scanned by an inner circle – echoing established urban phenomena such as graffiti and tags, as well as older hobo codes.

In the d-touch implementation, the creation of markers which are both machine- and human-readable is possible because the recognition is based on topological features of the markers rather than their geometry. Marker recognition is not based on shape, but on the relationship of dark and light regions. To validate the hypothesis that d-touch allows users to craft visually communicative markers, the dissertation reports a user study of the creation of markers, designed to understand how much people can visually express within the constraints imposed by the system. D-touch was initially developed for tracking objects in the context of tangible user interfaces and augmented reality (Costanza and Robinson 2003, Costanza et al. 2003a,b), and the visual appearance of early markers was optimised solely for resolution and recognition accuracy. The work presented in this dissertation is based on later developments that take advantage of topology-based recognition to allow a wider range of visual expression through the markers. The dissertation moves into exploring how designable markers can become part of mobile interfaces, presenting a number of design examples, a usability evaluation of a small variety of marker-based interfaces and more general field study of how markers support the association of digital information with physical items and places.

Practices of association of digital information with physical items and places, are a specific case of location-based systems[2]; they can be implemented through a variety of technologies, from RFID to 2D barcodes to even entering numerical codes found on signs into mobile devices, we refer to all of these as "physical tagging" systems. Examples include museum audio guides, that allow visitors to listen to pre-recorded commentary related to exhibits, as well as advertising campaigns in East Asia (Fowler 2005), and increasingly everywhere in the World, where consumers can scan 2D-barcodes embedded in street advertising posters with their camera-phones to retrieve information about the promoted product, or discount vouchers. However, these examples are rather far from current Internet practice: they are closed systems, where content is provided by authoritative actors – the museums or the advertising companies – and it is not part of a more general information ecology.

In contrast, this dissertation was driven by an interest for systems that allow end-users not only to passively access, but also to create, share and associate content to places and objects, and that address mobile Internet access alongside PC Internet. Given the role of user generated content (UGC) and social sharing in current Internet usage practice, we believe that addressing them in

[2]The systems under consideration are asynchronous, while there is also a group of location based services that implement synchronous communication, these aim at creating a shared awareness of users' location, for example for the purpose of coordination (Nova et al. 2006, Dearman et al. 2005) or for real-time artistic and entertainment applications (Tanaka and Gemeinboeck 2006, Vogiazou et al. 2006, Benford et al. 2009a), which are not considered here.

the design and study of systems for the distribution of digital information in physical space is fundamental. A combined mobile and PC internet approach is well demonstrated by commercial actors (such as Google, MySpace and Facebook) complementing their popular web interfaces with mobile applications that facilitate access to their services. Recently this trend also started also to be addressed by HCI researchers (Milic-Frayling et al. 2007). Understanding emerging practice derived from use of such systems, including the kinds of applications they support, is key to the design of future applications and services.

Visual markers enable the development of research probes for physical tagging systems: mobile applications running on consumer-grade phones, without hardware additions or modifications, working with symbols produced with common ink-jet printers (compared for example to RFID technology). Therefore, these relatively low-cost probes made it possible to study physical tagging, especially the aspect of content creation and dissemination which, as discussed in Chapter 2, still largely represents a research gap in the HCI literature.

1.1 Research Questions

In summary, this dissertation attempts to answer the question: **how can designable visual markers support mobile human-computer interaction, and especially the paradigm of associating digital information with physical objects and locations?** In particular the question is assessed through the use of the d-touch topology-based marker recogniton. The problem is decomposed in different smaller ones.

First, the hypothesis of "d-touch designable markers" is in itself verified: does topology-based recognition make it possible to define visual markers that can be visually expressive to humans, while being easy and efficient to identify by machines?
Second, if it does, can the marker design activity be supported by software tools?
Third, In which new ways can designable visual markers be integrated in mobile interfaces? How do different marker-based interface designs affect the users' performance and preferences?
Fourth, do visual markers support physical tagging? And what kind of practices emerge?

1.2 Methods

According to the ACM SIGCHI, *"Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them."* (Hewett et al. 1996). The multidisciplinary nature intrinsic in this definition is fully reflected in the methods adopted in this dissertation. Different methodological tools were adopted to address the specific research questions listed in the previous section, choosing each time a different balance between precise measurements and ecological validity.

Controlled or semi-controlled laboratory experiments, inspired by investigation methods from cognitive psychology and typical of the usability tradition (Shneiderman and Plaisant 2004, p.144-150), were conducted to verify that d-touch markers can be designed with minimal training, and to test the hypothesis that software tools can be crafted to support users in this design activity. Lab studies were also used to compare and characterise different marker-based mobile interfaces, in terms of task completion time and accuracy. In the case of marker design, the control is only partial, as the creative component of the task makes it highly subjective.

The integration of d-touch marker recognition into mobile applications is examined through actual design examples. These were partially produced by the author, and partially by others, for example within the context of a mobile service design workshop. The approach of analysing existing designs, or using design to conduct research is more typical of the *design discipline*, as it is evident, for instance, in the book "Designing Interactions" (Moggridge 2007).

Qualitative analysis methods were employed to study the phenomenon of physical tagging through d-touch markers. Field observations and their analysis through qualitative research methods, inspired by social science research (Miles and Huberman 1994, Silverman 2006, Corbin and Strauss 2008) are increasingly popular in the field of HCI (Preece et al. 2007, pp.373-385). Qualitative methods were chosen for this part of the work because its nature is fundamentally *exploratory*, and it was deemed not (yet) possible to formulate testable research hypotheses. Moreover, the very nature of the phenomenon under observation and the particular focus on user-generated content per-se requires interpretation. Finally, qualitative methods were found particularly suitable for working with a small number of participants, over a limited period of time.

Software engineering and development are the overarching means that enabled the research presented in this dissertation. Besides the marker recognition system itself, obviously at the centre of the work, interactive applications were designed and implemented around it, all had to be robust enough to withstand real usage. These systems were generally developed as research tools, integrating support to monitor their usage through the collection of interaction logs. Symbian S60 was the platform of choice for all mobile applications developed for this dissertation. This choice was determined by openness and availability: when this work started, S60 was one of the few mobile platforms that freely allowed the development of native applications[3] that could be installed on any commercially available compatible phone, through development tools openly available on the Internet, without the requirement to special agreements with the device manufacturers. Moreover, S60 was chosen because the form factor of most devices is similar to that of low-cost phones, which makes it easier to let users perceive the experimental marker-based applications as running on "normal" phones rather than special esoteric devices, and limit the *wow-effect* that may be related to it. Fore example, Fleck et al. (2002) report that during the evaluation of a mobile applications users were observed to pay more attention to *"the novelty of the hardware rather than system functionality."* However, it must be noted that the arrival on the market of the Apple iPhone in 2007 made touch screen phones widely more popular than they used to be.

1.3 Outline

Chapter 2 – Related Work – The chapter starts by reviewing other visual marker recognition systems. These are generally based on geometrical features, which makes the shape of the markers completely or heavily constrained. An overview of research related to mobile interaction techniques based on visual markers and RFID tags is then presented, followed by a discussion of

[3] as opposite to J2ME applications running in sand-boxed virtual machines

location-based and physical tagging applications and their study through field trials. The chapter is concluded with references to other forms of information dissemination in space, including signage design and street art.

Chapter 3 – Designable Visual Markers – The chapter describes the d-touch marker recognition system, discussing the design constraints it imposes on markers. The system implementation in multiple platforms is presented, followed by the description of the *d-touch analyser*: a desktop application to support the design of valid d-touch markers.

Chapter 4 – Markers Design Study – The chapter presents an evaluation of the creation of d-touch markers, designed to understand how much people can visually express within the constraints imposed by the system. During the study pairs of novice users generated between 3 and 27 valid markers within one hour of being introduced to the system, demonstrating its flexibility and ease of use.

Chapter 5 – Marker-based Interaction – The chapter reports the design and evaluation of a small variety of marker-based mobile interfaces. The interfaces allow the selection of objects and actions through the combination of marker-scanning, key presses and the use of touch screen. Prototype implementations were tested by 42 users to compare performance in terms of selection accuracy and task completion time, as well as user subjective preferences.

Chapter 6 – Marker-based Mobile Applications – The design of multiple mobile applications based on the d-touch marker recognition system is presented. The work includes design by the dissertation author as well as others, including examples generated in a workshop.

Chapter 7 – Physical Tagging – The chapter reports the analysis of two field trials where users created digital content and associated it with objects and locations using a research prototype that allows physical tagging based on d-touch marker recognition providing both mobile phone and web interfaces. The trials were organised to stimulate diverse conditions of use: the first one took the form of a collaborative annotation task, performed by architecture students, while the second was a user-generated campus tour guide.

Chapters 8 and 9 – Future Work and Conclusion – The last chapters summarises the contributions of the dissertation and outlines opportunities for future work.

1.4 Contributions

This dissertation introduces the idea of *designable markers*, visual markers that are both machine-readable and visually communicative to humans, and presents an investigation of the ways in which they can support mobile human-computer interaction. Specific contributions include:

- an experimental validation confirming that topology-based recognition allows for designable visual markers, and that markers that are both functional and visually expressive can be easily produced by a wide spectrum of users, without much training;
- the design, implementation and evaluation of a software tool that supports users in creating valid markers;

- a study-based comparison of a small variety of marker-based mobile interfaces, illustrating how interface variations affect usability;
- an overview and discussion of mobile application designs demonstrating the ways in which designable markers can be adopted for mobile interaction, including the design and implementation of prototypes enabling study of practices emerging around their use and the assessment of the interfaces' usability;
- the analysis of different physical tagging practices supported by visual markers, as observed in two field trials.

2. Related Work

This chapter provides an overview of related work in the areas of visual marker recognition, mobile interaction through visual markers and related technologies (e.g., RFID), location-based media, and in particular its creation. The chapter is concluded with references to existing practices of digital and non-digital forms of content dissemination in public space, from signage design to street art, to media art projects. Visual markers are also sometimes used for the implementation of Augmented Reality (AR) – a family of systems that let users interact with virtual three-dimensional objects precisely aligned to the physical environment, with the goal of giving the impression that virtual and physical objects co-exist in the same space (Azuma et al. 2001). The work presented in this dissertation does not share this goal, so the field of AR is deliberately not reviewed here. A review of AR can be found in the book "Mixed Reality: A Survey" (Costanza et al. 2009a).

2.1 Visual Marker Recognition

Visual marker recognition can be considered as a special case of *object recognition*, where the objects, i.e., the markers, are designed in conjunction with the recognition algorithm in order to optimise performance, achieve high recognition rate, low false positives and high processing speed. Barcodes are probably the earliest example of visual markers, dating back to the 1950s (J. and Silver 1952). They started to be used commercially at the end of the 1960s and are still used in most shops. Barcodes are designed to be read through high resolution scanners, and they are generally not easy to decode through standard consumer grade webcams and camera phones, at least without special macro lenses.

A more recent generation of visual markers, which can be more easily read by low cost imaging cameras has been presented over the last twenty years. Proposed applications include mixed reality systems (Rekimoto 1997, Kato and Billinghurst 1999), video post-production (Thomas et al. 1997, Johnston and Clark 2003) and, more recently, human-computer interfaces for cameraphones (Mohring et al. 2004, Nakamura et al. 2006, Rohs and Zweifel 2005). Because most systems rely on geometrical feature detection both to localise the markers within input images and for encoding unique identifiers in each marker, the markers' visual appearance is strongly constrained. In the vast majority of cases, the shape (i.e., the geometry) of the markers is generated algorithmically, following techniques derived from communication coding, without allowing human input on the visual design.

For example, ARToolkit markers (Kato and Billinghurst 1999) (Figure 2.1 f) are characterised by a thick black square frame containing arbitrary grayscale patterns. The system uses straight line detection and quadrilateral fitting to locate the markers and if a marker is found, its pattern is extracted and cross-correlated with all known patterns. As a consequence of this the system speed decreases the more patterns used and the more markers that are contained in the input image. The square deformations and asymmetries in the pattern are used to estimate the camera position in the marker's coordinate system. The patterns used for AR-Toolkit markers can be customised arbitrarily, however, later research (Fiala 2005) suggested to apply digital communication coding techniques to improve the system's performance, at the cost of customisation. The TRIP system (de Ipina et al. 2002) is based on edge detection followed by fast ellipse fitting to locate and track markers, known as *Spot Codes*, which are composed of concentric circles and arcs (Figure 2.1 e). Because ellipses are projection invariant, the system is robust to perspective distortion. The length and position of the arcs are used to recover camera pose estimation and the marker's ID.

A number of systems have common characteristics and they can be referred to as *2D barcodes*. In all of them bits of information are encoded in a matrix of black and white dots, generally with the protection of some error-correction coding. The markers also include lines to facilitate orientation recovery. Examples include Cybercode (Rekimoto and Ayatsuka 2000a), QR-codes (ISO 2000) (Figure 2.1 b), Data Matrix (ISO 2006) (Figure 2.1 a), and Visual Codes (Rohs and Zweifel 2005) (Figure 2.1 c). Algorithms to detect this type of markers are also available for mobile phones (ISO 2000, 2006, Rohs and Zweifel 2005), and several players in the mobile communication industry are promoting standards and a common infrastructure for them (Consortium February 2007). The markers are generated automatically through coding algorithms and they do not allow any aesthetic personalisation or tuning from human input.

The d-touch recognition system is based on the topology of the markers, rather than their geometry (Costanza and Robinson 2003). Initially the system was developed for tracking objects in tangible user interfaces and augmented reality applications (Costanza et al. 2003a,b), as the topology-based approach allows fast performance even when multiple markers are recognised in the same image. An example marker used in tangible interfaces is show in Figure 2.1 g. The shape is designed for compactness and to encode additional information in the position of the parts of the symbol. An initial attempt to design d-touch markers for a mobile application taking into account aesthetic aspects was reported in a position paper (Costanza and Leinss 2006), however, in that case markers were still mostly based on the rectangular grid of Figure 2.1 g, while the work presented here extends the expressive range considerably. Two systems other than d-touch, the VPS system (Johnston and Clark 2003) and reacTIVision (Bencina et al. 2005) (which was derived from d-touch – Figure 2.1 d), are known to use topology rather than geometry for marker detection, however, in both of them markers are generated through algorithmic techniques, with little or no input from the user regarding aesthetic aspects of the markers.

Fast watermark detection (Nakamura et al. 2006) can be an alternative approach to visual marker recognition. This technique encodes information such as an identifier over any 2D image, invisibly. However, the method still relies on geometry, and it requires images to either have a dark border on a white background or to be framed with a thin black rectangle. Moreover, the method is not as fast as other marker recognition techniques, as it is reported to take slightly more than one second to process an image of size 288×352 on a mobile phone.

An alternative to adding markers to objects is the real-time tracking of natural features present in the scene. Researchers in the area of computer vision proposed a variety of methods which are demonstrated to produce reliable results (Lowe 2004, Ozuysal et al. 2007, for example). While

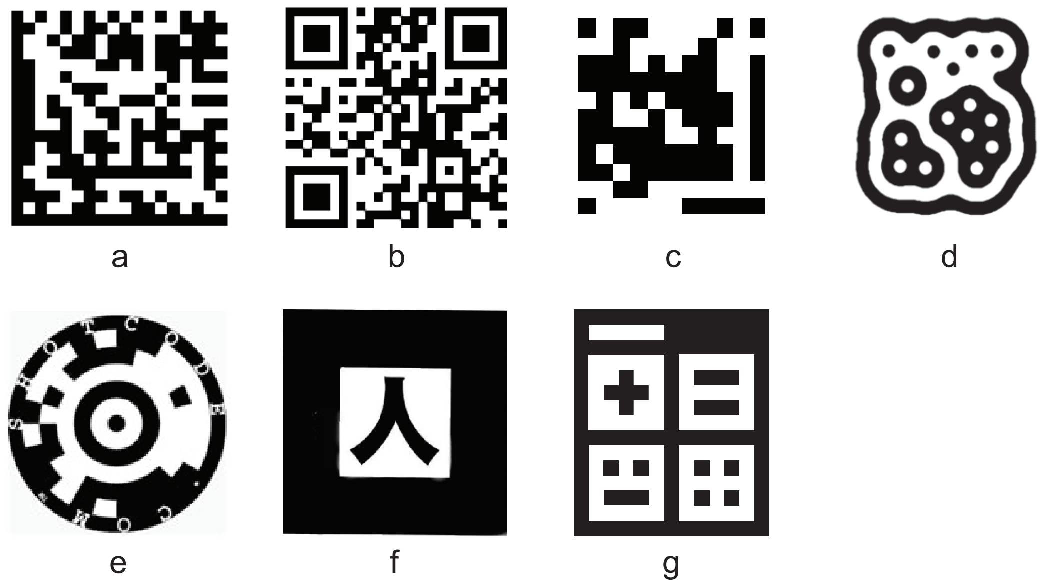


Figure 2.1: Example markers from the literature: a) Data Matrix(ISO 2006); b) QR-codes (ISO 2000); c) M. Rohs's visual codes(Rohs and Zweifel 2005); d) reacTIVision markers(Bencina et al. 2005); e) TRIP(de Ipina et al. 2002); f) AR Toolkit(Kato and Billinghurst 1999); g) early d-touch marker(Costanza et al. 2003a).

these methods are known to require too much processing power and memory compared to what is available on current mobile devices, Wagner et al. (2008) report a modification of state of the art techniques to allow them to run on smart phones.

Given that for most other marker recognition systems symbols are designed algorithmically, no precedents were found in the literature for user studies similar to the one proposed here.

2.2 Mobile Interaction with Visual Markers and RFID Tags

Despite the relative abundance of marker recognition systems and the proliferation of commercial systems, relatively few studies attempted to shine light on the advantages and limitations of marker-based interaction through camera phones.

In his doctoral work Rohs (2005) explored how these markers can support human interaction with camera phones. The exploration was carried out through the definition of a framework encompassing different ways to use the relative position and orientation of the camera phone with respect to the marker as an input parameters, and performing gestures – in this case the movement of the phone relative to the marker, rather than its position, is used as input. A total of 8 "interaction primitives" were defined. Rohs' dissertation reports a usability evaluation experiment with 8 subjects – a small sample compared to the number of different conditions – who were asked to experience the different interaction techniques and subjectively rate them according to their general preference. No other performance measure is reported, which leaves the usability of such interfaces unclear.

Parikh et al. (2006) used 2D barcodes to develop a mobile application that lets users associate data to printed documents. The work was done in the context of rural India, to support “book keeping” for micro-finance. Through a controlled study Parikh et al. demonstrated that the marker-based interface is easy to learn for rural users and it is as efficient, or better than a PC interface for someone who has never used a computer mouse before. These results were influenced by the use of a box that held the phone fixed while users moved the paper documents – this solution was effective in counteracting the problems that users had in holding the phone steadily in front of the marker while pressing keys on it.

Toye et al. (2007) report a laboratory study around user interaction with mobile phones and visual markers. The study is organised in two phases, the first one was a scenario-based evaluation – participants were provided with a working prototype but were asked to imagine real conditions of use despite the experiment taking place in a computer science lab. Eighteen subjects, who did not have previous experience of marker-based interfaces, tested a mobile application allowing them to select items by pointing a camera-phone to a visual marker and pressing one of the phone keys. The second phase of the study was setup as a controlled experiment in which the same subjects performed a repeated pointing task: two visual markers (start and goal) were displayed on a large computer display at a variety of sizes and distances and subjects had to point the camera phones to them and confirm the selection by pressing a key. Task completion time and accuracy were used as performance metrics and compared to arbitrarily defined acceptance threshold (task completion time $\leq 3s$ and error rate $\leq 10\%$). The first phase of the study revealed that the use of visual markers for a practical mobile application was understood by novice users, and no major usability issues were found. The second phase of the study revealed that all subjects had performance levels within the pre-defined thresholds at any target size and distance. The results were found not to follow Fitts’s law, this finding was attributed to the ability to aim at a target by rotating the phone about its vertical axis, which is not accounted by Fitts’s law.

Further work addressing the modelling of camera-phone pointing tasks through Fitts’s law was more recently presented by Rohs and Oulasvirta (2008). Their work reports that performance of target selection through camera-phones is not only affected by the device movement taking place in 3D space, but also by the users switching their attention between the phone display and the physical world, as well as delays, distance range and movement speed limitations imposed by the recognition software.

Visual markers can be often considered an alternative to RFID tags, even though the two technologies are not equivalent (as mentioned in Chapter 1, visual markers are *read-only*, whereas some RFID tags can be rewritten – yet visual markers can be produced through normal printers and accessed using standard photo cameras, while RFID requires special readers devoted only to this purpose, integrated only in a limited number of devices). Mäkelä et al. (2007) conducted a field trial to compare users’ perceptions of RFID tags and visual markers. They interviewed 50 participants on the street in two major cities in Finland showing them a poster augmented with a visual marker and one augmented with an RFID tag together with a phone able to read them. Participants mostly reported not to be familiar with either system, and did not find it generally obvious how to interact with them. About 60% of the interviewees expressed a preference for the RFID system, even though this may have been influenced by the specific visual marker reader used, which did not provide real-time feedback.

Geven et al. (2007) investigated user perceptions of NFC-enabled phones according to 4 different UI metaphors, using a small variety of existing applications. Through a mix of focus group, workshop, on-line survey, diary study and observing users interacting with the system in controlled conditions, Geven and colleagues observed that users were generally positive about the system, but inconsistencies and unclear affordances in different applications design caused confusion.

2.3 Location-based Media: Distributing Digital Information in Physical Space

Several projects in the last two decades have explored the association of digital media to specific locations or objects. A large portion of the literature has a technical focus: it reports systems, frameworks and architectures, generally supported by proof of concept implementations, with little or no emphasis on user evaluation.

Pioneering work in this domain was carried out in the early 1990s at Olivetti and Xerox research labs, with the development of the Olivetti Active Badge (Want et al. 1992) and the Xerox ParcTab (Schilit et al. 1993). The Active Badge system is able to track users in an office environment using badges emitting infrared signals which are received by infrared sensor stations installed in each room. The Active Badge project was designed to enable telephone operators to route calls to the phone closes to the users. The ParcTab is a touch-screen hand-held device which uses infrared transceivers not only to locate users but also to access a local area network. Through ParcTab users can access applications such as calendars, electronic mail and paging running on a local server, which can be augmented with the location information.

The idea of using visual markers to provide location-aware information was introduced with NaviCam (Rekimoto and Nagao 1995), a hand-held device that includes a camera and a visual display used as a viewfinder (similar to contemporary camera phones). NaviCam is able to detect and identify colour-based visual markers and present the user information related to them. The markers, printed on paper labels, can then be used as links between the physical and digital realms. The work was further developed with CyberCode (Rekimoto and Ayatsuka 2000b) using a black and white 2D-barcode, adding the possibility to perform selections through the relative position of the marker with respect to the viewfinder and proposing more application examples.

A more integrated example of location aware system was presented by Abowd et al. (1997), their Cyberguide is a location-aware tour-guide running on commercially available PDAs and tablet PCs augmented with a location sensing infrastructure based on infrared transceivers. The Cyberguide is meant to supersede tourist guide books: it allows users to access textual information based on their current position and their previous actions, as well as to get navigational guidance. Abowd and colleagues report a number of usage scenarios and discuss the technical implementation challenges they addressed. Actual user evaluation was carried out only informally, through demonstrations to lab visitors.

The Electronic Lens project (MIT Mobile Experience Lab 2006, Costanza and Leinss 2006) was designed as communication medium between citizens and the local government in Catalonia, Spain, with the high level objective of *"improving civic participation in the 'res publica."'* By scanning visual markers (using an earlier version of d-touch) users can access information published by the institution, as well as audio content previously uploaded by other users. A prototype was developed as a proof of concept on touch-screen smart-phones (Symbian UIQ phones) and its functionality briefly demonstrated to real users in Spain – unfortunately the project did not include actual observation of the system in use, yet its design influenced the development of the work presented in this dissertation.

Another area of application for location-based systems is narrative. Several projects have explored the potential of triggering the playback of digital content on mobile devices when users are in specific locations, as a form context-aware story telling. Examples include the "Media Portrait of the Liberties" (Nisi et al. 2007), which includes a collection of short movies based on folk stories set in the *Liberties* neighbourhood of Dublin, Ireland. The movies are delivered through PDAs equipped with GPS when users reach pre-defined locations in the neighbourhood, corresponding to the original setting. Similarly the M-Views project (Crow et al. 2003) uses Wifi to detect the position of users and deliver pieces of a story when they are at specific locations. In these kind of projects the narrative generally follows a non-linear path, determined by the physical path of users in space.

More recently, several projects addressed frameworks and infrastructures to create location-based systems that reflect the popularity of *social media* on the Internet (blogs, wikis), mostly still with a strong technical emphasis and little or no rigourous evaluation with users. For example, Hansen and Grønbæk (2008) describe a lightweight infrastructure for "urban and social computing applications", which allows the association of online resources and services with a variety of location techniques, including 2D barcodes, RFID tags, GPS and even a numeric code manually typed by users. Three urban applications based on this infrastructure are briefly described, the authors report that all were prototyped but no information is available about real usage. All 3 applications are implemented using 2D barcodes, while the question of how different technologies may impact the interaction is left open for future work. In a separate publication, Hansen et al. (2008) report in more detail about one of the 3 applications, which is defined as "location-based audio theatre" and it was developed in collaboration with a local theatre group. The system was deployed for 3 different plays, each requiring some technical adjustments, but all generally involving the delivery of multimedia content through mobile phones in conjunction with live acting. Once again, the emphasis of the paper is on technical issues rather than on documenting how the system was used by either the play writers, the actors or the audience, which is particularly disappointing considering the richness of documentation and HCI reflections around performances and experiences involving a combination of live acting and technology mediation (Crabtree et al. 2004, Benford et al. 2009b).

Three other projects are related especially to wikis, a format which is indeed interesting because of its flexibility. For both Semapedia (2005) and the work reported by Siira et al. (2009) the starting point seems to be the great success of Wikipedia: the two projects focus on the software infrastructure to support a collaborative encyclopaedia that can be accessed through 2D barcodes or RFID tags. Little attention is given to the presentation and information of the information on mobile devices, and no user evaluation is reported. Semapedia (2005) allows users to associate Wikipedia pages to 2D barcodes using a web interface, and to retrieve content using any compatible mobile barcode reader. Siira et al. (2009) report the design of a technical architecture to associate wiki pages to NFC tags, and the development of a mobile client prototype, a server component and a web interface. Information can be edited either from the mobile client or from the web interface.

CampusWiki (Schuler et al. 2007) is a wiki system where pages are associated to specific locations (building, floor and room number) on the campus of the New Jersey Institute of Technology. When users access the wiki via the campus Wifi network their location is approximated to the position of the wireless access point they are connected to. One interesting aspect of this approach is that users do not need a special client application: they can use any standard web browser, while the location sensing takes place at the server and infrastructure level. On the other hand, the deployment of such system requires low-level access to the network infrastructure, which may not always easy to obtain for those running the service. Based on the estimated position users are shown the wiki page associated with the closest place (e.g., the cafeteria) as a start page for their web navigation. On this page four sets of links are also provided: pages related to nearby locations, popular pages, recent pages and random pages. Usage of CampusWiki was monitored for about three and half months from 2 campus buildings which had the largest number of associated pages. Out of 104 clicks on the four sets of links 38% were on the location-aware set. Based on this result Schuler et al. suggest that adding location-awareness to a wiki add value for a co-located community. No interviews with users are reported, nor information about the location-based content and its creation.

Urban Tapestries (Angus et al. 2008) is a project about public authoring and the sharing of local knowledge which includes a family of location-aware mobile phone applications that evolved from 2002 to 2008. All prototypes use GPS to determine the users' location and a map-based interface to let them create and browse content; latest versions of the system include also a web-based interface to browse content from fixed computers. Angus et al. discuss the evolution and potential of the system, but the argumentation appears more speculative than grounded in real usage.

Among the location-based media systems for which user observation reports are available, several are focussed on content consumption (Cheverst et al. 2000b,a, Aoki et al. 2002, Grinter et al. 2002), and especially on "here and now" consumption, as noted also by O'Hara et al. (2007a,b). Cheverst et al. (2000b,a) report the development and deployment of "GUIDE", a mobile tour guide running on tablet computers connected through a Wi-fi network. The network access points are also used to detect the users' location, so the GUIDE can provide historical information relevant to the place where users are, and provide navigational guidance towards other site of touristic interest. The system was actually deployed in the city of Lancaster and 60 users were observed while interacting with the system – the observation revealed a very positive response from users in different age ranges and backgrounds. SottoVoce (Aoki et al. 2002, Grinter et al. 2002) is a system designed to augmented the visitors experience to a historic house without isolating them from other fellow visitors, given that visiting monuments is often a social activity. The system runs on commercially available PDAs and it uses a graphical interface based on image maps to let the user manually select their location and the specific exhibit they want to learn more about. User evaluation with 6 pairs of participants demonstrated that the design and implementation were successful both at the usability level and in supporting social interaction.

Isomursu et al. (2008, 2008, 2009) report a set of field trials where participants experienced access to mobile services using phones with NFC readers and tags distributed in the environment. The first set included 4 trials originally designed to test payment through NFC tags in a parking lot, in a theatre, in a restaurant and in a pub. One further trial was arranged with high school pupils, who used the same technology for a learning application. For all trials service access logs were automatically recorded, additionally for some trials participants' feedback was collected through interviews and for some through interviews. Some of the tags were related to service payment through the phones, however Isomursu et al. report only about access to web-pages (even though these were not always related to the specific tag localised, as it was negatively reported by some users). Overall users accessed most frequently tags associated with news from a local newspaper, especially because the content was frequently updated. Participants reported interest towards the tags and were generally positive about the ease of use and social acceptability of the system. Specific concerns were related to the playback of loud sound as a result of scanning a tag which may be inappropriate in some situations, and the negative impact that such system may have on social interaction – for example if in a pub clients get information about the beers through the system, rather than asking the waiters.

After observing the natural behaviour of visitors at the San Francisco Exploratorium, a science museum characterised by engaging interactive exhibits, and the experience of 35 participants experimenting with electronic guides running on PDAs augmented with infrared beacons and barcode readers, Fleck et al. (2002) concluded that a major point in visit enhancement is the collection of material that can be accessed after the museum experience. So they created a simpler system where visitors take with them a "smart card" (a card with an RFID tag) and by swiping it in readers placed near the exhibits they can capture information about them, which will be displayed on a personalised webpage that can be access after the visit. When the card is swept a sequence of 4 photos is also captured to try and portray the visitor interacting with the exhibit.

O'Hara et al. (2007a,b) report the design and evaluation of a system to enhance the experience of primary school children visiting the London Zoo by providing them media content about the animals through mobile phones. To access the content users have to scan barcodes available near some of the cages. Their emphasis is on the collection of location-based content, rather than their immediate consumption and disposal.

2.4 Location-based Content Creation

For the projects described in the previous section, little information is provided regarding the content creation and distribution in space, and generally it seems that content creation was considered a marginal aspect of the work, maybe even an unimportant part. For example Isomursu and Ervasti (2009) reports:

> *The content provided through tags was selected in a brainstorming session of the researchers, and the selection criteria used were probably very different from those that would be used if the tags were used for commercial, or for any purpose other than research. As a result, some tag content was obviously very poorly suited for the specific place it was offered in. For example, many users commented that a tag that helps you call a taxi when you are paying a parking fee was pretty useless.*

Fleck et al. (2002) report that *"only 6 exhibits were used due to equipment limitations and the man-hours required to develop content for each exhibit"*. O'Hara et al. (2007a,b) inform us that media content was provided for only 13 of the hundreds of animal species present in the zoo. The media content seems to be provided by the BBC (one of the partners on the research project), but there is no mention of how the content was selected and associated with markers.

Even though the main purpose of the "George Square" project (Brown et al. 2005) is to let a user visiting a physical location share their experience in real-time with someone remote, the system also involves an aspect of creation of location-based content. The person doing the physical visit uses a mobile tablet computer, equipped with camera, GPS, and Wi-fi connection, while the remote participant uses a fixed computer. An audio link between the two participants is provided, and most of the interaction is mediated through a shared map on which users' avatars are displayed. While the position of the mobile user is updated automatically via GPS, the one of the fixed user can be selected manually. The map can be annotated with photos and links to web pages. A field trial was conducted with 10 pairs of participants, who were asked to freely explore the functionality of the system as well as complete some specific tasks (e.g., find out information about statues present in the square). The main focus of this paper discussion is on the use of collaborative systems for entertainment (compared to work applications, more common in the literature) and the real-time nature of the communication. However, it is interesting to note that annotations from users are accumulated in the system, so users who did the trial later had access to content generated by earlier participants:

> *The map displayed where the users were in the square, and the photographs they had taken. In addition, recommendations of web pages, places in the square and photographs taken by others were generated by running the previous history of visits to the square through a collaborative filtering algorithm. [...]*
> *Our recommender made use of historical data to weave together online information with urban locations. Photographs taken and web pages browsed by users, such as the Wikipedia page on William Gladstone used near his statue, were stored as an archive of information about particular locations. Without this gradual adaptation to users' behaviour, a large amount of context would have had to be manually entered in the locations that we judged to be appropriate for visitors. Instead, our system made use of patterns of co-occurrence of location and browsing, to place information in contextually relevant locations on the map (Chalmers 2004). This suggests a broad method for making use of people's behaviour to connect together information and locations, complementary to the pre-authored content.*

Other research projects have specifically investigated the topic of location-based user-generated content, often in the form of collaborative text annotations. An early example in provided by Burrell and Gay (2002) who report a user study of E-graffiti, an application running on laptops that allows users to post and access text messages that are geo-located, and that can be either public or addressed to individuals. The users' location is approximated by the Wi-fi access point to which their laptop is connected. E-graffiti limits the content access to take place in the physical location where users are. Its authors report this as a design decision *"to force users to confront the location-specific features of the system. We hoped this would prevent them, as much as possible, from using E-graffiti like a traditional bulletin board or email system."* Conversely, message creation in can take place from anywhere: the destination location of messages can be selected manually. Burrell and Gay report a field trial on the Cornell University campus where 57 undergraduate students used the system on Wi-fi enabled laptops borrowed for duration of the trial from the university. The main result reported that the system was under-used compared to the expectations of its creators: *"[the users] reported in large number that they did not really think about information in terms of location, did not know what notes to write, and did not really have anything to share with others at a location."* It is worth underlining that the experiments did not provide or suggest any specific usage application for the system and that participants were simply expected to create content of some sort. Most of the system's actual usage took the form of a private synchronous chat.

GeoNotes (Persson et al. 2001, Persson and Fagerberg 2002) also aims at letting users share geo-located text notes, using Wi-fi for location sensing. Its design was influenced by E-graffiti, however its relation to location is even more restrictive: *"In contrast to E-graffiti, which allowed remote authoring but not remote reading, GeoNotes permitted neither. If users were allowed to read and author notes from remote positions – we reasoned – the connection between the note and its spatial context would be endangered (and thereby the whole concept)."* Differently from E-graffiti, GeoNotes supported only public messages and it allowed users to sign with their real name, with a pseudonym or to stay anonymous. A user study is reported (Persson and Fagerberg 2002), where 78 engineering students interacted with the system for one month on a university campus, using their own laptops. The trial was observed through access logs and questionnaires from a subset of 14 users. Similarly to E-graffiti Persson and Fagerberg report that *"182 notes plus 101 comments attached to 28 access points were below our expectations"* and during the trial most participants used GeoNotes as a synchronous chat tool, rather than for sharing location-based information. The system allowed users to label the locations where they attached notes, the analysis of these labels revealed that in most cases users referred to rooms, while in some cases to larger areas, objects, or events temporarily happening in a location.

The ActiveCampus project (Griswold et al. 2004), and in particular the "Explorer application" within it, runs on PDAs connected to the wireless network of UCSD campus. Using Wi-Fi to sense the clients location, the system enables students to see on a map the location of other students who are in their physical proximity and to contact them using an instant messaging application. It also allows them to associate "graffiti" text messages in a permanent fashion to specific locations. A field trial was conducted to test ActiveCampus Explorer over one semester with 300 students which revealed that the instant messaging functionality was used much more frequently than the graffiti feature, and that the location-awareness seems to be understood and adopted by users (based on the fact that users send messages to each other more frequently when they are in physical proximity). It is also reported that the system deployment was not as successful as they expected because of a mix of technical (e.g., limited battery life of the PDAs) and design issues (e.g., lack of a delete function).

Cherubini et al. (2007) developed STAMPS, a mobile application running on S60 phones that enables users to associate text notes to specific points on a map displayed on the phone's screen. The system centres the map on the user's current location (estimated through the mobile phone cell-ID), but leaves them free to browse away from this location and post and read notes anywhere on the map. A field study is reported in which 21 participants, recruited from a variety of contexts, were provided with compatible phones and used the application for 3 months. User interaction was logged and post-trial questionnaires were used to gather feedback. Similarly to E-graffiti, GeoNotes and ActiveCampus projects, in the STAMPS trial no specific application of the system was suggested to the participants, with the idea that the trial would reveal what kind of annotations people would share on a map. Cherubini et al. report that *"only 150 messages"* were produced, which is below their expectations, and they attribute this result to lack of structured scenario and to the network effect (not enough people using the system) made even more severe by the fact that participants were not all parts of the same social group. In his PhD dissertation (Cherubini 2008) Cherubini reports a follow-up trial with architecture students, who were asked to perform over a semester a collaborative analysis of an urban area in Lausanne using STAMPS. The activity was part of a university class. Despite the more structured task the trial, very few messages were produced. Post-trial interviews highlighted some usability issues and the lack of accurate location sensing as causes for the students not using the system. In private communication Cherubini attributed the low level of activity in the second trial also to an inadequate integration of the technology in the course structure, most notably the annotation activity performed with STAMPS did not have an influence on the students grade, so they preferred investing time on completing the compulsory assignments instead.

The main result produced by the 4 projects (E-graffiti, GeoNotes, ActiveCampus and STAMPS) that attempted to observe the production of user-generated location-based content is that systems were not used or used below researchers' expectations. This result is attributed to different possible causes, ranging from network effect to implementation or design limitations. However, it must also be underlined that all trials did not provide specific context or motivation for using the system, and often selected participants through random sampling, rather than focussing on covering a tight social group, as it was done elsewhere (Jung et al. 2005). While the high expectations may have been biassed by the popularity of user generated content on the Internet, it must be emphasised that social media, including for example tagging, is motivated by the wide exposure potential of the web – as it was confirmed, for example, by Ames and Naaman (2007) who report and discuss a qualitative study of photo tagging. In contrast, the location based systems discussed above represent closed systems, accessible only by the participants of the trial, who may not always part of the same social circles, therefore they lack mechanisms for motivating or rewarding content creation.

The design of these last 4 systems, and more in general also of the other location-based projects discussed earlier, has a tendency to restrict the information access to specific modalities (text) or interfaces (e.g., from mobile devices only, on-location only), practically discouraging the integration of the systems with other technologies or communication practices. Few projects support mobile plus web access, and generally this is only for post-experience purpose (O'Hara et al. 2007a,b, Fleck et al. 2002), in specific formats (for example Cherubini (2008) reports that *"participants could download the messages posted by their group to their computer in a format compatible with mapping applications"*), or for synchronous communication (Brown et al. 2005). Yet, Internet social media is generally associated to open APIs and letting the users combining different services (commonly referred to as *mash-ups*). It is refreshing from this point of view, to find that Milic-Frayling et al. (2007) argue for a different approach: to augment mobile location-based applications by combining them with web-interfaces and web-service architectures to allow the integration of different on-line resources. Even though the mGuide prototype, proposed by Milic-Frayling et al., seems to actually fall more in the category of post-experience web access (and evaluation is limited to usability and scenario levels), combined mobile and PC internet approach is well demonstrated by commercial players (such as Google, MySpace and Facebook) who complement their popular web interfaces with mobile applications that facilitate access to their services (yet not generally location-based).

2.5 Other Forms of Information in Public Space

The distribution of information in physical, public space certainly pre-dates interactive technology. Location-specific information has been presented for centuries through signs and signing systems. Wildbur and Burke (1998) present various case studies of design of signing systems, including examples from the domain of transportation, such as airports, or bus stations, monuments and city neighbourhoods. With the support of several visual examples, the authors illustrate how different elements of graphic design, such as typeface and colours, play an important role in making signs clear and legible even from distance, as well as convey a strong sense of identity.

While signage is generally rather permanent, other forms of communication of a transient nature include advertising and street art. Both of these share the main goal of attracting the viewer attention over competing pieces. Of course, advertisers and street artists follow different rules, the ones moving within the limits defined by public authorities and the large budgets provided by their customers, and the others *appropriating* or *invading* (depending on the point of view) space that is public or owned by someone else (DROPDROP Agency 2006, contribution by Buro Destruct, a commercial graphic design studio). Several books cover the topic of street art (Manco 2004, Andreas Ullrich 2006, DROPDROP Agency 2006, Burnham 2008, among many more), illustrating different types of intervention, ranging from graphics sprayed on walls (freely or through stencils), to stickers of different sized and materials, to three-dimensional installations. A more academic viewpoint on the subject is provided by Schacter (2008), who reports a social science investigation on the production and destruction of street art in London.

Some street art projects take an explicitly organized viral approach. The Space Invaders project[1] features relatively small (22-by-16 cm) mosaics portraying artwork inspired to the eponymous computer game, these have been stuck up on walls of cities Worldwide. “Invasion kits” to be applied by the customer are for sale from the project website. Similarly, the artist Shepard Fairey – recently famous for donating US president Obama a portrait that became the icon for his presidential campaign (Shepard Fairey & The Institute of Contemporary Art / Boston 2009) – started as a street artist, attaching stickers with a high contrast photograph of a wrestler and the lettering “obey” which became the artists’ own *brand*. The same stickers are for sale Fairey’s website.

A number of media art projects are situated at the intersection of interactive technology and street art. For example, the Yellow Arrow project[2] , defined by its creators as *“a global public art project of local experiences”*, is based on yellow arrow-shaped stickers, approximately 10-by-10 cm in size, each containing a unique alphanumeric identifier. By sending a text message (sms) with this identifier to the project’s server it is possible to associate content with the specific sticker, or retrieve content created by others. The stickers have all the same visual look, designed to attract the attention of those passing by.

[1] http://www.space-invaders.com
[2] http://yellowarrow.net

3. D-touch marker recognition

This chapter presents *d-touch*, a flexible system for the recognition of visual markers that can be designed according to aesthetic criteria to fit particular applications. D-touch is the enabling technology at the basis of this entire dissertation. The recognition algorithm is based on image topology, rather than geometry, in contrast to most other similar systems. The topology-based approach was inspired by the work of Johnston and Clark (2003), and originally developed while I was at the Electronics department of the University of York, under the supervision of John Robinson (Costanza and Robinson 2003). This chapter presents a formalisation and generalisation of the work, undertaken at EPFL.

3.1 Background: Region Adjacency Graphs and Trees

The d-touch recognition algorithm is based on the topological structure of the markers and of the image in which they are searched, in particular on the adjacency of connected regions. The adjacency information is stored in a region adjacency graph: an undirected graph $G(N, A)$ where each node $n \in N$ corresponds to a connected region of the image, and two nodes are connected by an arc (or edge) $a \in A$ if and only if the corresponding regions are neighbouring, i.e., they share a border (or part of it). If the image is binary, the region adjacency graph is in fact a *tree* (Rosenfeld 1974), in the sense of a "connected acyclic graph". Moreover, the tree is 2-coloured and bipartite: black and white nodes correspond to black and white regions and nodes of one colour can only be connected to nodes of the other colour. The region adjacency tree of an example binary image is illustrated in Figure 3.1. The white background region a contains 3 black regions: b, e and f – note that all the black pixels in the scissors are connected (region b). Region b, in turn, contains two white regions: c and d. In principle all region adjacency trees can be rooted in the image background region, however in the context of d-touch the trees of input images are left un-rooted.

From a computational point of view, dealing with a tree is significantly less expensive than dealing with a graph as navigation does not require checking for loops. For this reason, d-touch works on binary images, obtained through application of an adaptive threshold.

3.2 D-touch Markers

At the most general level, d-touch markers are defined solely in terms of their adjacency structure, and the definition spans a broad class of possible region adjacency trees. At the application level it

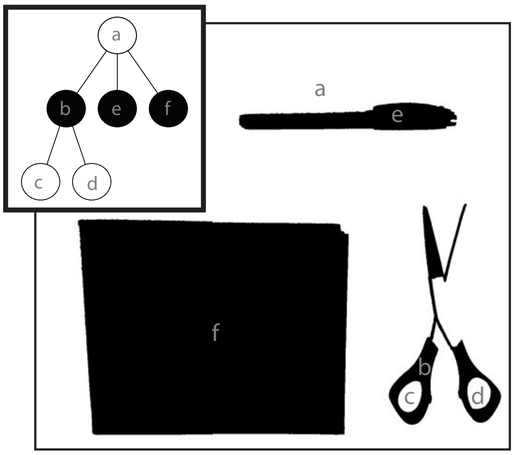


Figure 3.1: An example of binary image and its region adjacency tree.

is possible to work with only a subset of markers, to accommodate specific requirements on a case by case basis. For example, it is possible to use a subset of markers that include particular geometrical features allowing recovery of the marker rotation with respect to the image's horizontal axis or even texture mapping, or to use a subset that guarantees high resilience to misclassification.

3.2.1 Markers' Topological Structure

The adjacency structure of the markers is limited by 3 constraints to simplify this search from a computational point of view, and to reduce the likelihood of false recognition. The first constraint is that markers can have only 3 levels of nesting, named *root*, *branches* and *leaves* – the *root* is the outermost level of the marker, the region enclosing it, any region of the marker connected to the root is named *branch*, and regions connected to *branches* different than the root are *leaves*. Regions of the marker that are connected to the root and do not contain any leaves are called *empty branches*. The second constraint is that markers must have at least 3 branches and the third that at least half of the branches must contain leaves. In other words, a valid marker can be composed of a black region containing 3 or more white regions, and at least half of these white regions must contain one or more black regions – however the colours can be inverted. Because the first step of the recognition process is to convert the images into pure black and white, the markers can actually be drawn in any colour, as long as they present reasonable contrast. Four examples of the "minimal" valid marker structure are illustrated in Figure 3.2, other examples are shown in Figure 3.3.

In more formal terms, the adjacency structure of d-touch markers can be constructed through the following definitions, where $deg(n) \geq 0$ indicates the *degree* of node n, that is the number of arcs connected to it:

A node $l \in N$ is said to be a *leaf* if it has only one edge, i.e. $deg(l) = 1$, and the image region corresponding to l does not contain pixels that are part of the border of the image. The second condition ensures that each leaf region is contained in the region it is connected to.

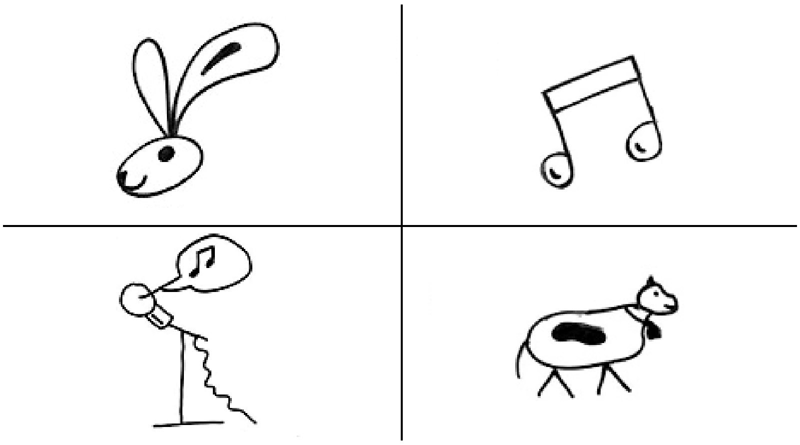


Figure 3.2: Four example "minimal" valid d-touch markers: each is a black region containing 3 white regions (the minimum number of regions allowed); 2 of the white regions contain 1 black region, one contains none. These examples were generated by subjects in the user study reported in Chapter 4.

$leaves(n) \geq 0$ denotes the number of leaves *contained* in (connected to) the node $n \in N$.

A node $b \in N$ is said to be a *branch* if all of its edges except for one connect it to leaves, and it is connected to at least one leaf, i.e. $leaves(b) \geq 1$ and $deg(b) = leaves(b) + 1$. From the definition of leaf it follows that each branch region is contained by the only non-leaf region to which it is connected.

$branches(n) \geq 0$ denotes the number of branches *contained* in (connected to) the node $n \in N$.

Following this terminology, d-touch markers can be defined in terms of their root:

A node $r \in N$ is defined as a *d-touch marker root*, from here on just a *root*, if it is connected to at least 3 branches and all of its edges except for one connect it to branches or leaves, i.e. $branches(r) \geq 1$ and $deg(r) = branches(r) + leaves(r) + 1$. This definition constrains the depth of marker trees to be exactly 3, a characteristic that simplifies their recognition. Note that, in contrast, no constraints are posed on the breadth, allowing, in principle, an infinite number of different markers. Example marker trees are shown in the third row of Figure 3.3.

A valid marker with root node r can be uniquely identified by the colour of the root – 'black' or 'white' – and an unordered sequence of non-negative integer numbers $\{s_i\}_{i=1..k}$, where each s_i is equal to the number of leaves in each branch connected to the root, or to zero for each leaf directly connected to the root and $k = leaves(r) + branches(r)$. The root colour and sequence of each of the markers in Figure 3.3 are shown on the bottom row of the same image. This information is adopted as the *marker ID*, it can be easily converted to a text string and used for database look-up operations or lexicographic comparison (provided the unordered sequence is consistently sorted in an arbitrary direction).

3.2.2 From Topology to Form

The structure defined in the previous subsection corresponds to graphical symbols that include nested dark and light regions, such as a black region (the *root*) containing a number of white

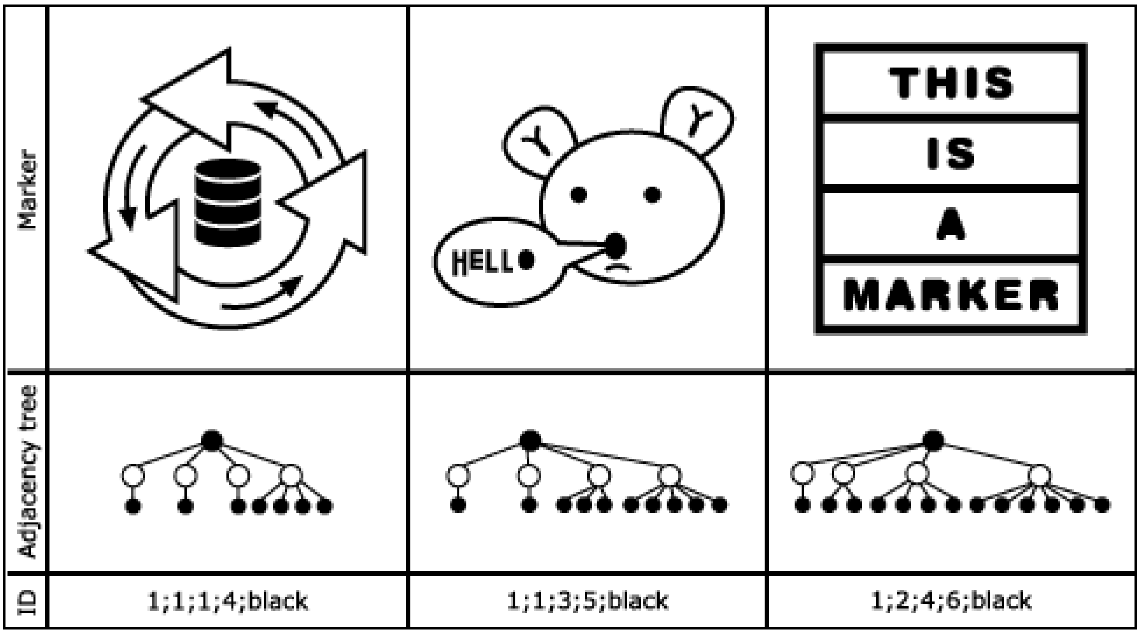


Figure 3.3: Five example d-touch markers, with binary versions, region adjacency trees and IDs. The grey arrows between the second and third row show the correspondence of root regions and root nodes.

regions (*branches*), which in turn contain other black regions (*leaves*). Examples are shown in Figure 3.3. It must be underlined that any geometric distortion that does not modify the topological structure of the markers will not affect the recognition. As demonstrated in Figure 3.4, markers can be recognised even when folded or stretched.

3.2.3 Geometrically Constrained Markers

The geometry of the markers is transparent to the topological recognition: not only does this property provide visual design freedom, but it also allows to encode extra information in the shape of the markers. Geometrical features can be used to add functionality to satisfy application-specific requirements: by defining an axis (or vector) in the marker, it is possible to recover its rotation with respect to the image's horizontal axis, and by identifying the corners of the marker it is possible to apply texture mapping or even calculate pose estimation. Moreover, geometry can be used to distinguish among markers that have the same topological structure, a convenient way to expand the number of different identifiers.

Geometrical features can be processed after markers have been identified through their topology, in this way the processing can be limited to restricted regions of interest and be computationally very lightweight. The features can be chosen to be projection invariant, for example, by using points collinearity and the ratio of their distances.

A class of geometrically constrained markers is illustrated in Figure 3.5. Each branch region contains a different number of leaves, so it can be immediately identified within the marker through topological processing. One branch region always appears in the same position, acting as a *pivot*:

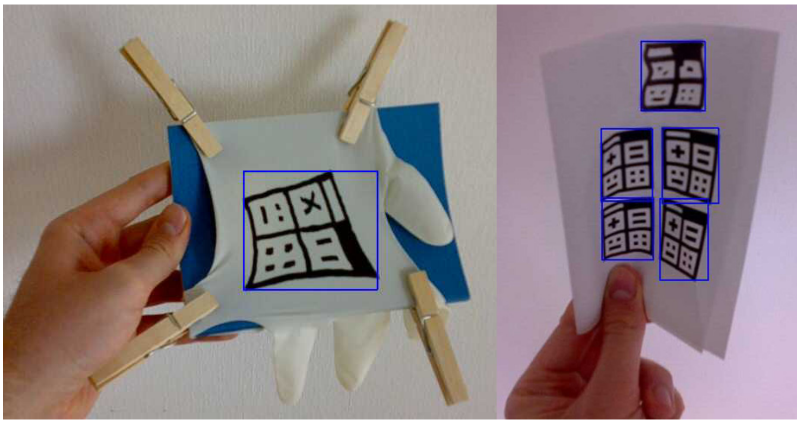


Figure 3.4: Example of markers being recognised by the system despite geometrical distortion. Left drawn on latex glove, right printed and drawn on paper.

the branch with only one leaf. The other branch regions are permutated, generating six markers that share the same topological structure. The branch regions can be sorted based on their centres of gravity, because these are collinear points, their order is projection invariant. The vector defined by the centre of gravity of the pivot region and any other branch region can be used to recover rotation of the marker with respect to the image horizontal axis.

Another, class of geometrically constrained markers is illustrated in Figure 3.6. In this case the empty branch (i.e., the leaf directly connected to the root) acts as a pivot, and 24 markers can be defined by the permutation of the other four branches. In this case the marker is read in two rows, Left to right and top to bottom, assuming the pivot is on the upper left. Collinearity is used again on the centres of gravity to find the two branch regions on the same column as the pivot, which are then ordered based on the ratio of their distances from the pivot. The remaining two branch regions are also sorted based on the distance from the pivot – even if this distance is not projection invariant, because the points are not collinear, it proved to be a fairly robust strategy. The vector defined by the gravity centres of the two branch regions on the top of the two rows is used to recover rotation. Once their order has been recovered, the centres of the four (non-empty) branches are used to calculate an estimate of the corners of the root region of the marker, which are then used to texture map images and video sequences in mixed reality applications.

It must be underlined that geometrical constraints can be introduced, still without completely impairing the visual design freedom. For example, if constraints are defined on the branch regions relative positions, as described above, the form of the leaf regions and root region are free, as illustrated in Figure 3.7.

A valid alternative strategy is to define constraints for the positions of the leaf regions, as suggested by Bencina et al. for their reacTIVision markers (Bencina et al. 2005) (Figure 2.1 d). In fact reacTIVision markers can be seen as a special case of d-touch markers, and the system was indeed derived from an earlier version of d-touch.

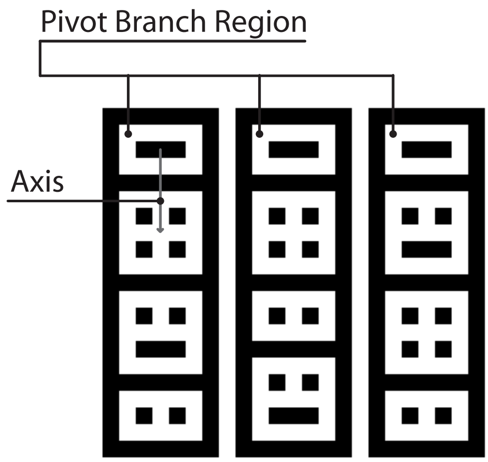


Figure 3.5: An example of geometrically constrained d-touch markers with projection invariant features. A subset of the 6 different markers generated by the permutation of branch regions.

3.3 Algorithm Description

3.3.1 General Overview

The recognition consists in essence of searching for the sub-trees of the markers within the region adjacency tree of an input grayscale image, referred to as the *scene image*. Scene images are generally acquired from a webcam or camera phone. The outline of the algorithm is as follows:

1. Convert the scene image to binary using an adaptive threshold;
2. Extract connected regions and construct the scene region adjacency tree;
3. Search for marker sub-trees in the scene region adjacency tree;
4. (Optional) process geometric features to calculate additional information (e.g., orientation, scale, vertexes).

The performance of the binarisation step is critical for the success of the following stages. The most intensive step from the computational point of view is the scene tree construction, while the sub-tree search is relatively light-weight. The last step is relevant only if the markers contain geometrical constraints, as described in the previous section. Each of the steps is detailed in the following subsections.

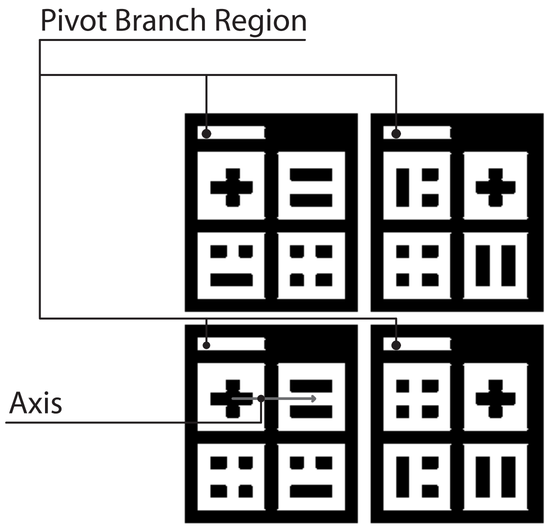


Figure 3.6: An example of a class of geometrically constrained d-touch markers. A subset of the 24 different markers generated by the permutation of branch regions.

3.3.2 Image Binarisation

Input images are converted to binary using an adaptive thresholding algorithm derived from the method proposed by Bernsen (1986). The image is divided into square regions $\{W_i\}$ of dimension s, overlapping for 50% both horizontally and vertically (the regions are defined on a $s/2$ spaced grid). For each W_i the contrast is defined as the difference between the maximum and minimum intensity levels $c_i = max(W_i) - min(W_i)$, and the mid-point $m_i = (max(W_i) - min(W_i))/2$ is calculated as the average of maximum and minimum. If c_i is higher than a fixed value k, then the value m_i is used as a local threshold for the central 25% of the pixels of R_i, i.e. those in the square of side $s/2$ at the centre of W_i, otherwise, m_i is compared to a global threshold T, and the central 25% of the pixels of W_i are all set to black or white accordingly. The global threshold T is calculated on the entire image using the method proposed by Otsu (1979).

To minimise processing, the image is initially partitioned in non-overlapping squares $\{C_j\}$ of side $s/2$ pixels, such that each W_i corresponds to the union of four C_j. Maximum and minimum intensity values are computed for each C_j, and these values are used to calculate the maximum and minimum for each of the W_j.

A sample of 486 images was used to optimise the values of s and k, and to compare the modified Bernsen threshold to a comparable modification of the method by Sauvola and Pietikäinen (2000). The images contained d-touch markers and were captured under a variety of illumination conditions, 246 were captured through camera phones (four models: Nokia 3230, Sony Ericsson Z550, Nokia N73, Nokia N80 – these phones were only used to capture the images, the processed off-line) indoor and in an outdoor urban environment, and contained the markers shown in Figure 3.8. The remaining 240 images were captured in a desktop setting (through a Philips Webcam SPC900NC) and contained only the type of markers shown in Figure 3.6. The marker recogni-

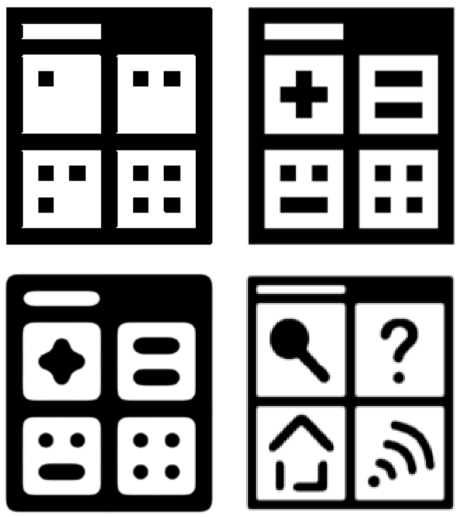


Figure 3.7: An example of geometrically constrained d-touch markers. Four different valid variations of the same topological and geometrical structure: even when using geometrical information for decoding markers their shape is not completely fixed.

tion results obtained from a swipe of values for s and k show that optimal settings are 32 and 44, respectively.

The threshold method proposed by Sauvola and Pietikäinen was reported to outperform the one by Bernsen for document binarisation (Sauvola and Pietikäinen 2000), at least when the threshold value is calculated individually for each pixel in the image based on its neighbourhood. The comparison between the two methods in the context described above indicates that Sauvola and Pietikäinen's method can outperform the one by Bernsen on some subsets of the images sample, but it performs much worse on others.

3.3.3 Connected Components Extraction and Tree Construction

The adjacency information is derived from the segmentation of connected regions, or *connected components*, so the two operations are conducted at the same time in a single image pass. The connected components extraction is inspired by the one described by Rahimi (2001). The binary image is scanned from the top-left corner, keeping track of the left, above and above-left neighbours of the current pixel. Pixels of the same colour are grouped in regions, and every time a new region is found a corresponding node is created in the tree and this is linked to the neighbour pixel's node. To keep track of which region each pixel belongs to, a *map of region labels* is created using an array of the same size of the image. Because the image is processed locally, sometimes regions need to be merged, as well as the corresponding adjacency tree nodes. This in turn requires enumerating all the arcs of a node (*arcs enumeration*), checking for nodes connections (*arc check*) and linking nodes in the tree (*arc insertion*). For this reason it was critical for fast performance to represent the tree using hashed linked lists, which is in fact equivalent to keeping

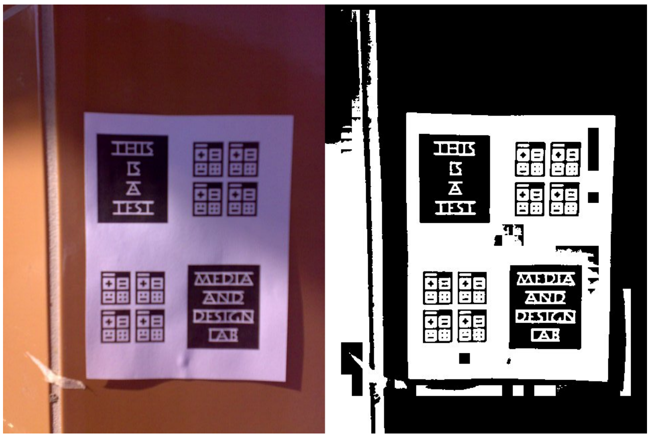


Figure 3.8: One of the images used for the training of the threshold parameters (left) with the binary image produced using the values selected from the training (right).

both an adjacency list representation (each node contains a list of edges) and an adjacency matrix representation (each element of this matrix m[i,j] is either equal to 1 or 0 depending upon whether nodes i and j are connected by an arc or not). Adjacency lists guarantee faster performance for *arc enumeration*, while an adjacency matrix is more efficient for *arc check*. The overhead of keeping the hashed linked list is an extra pointer assignment for each *arc insertion*. Every time two regions are merged, the pixels of one of the two need to be re-labelled; to make this operation fast pixels are related to labels using pointers, in this way re-labelling is equivalent to simply changing the pointed value's value.

Pixels on the border of each region are stored during the connected component extraction to facilitate geometrical post-processing. Linked lists are used for this purpose to minimise the cost of merging. The *map of region labels* also provides an efficient way to access all the pixels of a component.

To optimise the process, data structures for the tree and the regions were iteratively designed with careful consideration of the trade-off between memory consumption and processing speed (Skiena 1998, Mehlhorn and Näher 1999). To avoid memory allocation and de-allocation at each frame, all data structures are based on pools: a large number of elements are allocated at start-up, recycled from frame to frame, and de-allocated only when the application terminates. Hashed linked lists and pools require reallocation if their size grows beyond initial reserved space, significantly slowing down the processing. However, it was empirically noticed that for a $w \times h$ pixels image, the tree size is generally smaller than $l = w \times h/100$. Larger adjacency structures generally correspond to noisy images, generated for example by lag in the camera automatic gain control when there is a change in the scene illumination. For this reason, if the graph construction grows beyond l the current frame is dropped.

3.3.4 Detection

Two strategies are possible for the sub-tree search: one is referred to as *specific detection* and it requires *a priori* knowledge about all adjacency trees used within the application, while the other, referred to as *generalised detection*, searches for any sub-tree that conforms to the d-touch marker definition (as described in Section 3.2.1) and checks whether any instance found corresponds to known markers stored in a look up table. The *specific detection* is suitable when only a small number of marker adjacency trees is considered, such as when a single topological structure corresponds to multiple markers through geometrical features. Conversely, the *generalised detection* is convenient for applications in which a large number of adjacency structures are used, these might be stored in a central repository, accessed by several distributed clients.

3.3.4.1 Specific Detection

Each node in the scene tree with the same degree as a marker root node (known *a priori*) is considered a candidate marker root. The tree neighbourhood of each candidate is then compared with the marker model to determine matches. The comparison reduces to calculating the intersection of two sets of integers, one representing the sub-nodes of the candidate root node and the other the sub-nodes of the model: if the result has maximum cardinality there is a match.

3.3.4.2 Generalised Detection

The generalised detection follows directly form the marker definition introduced earlier in this paper. In a first pass through the tree all single-connected nodes of the tree are marked as leaves, and the number of contained leaves is stored in each leaf-parent. In a second pass all nodes for which $leaves(b) > 1$ and $deg(b) = leaves(b) + 1$ are marked as branches, and the number of contained leaves is stored in each branch-parent. In the last pass, all nodes for which $branches(r) > 0$ and $deg(r) = branches(r) + leaves(r) + 1$ are identified as candidate marker roots and checked against a known markers look-up table.

Even though the generalised detection is slower than the specific detection, it must be underlined that the difference is negligible as compared to the duration of the graph construction. Moreover, if the look-up table check is performed remotely, network delays will probably constitute a bottle neck.

3.3.5 Geometrical Features Detection

The most basic geometric operation that is performed after the topology-based recognition of the markers is the calculation of their position within the image. This can be expressed as the centre of gravity of the root region, or as the average of the centres of branch and leaf regions. The centre of gravity can either be calculated as the average of the coordinates of all pixels in the region, or it can be approximated considering only the border pixels for faster processing. Information about both border pixels and internal pixels of all the regions in each marker is available from the topology-based recognition.

Optional geometrical features that markers may include, as described in the previous section, are detected after the topological detection. This information is used to calculate the centre of gravity of each region, branch centres are then tested for collinearity and sorted on relative distance from the pivot region.

3.4 Implementation and Performance

D-touch runs in real-time on PCs (Linux, MS Windows and Mac OS) and on Symbian OS mobile phones. The topology-based approach does not involve floating point operations, making the system particularly suitable for embedded devices. The source code, written in ANSI C++ (using only the core C++ language), is platform independent and it is publicly available under the terms of the GNU Public License (GPL).[1] All the data structures were developed from the ground up as template classes, rather than using existing libraries, such as the Standard Template Library. This allowed to improve the portability of the code and optimise its performance, as mentioned above. For compatibility with Symbian C++, which lacks part of the standard C++ exception handling mechanism, in d-touch the memory allocation follows the 2-phase construction pattern.

The processing speed depends on the number of connected components in the input image (this is directly reflected in the size of the tree and on the number of operations required for its construction). However, it is independent of the number of markers contained in the image. On a *desktop PC* powered by an Intel Pentium D processor at 2.8 GHz with 2MBytes of cache and 1GByte of RAM, running Ubuntu Linux 6.10, the average time required to process a 640×480 frame was 16.45 ms (std. dev 0.83, N=65000 frames), corresponding to about 60 frames per second. On a *Nokia N73* mobile phone the time to process a frame is 86.0 ms (std. dev 18.2, N=144), i.e., about 12 frames per second in viewfinder mode with a resolution of 240×180 pixels, and it goes down to an average of 347.3 ms (std. dev 85.6, N=14) when processing 640×480 images. Code profiling on the PC revealed that the construction of the scene region adjacency tree (including connected components extraction) accounts for at least 76% of the processing time, the calculation of the adaptive threshold accounts for about 22%, while the actual search for marker sub-trees in the image tree is responsible for less than 2%.

This level of performance has a strong impact on usability as it allows the creation of an effect similar to *mouse hover* when a camera phone is pointed to a d-touch marker. The marker can be highlighted in the viewfinder to show users that it can be selected.

D-touch is available in the form of a class library that can be statically linked into C++ applications – the higher level API is rather simple, in that it offers a function, *process*, that receives as argument an image where to look for markers are returns a list of markers found in it (if any). For each marker the library provides ID, position and size in pixels within the image and rotation with respect to the horizontal direction (even though, as discussed above, depending on the marker geometry the rotation information may or may not be meaningful).

The C++ library is also wrapped into a native module for the Python programming language, *dtrecognition*, available also for PyS60,[2] the port of Python to the S60 mobile platform (running on most Nokia smart phones and few other devices), to enable the rapid prototyping of mobile applications. In PyS60, some additional modules, written in Python, extend the functionality of the core library. The *buffdtrecognition* module wraps dtrecognition and transparently provides some simple tracking on top of the frame-by-frame recognition. The *dtevents* module in turn wraps the buffdtrecognition, it handles the acquisition of images from the viewfinder and it provides an API mimicking the mechanism built in PyS60 to handle the phone keys and the touch-screen. The

[1] The d-touch source code is available from http://sourceforge.net/projects/libdtouch/ and http://d-touch.org/

[2] http://wiki.opensource.nokia.com/projects/PyS60

dtevents module generates events corresponding to: a marker entering the viewfinder (corresponding to the EEventKeyDown event defined in PyS60 for the phone keys and touch-screen), a marker being in the viewfinder (corresponding to EEventKey) and a marker exiting the viewfinder (corresponding to EEventKeyUp). This API attempts to make the use of d-touch on phones as simple as the use of the phone keys and touch-screen, and it makes it easy to handle d-touch events together with other phone UI events.

On desktop computers, to allow development in other programming languages (such as Java, Adobe Flash, Pure Data or MAX), d-touch is also available through the *d-touch server*, a multi-platform application for Windows, Mac OS and Linux. The d-touch server handles the image acquisition from a camera connected to the computer, it processes them and sends the results in XML format through a network socket. Applications coded in any language can connect to this socket even on the local host, and use the information about the found markers. Similarly to the PyS60 module the d-touch server also provides simple marker tracking to alleviate occlusion issues.

3.5 D-touch Analyser

The *d-touch analyser* (formerly DTAnalyser) is a desktop application developed to help users design valid markers. It is GUI-based and runs on Windows, OS X and Linux. Users can import candidate markers as bitmap files or through copy and paste from other applications. A screen shot is shown in Figure 3.9. The d-touch analyser does not provide any drawing functionality, it is designed to be used in conjunction with existing graphic applications, given their availability under both commercial and open source licenses. It was preferred to develop the application as stand-alone rather than as a plug-in and avoid ties to a specific platform. For any imported graphics, the d-touch analyser shows how the image is transformed into black and white and how it is segmented into connected components – this information is rendered through a coloured map. The application analyses the imported graphics and checks whether it complies with the structure of d-touch markers; if they do not, it attempts to detect how the proposed symbol violates the d-touch rules and presents this information to the user.

In the case of valid markers, the application displays the marker ID, and performs a robustness analysis. Low resolution scanning of the marker is simulated by resizing the image with a low-fidelity method (pixel replication). The image is iteratively scaled down until its topology becomes different from the original. The distorted image is compared with the original to detect which elements are most likely to be corrupted, normally corresponding to smaller details of the symbol. The elements that make the marker *weak* to low resolution scanning are displayed to the user, as illustrated in Figure 3.9, together with the black and white and segmented representations of the distorted image. The minimum (simulated) resolution at which the marker can be successfully read is also displayed. Given that most mobile phones on which the d-touch recognition runs have a viewfinder resolution of 240×180 pixels, markers which are readable at a resolution under 100×100 pixels are suggested to be “good for a mobile phone”, those readable below 200×200 pixels “not easy for a mobile phone” and the others as “high-resolution cameras only”.

As shown in Figure 3.9, all the available views of the candidate markers are scaled down and displayed as thumbnails at the top of the application window. Each view can be selected by

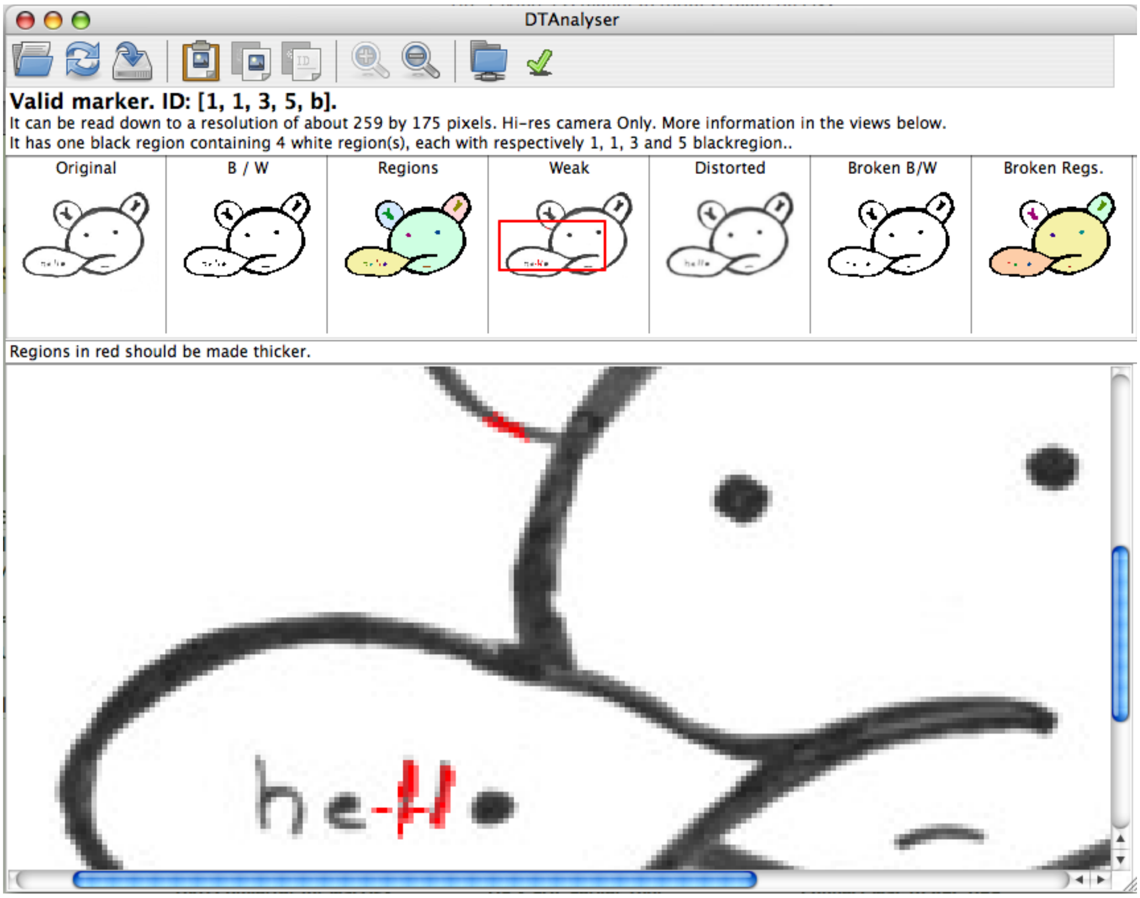


Figure 3.9: A screenshot of the *d-touch analyser* application, highlighting in red the elements of the marker that make it unreadable when scanned at low resolution.

clicking on it with the mouse and in this way displayed on the main panel of the application, where it is possible to zoom in on details. It is worth underlining that the d-touch analyser does not attempt to automatically fix the imported graphics, it rather points out problems and let the users modify their design (back in the original drawing application) to solve them. This choice was made to leave the user in control of the design trade-offs of their markers.

The d-touch analyser also provides access to an online repository of markers – a data base on a remote server where users can "register" marker images and corresponding IDs. If a marker is valid the application allows users to check if its ID has already been registered, and, if not, to register it. Different independent repositories can be defined, each corresponding to a different mobile application, to minimise marker ID collisions. The d-touch analyser can be configured to point to a specific repository through a URL.

To allow the observation of the application usage, and attempt to understand how people start and adapt the design of markers that comply to the d-touch rules, the application usage was logged to a remote server, the same one used for the marker registration. The logging recorded all images imported in the application and all operations users performed on the UI, including for examples switching to a specific view. The recorded information was used both in the context of a laboratory experiment run at EPFL and to remotely observe people who downloaded the application from the

Internet. Users of the d-touch analyser are required to register through the web and log-in when they start the application for the first time; the credentials are used both for the marker registration and the usage logging.

3.5.1 Implementation

The d-touch library was extended to support analysis of candidate markers, in addition to recognition. The starting assumption is that the input image would contain just a candidate marker symbol. The region adjacency graph is constructed following the same procedure used for the recognition, and its root is defined as the largest region adjacent to the border of the image. A graph transversal is then started from the root, searching for a candidate marker root region, i.e., a region that has more than 1 nested regions. If such region is not found the candidate is rejected. If a candidate marker root region is found, the algorithm analyses the regions contained in it, searching for possible problems: having less than 3 branch regions, and having more than 3 levels of nesting.

While the marker analysis was implemented in C++ extending the core d-touch library, the rest of the application was implemented in Python, including the networking and the user interface (using the WxPython library).

The server side of the system was implemented in PHP, as an extension of the TikiWiki[3] open source project. Both the marker repositories and the usage logging are part of the more general uWiki server, which is described more in detail in Chapter 6.

[3]http://tikiwiki.org version 1.9.11

4. Evaluation: Drawing Functional Markers

The rules that define valid d-touch markers, described in Chapter 3, allow the creation of symbols that can both be read at the topological level by the recognition algorithm and have an iconic meaning for people. While the rules are in principle very flexible in this sense, humans usually refer to graphics in terms of shapes and composition, rather than nesting of connected components. Thus it was not obvious how easily people would be able to create symbols that carry expressive meaning while complying with the d-touch rules. A user study was designed and run to address this question. At a more general level, the study also aimed to explore the design space of d-touch markers, in terms of topological complexity of the symbols drawn, number of unique IDs generated and "collisions" of marker IDs. Two experiments were designed: the first one examined the ability of users to draw valid markers, evaluating also the effect of information provided by the d-touch analyser application, while the second experiment focussed on the creation of markers which are not only valid, but can also be scanned at low resolution.

More formally, our study aimed to test the following hypotheses:

1. people can, with minimal training, draw markers both recognisable by the system and expressing a concept or message which can be understood by others;
2. the information about which features of a symbol violate the d-touch constraints, provided by the *d-touch analyser* application, aids in the creation of functional markers;
3. taking into account limitations due to scanning resolution *robustness* of the markers does not reduce significantly the expressive ability.

4.1 Experiment 1: Valid Markers

4.1.1 Experimental Design

The first experiment was designed as between-groups with two conditions: a "Feedback" condition in which participants were given detailed information if their drawing violated the d-touch rules, and a "No Feedback" condition in which they were just told whether or not their drawing was a valid marker, without explanations. The experiment was carried out in pairs and all sessions were video recorded to allow the analysis of conversation between the participants. Participants were given a total of 1 hour to both study written instructions and to draw as many valid markers as possible.

The written instructions introduced the d-touch system and its rules to define valid markers, illustrated through a number of examples, and described the drawing task. The instructions are available in Appendix A. Only markers with black root were considered, so the rules could be simplified as follows: "a valid marker can be composed of a black region containing 3 or more white regions, and the majority (i.e. more than half) of these white regions must contain one or more black regions. This makes exactly 3 levels of nesting – it must be no more and no less. However, there is no limit in the number and shape of the regions." No mention was made of scanning resolution issues, as this aspect was not covered by the first experiment. The instructions subsequently briefed the participants to draw as many markers as possible that could be placed in a public space to attract attention to any of the following topics: "Music", "Animals", "Pollution/Energy Consumption" and "Children". The topics were chosen to stimulate participants in thinking of familiar and easy-to-visualize items. It was made clear that markers could be distributed in any way subjects liked, from having all of them belong to one topic to an even distribution. Participants were instructed that the level of detail and accuracy should be just enough for someone else to guess which of the four topics each marker is related to.

The experiment was designed for subjects without specific drawing skills nor familiarity with graphic software applications. For this reason, the drawing took place on a white dry-erase board with a black pen: an informal medium that affords playful drawing and easy modifications. While the ability to hand-draw markers can be advantageous in some situations, we imagine marker design to take place mostly through graphic design software tools – we choose to use the whiteboard in the experiment to avoid effects related to familiarity and learning curves of specific software applications or digital drawing devices (e.g., tablet computers), we expected participants from the university to be already familiar with this medium.

To check the validity of the drawings as markers an early version of the d-touch analyser was used, running on a standard desktop computer running MS Windows XP. This version differs from the one described above in that it does not provide information about the segmentation in connected components, it does not include zoom functionality and the interface is based on tabs, rather than thumbnails. The application was connected to a webcam pointed at the board, firmly attached to a desk (so it could not be misplaced). When participants wanted to check their drawings they had to explicitly import them by clicking on a button. The imported images had a resolution of 320×240 pixels. A separate window showed the continuous video feed of images from the webcam, displayed in normal or thresholded (black and white) mode according to participant preference. Particular care was taken in making sure that the image acquired by the webcam contained only the candidate marker, and no other additional object. Because the whiteboard is relatively reflexive, it was necessary to adjust the illumination in the room where the experiment took place to avoid glare. All the information displayed to the users, all the images they checked as well as their actions within the d-touch analyser were logged with accurate time-stamps. The experimental setup is illustrated in Figure 4.1.

In both experimental conditions the d-touch analyser would inform users whether or not the proposed drawing is a valid d-touch marker. In the case of a drawing that is not a valid d-touch marker

Figure 4.1: The experimental setup.

the application behaviour varied depending on the experimental condition. In the "Feedback" condition the d-touch analyser provided information about detected violations of the d-touch rules, namely the presence of nesting beyond three levels, and less than half of the white areas containing black regions. In both conditions the application did not provide any information about scanning resolution or robustness of the markers.

Participants were asked to save their drawings as soon as they were valid and the participants were satisfied with them. A "save" function was included in the application. This required participants to name their drawings according to the relative theme, and it was made available only for valid markers. When attempting to save the system would also inform the users if a marker with the same ID had already been registered within the same experimental session – in such a case participants were asked to modify the current marker to avoid the ID collision. Subjects were asked to swap chairs and roles after drawing each marker, so that the person drawing always sat in front of the whiteboard and the other person in front of the keyboard and mouse. The computer monitor was visible to both participants.

4.1.2 Participants

Thirteen pairs of volunteers were recruited from our university population, through posters and mailing lists, reported in Appendix B and Appendix C. A total of 6 females and 20 males, age range between 19 and 38 (mean: 24.7, std. dev. 4.6). All subjects expressed interest in participating via email, showing familiarity with computers, and received CHF 20 for their time. Anyone who expressed interest and was above 18 years of age was included in the study, no specific drawing skills were required to participate. Advertising and instructions for the experiment were given in English, in which all of the subjects were fluent. All participants provided informed consent through a standard consent form (Appendix D).

4.1.3 Results

The participants in one of the sessions in the "Feedback" condition drew only 3 markers, but these were considerably more complex[1] than the ones produced in the other sessions, therefore data for

[1]More complex in amount of details and consequently in number of branches and leaves

this session was excluded from the analysis.

An average of 13.4 valid markers (std. dev. 5.7) were saved in each of the 1-hour long sessions, ranging from a minimum of 6 to a maximum of 27, median 13, and an overall total of 161. A selection of markers produced in the first experiment is reported in Figure 4.2. In the "Feedback" condition the average per session was 14.7 (std. dev 3.8) while in the "No feedback" condition 12.2 (std. dev. 6.9) and a one-way ANOVA test shows no significant difference. However, removing one outlier (distance from the average is twice std. dev) from the "No feedback" condition, the average for "No feedback" is 9.2 and std. dev 2.28, and one-way ANOVA test indicates that this difference is significant ($p < 0.05$).

In 6 of the sessions participants drew markers with the same ID as one of their previous entries. This happened between 1 and 3 times per session; in all cases subjects modified their marker and solved the ID conflict within seconds. Out of a total of 161 valid markers submitted, 116 had different IDs, corresponding to an ID collision rate of 27.9%. It must be underlined that subjects were blind to the IDs generated by other participants.

The number of inner regions, or *branches* and *leaves* in the terminology defined above, was taken as an indicator of the complexity of the markers; markers had on average 5.2 branches (std. dev. 3.1) with a minimum of 3 (the minimum accepted by the system) to a maximum of 19, median 4. Figure 4.3 shows the distribution of these values. The number of leaves was on average 8.8 (std. dev. 7.0), minimum 2, maximum 49, median 7. One-way ANOVA showed no significant differences either in the average number of branches or in the average number of leaves per marker between the two conditions.

4.1.3.1 Application Logs and Video Recordings

The log files collected by the d-touch analyser application show that participants tested their drawings 482 times over the entire experiment. 195 times (40.5%) the candidates were valid markers, while 287 times (59.5%) they were not. The invalid candidates were manually categorised according to the reason for being invalid; this analysis revealed that in 153 of the 287 times (53.3%) an input drawing was not valid, it was because of "artefacts" specific to the whiteboard, such as incomplete pen strokes and gaps in large filled areas, which modified the topology of the drawings. Analysis of the video recordings showed that these artefacts often caused frustration, and that participants often attempted, and succeeded, to detect problems in the drawings of their partners before checking through the d-touch analyser.

The number of explanations about why candidate drawings are or are not valid markers was counted for each session. Over the entire experiment 208 explanations were formulated by participants, corresponding to an average of 17.33 per session (st. dev. 7.57). Participants in the "Feedback" condition formulated 128 explanations (on average 21.33 per session, st. dev. 8.26) while those in the "No feedback" conditions a total of 80 (on average 13.33 per session, st. dev. 4.41).

Generally participants were observed to enjoy the experimental task, several of them self reporting that they had fun. The creative and collaborative nature of the task, together with the physical interface and the computer providing feedback about success or failure of the attempts, probably made the activity be perceived as playful, in most of the cases.

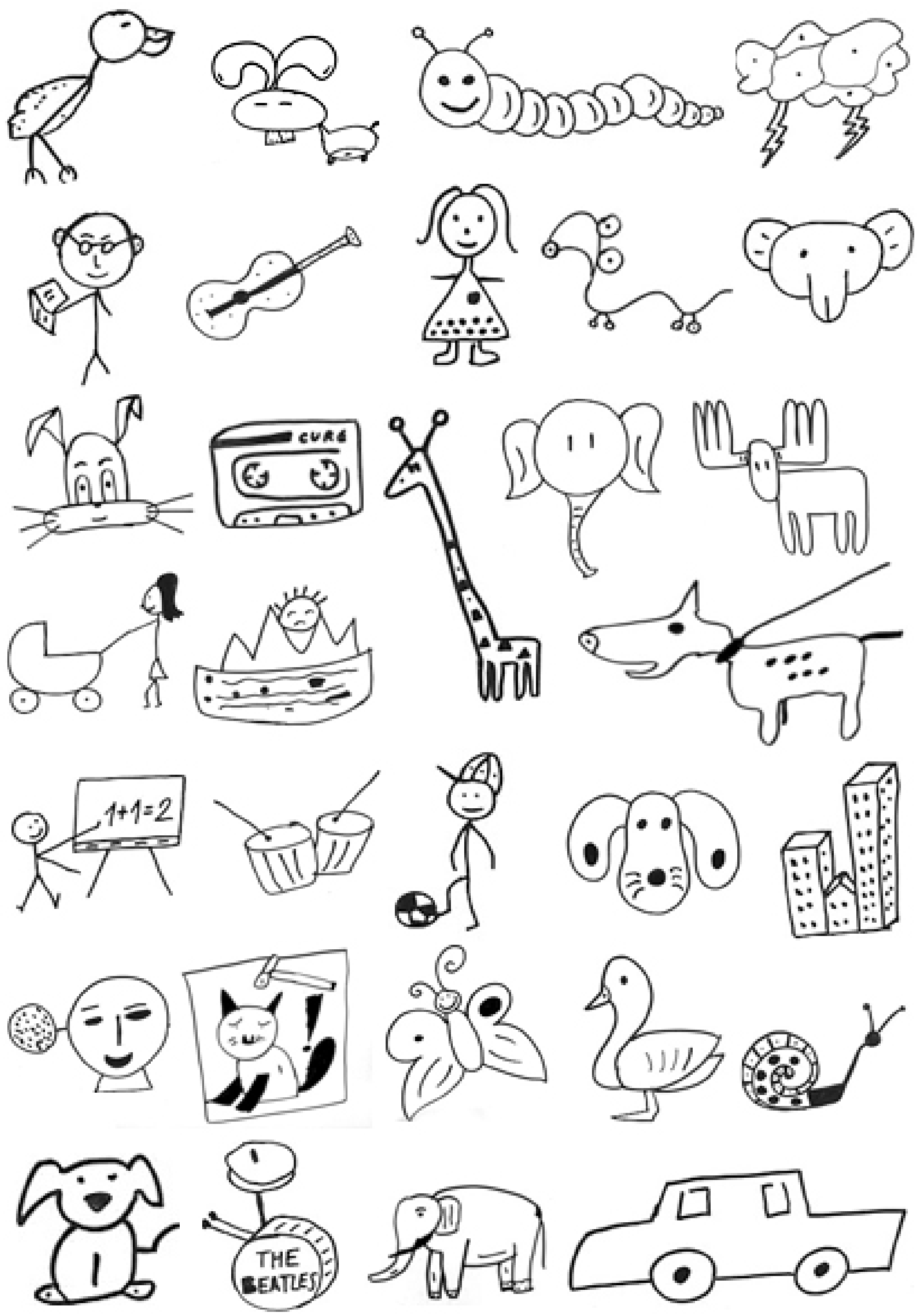


Figure 4.2: A selection of valid markers produced in the first experiment.

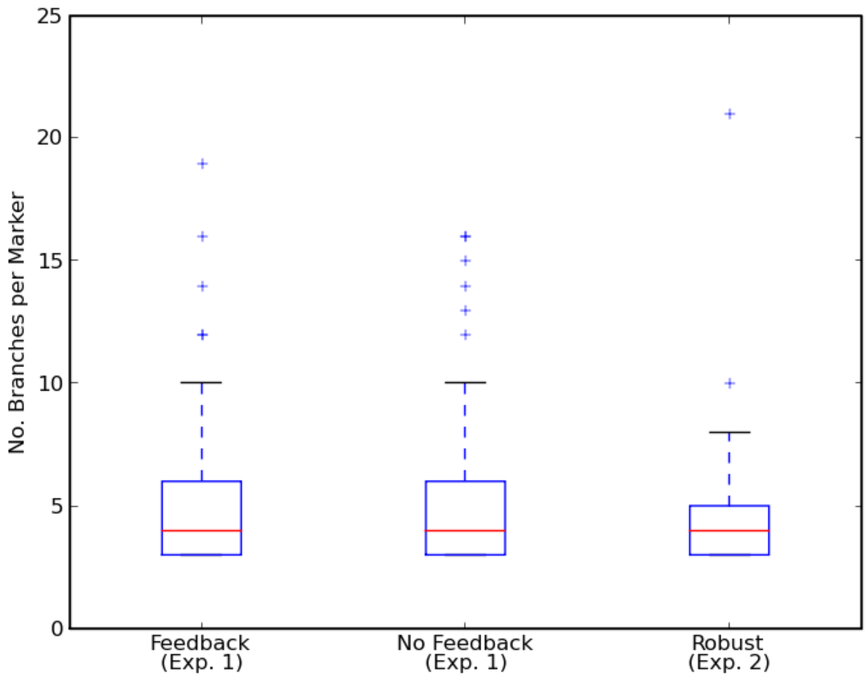


Figure 4.3: Boxplot of the number of branches (white regions, in this case) per marker, for the entire study. The red horizontal line indicates the median value, the box displays the inter-quartile range containing the central 50% of the data, while the whiskers indicate the extension of the data without outliers (these are shown by the crosses).

4.1.3.2 Marker Coding

Two volunteers, also recruited from the university population, served as independent coders of the symbols drawn in the study. The coding took place through a simple web application where the markers were shown one at the time, and for each one the coder was asked to answer: *“ Which theme do you think this drawing is related to?*
1. Music 2. Animals 3. Pollution 4. Children
5. None of the above ” and to also enter a freeform short description of the marker.

In 133 of 161 cases (82.6%) the coders agreed with each other. In 123 of the cases (76.4% of the total) both coders’ choice was in agreement with that declared by the creator of the marker. For 148 of the markers (91.9%) at least one of the two coders guessed the intention of the creator of the marker.

4.1.4 Discussion

The fact that in all sessions participants managed to draw at least six valid markers, together with the high recognition rate of the marker meaning by coders, confirms the first hypothesis: people

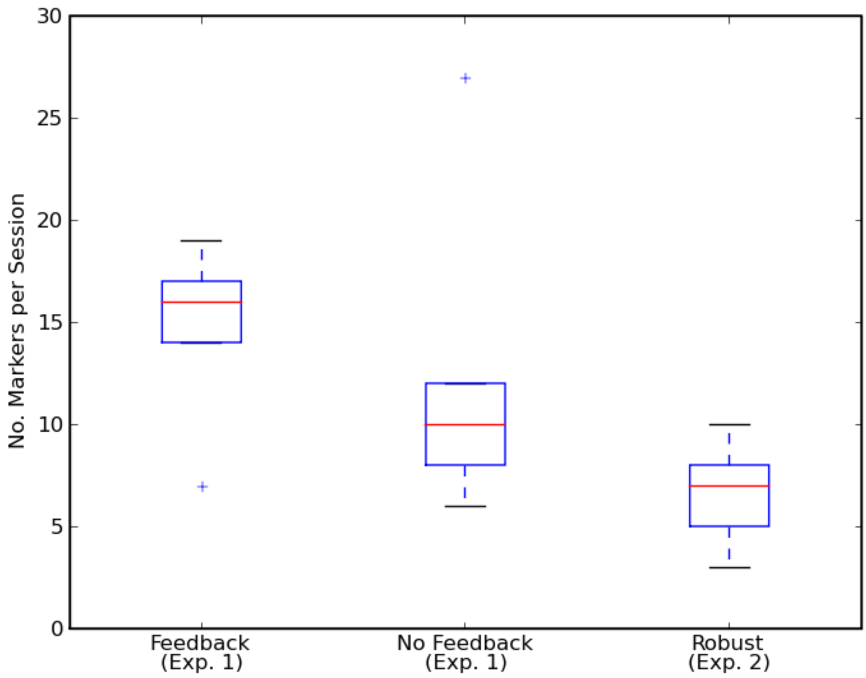


Figure 4.4: Boxplot of the number of valid markers created per session.

can, with minimal training, draw markers recognisable by the system and that at the same time express a concept or message which can be understood by others. The difference in the number of markers produced per sessions suggests that the d-touch analyser application can help in creating valid markers. The distribution of branches indicates that participants were able to create complex markers, but generally drew simple ones – indeed in the experiment there was no requirement nor incentive to favour complexity.

Even though participants were warned in the instructions, artefacts specific to the whiteboard were often cause of frustration. The version of the d-touch analyser application used in this first experiment was unable to help with this kind of artefacts, so their presence somehow "disturbed" the comparison of the two conditions. The visualisation of segmented components was added to the d-touch analyser to remedy to this problem. Automatic closure of open contours was considered as a possible extension of the software, however, it was noticed that sometimes users deliberately leave small gaps in the drawings as a way to reduce the levels of nesting – for example this is the case for the windows of the car on the bottom right of Figure 4.2 – making it very difficult to distinguish artefacts from deliberate choices.

The higher number of explanations formulated in the "Feedback" condition may be interpreted as the fact that additional feedback triggers more discussion. In fact, the additional information available in the "Feedback" condition might have been used to reason about the artefacts.

4.2 Experiment 2: Robust Markers

4.2.1 Experimental Design

Based on the results from the first experiment, the second experiment aimed at assessing the creation of markers which are not only valid, but also robust to blur and low resolution scanning. The experimental setup for the new experiment was nearly identical to the first one, except that this time the instructions covered the issues of marker robustness (Appendix A), and subjects were asked to draw markers which could be read at a (simulated) resolution of no more than 200×200 pixels or preferably no more than 100×100. The d-touch analyser application was used in its current version, as described in the “d-touch Technical Overview” section.

4.2.2 Participants

Eight new pairs of volunteers took part in the second experiment, 9 females and 9 males, age range between 20 and 32 (mean: 24.6, std. dev. 2.8). Subjects were recruited and compensated as in the first experiment. All participants provided informed consent through a standard consent form.

4.2.3 Results

Of the markers submitted by the participants only one did not comply with the resolution requested in the instructions and this was excluded from the following analysis. A total of 58 valid markers were submitted over the 9 1-hour long sessions, on average 6.4 valid markers in each of them (std. dev. 2.0), ranging from a minimum of 3 to a maximum of 10, median 7. A selection of the markers produced in the second experiment is reported in Figure 4.5. Out of the 58, 36 markers (61.1%) satisfied the more stringent requirement of 100×100 pixels. The number of distinct IDs produced was 44 (75.9%).

The average number of *branches* was 4.6 (std. dev. 2.6), ranging from 3 to 21, median 4. The number of leaves per marker was on average 6.26 (std. dev. 4.51), minimum 2 and maximum 25, median 5. Comparing these values with the results from the “Feedback” condition in Experiment 1 through one-way ANOVA reveals that the difference in number of markers produced per session is significant ($p<0.01$), as well as the difference in number of leaves per marker ($p<0.01$), while no significant differences were found for the number of branches per marker.

4.2.3.1 Marker Coding

The same volunteers who coded the first experiment coded the second one as well, following the very same procedure. In 47 of 58 cases (81.0%) the coders agreed with each other. In 44 of the cases (75.9% of the total) the choice of both coders was in agreement with what declared by the creator of the marker. For 54 of the markers (93.1%) at least one of the two coders guessed the intention of the creator of the marker.

Figure 4.5: A selection of valid markers produced in the second experiment.

4.2.4 Discussion

The second experiment demonstrates that novice users can create robust valid markers with minimal training. Compared to the first experiment, the increased complexity of the task resulted in fewer markers created in each session, and in lower complexity, at least taking the number of leaves as an indicator.

4.2.4.1 Limitations

While the whiteboard proved to be easy to use by our subjects, the artefacts suggest that the use of a digital medium, such as a tablet computer should be reconsidered. The performance recorded in the experiments is relative to a set initially empty: the difficulty of the task can be expected to increase if subjects had to add markers with unique IDs to a set already populated with hundreds of markers. More generally, the relatively small size of the ID address space (especially when compared to 2D barcodes) implies that markers cannot encode URLs directly, so the IDs must be used as keys to a database (on-line or residing on the phone itself). Similarly, it would be difficult to encode checksum information directly in the topology-based IDs, so redundancy is more likely to be included through geometrical features.

4.2.4.2 Distinct IDs

Across both experiments a total of 219 valid markers were created, 149 of them (64.8%) corresponding to distinct IDs, while 34 IDs were repeated between 2 and 9 times. It must be emphasised that each pair of users was only aware of the IDs that they produced during the experiment and that they were "blind" to the ones produced by other groups. While further investigation is needed to assess how many distinct IDs can be supported by the system, we expect that the number of IDs will be large enough to cover many mobile applications. We envision markers and IDs to be application-specific, rather than having one central repository: with each application maintaining its own database of d-touch markers. Moreover, multiple markers can be combined together to increase the ID space; markers can be placed next to each other and scanned simultaneously by the client. As an example, using 2 markers from a set of 44 (as those produced in the second experiment) yields 990 combinations.

4.3 Probing Different IDs: a Simple Marker Design Simulation

To further asses the limits in the number of distinct marker IDs can be supported by the d-touch recognition system, a simple software simulation was carried out. Rather than attempting to mimic the creative process of human drawing or graphic design, the simulation was based on a simple mechanism to automatically and randomly generate markers to evaluate the number of "ID collisions."

As shown in some of the examples in Chapter 3, valid d-touch markers can be formed from text. This process can be automated by filling up holes in the letters (e.g. in the letters 'a','b','d','e'),

Figure 4.6: A d-touch marker generated from text.

by creating a white contour around each word, and by creating a black contour around a group of words. An example marker generated through this process is show in Figure 4.6. This simple algorithm simply allows to generate markers in an automatic way by randomly selecting words and grouping them.

To more easily interpret the results of the simulation, a basic distance metric was defined on marker IDs, based on informal observations around the misclassification of markers. It was noticed that misclassification generally involves markers with the same number of branches, because one or more small leaves vanish after the thresholding step (for example, because of blur, over-exposure, or under-exposure), or one leaf is read as two separate ones. These errors result in a distortion of the marker ID, where one or more digits are off by one or more units. To reflect this observations, the distance between two markers m_a and m_b with IDs respectively $ID_a = a_1, a_2, ..., a_n$ and $ID_b = b_1, b_2, ..., b_n$, where a_i and b_i are the number of leaves in branch i (and branches are sorted in non-decreasing order), is defined as: $d(m_a, m_b) = \sum_{n}^{i=1} |a_i - b_i|$ The larger the distance between two markers, the more unlikely it is that they will be misrecognized for each other. A simple way to reduce instances of misclassification then is to work with a set of markers that are well distanced from each other.

A total of 11340 markers were generated starting from text randomly selected from Wikipedia pages. These markers were all "robust" according to the d-touch analyser application, and were all different lexicographically. The graph in Figure 4.7 reports the number of distinct IDs plotted against the number of generated markers. On the same graph it is also plotted the number of markers having distance larger than 1 and larger than 2. If the random generation is interpreted as a *brute force* approach to resolve ID duplication, the simulation's results suggest that it is relatively easy to generate sets of individual markers with few hundreds with different IDs, but it is difficult to go above this amount. Of course the approach of combining multiple markers to produced different IDs can still provide a good solution to increse the number of markers.

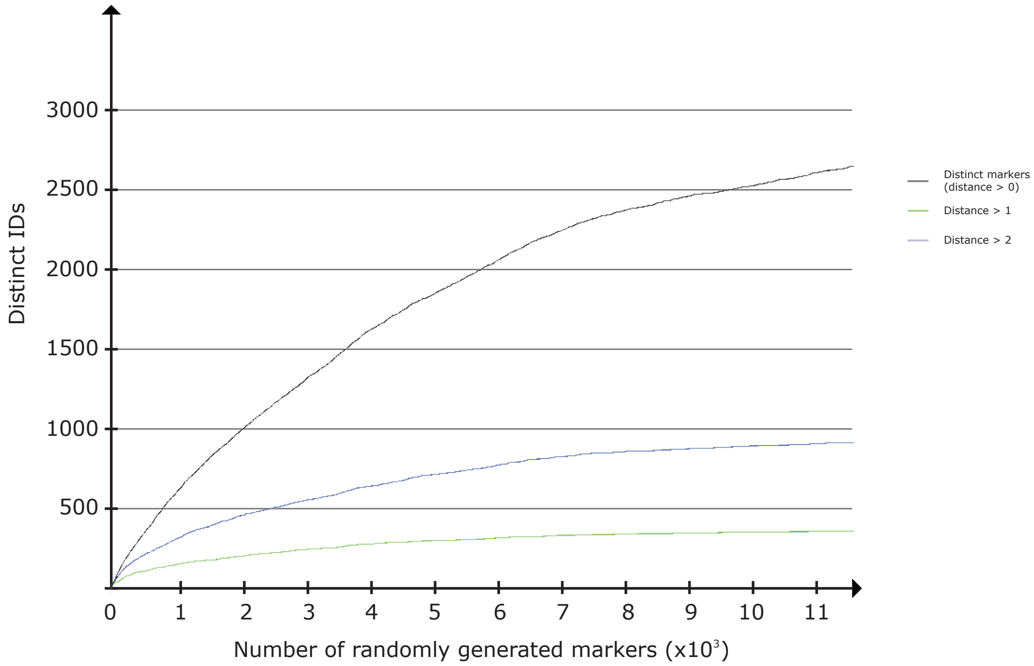


Figure 4.7: A plot of the results of the marker generation simulation.

4.4 Conclusion

The results reported in this chapter provide an answer to the first two research questions formulated in the introduction. The large number of valid markers created by participants in the user study demonstrates that the topology-based recognition adopted in d-touch makes it possible to define visual markers that can be visually expressive to humans, while being easy and efficient to identify automatically. The comparison between the two experimental conditions in the first experiment shows that the d-touch analyser is an effective tool that supports users in the marker design activity. More examples of d-touch markers designed by users outside of a laboratory situation, and taking advantage of the d-touch analyser, are reported in Chaper 6 and Chapter 7 providing stronger ecological validity for both these research questions.

5. Usability of Marker-based Mobile Interfaces

Visual markers can be used within mobile interfaces to select physical items and trigger interaction related to them. Marker scanning takes placed through an augmented viewfinder, which highlights markers when they are in view, implementing a *magic lens* metaphor. Once a marker is scanned, the viewfinder can be closed and the interaction flow shifts away from the marker to a standard phone GUI, for example based on lists. In contrast to this scenario, this chapter explores ways to achieve a tighter integration of the markers in the mobile interaction: markers are continuously at the centre of the interaction flow. Example applications of these interfaces will be described in Chapter 6.

To understand the potential of visual markers for the design and development of mobile user interfaces, a small variety of mapping strategies is explored. All of them are based on the *object-action interface paradigm* (Shneiderman and Plaisant 2004, p. 95), which states that any interface can be decomposed in a set of objects and a set of actions. For example, for a word processing application objects include *words* and *documents*, and actions include *open*, *save*, *select*, *copy*. On graphical user interfaces actions and objects are normally associated with icons, buttons or other widgets, but what are appropriate representations for them within marker-based mobile interfaces? To attempt to answer this question different mapping strategies were designed, prototyped and evaluated through a user study. The usability evaluation presented in this chapter aims at providing a basic indication of the performance of different ways to integrate markers in mobile interfaces, to inform future design.

The work was carried out first on keypad phones, and then further developed and adapted to the different requirements of touch-screen devices. Even though touch-screen mobile phones have been available since at least 2001 through the Microsoft Pocket PC and Symbian UIQ platforms, it was probably the launch of the Apple iPhone in 2007 that made touch-screen truly popular on mobile phones. There is therefore an interest in extending the marker-based interfaces to touch screen devices. The availability of touch-screen phones also within the S60 platform made it relatively easy to prototype marker-based interfaces on this type of devices, in particular on the Nokia 5800 model.

The following section describes 4 interface variations on keypad phones, followed by a brief overview of their implementation. A first experiment comparing them is then presented and discussed. Following, the chapter illustrates how the interfaces had to be modified to work on touch-screen phones, and then presents and discusses a second experiment.

5.1 Marker-based Mobile Interfaces on Keypad Phones

For all interface variations objects are represented by markers, these markers are referred to as *object-markers*. What distinguishes the various interfaces is the representation of actions.

Action-marker . Each action is mapped to a marker: to perform an action users need to point the phone to an action-marker first and then point it to an object-marker, without pressing any key, as illustrated in Figure 5.1. The choice of avoiding key-presses was inspired by the results by Parikh et al. (2006), who reported that users not familiar with technology found it difficult to aim the camera phone at a target and press a key at the same time.

Action-key . Each action is mapped to a key on the phone keypad. To apply an action users need to press the corresponding key while the phone is pointed to an object-marker. Visual cues are shown on the phone display to indicate which key corresponds to which action, as illustrated in Figure 5.2. If a key is pressed while the phone is not pointed to a marker nothing happens. Labels for keys are displayed all the time (not just when marker is in view).

Action-position . Actions are mapped to the relative position of phone and marker, when the marker is within the phone viewfinder. Two variants are defined: **Action-Position Finder** and **Action-Position Print**. The first one uses the phone's viewfinder as a frame of reference: this is divided in areas corresponding different actions, as illustrated in Figure 5.3 left. To perform an action users need to let an object-marker be in the corresponding area. The centre of the finder is left empty and it corresponds to no action. The second variant (action-position print) uses the printed marker as a frame of reference, the physical surface around the marker is annotated with text labels (or icons) indicating the actions, as illustrated in Figure 5.3 right. To perform an action, users need to align the centre of the viewfinder, marked with crosshairs, with the desired action label, while keeping the marker in view. Similarly to the previous case, the central area of

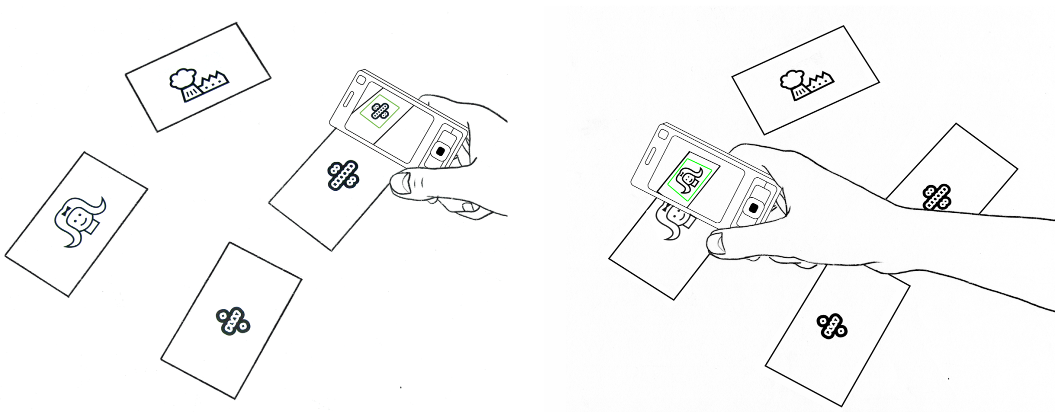


Figure 5.1: An illustration of the *action-marker* mapping strategy. The two iconic markers represent objects, while the two text-based markers represent actions. An action is selected first (left) and then an object (right).

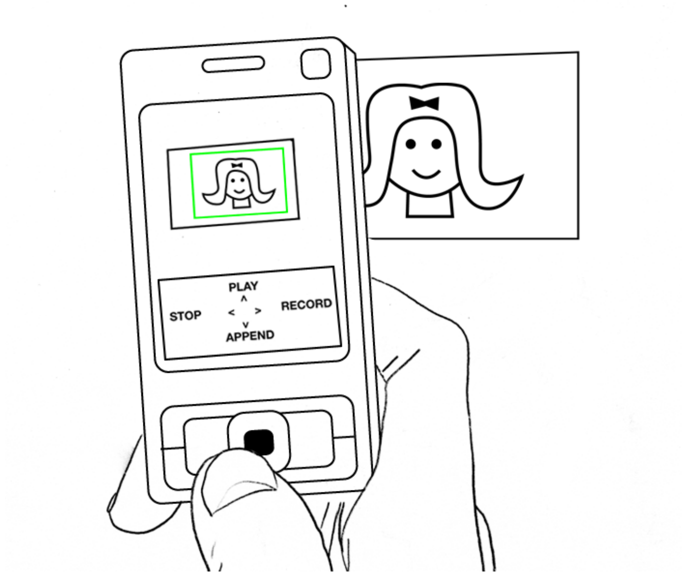


Figure 5.2: An illustration of the *action-key* mapping strategy: each action corresponds to one of the keys on the phone.

the interface (in this case the marker) corresponds to no action. Like in the action-marker case, the interface was designed to avoid key-presses; to avoid involuntary selections, in both variants users need to keep the phone in the desired position for 4 subsequent frames. This mapping was informed by the work of Rekimoto and Ayatsuka (2000b) and Rohs (2005).

The 3 strategies are not equivalent. In the action-position interface the number of action that can be presented is limited by the resolution of the viewfinder, and in the Action-Key it is limited by the number of available physical keys on the phone, while the action-marker allows as many actions as the number of different markers. In contrast, the Action-Marker interface has the disadvantage of

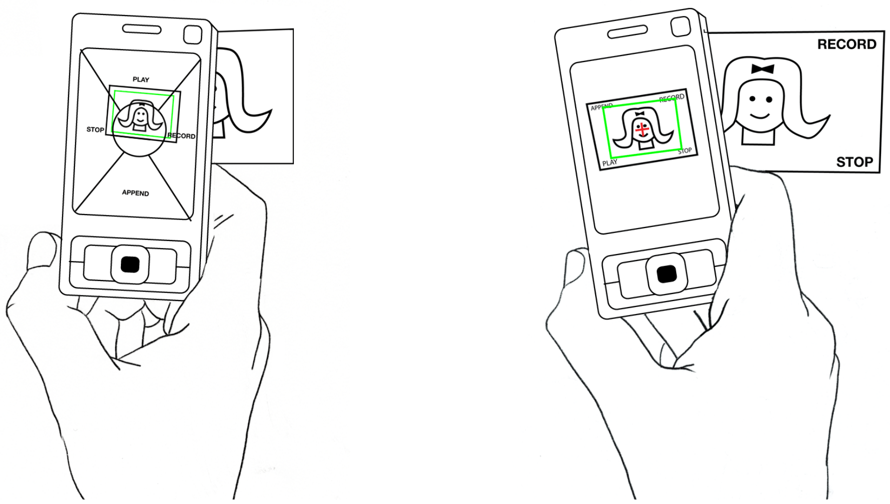


Figure 5.3: On the left: an illustration of the *action-position finder* mapping. On the right: an illustration of the *action-position print* mapping.

requiring more markers than the alternatives, and to use the same form of representation for both objects and actions, while in the other two interfaces the two categories are more clearly separated. In the action-key and action-position strategies, actions can be defined "'contextually" depending on the marker; the action-marker interface does not allow this option.

5.1.1 Implementation

Prototypes for the 3 interfaces were implemented using Python for S60 and the dtevents module described in Section 3.4. Events related to marker recognition are handled in combination with events from the phone keypad using a simple finite state machine. To make the code modular, a generic EventManager class was defined as a virtual super-class. Subclasses implementing the specific mapping strategies were derived from EventManager.

The following are the main classes used in the application:
The **EventManager** class is an abstract class representing the interface mapping strategy. Its concrete subclasses implement specific mapping strategies.
The **Screen** class is responsible for what is displayed on the phone screen, and offering an API for displaying text or icons corresponding to actions such as play and record.
The **Application** class contains instances of all other classes and incorporates the application's logic, essentially integrating all pieces together.

5.2 Experiment 1

A user study was conducted to compare the different marker-based interfaces. Of the two action-position variations, only the finder version was tested, to limit the number of experimental conditions, given also the similarity of the two. Performance was defined in terms of error rate and task completion time, to provide a simple and consistent measure for the comparison of the alternatives. Task completion time was used as an intermediate metric to compare the performance of the different interfaces. This choice was based on the assumption that a more comfortable and effective interface, everything else being the same, would lead to a faster task completion. At the end of the experiment subjective preferences of the participants were also collected.

5.2.1 Experimental Design

To provide context and add ecological validity to the experiment, participants were described a simple but credible application scenario: the mobile phone and a set of cards with markers printed on them would be used to record, playback and arrange sounds in sequences. This application was only briefly described and not demonstrated to avoid confusion, and it is based on the Mobile Audio Cards application described in Chapter 6.

Based on this scenario, participants were asked to complete 18 selection tasks with each interface. The tasks were defined as the combination of 3 actions, *play*, *record* and *append*, with 3 iconic object-markers representing a *dog*, a *girl* and a *factory*. The sequence of tasks was randomized

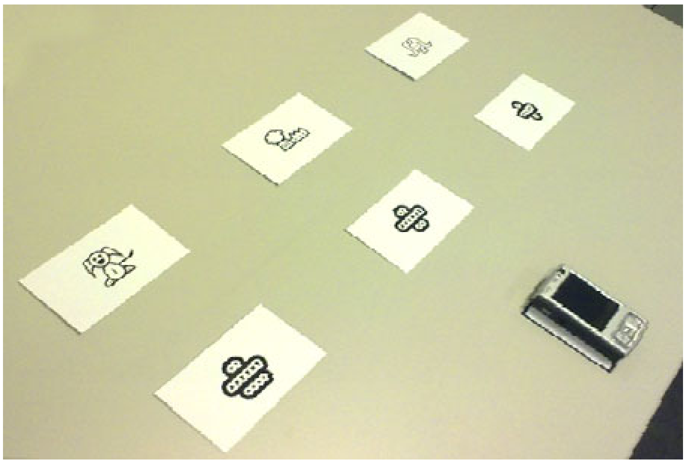


Figure 5.4: The experimental setup used in both Experiment 1 and Experiment 2.

and it included each of the 9 action-object combinations twice. The experiment took place on a desk, and participants were asked to stand next to it. To provide a consistent setup across all trials, markers were printed on 6 cards of size 12 by 7.5cm and these were affixed to the desk, as illustrated in Figure 5.4. The rest position for the mobile phone was also clearly marked, and subjects were asked to place the device on it after the completion of each task. When the Action-Key or Action-Position interfaces were in use, the cards representing the actions were covered with white pieces of paper. A Nokia N95 phone was used for the experiment. It is worth emphasizing that these tasks are not equivalent to Fitts' Law tasks, as the experiment is not assessing purely pointing performance.

All subjects used all 3 interfaces: within-subjects design. The order of the interfaces was fully counterbalanced, to minimise a possible learning effect. Task completion time was measured starting from the lifting instant until the action and object selection was completed. To recognize this lifting gesture, the accelerometer built into the phone was used – detection was based on the acceleration along any of the 3 axes exceeding a threshold. The camera was used to detect when the phone had been placed back on the desk. The target object and action were shown only when the device was placed back on the desk and until it was lifted. After each task's completion, feedback about correct or incorrect selection was also provided on the phone display. The selection was limited to 15 seconds, after this time the trial was counted as a miss and the experiment proceeded to the next task.

Before performing each group of 18 selection tasks, participants were given written instructions describing the interface they were about to use and they were asked to familiarise with it for about 2 minute. The instructions are reported in Appendix E. They were asked to complete the task as quickly and as accurately as possible, and informed that the task completion time was measured starting from when the phone was lifted from the table.

At the end of the experiment participants were asked to complete a paper questionnaire, including questions about the subjects personal data and general background, as well as subjective evaluation of the interfaces they had just tried. In particular subjects were asked which interface they found easier to use and why; which one they found most accurate and why; which one they found most

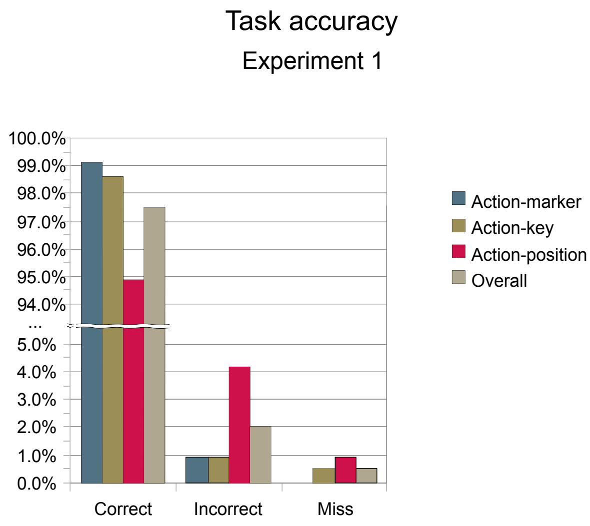


Figure 5.5: Task completion accuracy, results from Experiment 1.

enjoyable and why; and finally which one they overall preferred and why. The entire experiment took less than 30 minutes, including the 3 groups of selection tasks plus the time to read the instructions and fill out the questionnaire.

5.2.2 Participants

Participants were 24 adults: 2 women and 22 men. All were volunteers recruited from the university through mailing lists, with the incentive of entering a lottery to win CHF 100 (the email used for recruitment is reported in Appendix F). Nineteen participants reported not having used similar interfaces before, while 5 had experienced mobile interfaces based on visual markers. Participants average age was 25.6 years (st. dev 4.0), all of them had a background in computer science, except 2 who were studying art history and mechanical engineering. Advertising and instructions for the experiment were given in English, in which all of the subjects were fluent. All participants provided informed consent through a standard consent form (Appendix G).

5.2.3 Results

5.2.3.1 Performance Measures

Overall subjects performed correct selections in 97.5% of the tasks (1264 out of 1296 times), incorrect selections occurred in 2.0% of the cases (26 times), while in 0.5% of the cases (6 times) no selection was performed within 15 seconds. In the *action-card* condition correct selections were performed in 99.1% of the cases (428 out of 432 times), while in the remaining 0.9% wrong selections were performed, and no misses. In the *action-key* condition correct selections were

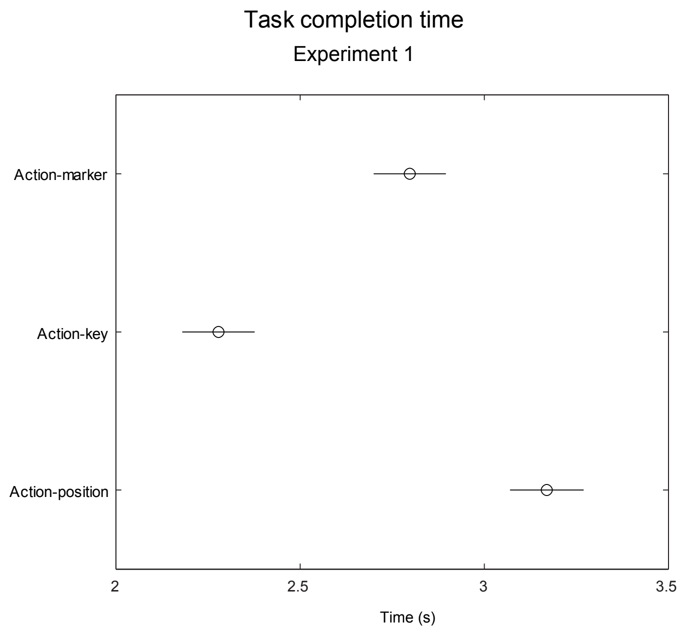


Figure 5.6: Task completion time, results from Experiment 1. The circles indicate the averages, while the bars indicate the 95% ANOVA confidence intervals.

performed in 98.6% of the cases (426 out of 432 times), wrong selections in 0.9% of the cases (4 times), and no selections in 0.5% (2 times). Finally, in the *action-position* condition correct selections were performed in 94.9% of the cases (410 out of 432 times), wrong selections in 4.2% (18 times) and no selection in 0.9% (4 times). These results are summarised in Figure 5.5.

The general average task completion time for the successful selections was 2.74 seconds (STD=1.27). Averages per each interface type are reported in Table 5.1 and in Figure 5.6. A 1-way ANOVA analysis of the task completion time showed that the pair-wise differences determined by the interfaces have a statistically significant effect ($p < .01, F = 56.2$). Using a t-test no significant differences in task completion time were found between participants who reported previous experience with marker-based interfaces and those who used them for the first time during the experiment.

5.2.3.2 Interface Preferences

The subjective ratings are reported through the histograms in Figure 5.7, expressed as percentages of preferences given that two subjects expressed more than one preference for some of the

Interface	Average	St. dev.
Action-marker	2.80	0.90
Action-key	2.28	0.91
Action-position	3.17	1.70

Table 5.1: Task completion time, in seconds, for Experiment 1.

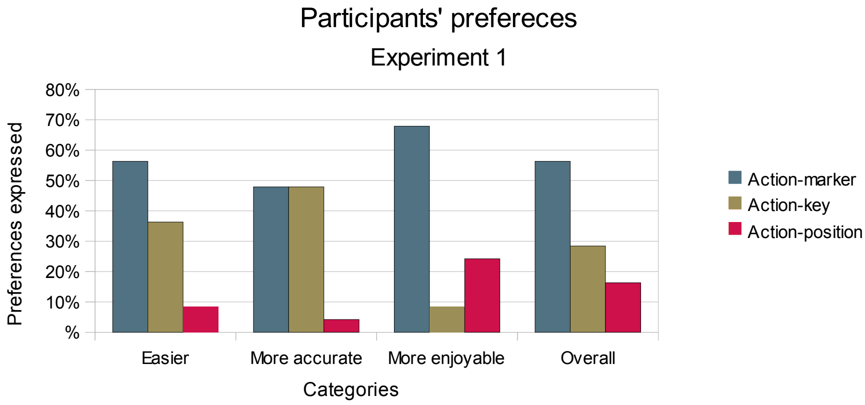


Figure 5.7: Preferences expressed by participants of Experiment 1.

questions. In summary, the majority of participants found the action-marker interface preferable in all categories except for accuracy, where the same number of preferences were expressed for the action-marker and action-key interfaces. The action-position interface scored the lowest for all categories except one: it was found to be more enjoyable than the action-key interface.

The reasons expressed by participants to prefer one interface over the other were coded and grouped into broad categories. The results were quite varied.

- The action-marker interface was often preferred because it does not require accurate aiming (8 participants), it does not involve the use of keys (7 participants) and because of the clear visibility of the action labels so that there is no need to memorise the key associations, so not much visual attention needs to be devoted to the screen (7 users). Four participants also appreciated the sequential nature of the action-marker interface, and one of them pointed out that the interface is spatial, and therefore more easily memorable than for example keys.
- The main reasons to prefer the action-key interface was that it provided more control (7 users) and that it felt familiar because mobile phone interfaces are normally based on keys (5 users). Three users reported to like the action-key interface because it only required one step, one pointed out that it does not require accurate aiming and another one appreciated the separate representations (keys and markers) for actions and objects.
- Among the participants who preferred the action-position interface, three liked it because it required more precision and the challenge made it feel similar to a game. One participant liked its novelty and another one that it requires only one action (compared to the two scans of the action-marker and the scan plus key press of the action-key).

5.2.4 Discussion

The high correct selection rates for all of the interface variations demonstrates that all interfaces are usable: through any of them users were generally successful in performing the required tasks.

The *action-key* mapping strategy resulted in the fastest selections, and accuracy just below the *action-marker* mapping. The comments expressed by participants suggest that the key-based interface was deemed more accurate than the others because of the greater control that it provides: indeed the explicit key-press confirms the selection. Therefore, even though the measured performance level is just below that of the *action-marker* mapping, the *action-key* mapping could be preferred when incorrect selections have a cost for the users, and their subjective perception is more important than the actual accuracy.

The *action-marker* mapping strategy produced highest accuracy but it was, on average, 23% slower than the *action-key* one. Despite being slower, it was the most preferred of the 3 options. Parikh et al. (2006) report that users who are not familiar with technology, such as those in rural India, find it difficult to aim the camera phone at a marker and at the same time press its keys. The results produced by this experiment suggest that interfaces that do not require aiming and pressing at the same time are also preferred by those familiar with technology, such as the participants of our experiment. The *action-marker* strategy then should be preferred, if the task completion time is not critical, and it is not problematic to have additional markers to represent the actions.

The longer time required to complete tasks with the *action-position* interface, as well as its higher error rate, show that this option was less efficient and effective than the alternatives. Such observation is reinforced also by the participants subjective ratings. The task completion time resulting from the *action-position* condition, compared to the *action-marker* one, shows that the higher precision aiming required by the action-position mapping is more time consuming than even aiming at 2 different targets with coarser precision.

5.3 Marker-based Mobile Interfaces on Touch-screen Phones

While both the action-marker and action-position mapping strategies can be employed on touch screen phones without any modification, the action-key mapping needs to be adapted to use touch-sensitive virtual buttons. To make the buttons more realistic, as it is commonly found on this type of interfaces, haptic feedback is produced through the phone built-in vibrating motor when the button is touched and the button is highlighted in a different colour while pressed. As illustrated in Figure 5.8, the touch buttons are positioned at the bottom of the screen, to make the interface layout similar to the one used for the action-key interface.

The touch-screen allows the definition of virtual buttons anywhere on screen, so they can be placed close to the marker they refer to, as shown in Figure 5.9. This may reduce the cognitive distance between the marker and the buttons. At the same time, because the position of the buttons is not static with respect to the screen, they may become more difficult to trigger.

The first mapping strategy, where touch buttons have a fixed position with respect to the device is named **action-touch fixed**, while the second one, where buttons are displayed next to the marker they refer to, is named **action-touch moving**.

The *action-touch-moving* strategy has the advantage of allowing, in principle, to have multiple markers in view at the same time, each with its own set of buttons. It must be noted, however, that in practice with the current generation of phones it may be difficult to accommodate multiple markers and multiple sets of buttons at the same time, because of the limitations on the viewfinder resolution imposed by the display technology as well as the processing speed.

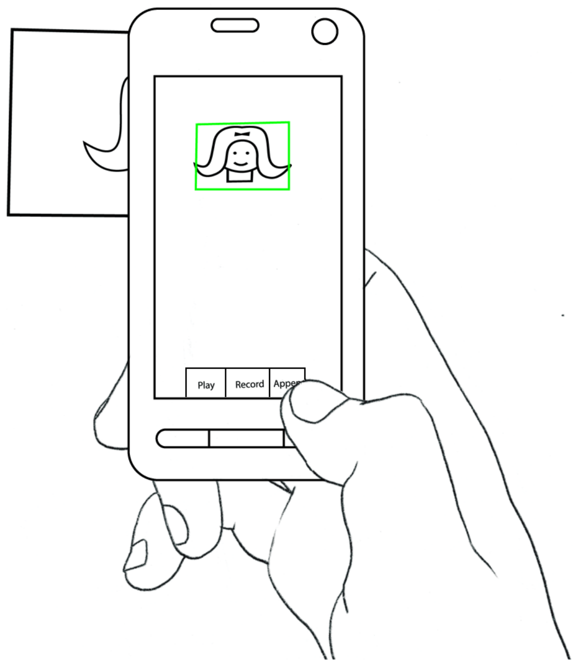


Figure 5.8: An illustration of the *action-touch fixed* mapping strategy: touch-screen buttons are at the bottom of the screen.

5.3.1 Implementation

The software architecture described earlier was extended to accommodate the new mapping strategies. In Python for S60 touch events are handled similarly to key-press events, so it was particularly easy to extend the previous work.

One technical issue arose from the fact that the 5800 screen has a higher resolution than N95, so when the camera finder fills the screen horizontally the image resolution is 360x240 pixels, rather than 240x160. The number of pixels is larger by a factor of 2.25, which made the processing of each frame noticeably slower than before. While the system might still be usable at the lower frame rate, it was desirable to make the two experiment as similar as possible, to help comparing their results. Using a smaller viewfinder was found undesirable because the physical size of the image made it difficult to display touch-buttons. The solution was then to acquire images from the camera and process them at a low resolution (240x160), and once processed, interpolate them to fill the screen horizontally.

5.4 Experiment 2

5.4.1 Experimental Design

The design was almost identical to that of the first experiment. A screen-based questionnaire was used in place of the paper-based one used earlier, to simplify the processing of the answers. An extra question was added, asking users about their experience with touch-screen phones.

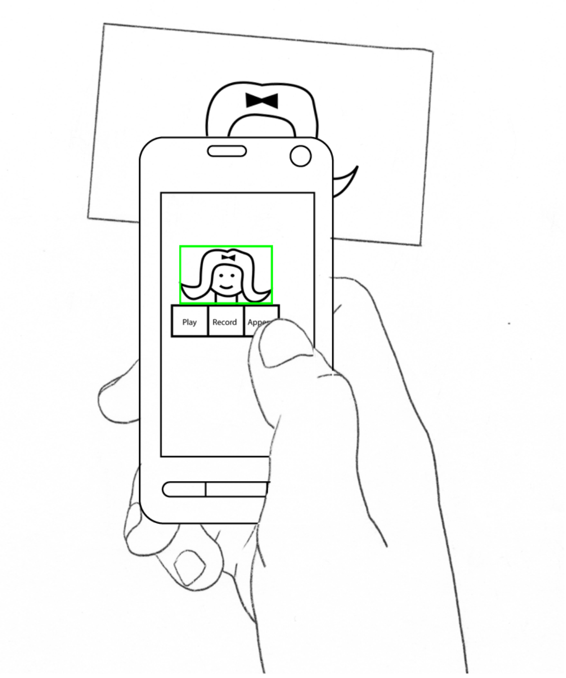


Figure 5.9: An illustration of the *action-touch moving* mapping strategy: touch-screen buttons are displayed below the object-marker they refer to.

5.4.2 Participants

Participants were 18 adults: 6 women and 12 men. All were volunteers recruited from the university through mailing lists, with the incentive of entering a lottery to win CHF 100. None of them took part in the previous experiment. Regarding prior experience with touch-screen devices, 2 participants reported to have no prior experience, 4 reported to own a touch-screen device, while 12 – the vast majority – reported having tried such devices sporadically. The average age was 21.5 years (st. dev 2.1), all participants had a background in computer science. Advertising and instructions for the experiment were given in English, in which all of the subjects were fluent. All participants provided informed consent through a standard consent form.

5.4.3 Results

5.4.3.1 Performance Measures

Overall subjects performed correct selections in 96.9% of the tasks (1046 out of 1080 times), incorrect selections occurred in 2.9% of the cases (31 times), while in 0.3% of the cases (3 times) no selection was performed within 15 seconds. In the *action-card* condition correct selections were performed in 98.1% of the cases (353 out of 360 times), while in the remaining 1.9% (7 times) wrong selections were performed, and no misses. In the *action-touch fixed* condition correct selections were performed in 96.4% of the cases (347 out of 360 times), wrong selections in 3.1% of the cases (11 times), and no selections in 0.6% (2 times). Finally, in the *action-touch moving*

Interface	Average	St. dev.
Action-marker	3.59	1.23
Action-touch fixed	3.12	1.61
Action-touch moving	3.37	1.42

Table 5.2: Task completion time, in seconds, for Experiment 2.

condition correct selections were performed in 96.1% of the cases (346 out of 360 times), wrong selections in 3.6% (13 times) and no selection in 0.3% (1 time). These results are summarised in Figure 5.10.

The general average task completion time for the successful selections was 3.36 seconds (STD=1.44). Averages per each interface type are reported in Table 5.2. A 1-way ANOVA analysis of the task completion time showed that the difference between the *action-marker* and *action-touch fixed* interfaces is significant ($p < .01, F = 9.35$), while no significance was found for the difference between either of these and the *action-touch moving results*. The results are displayed in Figure 5.11.

A 1-way ANOVA analysis of the task completion time between the first and second experiment revealed that the difference between the two is significant ($p < .01, F = 119.15$). The same analysis performed to compare the results produced by the *action-marker* interface between the two experiments showed that the difference is also significant ($p < .01, F = 106.17$).

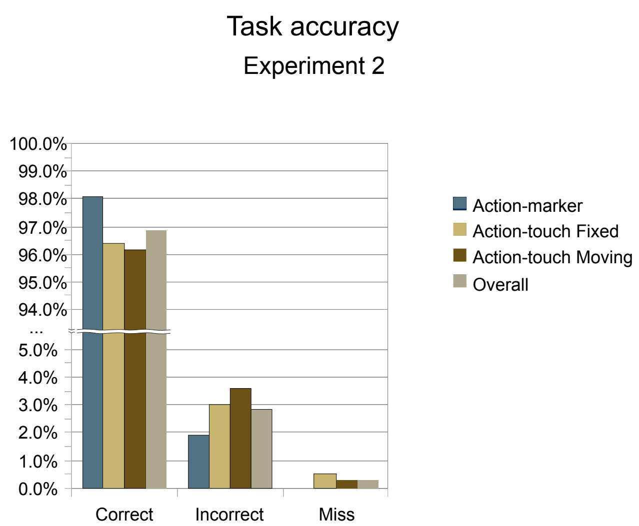


Figure 5.10: Task completion accuracy, results from Experiment 2.

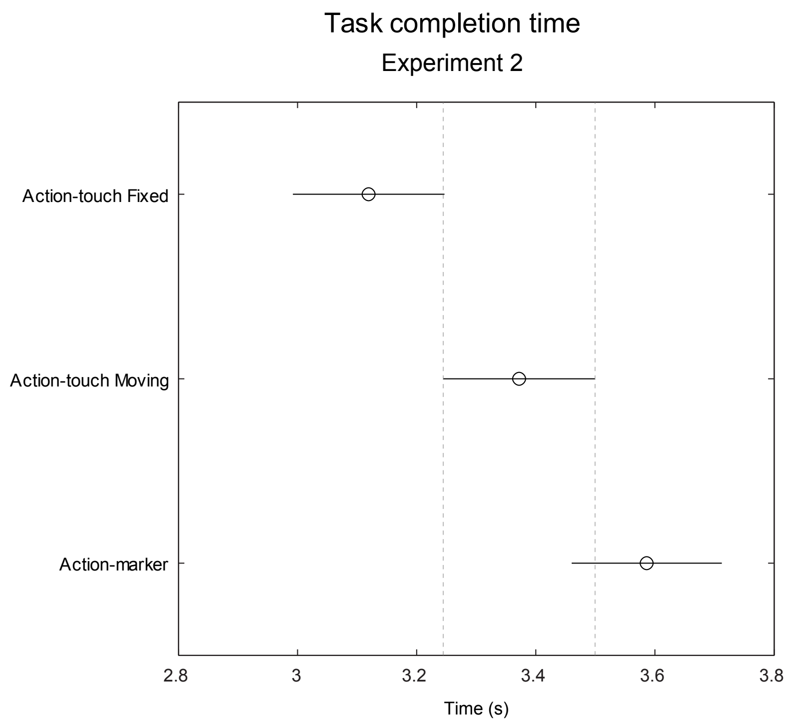


Figure 5.11: Task completion time, results from Experiment 2. The circles indicate the averages, while the bars indicate the 95% ANOVA confidence intervals.

5.4.3.2 Interface Preferences

The subjective ratings are reported through the histograms in Figure 5.12, expressed as percentages of preferences. In summary, the majority of participants found the action-marker interface preferable in all categories. The action-touch moving interface scored the lowest for all categories except one: it was found to be more enjoyable than the action-touch fixed interface.

The reasons expressed by participants to prefer one interface over the other were coded and grouped into broad categories. The results were quite varied. The *action-marker* interface was often preferred because it did not require to touch the screen (11 users) and because users felt that it was more accurate (8 users). Other reasons to prefer the *action-marker* interface included its novelty (5 users) and the fact that compared to the other two interfaces it does not require synchronisation between scanning a marker and touching the desired key (5 users). Users who expressed a preference for the *action-touch fixed* interface found it more accurate (5 users) or faster because it required less movement than the alternatives (4 users). Finally, reasons to prefer the *action-touch moving* interface included the challenge involved to operate it (2 users, in relation to the interface involvement), the fact that it required only one scan and its potential for context-sensitive options (*"It could be that specific buttons come with different cards and that would be interesting"*).

5.4.4 Discussion

The minimum correct selection rate for the 3 interfaces was 96% suggesting that all of them can be successfully used for the tasks defined. The *action-marker* mapping was, once again, the one producing most accurate selections. No significant difference was found for task completion time between the *action-touch moving* mapping strategy and the *action-touch fixed*, and the correct selection rate for the two is very close. These results suggest that *action-touch moving* could provide a valuable alternative, especially when multiple markers may be within the viewfinder at the same time.

The difference in task completion time between the first and second experiments could be attributed to different factors, and it is difficult to draw conclusions regarding this observation. Even though the d-touch frame processing speed was made comparable between the two experiment, the second experiment used a different version of Python for S60, which may have been slower than the one used in the first experiment. The different form factor of the device might have also influenced how easy it is to grasp it; the different time of the year – June for the first experiment, after the end of the semester, December, during the last part of the semester for the second experiment – perhaps had an influence on how tired participants were.

Compared to the preferences expressed after the first experiment, this time the preference for the action-card mapping over the alternative was more marked. This may be because touch-screen buttons were found less comfortable than standard keys, perhaps also because most participants reported not to be very familiar with touch-screen devices. However, it should be emphasised that because different subjects took part in the two experiments, and they were performed at different times of the year, a direct comparison is not appropriate.

5.5 Conclusion

The material presented in this chapter provides a partial answer to the third research question formulated in the introduction (*In which new ways can designable visual markers be integrated in*

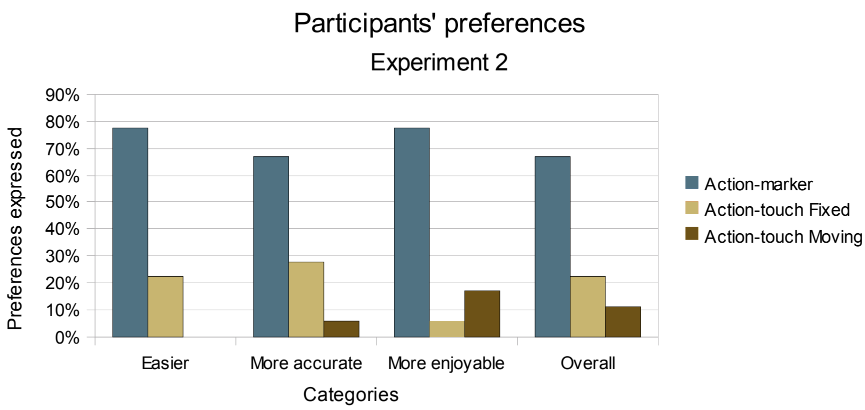


Figure 5.12: Preferences expressed by participants of Experiment 2.

mobile interfaces? How do different marker-based interface designs affect the users performance and preference?). The design variations presented show different ways in which designable markers can be integrated in mobile interfaces, illustrating the use of both iconic and text-based d-touch markers. The results of the usability experiments show that all marker-based mapping strategies were usable: in the worse cases the correct selection rate was 94.9% (experiment 1, action-position mapping) and the task completion time was 3.6s (st. dev. 1.42s; experiment 2, action-marker mapping). The experimental results provide a characterization of the different interfaces, also in terms of user preferences. More interface design examples based on designable visual markers will be presented in Chaper 6 to further answer the third research question.

6. Marker-based Mobile Applications

This chapter presents mobile applications based on the d-touch marker recognition system. Their design is presented here to provide concrete examples of how d-touch markers can be used within mobile interfaces. At a general level, the application of visual markers to mobile interfaces is to create links between digital information and physical items or location. Such links can be cast to augment existing physical objects (or locations) with digital information, or, alternatively, to create user interfaces where physical tokens are used to represent and control digital information, in the fashion of tangible user interfaces (Costanza et al. 2003b). In the first case, the augmentation of existing physical items, digital information may provide information not normally available about the physical world, such as the origin of commercial goods, the history of buildings or narratives set in a specific physical location. In the second case, using physical objects to operate a mobile devices, or more in general a computer interface, may make the interaction easier or more playful to use for specific groups of users (e.g. elderly, as proposed by Häikiö et al. (2007)) or in specific circumstances (e.g. group interaction). Markers can support applications where digital content refers to specific *locations* –in such cases the markers, and their content, would be geographically stationary and refer to places or buildings– specific *objects*, or even *parts of objects* – such as different pages of a book.

Marker-based mobile applications can be classified based how the markers and the digital content are created and associated with physical items. The markers can be placed by the *application designers*, by the *end-users* or they could be *already in place* as part of an existing infrastructure. Similarly, content can be associated with markers by the application designers or by the end-users, following the very popular paradigm of user-generated content. For example, barcodes are placed on most commercial goods by manufacturers and merchants, a mobile application may be developed by those who manufacture the goodsto deliver extra information about their own products to potential consumers. Different applications may be developed by third parties to associate with the barcodes information generated by consumers, such as reviews (Brush et al. 2005), or by independent organisations, such as human rights organisations providing information about the conditions of the workers producing goods (Patten 2005). Because the placement of markers happens in the physical world, it is limited by the constraints of physical access, including ownership.

A further distinction can be made between applications where digital content associated with markers is *static*, i.e the information does not change over time, and those where the content is *dynamic*. Finally another division can be drawn between *real-time* applications, where digital information is displayed on the markers as they appear in the viewfinder, and *near real-time* applications, those where the information is presented only after a marker is selected or scanned. The categorisation is summarised in Table 6.1

What do markers and content refer to?	
Existing objects or locations.	Especially designed tokens.

What degree of granularity do marker references have?		
Places or buildings: location-based.	Objects: object-based.	Parts of objects: sub-object-based.

Who places the markers?		
The application designers.	The end-users.	Markers already in place.

Who associates digital content with the markers?	
The application designers.	The end-users.

Does the digital content change over time?	
No: static content.	Yes: dynamic content.

When is information presented?	
While markers are scanned: real-time.	Immediately after markers are scanned: near-real-time.

Table 6.1: Categorisation of marker-based mobile applications.

The following sections describe each specific application. The first one, *Mobile Audio Cards*, was designed by me and implemented with the support of master students at the EPFL Media & Design Lab. The rest of the applications were designed by others, with whom I collaborated to various degrees. These are: *d-touch video player*, which I developed in collaboration with an interaction design student at the Central Saint Martins College of Art & Design in London; *FoodTracer* which was later designed and implemented by the same student using the d-touch library; and, finally, 3 application concepts generated by participants of the "Near Futures Workshop" a 5-day workshop about service design that I taught, in collaboration with 4 colleagues, in the summer of 2008 at CMU/Portugal in Madeira.

The first 3 applications were implemented through prototypes running on actual phones. These were written in Python, using the native C++ dtrecognition module described in Section 3.4. They run on S60 3rd edition phones – that is keypad-based smart phones (such as the Nokia N95, N73 and E51).

6.1 Mobile Audio Cards

Mobile Audio Cards (MAC) allows users to associate audio clips with printed cards containing d-touch markers. Its design was inspired by the work of Shankar (2004), who showed how digital tools for audio editing can help children learn editing skills transferable to writing. Shankar developed a desktop software application that lets children record and edit audio using a graphical user interface operated with the mouse. Similarly, the aim of MAC is to let children record and edit stories. Rather than using a PC graphical user interface, the system runs on camera phones and the interface is based on a set of cards with figurative d-touch markers printed on them. For example, as shown in Figure 6.2 left, markers visually represent a house, a girl, a book, a ship, and so on. Supporting children story telling through visual cards is at the basis of several commercially available toys, such as "+ e -" (plus and minus) by Bruno Munari[1] and "Il gioco delle favole" ("the fable game") by Enzo Mari[2].

Two versions of the applications were developed, with slightly different emphases. Each version underwent informal qualitative evaluation in the form of field trials. In terms of the classification outlined above, in MAC the cards with markers printed on them act as *tokens* – manipulating them users can interact with digital content. The content is *generated by users* and it can *dynamically* change during the interaction. In the first version content is displayed *after* markers are scanned (*near real-time*), whereas in the second version the digital content is reproduced *while* the markers are scanned (*real-time*).

6.1.1 First Version

6.1.1.1 Design

In the first version, referred to as *MAC1*, adhering strictly to the prior work by Shankar (2004), the focus was on verbal audio: snippets of speech would be associated through the phone with the cards and assembled in sequences to form a story. For example, ["in the house" – house-shaped marker] + ["the girl" – girl-shaped marker] + ["reads" – book-shaped marker] (this example was actually created by one of the children in the field trial reported below). The application allows children to record sounds "on the cards", to playback the cards audio content and to append the cards audio content to an audio sequence stored on the phone. This sequence does not have a physical representation. The application also allows to reset the sequence, i.e., to make it empty.

The application supports 6 actions: "record", "play", "stop", "add-to-sequence", "play-sequence" and "erase-sequence", all of them are represented through physical cards which have d-touch markers on them. These cards are then named *action-cards*, while the ones that can be associated with audio clips are referred to as *object-cards*[3]. The d-touch markers used for the action-cards are based on text, while the object-cards are iconic. The set of all markers used with MAC1 is shown in Figure 6.1.

To execute an action users need to point the camera phone to the corresponding action-card and then to the desired object-card (except in the cases of the "erase-sequence" and "stop" actions which do not refer to specific object-cards). Recording continues until users scan the "stop" card. The same card can also be used to interrupt the playback which otherwise runs until the end of the audio clip. Icons are displayed on the phone screen to provide visual feedback about the action being performed. The interface does not require pressing any key on the phone, all operations are just based on scanning markers. This design choice was informed by the results of Parikh et al. (2006) who report that novice mobile phone users found it difficult to press the camera phone keys while aiming at a target – it was assumed that young children may also find it difficult.

The emphasis of MAC1 is on *sequencing* verbal sounds, so the application does not support layering or mixing, i.e., reproducing different audio clips at the same time. Therefore it allows markers to be scanned only one at the time.

[1]http://www.corraini.com/scheda_libro.php?id=337&lang=eng
[2]http://www.corraini.com/scheda_libro.php?id=182&lang=eng
[3]See also the discussion of alternative mapping strategies in Chapter 5

6.1.1.2 Implementation

The implementation was based on the one described in Section 5.1.1, with the addition of an **AudioPlayer** class. This class wraps the PyS60 built-in audio module, it handles audio clips and it offers a simple API for recording and playing back sounds.

6.1.1.3 Informal Evaluation

A brief field trial was conducted in Switzerland with a primary school class of 19 pupils (7-8 years old), who interacted with the system for about 30 minutes in groups of 2 or 3. Each group was given a phone running the application, the 6 action cards and the 8 object-cards shown in Figure 6.1, and was asked to use the system to create stories combining the cards. The field trial was arranged with the collaboration of one of the class teachers and of the school principal, through the support of the Département de la Formation et de la Jeunesse of Canton Vaud. Permission was asked to the parents of all children who signed a customary consent form.

The positive outcome was that children were very comfortable with handling the phones and scanning the cards to associate sounds to the cards and play them back. The cards, even though they were available in very limited number and their design was extemporaneous, were successful in supporting the storytelling activity, and children were enthusiastic about associating sounds with them and generally about experiencing a new technology.

In contrast, the interface for creating sequences was unsuccessful: children often mixed up recording and sequencing, or, more precisely, how these two actions were supported by the MAC1 application. It was often observed that children started recording and then pointed the phone sequentially to different cards, without stopping the recording, expecting that the phone would automatically segment the audio and associate different chunks to different cards.

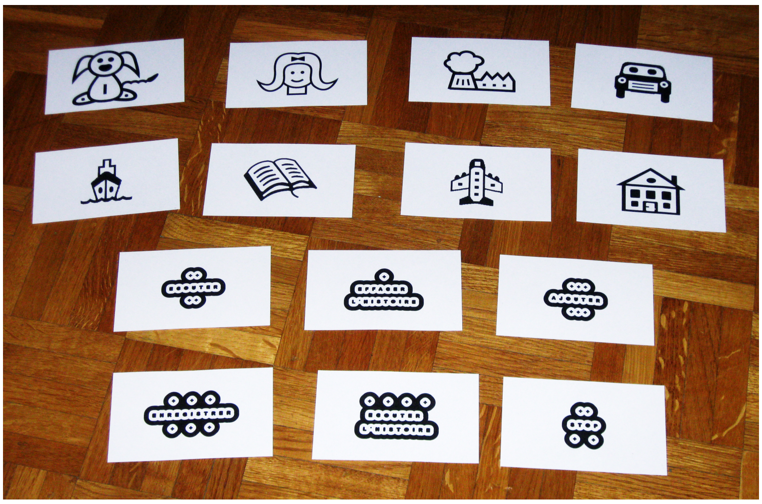


Figure 6.1: D-touch markers printed on cards, used with in the first version of the Mobile Audio Cards application. Iconic markers (top) represent interface objects, while word-based markers (bottom) represent interface actions.

6.1.2 Second Version

6.1.2.1 Design

In the second version of Mobile Audio Cards, referred to as MAC2, the focus shifted to using the mobile phone for recording and playing back non-verbal sounds which children would use to augment the stories they tell with their voice. This change in focus was partially driven by the outcome of the field observation of MAC1 and partially because of a collaboration with colleagues[4] interested in the use of voice and technology to create, or "sketch", non verbal sounds. As detailed in the Acknowledgement section, this collaboration was supported by an EU COST grant and took place at the Holon Institute of Technology, in Israel. The basic setup for MAC2 was the same as for the first version: the application allows to associate sounds with cards containing d-touch markers and reproduce them. MAC2 was designed to let children record one ambient sound or sound effect on each card, and then *layer* and *sequence* the sounds by physically laying out the cards and scanning them with the phone. Arranging multiple cards together children should be able to create soundscapes: groups of ambient sounds that enrich one part or "scene" of the story (some soundscapes may include only one sound). The set of all soundscapes creates the soundtrack of the story.

Like in MAC1 the cards that can be associated with sounds have figurative design, so that they can visually support the storytelling activity. Children should start by creating a story based on the cards' visual content. Once the story is defined, they should record sounds and associate them with the cards, to create a soundtrack for it, and finally enact the story augmented with the soundtrack.

In an attempt to make the interaction more immediate and playful, and overcome the problems observed with MAC1, a new mapping strategy was explored. Because the sequencing of sounds

[4] Michal Rinott, lecturer at Holon Institute of Technology, and Tal Drori, independent interaction and graphic designer

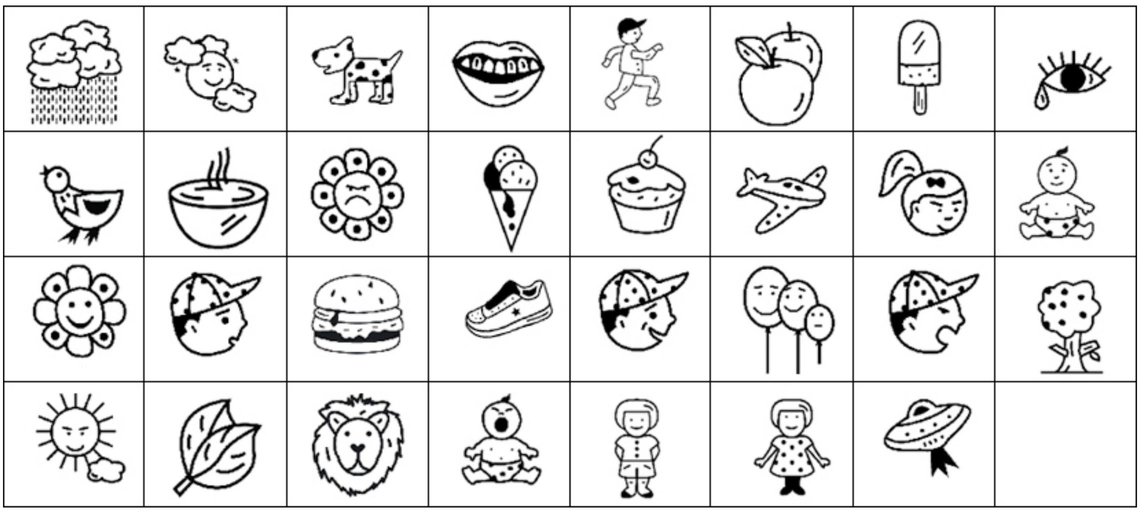


Figure 6.2: D-touch markers used in the second version of the Mobile Audio Cards application.

in MAC2 is controlled in real time during the story-telling performance (rather than by adding or removing items to a sequence), the number of actions that the interface needs to support is reduced to simply "record" and "play". Play is expected to be performed much more frequently than recording, so the simple scanning of a marker was assigned to this action (one can think of it as an implicit play action). It must be noted that playback is a "safe" action: even if it is involuntarily triggered, it does not cause changes in the recorded content. The phone vibrating motor is activated every time a marker is recognised, this haptic feedback is designed to make users more aware of this event and prepare them for the playback of the related sound. To facilitate the creation of continuous soundscapes, the playback of sounds was looped as long as the phone is kept in front of the markers.

Conversely, record needs more precise control, both for starting and stopping, so it was mapped it to the "select" key, the central key of the phone directional keypad. To record a sound "on a marker" users need to point the phone to the marker and press the central key, or equivalently press the central key and point the phone to the marker. Recording will take place as long as the key is pressed, with the hope that this will make it very precise yet easy to control. Similarly to MAC1, visual icons displayed on the phone screen provide additional feedback for the playback and recording operations.

Thanks to the involvement of a graphic designer, in this second phase of the project, it was possible to design 31 cards; the entire set is illustrated in Figure 6.2. Subjects for the cards were selected as a combination of items considered obvious to sonify (e.g. dog, lion), items considered more challenging to sonify (e.g flower, sun), and actions (e.g., laughing and running) to see how children would deal with the different types. Some of the cards subjects were based on words extracted from a children story book. The visual language was iconic and simple. The d-touch analyser application was used to facilitate the design of the markers complying to the d-touch design constraints; the entire set was further analysed manually to ensure that the 31 marker IDs were clearly distinct from each other, to minimise the chances of misclassification. The markers were printed using a standard ink-jet printer, and glued to square cards cut from tick cardboard.

6.1.2.2 Implementation

The MAC1 prototype was modified to support the new design. The main change was to support the concurrent recognition of multiple markers and the audio mixing of the corresponding audio clips. While concurrent recognition of multiple markers was already supported by the d-touch library, MAC1 was optimised based on the assumption that only one marker would be scanned at the time. The new mapping strategy described above (based on the select key) was implemented by simply creating a new subclass of the EventManager abstract class.

6.1.2.3 Informal Evaluation

Three informal trials were arranged, this time through personal connections of the researchers collaborating on the project; they all took place in Israel. A 7-years old boy, a 5-years old girl and a pair of 6-years old children (a boy and a girl) tried the MAC2 application. The trials took places in their own homes and lasted approximately between 20 and 40 minutes. Again the consent of the parents was asked, this time informally, before involving the children.

Similarly to the first trial, all children were excited about trying a new “technological toy”. The new cards supported well the storytelling activity and, again, children were enthusiastic about the possibility of recording their own voice and associate it, through the phone, with the cards. However, the interface proved to be sometimes problematic. While 2 of the 4 children were very comfortable with using the system, the other 2 had difficulties in coordinating the press of the record button with pointing the phone to the marker and producing sounds. These difficulties were in one case overcome with some extra guidance. Two of the children who managed to associate sounds with the cards used the interface to augment their stories, recording both verbal and non-verbal sounds.

6.2 D-touch Video Player

D-touch video player is a mobile application to play video clips by scanning d-touch markers. It was designed to augment “Zooming out from the Desktop”[5] , a booklet about computer interfaces, produced as part of the coursework of Giuseppe Costanza for the MA in Communication Design, Digital Media at the Central Saint Martins College of Art & Design, London. By scanning with a mobile phone the d-touch markers found on the pages of the book, readers can play video content that complements the printed text and images. Similarly to uWiki, when the phone is pointed to a marker, this is highlighted in the viewfinder and a text label is displayed to inform users that they can press the left soft-key to play the video associated with it, as shown in Figure 6.3 Left. The project was conceived as a demonstrator, so the video content was stored locally on the phone’s memory.

[5] the booklet is fully available from http://www.giuseppecostanza.it/research_hci_intro.htm

Figure 6.3: On the left, *d-touch video player*in use. On the right, the set of 12 markers designed by Giuseppe Costanza to augment his “Zooming out from the Desktop” booklet.

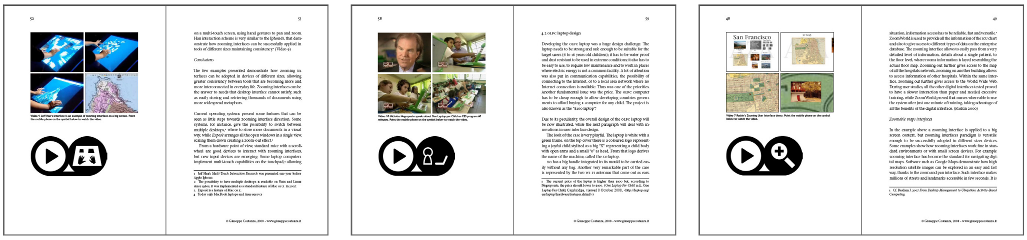


Figure 6.4: Pages from the "Zooming out from the Desktop" booklet. D-touch markers are consistently laid out on the left page of double spreads, under groups of pictures.

The d-touch video player displays content that was selected by the *application designer*, who also designed and placed the markers. The digital content is displayed after a marker is scanned (*near real-time*), and it does not change over time (the content is *static*). Each piece of video is related to a specific book page, a specific part of an object.

The graphic design of the markers was performed by Giuseppe Costanza with the support of the d-touch analyser application (Chapter 3). The design of the markers and their layout in the book provide an interesting example of how the d-touch technology can be integrated in printed material. A total of 12 d-touch markers, shown in Figure 6.3 Right, were included in the book, their uniform visual design facilitates their identification: once users have seen one marker, any other marker will be easily recognised. Indeed in the booklet the first marker appears in the preface, accompanied by the following explanation: *"Every time on a page appears a rounded black symbol similar to the one at the end of this paragraph, just point the mobile phone camera on it and the screen will display the related content."*

The universal "play" symbol on the left of each marker provides a visual cue about its interactive function: to play a video – markers, then, are visually used as part of the interface. The icon on the right of each marker creates a visual link with the specific content that the marker refers to (as can be noticed in Figure 6.4). Throughout the booklet markers were consistently placed in the same relative position: on the left page of a double spread, under a group of pictures and with blank space under it, as illustrated by the examples in Figure 6.4. To further facilitate the human recognition of the markers, the following sentence was included in all the captions of the images above the markers: *"Point the mobile phone on the symbol below to watch the video."*

6.3 Foodtracer

FoodTracer was designed and implemented using d-touch by Giuseppe Costanza as part of the coursework for the MA in Communication Design, Digital Media at the Central Saint Martins College of Art & Design, London. In the words of its author[6] :

> *FoodTracer is a mobile phone application that visualises information about the origin, production and distribution of food, including the amount of natural resources*

[6] http://vimeo.com/5373197

Figure 6.5: The FoodTracer application in use. Digital information about the two products framed in the viewfinder is displayed in the form of the two circular icons (red on the left and green on the right).

> *consumed in the process. FoodTracer is designed to be used in supermarkets, to enable its users to make better informed decisions while they shop. The application empowers consumers by providing them with information in context and allowing them to compare different products. On a wide scale consumer choices can determine changes in food industry practices and trades.*

D-touch markers are integrated, or rather *hidden*, in the labels of 12 fictitious food products. The left hand of Figure 6.6 shows how 3 wine labels were modified to embed markers that fulfil the d-touch constraints. Generally just one portion of the label is the actual marker (e.g., in the leftmost label of Figure 6.6 only the dark oval containing '2009'). The right hand of the same figure shows the rest of the labels designed for the application. It is interesting to note that, in contrast with the previous example, in this case the visual design of the markers does not communicate at all their interactive function. The system is used almost as a substitute for a natural features visual recognition system.

The interface of FoodTracer follows the augmented reality paradigm: virtual information about the products is overlaid on the products themselves, in the viewfinder. A GUI menu on the phone allows users to choose what type of information to display about the products. As illustrated in Figure 6.5, the characteristics of products can be compared by physically placing them next to each other and framing both of them in the viewfinder. An interesting feature of the interface is a *split-screen* mode: in order to compare products that are not physically next to each other,

Figure 6.6: Fictitious food product labels designed by Giuseppe Costanza for the FoodTracer application. On the left an illustration of how wine labels (originally designed by Maurizio Schifano) were modified to include d-touch markers.

one portion of the viewfinder can be temporarily "frozen", while keeping the rest of it live. Even though from a functional point of view this is simply a way to store the digital information, this interface metaphor is coherent with the physical action of "placing two product next to each other" to compare them. As it can be noticed in Figure 6.5, because in each product a different portion of the label acts as a marker, the relative position of the digital information with respect to the package is not necessarily consistent.

A working prototype of FoodTracer was implemented. All the product information was stored locally on the phone memory, even though in principle the application could retrieve data from a server. Of course, one challenge that was beyond the scope of this project is to gather appropriate and accurate information for the different products.

In FoodTracer the digital content is related to specific *objects*, and it is selected by the application designer. Information is displayed in *real-time*, while markers are scanned. For the project demonstrator markers were designed and placed by the application designers, who also selected the digital content. However, in an hypothetical deployment version of the system may relay on markers placed by the product manufacturers. The information associated with each marker is static.

6.4 Near Futures Workshop

The *Near Futures Workshop* was a hands-on, 5-day workshop about interaction and service design that I taught, in collaboration with 4 colleagues, in summer of 2008 at CMU/Portugal, in Madeira[7]. Ten participants from local industry and academia attended the workshop, their backgrounds ranged from engineering to graphic design. During the course of the week participants were guided through a number of design exercises aimed at creating new mobile services or applications. In one of them participants were asked to *"design a service or experience to enrich tourism in Madeira which leverages the d-touch visual markers technology."* This exercise built on some field observations and concept development that participants had performed earlier that week.

Participants worked in 3 groups. The designs were developed only up to a conceptual level and presented through scenarios. Even though the results are at a rather early stage, they are briefly reported here because they provide additional examples of how d-touch markers can be blended, both physically and visually, into the environment. In all 3 cases, the digital content and the markers are selected and placed by the application designers. The content is *location-based*, it is static and it is displayed after markers are scanned (*near real-time*). The following subsections report the projects presented by each group.

6.4.1 Educational Treasure Hunt

The first group, composed of Filipa Jervis (Universidade da Madeira), José Luis Malaquias (ISA - Intelligent Sensing Anywhere) and Rita Tavares Katzenstain (Portugal Telecom, Sistemas de Informação), proposed an interactive and educational treasure-hunting experience for children of different age ranges. The system runs on a device that can be borrowed from the tourist information kiosk. D-touch visual markers are visually integrated in the environment, almost hidden, so that finding them is part of the game. The game is based on a tale related to the history of Madeira, when a marker is scanned the system presents a piece of the story which contains a hint for the localisation of the next marker. The information is presented in audio format, to facilitate the interaction of young children who may not yet be comfortable with reading, but it is also summarised

[7] The workshop was co-organised by Valentina Nisi (CMU/Portugal in Madeira), Arianna Bassoli (at the time at the London School of Economics), Johanna Brewer (at the time at University of California Irvine) and Tal Drori (independent interaction and graphic designer). More information is available on http://design.epfl.ch/nf

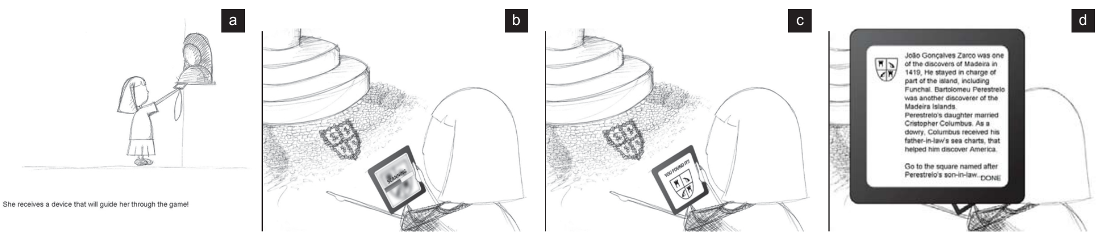


Figure 6.7: Illustrations by the first group in the "Near Futures" workshop of the *Educational Treasure Hunt* project.

Figure 6.8: Illustrations by the second group in the "Near Futures" workshop of the *Culway* project.

in text form, so that parents can help. At the end of the experience participants receive a small present, a souvenir form the island. Selected slides from this group's presentation are reported in Figure 6.7. The highlights of this example are the marker embedded in the mosaic pavements, which is a typical feature of the streets of Madeira, and the delivery of personalised information according to the participants age.

6.4.2 Culway

The second group was composed of Ana Isabel (ALERT Life Sciences Computing, S.A.), Catarina Pereira (Universidade da Madeira), Patricia Fernandes (Expedita) and Rui Henriques (Empresa de Electricidade da Madeira). The design proposed by this group, named "Culway," is an interactive guide of the city of Funchal, which provides navigation directions as well as historical and cultural information related to points of interest in the city. Two series of d-touch visual markers support the interaction: directional markers and localised markers. Directional markers are integrated in the typography of street name plates, as illustrated in Figure 6.8 a and b, so that they can be read by people as well as by the d-touch recognition system. When users scan one of these markers with their mobile device, they receive information about the points of interest closest to their current location, directions for how to reach them as well as a teaser to invite them there, such as an open question. Localised markers, shown in Figure 6.8 c and d, are displayed in front of buildings and places of interest, such as museums and churches. When these markers are scanned information about the location is presented, answering the teaser presented earlier, and inviting the visitors to discover by, for example, entering the museum. Other localised markers inside the buildings provide information about specific items.

6.4.3 TagLand

The third group included Adriana Pereira (ALERT Life Sciences Computing, S.A.), Emanuel Fernandes (Universidade da Madeira) and Maria Batuca Toureiro (Portugal Telecom, Sistemas de Informação). Their design integrates a mobile experience with a web-community: members of the community can send challenges to others inviting them to discover specific locations in a holiday destination. The challenge starts when receivers are still in their home countries, the initial phase of the experience is delivered through a website, where photos of the holiday destination are shown. As represented in Figure 6.9 a, these photos may actually portray markers (in this case a marker

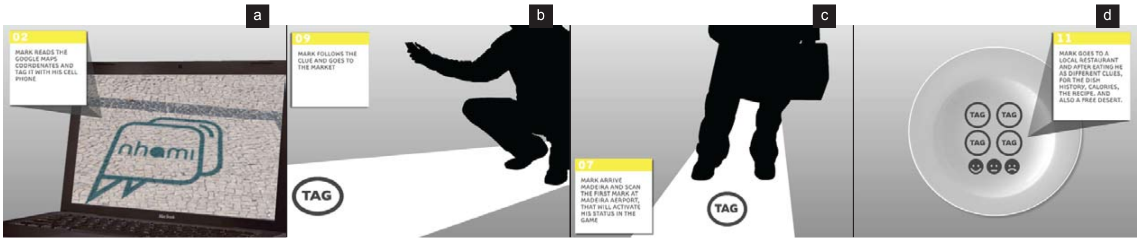


Figure 6.9: Illustrations by the third group in the "Near Futures" workshop of the *Tagland* project.

embedded in the mosaic pavement, as for the first group). Once the first challenge is solved and participants arrive to their holiday destination more challenges are presented, guiding them in the discovery of the local culture and habits. These challenges on location are experienced by scanning markers, which are hidden in the environment, for example on the airport floor (Figure 6.9 b, c), or on restaurants' plates (Figure 6.9 d) Chains of challenges are linked by thematic areas, selected on the basis of the participants profiles as well as on the choice of those who invite them.

6.5 Discussion

In general the example designs show how designable markers can be integrated in mobile applications. Table 6.2, a modified version of Table 6.1, summarises how the project described in this chapter fall into the different categories outlined at the beginning of this chapter.

All projects demonstrate how the designable quality of the d-touch markers can be used to make them better fit in the context where they are placed (FoodTracer, Near Futures), or to convey visual cues related to the interaction (MAC, d-touch video player). Most of the example applications can be easily supported by a limited number of unique markers, with the exception of FoodTracer, which in the case of a real deployment would require hundreds, if not thousands of unique identifiers, clearly posing a challenge for the d-touch system.

Generally all applications take advantage of the real-time processing, in that they use the interactive viewfinder to provide rich visual feedback to users about the markers being recognised. However in most cases the actual digital content is displayed only after a marker is scanned. Exceptions to this are FoodTracer, where part of the information is displayed through the viewfinder, and the second version of Mobile Audio Cards, where audio clips are reproduced while markers are in view in the finder.

For some applications d-touch marker recognition could be replaced by GPS or RFID tags. For example, for the projects created in the Near Futures workshop GPS could be an appropriate alternative, given that most interaction is around specific geographic positions. Mobile Audio Cards and FoodTracer could be re-implemented using RFID tags and an RFID reader embedded in a phone. In contrast, the D-touch Video Player is an example where the use of RFID tags may be unfeasible, because RFID readers would tend to scan multiple pages of the book at the same time, given that radio can propagate through paper. It should also be noted that for some of the applications presented in this chapter the markers are scanned from a distance: in FoodTracer this form of interaction allows to compare different physical items by framing both of them in the viewfinder; in CULWAY street name plates are scanned without having to reach them. Replacing visual markers with RFID tags would require a re-defintion of such interactions.

6.6 Conclusion

The material presented in this chapter complements the content of Chapter 5 in answering the third research question formulated in the introduction: the mobile application designs reported above illustrate different ways in which designable visual markers can be integrated in mobile interfaces. The visual design of markers can take different roles: in some instances the visual aspect of the markers conveys cues about their interactive role (action-cards in the first version of *mobile audio cards*, Section 6.1, and *d-touch video player*, Section 6.2); in other instances iconic markers are blended into the physical world, still recognizable as interactive elements (the object-cards of *mobile audio cards*, Section 6.1, and the semi-hidden markers in the *educational treasure hunt* and *TagLand*, Section 6.4); finally markers can also be made *disappear*, hiding their design into existing graphic elements (the product packages of *FoodTracer*, Section 6.3, and the street name plates in *Culway*, Section 6.4). The examples were also varied with respect to the functional role of markers: from being complementary to existing printed information – in the case of the augmented book, the product packages and the street name plates – to acting as central or primary elements in the interactive experience – as in the case of the cards and of the markers to be found in the treasure hunt.

What do markers and content refer to?	
Existing objects or locations.	Especially designed tokens.
D-touch Video Player; FoodTracer; Near Futures Workshop.	Mobile Audio Cards.

What degree of granularity do marker references have?		
Places or buildings: location-based.	Objects: object-based.	Parts of objects: sub-object-based.
Near Futures Workshop.	Mobile Audio Cards; FoodTracer.	D-touch Video Player.

Who places the markers?		
The application designer.	The user.	Markers already in place.
Mobile Audio Cards; D-touch Video Player; Near Futures Workshop.		FoodTracer.

Who associates digital content with the markers?	
The application designer	The user.
D-touch Video Player; FoodTracer; Near Futures Workshop.	Mobile Audio Cards.

Does the digital content change over time?	
No: static content.	Yes: dynamic content.
D-touch Video Player; FoodTracer; Near Futures Workshop.	Mobile Audio Cards.

When is information presented?	
While markers are scanned: real-time.	Immediately after markers are scanned: near-real-time.
Mobile Audio Cards (late version); FoodTracer (partially).	Mobile Audio Cards (early version); D-touch Video Player; FoodTracer (partially); Near Futures Workshop.

Table 6.2: Categorisation of marker-based mobile applications with examples from this chapter.

7. Markers in the Real World: Two Field Studies of Physical Tagging

In the context of user interfaces for mobile applications and services, visual Marker recognition can be used to implement "physical tagging" systems – systems where digital information is associated with physical items. Physical tagging systems are a specific case of location-based systems; they can be implemented through a variety of technologies, besides visual markers: from RFID to 2D barcodes to even entering numerical codes found on signs into mobile devices. Examples of existing practices of physical tagging include museum audio guides, that allow visitors to listen to pre-recorded commentary related an exhibits, as well as advertising campaigns in East Asia (Fowler 2005), and increasingly in the rest of the World, where consumers can scan 2D-barcodes embedded in street advertising posters with their camera-phones to retrieve information about the promoted product, or discount vouchers. However, these examples are rather far from current Internet practice: they are closed systems, where content is provided by authoritative actors – the museums or the advertising companies – and it is not part of a more general information ecology.

This chapter reports two field trials where users created digital content and associated it with objects and locations using uWiki, a research prototype that allows physical tagging. UWiki offers both mobile phone and web interfaces and it is based on d-touch marker recognition. Both trials focussed on the creation, sharing and distribution of content by users – in some cases including also the visual design of the markers. Given the role of user generated content and social sharing in current Internet usage practice, addressing them in the design and study of systems for the distribution of digital information in physical space is fundamental. The aim of the trials and their analysis, then, is to investigate what kind of everyday practices can emerge around the use of physical tagging systems.

The trials were organised to stimulate diverse conditions of use, and they allowed me to address a variety of questions including: How would the behaviours which emerge through the use of the system related to its physicality? How are they affected by the social role which the technology plays? And which behaviours arise from the ways in which the uWiki system is embedded in a larger technological landscape?

7.1 uWiki

7.1.1 Design

UWiki is a system that allows users to associate digital content with d-touch visual markers. It includes a server component, a web interface, and a mobile interface, in the form of a stand-alone mobile application running on S60 camera-phones. The digital content is organised as a collection of single media items (text, images, audio and video clips) stored on the server, and URLs pointing to user-selected web pages external to the system. Each item is labelled by a title and must be associated with one marker. Any number of items can be associated with a marker, arranged in a list or in threads (similarly to a web forum). A single log-in is used to identify users from both mobile and web interfaces.

The system was designed as a research tool, to allow the observation and analysis of users interaction with location-based services, in particular marker-based media creation. The system was designed to be as simple and neutral as possible, in order to provide enough flexibility to support a variety of application scenarios: from collaborative note-taking to authoring and consumption of interactive mobile tour guides and narratives. Users can take field notes through the mobile application and later retrieve and expand them using the web interface from a computer. Authors of tour guides or other interactive mobile experiences can edit content using a standard web interface, while the audience can access the same content from the mobile devices.

7.1.1.1 Mobile Application

The mobile application was designed to include minimal amount of features aiming at being simple to use. The flow of the interaction always starts by scanning a marker, the only action available in the initial screen of the application, beside changing the user settings, as illustrated in Figure 7.1 a. Marker scanning takes place through an interactive real-time viewfinder: as soon as the phone's camera is pointed to a marker, it is highlighted by a green rectangle and an overlaid text label informing the user that the marker can be selected by pressing one of the phone's keys (Figure 7.1 b). Once a marker is selected the application retrieves from the server the list of items associated with it and displays it together with the option to create a new item and associate it with the marker (Figure 7.1 c). This list view displays the media type, title and author of each item. Selecting one of the items from the list brings up a detail view, where the full content is displayed. The list of items associated with a marker may be empty, in such cases only the "create new item" is available. The "create new item" screen, shown in Figure 7.1 d, prompts users to select one media type and opens the corresponding media creation screen, which let them compose a piece of text, capture a photo or video using the phone's camera or capture audio using the phone's microphone. In order to save the new item users are requested to provide a title for it. The new item is then uploaded in background to the server, while the application returns to the list of items associated with the current marker. To provide users feedback about the fact that the items they just created are being saved to the server, uploading items are immediately added to the list and marked with an asterisk (as illustrated by the 2 items at the bottom of the list in Figure 7.1 c).

Handling media types separately, rather than as a hypertext, was a design choice, made to simplify the interaction, especially the editing, from standard keypad phones. One of the items for each marker can be optionally designated as *autoplay*, so that when the marker is selected the detail view of such item is directly activated, rather than the list. Users interaction with the application is recorded on logs which are automatically uploaded to the uWiki server, to allow remote observation of the system usage.

7.1.1.2 Web Interface

The web interface is designed to *augment* (as defined by Milic-Frayling et al. (2007)) and complement the mobile application. Through a standard web browser users can log-in to the uWiki service from a computer and access a list of all markers they scanned using the mobile application. Selecting one marker brings up a separate page showing the list of items associated with it, illustrated in Figure 7.2. From this page users can edit or remove items that they created, add new items by uploading media files or by adding text through a web form. Content from uWiki can also be easily exported to any other computer application through copy and paste or by saving the page, as from any standard website.

7.1.2 Implementation

From the technical point of view, uWiki is based on open source software: the server is written in PHP as an extension of the TikiWiki[1] system, it runs on an Ubuntu Linux server and uses MySQL for data storage. TikiWiki is a content management system including wiki and file storage functionality (among many others) which were leveraged to store uWiki items: text was stored as wiki pages and media items as files. The TikiWiki database was extended to store information about the d-touch markers, and their association with items. The web interface of the system was modified to allow the display and manipulation of items and markers, as described above. The server was also augmented with an additional HTTP interface for communication with the mobile client. The mobile client was developed from scratch following a standard object-oriented approach.

[1]http://tikiwiki.org

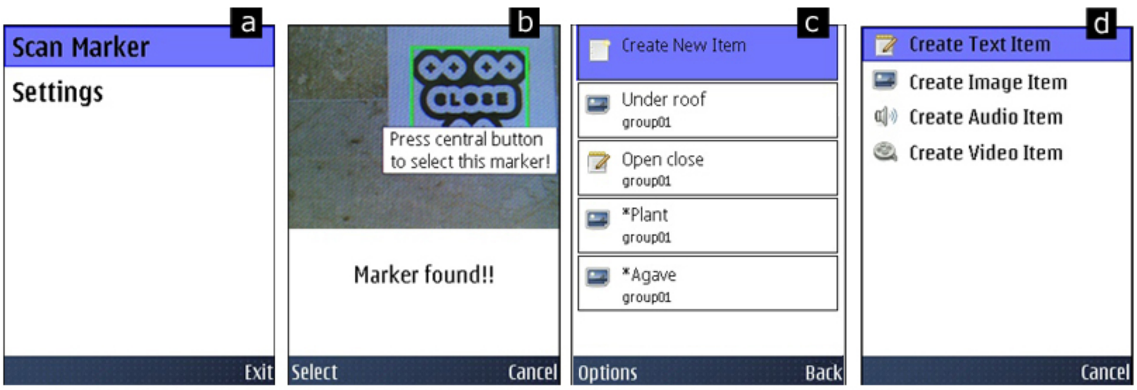


Figure 7.1: Screenshots of the uWiki mobile application: (a) initial screen; (b) marker scanning; (c) list of items associated to marker; (d) the "create new item" screen.

7.2 Method

The work reported in this chapter follows a qualitative research approach. The field trials were documented through photographs of the markers as participants affixed them, as well as through the multimedia content that participant produced: photograph, text, video and audio. Participants were observed and sometimes shadowed while they interacted with the system. Semi-structured interviews were carried out after the trials, and documented through audio recordings and notes. All this material was coded and analysed during and after the collection to extract from it categories and concepts, as it is common in qualitative data analysis (Miles and Huberman 1994, Corbin and Strauss 2008). The method was in particular inspired by *grounded theory*, as de-

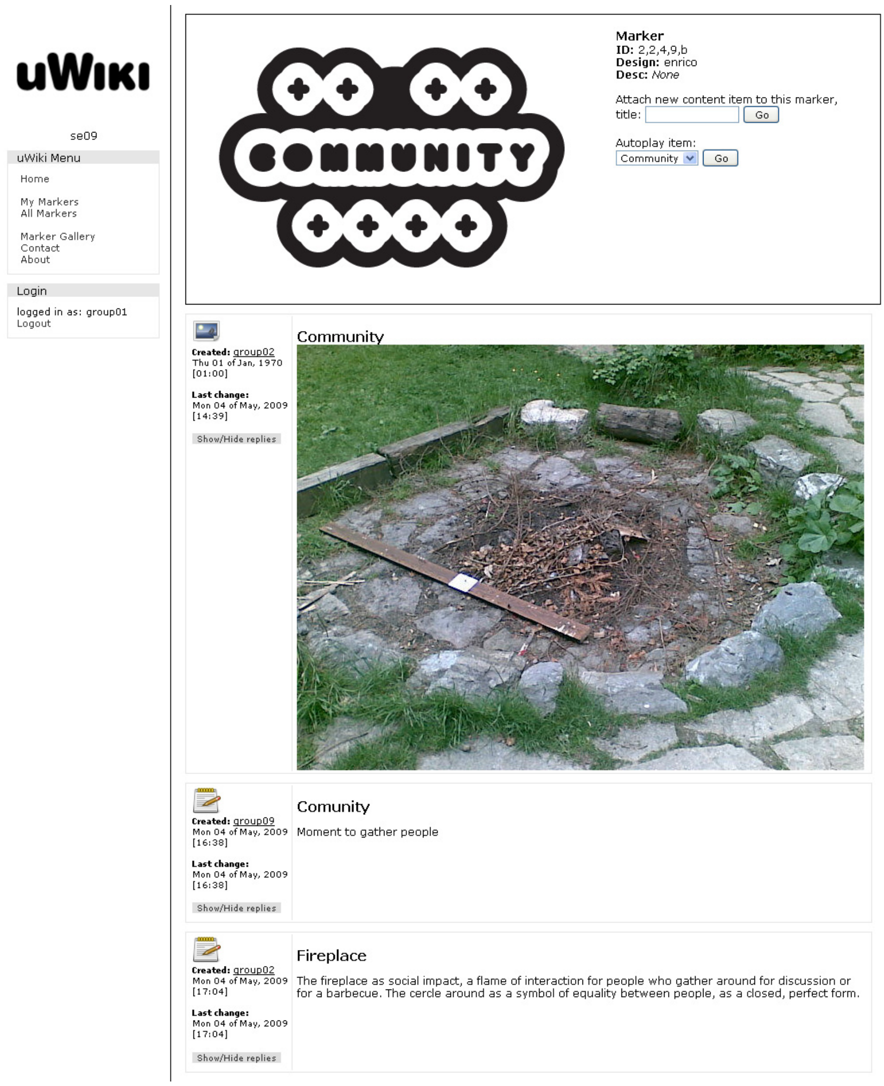


Figure 7.2: Screenshot of the uWiki web interface showing the list of items associated to a marker. The marker is pictured on top of the page, next to it, on the top right, a simple interface to associate new items with it, and a drop-down menu to set one of the items to autoplay.

scribed by Preece et al. (2007) and Corbin and Strauss (2008). However, given the exploratory nature and the limited extension of the observation the analysis did not attempt to fully develop a theory, but it simply aimed at extracting descriptive categories and exposing relationships between them.

The qualitative approach was chosen over a quantitative one for multiple reasons. First and foremost, the nature of this research is *exploratory*: observational studies of user generated content for location based systems, and especially for physical tagging systems, are at this point so rare that it was not found possible, just yet, to formulate research hypotheses that could be tested experimentally. This lack of testable hypotheses, and therefore of comparable conditions, makes it impossible to approach the problem with quantitative methods. Moreover, the very nature of the phenomenon under observation, user-generated content, requires interpretation. Any quantitative analysis attempt would require the definition of interpretive categories that could be used to count occurrences of specific patterns. Thirdly, qualitative methods were found to be particularly suitable for working with a small number of participants, and over a limited period of time.

The field trials reported in this chapter were designed to provide users a clear context and motivation of use: the focus was on *what kind of practices emerge* from the use of such systems, given these defined conditions, rather than trying to understand *why* users may adopt physical tagging systems. As discussed in Chapter 2, other researchers attempted to observe how users generate content for location-based systems, setting up completely open-ended trials, but the main outcome reported was always that users created little content compared to the researchers' expectations. Location-based systems are, at a general level, communication systems, therefore they suffer from the *network effect*: their value is low until they become available to a large amount of users, which may be unfeasible in research setting.

7.3 Architecture Class Trial

The first trial (the "Class trial") took place in May 2009, in the first day of a week-long intensive elective course about innovative residential housing offered for academic credit to 2nd year architecture students at EPFL. As part of the class activity students had to perform a *site analysis* of the "Îlot 13", an urban residential area in the city of Geneva, Switzerland. A site analysis is a common initial stage of an architectural process, often a critical one conditioning the success of a project bid. It is normally performed by architecture students and professionals using paper notes and photographs.

The Îlot 13 site was of interest to the class because of the strong sense of community of its inhabitants which led to a process of participatory design to achieve "interesting architectural solutions and affordable rents, as well as low-cost renewable energy production" while "preserving the existing social and economic structure" (Gisselbaek et al. 2006). The area is contained in a city block of approximately 440 meters of perimeter and 9500 square metres of area. It includes several building blocks situated around a system of 4 inner courtyards. One of the buildings hosts a student residence for one of the local universities. On the ground level many of the buildings house private business and communal spaces, ranging from a second hand bookstore, to craft workshops, to private offices and shared rehearsal rooms. Several of these have large windows which visually connect them with the outdoor areas. Some of the walls in the inner courtyard are covered with

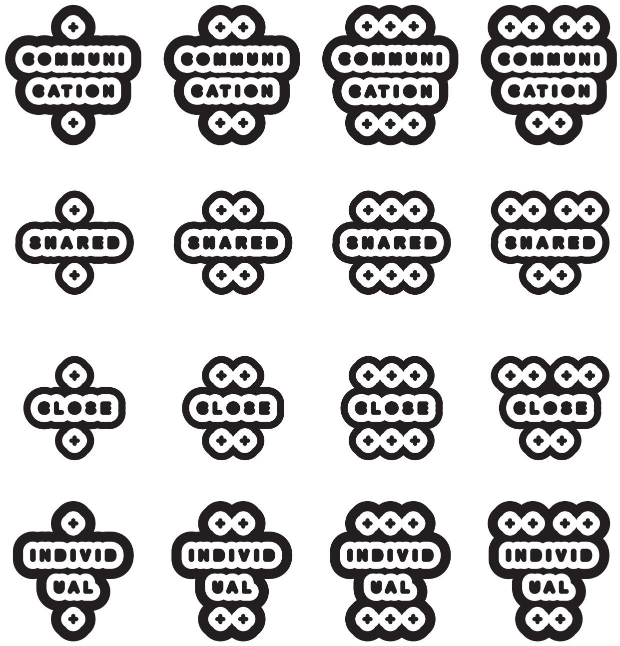


Figure 7.3: Example markers from the set that was provided to participants in the Class trial. Here 4 words are augmented in 4 different ways – the full set included 10 words augmented in 12 ways.

graffiti. One peculiar space that attracted the attention of some of the trial's participants was a communal space on the 1st floor of a building near the main entrance – even though raised this terrace is freely accessible, but it is contiguous to several bedrooms of the student residence.

A total of 19 students took part in the trial, divided in 8 groups. To avoid crowding the site, half of the students visited it for 1.5 hours and the other half went after them for the same amount of time. Each group was given a phone running the uWiki application, a deck of 10 labels (or more as requested) with markers on them and some semi-permanent adhesive ("Blu Tack") which they could use to affix the markers. Participants were informed about the site and why it was considered interesting for the class and instructed to analyse it with the support of the uWiki system. The system functionality was briefly demonstrated to them and they were all asked to test the system before starting the task. The participants were told that the markers were going to be removed at the end of the week, but they were expected to be available throughout the duration of the class. They were all familiar with wikis but none of them had used system similar to uWiki before the trial.

The markers were printed on water-resistant vinyl scrim cut into pieces of approximately 10-by-10 cm. These dimensions were chosen as a trade-off between having markers that are large enough to be easy to notice, and small enough to be comfortable handle and affix. A set of 120 markers were produced based on 10 words: public, private, open, close, shared, communication, community, retrofit, functional, individual. The terms were selected based on literature about the site (Gisselbaek et al. 2006) and on the technical constraints of the d-touch system (in particular having words that would result in different configurations of strokes, i.e. regions). The markers were generated semi-automatically by manipulating the typography of each word and adding graphic elements (see Figure 7.3 for some examples). Each word was augmented in 12 ways – the set of markers was shuffled so that each group was given 10 markers having all the different words and a random mixture of graphic patterns. This was done to avoid identifying groups by the visual patterns of the markers, in the attempt to reduce possible feelings of ownership which might have hindered one group to add content to the markers of another group. Because a limited number of marker-words were provided to students, it was emphasised to them that they could also choose not to take the markers literally, in other words not to adhere to the words printed on the markers but just use them as placeholders for the digital content.

7.3.1 Data Collection: User Generated Content, Logs and Interviews

The class trial was documented through the content that users created on the uWiki server, through interaction logs automatically recorded by the mobile application, and also through photographs of the markers as they were affixed by participants. Class reports submitted by the students for grading at the end of the week were also examined. Based on initial analysis of this material, one individual and one group semi-structured interviews were carried out with participants after the completion of the trial. Audio from the interviews was recorded, and later selectively transcribed and coded. The interviews centred around the ways in which they used the uWiki system to complete the task at hand. Participants were asked to expound on the ways in which they choose to place the markers, what kind of content they decided to attach, and how they used (or intended to use) this content after the trial was completed, and how they interacted with the other groups in the trial. The breakdown and categorisation of the results presented in this section is not meant to be assessed statistically but to be indicative of a variety and general frequency of behaviours observed with this kind of technology. This data is further explored through the interviews and pursuant qualitative analysis.

	markers	items	from phone	from web	image	text	video	audio
Group 1	10	25	16	9	16	9	0	0
Group 2	8	14	8	6	8	6	0	0
Group 3	16	27	11	16	19	5	3	0
Group 4	9	10	9	1	9	1	0	0
Group 5	3	3	3	0	3	0	0	0
Group 6	5	5	5	0	5	0	0	0
Group 7	7	13	8	5	8	5	0	0
Group 8	2	3	2	1	2	1	0	0

Table 7.1: Overview of content generated by groups in the Class trial.

A total of 100 digital items were created and attached to 57 markers, an average of 1.75 (1.02) items per marker, min. 1, max. 6, median 1. Of the total, 61 items were created on the field using the mobile client, while 39 were added later through the web interface. Different groups contributed in different measures, as reported in Table 7.1. Almost all of the field generated content is in the form of photographs (58), with the exception of 3 videos. Of the content added from the web interface, 27 items (69%) are in the form of text while 12 (31%) are in the form of photographs. The titles of the 70 images were coded as follows: for 42 (60%) the title is descriptive of the content (e.g. "Open covered space"), for 18 (26%) it is the same text as the marker, for 7 (10%) it is a comment (e.g. "It's mine, don't touch it"), while for 3 (4%) it is generic (e.g. "photo"). In 50 of the photographs (71% of all photos) the marker was portrayed in the context where it was affixed. The summary of this classification is summarised in Figure 7.4.

All the digital content associated with the markers was created by the same groups who affixed them, with the exception of 3 cases, where one group associated digital items with markers affixed by someone else. In at least 2 cases markers were placed nearby ones which were already there; in all these cases the word on the markers were different (e.g. 'individual' and 'community'). Five students reports were produced in groups by the end of the week, summarising their analysis of the Îlot 13 site a sketching an architectural intervention inspired to the analysis. In 4 of the 5 reports images from the uWiki system (between 2 and 8) were used.

Marker Placement & Referring

The position of markers was coded through their photographs along 4 dimensions. The results are summarised in Figure 7.5. The first dimension is the type of element each marker was attached

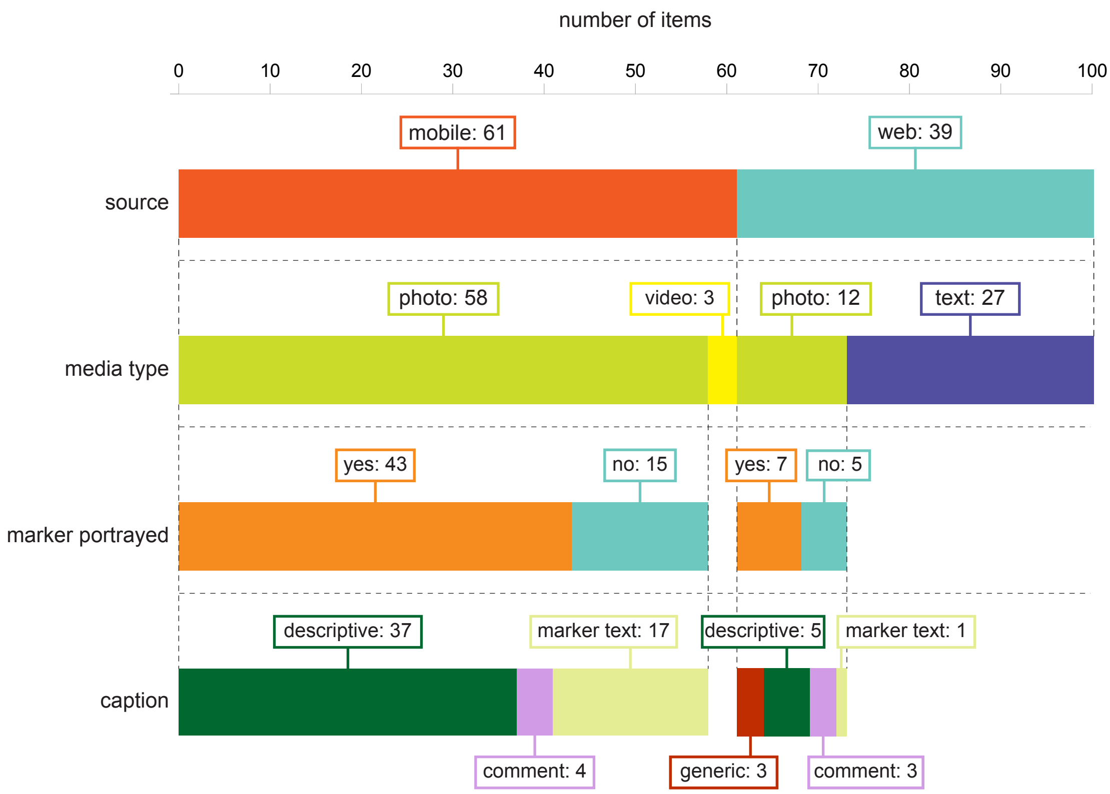


Figure 7.4: Overview of the items that participants created during the Class trial.

to. The second dimension is the stability of the marker positions: their placement was static (e.g. firmly attached to a stationary item such as a wall) in 43 cases (75%) while it was mobile or unstable (e.g. attached to a moving object such as a vehicle, or loosely attached to the leaf of a plant as in Figure 7.11 left) in 14 cases (25%). The third dimension was the level of obtrusiveness of markers: 10 markers (18%) were attached where other content was present (e.g. bulletin boards or walls with graffiti) which generally gave legitimacy to their placement, 16 (28%) were placed in unobtrusive positions (e.g. on an empty wall), 24 (42%) were placed in intrusive ways (for example on windows blocking their opening), while 7 (12%) covered other information posted in space (e.g. a marker covering opening hours displayed on a door). The final coding dimension was whether each marker was attached by itself or in a group, in the first trial all markers were placed by themselves.

In the first trial all of the markers were used in a *deictic* way: to refer the digital content associated with them to something (an object or a space) on the site where the trial was conducted. The markers and the content were therefore jointly coded according to what they referred to; a few times disambiguation was derived from the participants interviews. For example, in one instance a marker was attached to a table in a public area, a photo and a text note associated with the marker referred to the space around the table – in this case the marker was coded as casting a reference to the space immediately around the marked object. References were made with various degrees of proximity: in 29 cases (51%) the marker was placed directly on the referred item; in 9 cases (16%) the reference was to the space immediately around the marked object; in 8 (14%) cases to something near the marker; in 7 (12%) cases the reference was to an inaccessible space behind a marked window or wall; in 3 cases (5%) the reference was to large areas around the marker, such as the entire block. In 1 case the marker was deliberately hidden by the students, inside the motorbike it referred to.

7.3.2 Initial Analysis

The most salient phenomenon emerging from the direct observation of the first trial was the placement of markers in *non-permanent*, *fragile* positions, such as on mobile objects (e.g. on a bicycle or moped) or in situations where they would easily fall. It was also prominent to see markers in *obtrusive* positions, being *in the way* of inhabitants of the space, for example on a letter box or on a window, from where they would likely be removed. The placement of these markers appeared spontaneous, almost "careless", and the result of a reflexive action, placed more for one's own sake rather than thinking of others to find them and scan them at some later moment. Often markers (permanent or non-permanent) were placed directly on the objects they referred to, in few occasions even if the object was not easy to reach - for example on a street name plaque, which is at about 3 metres from the ground.

Even though the participants of the class trial undoubtedly understood the interactive features of the system, as all of them scanned at least 2 markers and associated content with them, in several instances the marker placement did not afford them to be easily scanned by others to retrieve the content. Interestingly, these practices are in clear contrast with existing practices of physical tagging. For example in museum audio guides, or print advertising augmented with 2D barcodes, the tags are placed *next to* the items they refer to, and they are generally placed in positions that are permanent and easy to reach by users. It is important to notice that such unexpected practices were not "universal" within the trial. A good number of markers were affixed in stable and unobtrusive positions. Even though the specific conditions under which the trial was run (semi-permanent adhesive, content creators were also the only content consumers) unsurprisingly had an influence on participants behaviour, it is interesting that this influence did not lead to a single style of physical tagging.

Participants generally associated photographs with the markers through the mobile application and later added textual comments using the web interface. This pattern confirms the expectations of using the mobile system to take quick notes while in the field and expanding them using the full-featured web interface. This usage pattern may also explain the lack of audio content and the very few videos: even though also quick to create, audio and video are more difficult to elaborate

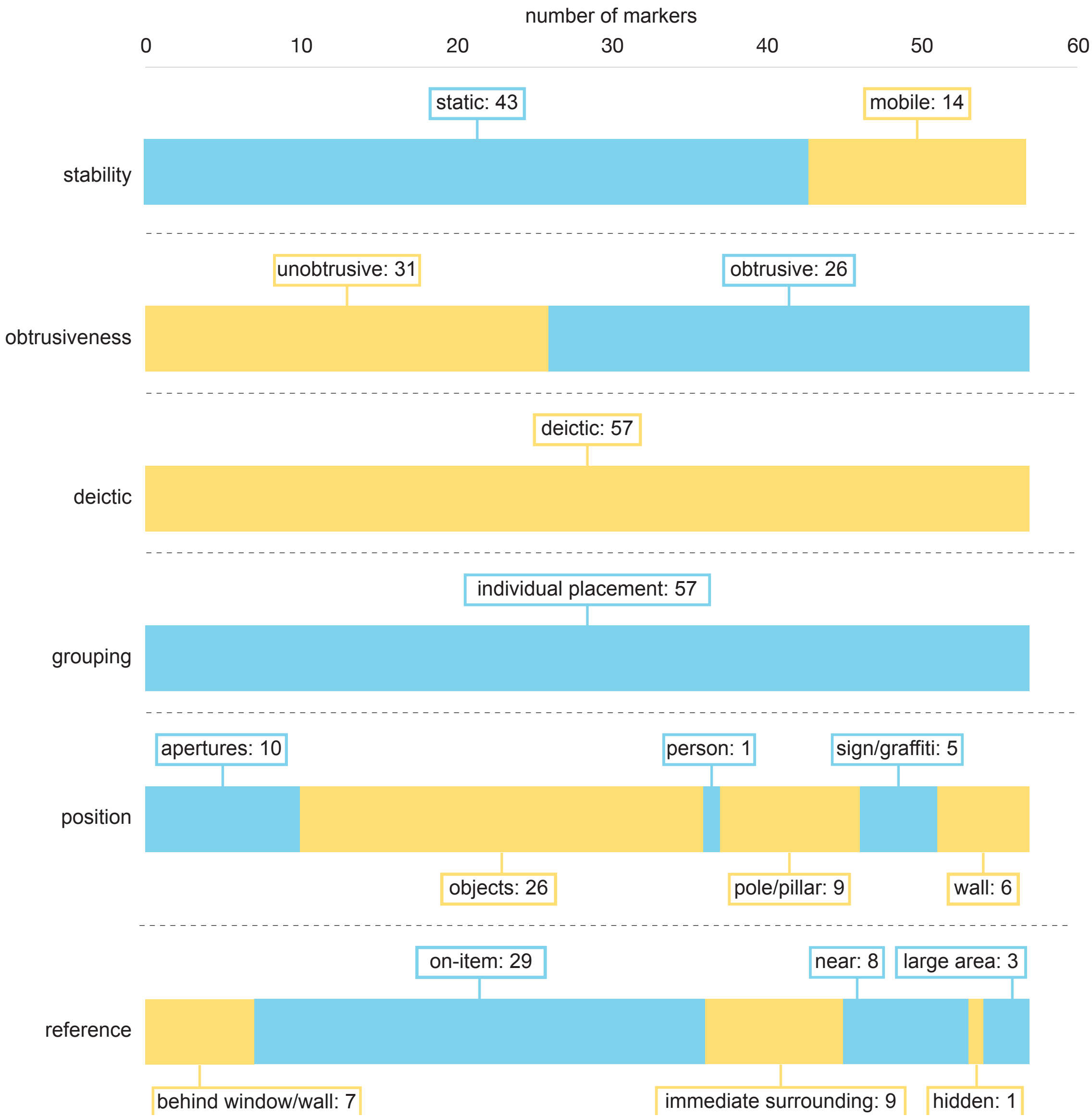


Figure 7.5: Overview of the markers that participants placed during the Class trial.

and comment than photographs. This initial analysis informed the semi-structured interviews and the design of the second trial. The data and the interviews from the first trial were further analysed after the second trial, as reported below.

7.4 Mobile Media Contest Trial

The second trial (the "Contest trial") took place between June and July 2009. It was designed on the basis of the initial analysis of the class trial, to stimulate diverse conditions of use. While in the first trial the same group had the role of content producers as well as consumers, the second trial aimed at defining a clear audience that participants were asked to address. In contrast to the semi-permanent placement of markers in the first trial, in the second trial participants were asked to take into account permanent positions for them. Rather than letting participants work with a pre-defined set of markers, they had complete freedom for the quantity and design of the markers. To implement these conditions, at the end of June 2009 a creative contest among EPFL students was launched, advertising it through several EPFL mailing lists (see Appendix H for the promotional email used). Participants were asked to create uWiki content that would entertain or engage new students or visitors to explore the campus. The content was suggested to be a story telling experience, but other options were left open. Participants were told that, after the contest, the organisers would arrange with the university administration a permanent setup: phones running the uWiki application would be made available for visitors to borrow from an existing information desk (to counteract that the system currently runs only on a limited number of phones), similarly to how one can borrow audio guides in museums. Markers were initially attached with the same semi-permanent adhesive of the previous trial, and participants were told that later they would be attached permanently. Therefore, participants were asked to place markers where they considered it plausible to have them permanently affixed.

To motivate submissions a prize of CHF 300 was offered to the winner and CHF 150 to the runner-up. Entries were judged by a panel of mobile media and media art experts. 3 teams, pairs and individuals, completed the contest. They will be referred to as Team A, B and C. The 3 teams were made up of 5 participants, all of them are students of the university who were familiar with computers and mobile phones, but not frequent users of mobile internet.

In a number of initial individual meetings participants were shown the uWiki system, including a number of markers, printed on different sizes, and discussed doubts and ideas about the competition. It was emphasised that the entries would be judged mostly on the digital content and its relation to the physical location, rather than the visual aspect of the markers; participants were asked to at least *define* what kind of markers they wanted and they were offered to have simple markers designed by the organisers according to their specification. The option of directly designing the markers was left open, saying that it would have been regarded as a bonus—participants were provided with documentation about marker design and the *d-touch analyser* application (see Chapter 3). All markers were printed by the organisers on water-resistant vinyl scrim, in a size defined by the participants. A text note in French was added to all markers using a small font, asking not to remove the marker as it was part of a scientific experiment; the name of the researcher was also stated, suggesting to contact him for any question.

7.4.1 Data Collection: User Generated Content, Logs, Interviews and Shadowing

In addition to the same methods of observation employed in the first trial, in the second trial participants placing markers around campus were shadowed by a researcher who interviewed them in situ as they completed the task. This process was documented through photographs and by audio recording. After the trial interviews were conducted with 2 of the groups. Audio from all interviews was recorded and later selectively transcribed for analysis.

Participants in the second trial placed a total of 41 markers, respectively 17, 14 and 10 per team. All works are defined to a level of being a draft. Because of a technical problem, after placing markers around campus one of the groups could not update their digital content, however, they provided a detailed description of the content while placing markers, so that information is used in this analysis. A total of 73 digital items were created or defined in association with the markers, all of them through the web interface. The same coding strategy employed for the first trial was employed for the second one, results are visualised in Figure 7.6. In this case the photographs created by participants never portrayed the marker they were associated with, and the captions were always descriptive, so these two dimensions are not reported in the visual summary. Given the limited number of teams, and the considerable differences among them their outputs are individually described in the following paragraphs.

The work of Team A, a group of 2 students, can be described as a combination of a tour of the campus and a treasure hunt. It included 17 markers associated with a total of 45 digital items, a combination of 25 text items, 11 photographs, 2 videos and 7 URLs. All markers were designed by the group without help from the organisers, and they generally represented logos, icons and high contrast photographs modified to fulfil the d-touch technical constraints A starting point was defined by 2 markers just outside the information desk at the centre of campus where phones would be borrowed by visitors. The 2 starting markers, one for English and one for French, are associated with an introductory text explaining that the audience should try to solve a riddle. The experience also had clear destination point, also defined by a marker, where the solution to the riddle was. The rest of the markers contain information about campus events, shops, student associations, in

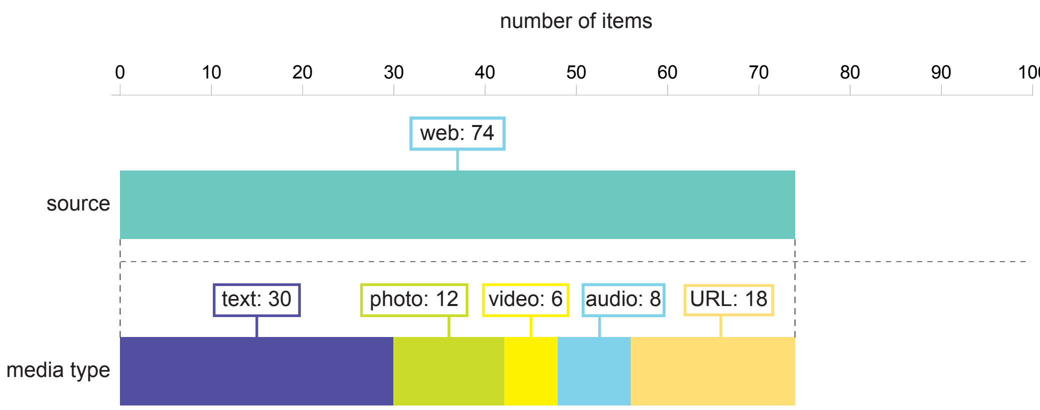


Figure 7.6: Overview of the items that participants created during the Contest trial.

a form that may be useful for incoming students—this information is mixed with clues guiding the visitors towards the destination. All photographs associated with the markers portrayed events that took place where the marker was affixed, except for one photo which portrayed the President of EPFL and was associated to a marker affixed near his office.

Team B, an individual participant, produced a distributed crime story partially set on campus. It included 14 markers related to 18 digital items, 5 text, 8 audio clips, 4 video clips and one image. 12 markers support the core of the story and the team divided these into 3 groups: *beginning*, *middle* and *end*; each group contains 4 markers numbered from 1 to 4—visually these markers are defined by text augmented with graphic elements, added to fulfil the d-touch marker design constraints. Each group is consistent in terms of media type used: *beginning* markers have text attached to them, *middle* have video and *end* have audio, all are associated with just one media

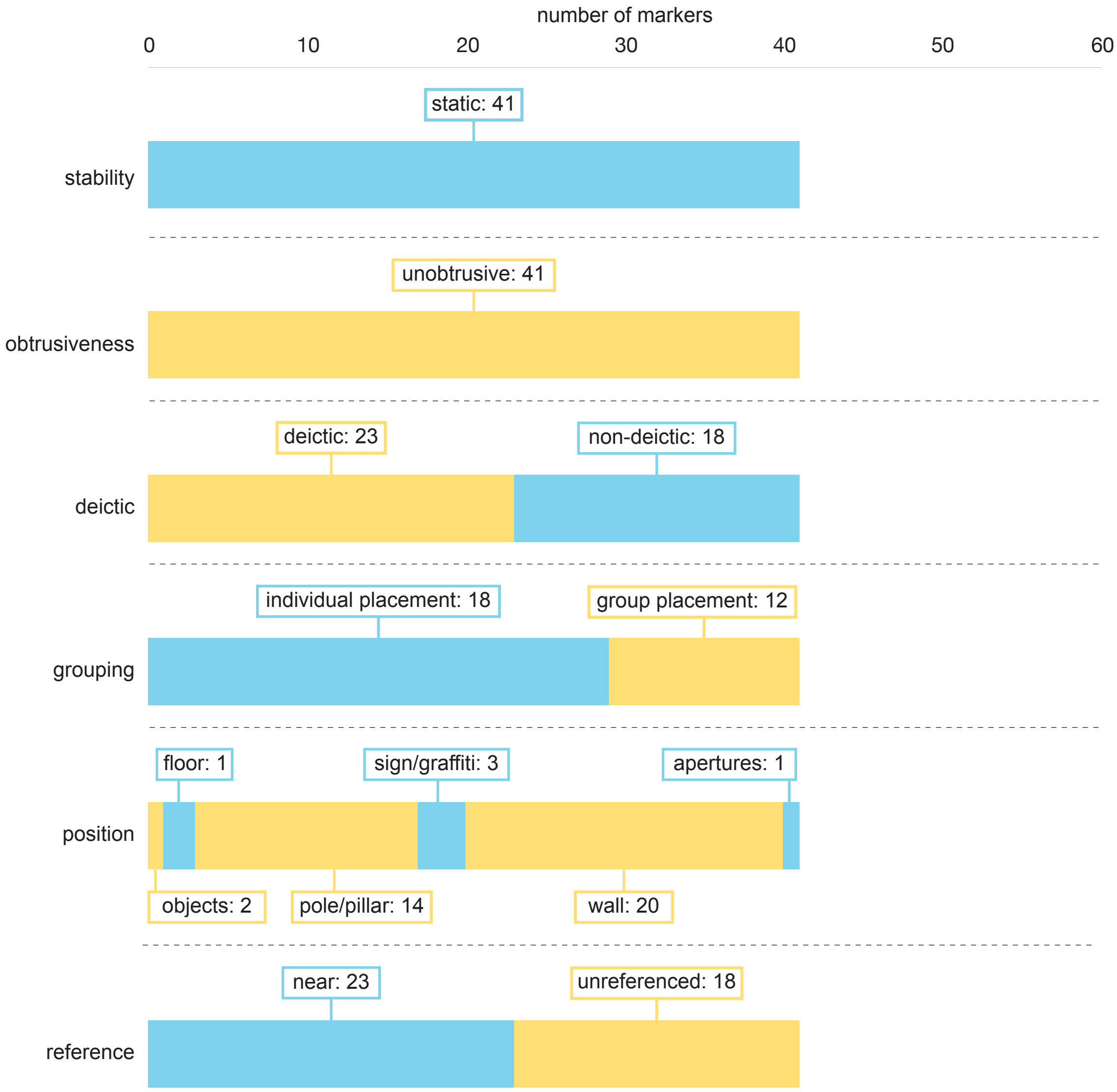


Figure 7.7: Overview of the markers that participants placed during the Class trial.

Figure 7.8: The 'Answering Machine' marker by Team B of the Contest trial. The black magnets can be moved in and out of the marker to change its topological structure and hence its ID.

item. The remaining two markers are related to additional contextual material, they can be distinguished by a different visual design. One is a *dynamic marker*: a marker placed on a metallic surface, with 3 small black magnets left next to it that could be placed by users inside or outside the marker, to modify its structure and, as a consequence, its identifier—in this way users could access different digital items. The marker, shown in Figure 7.8, represented a phone answering machine, modifying the marker with the magnets users can listen to different messages. The last marker represents the front of a postcard, with a black and white photo, attached to it is an image representing the back of the photo with an hand-written message. The markers, including the concept of an interactive one augmented with black magnets, were ideated by the participant and designed with the help of the organisers.

The work of Team C, a group of two, can be described as a collection of practical information related to campus life. It included 11 markers: 4 related to languages that can be learned at the campus language school are all placed outside it; 2 related are placed outside a bar and are related to events organised there; one is in front of a bike repair shop and the remaining 4 are all placed in the campus central square, 2 of them are related to campus restaurants and the others are related to the local meteo forecast and to the metro timetable. The content could not be updated because of the technical issues mentioned above, but it was described while the markers were physically attached around campus.

Marker Placement & Referencing

The position of the markers was coded along the same 4 dimensions used for the first trial. The summary of this coding is visualised in Figure 7.7. All markers in the second trial were placed in static and unobtrusive positions. Regarding grouping, 12 markers were placed in groups of 2 or 4, while the remaining 29 were placed individually. In the second trial markers were not always used deictically: while 23 of them where used to reference objects, buildings, or spaces on campus, for the remaining 18 the content was not related to the specific location where the marker was affixed.

Figure 7.9: Letterboxes at Îlot 13. Photograph taken by Group 1 and associated with a marker displaying the text 'individual', visible in the photo itself: the physical marker is rendered as digital content.

7.4.2 Initial Analysis

In the second trial markers were always placed in permanent and unobtrusive positions. This confirms that the trial design was successful in that it stimulated a different approach to physical tagging. Questions related to the permanency and legitimacy of markers were often raised. Participants gave to the marker visual design more importance than it was originally expected, suggesting that the designable nature of d-touch markers was perceived as a strong feature. The differences between the trials provided a rich and varied approach towards understanding the potential of physical tagging systems as they are used in a variety of situated practices.

7.5 Qualitative Analysis and Discussion

Seven different categories were identified through several analytic coding passes on the data, a procedure common in qualitative data analysis (Miles and Huberman 1994, Corbin and Strauss 2008). The categories are described in the following subsections, each of them illustrated through multiple examples from the data. Relationships between different categories are discussed at the end of this section.

7.5.1 Digital Content and Markers Position: Different Physical Tagging Strategies

The relationship between digital content and position of the associated marker varied. In the Class trial the content associated with markers always referred directly to real-world items in the physical proximity of the markers. The same was not true in the contest trial, where often the digital content was not directly related to the immediate physical context of the markers. The following examples

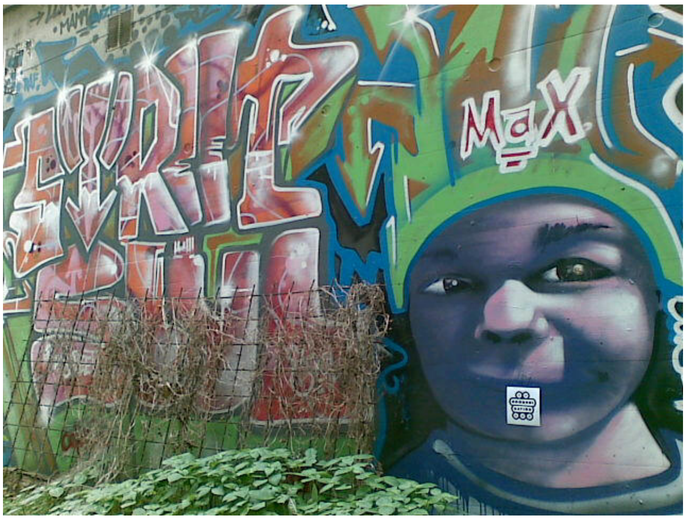


Figure 7.10: A piece of street art on the walls of the inner courtyard of Îlot 13, photographed by Group 1. The photograph was associated with the marker attached on the mouth of the character in the graffiti.

illustrate how the relationship between digital content and physical position of the markers varied even within the same trial and even within the instances produced by the same participants.

Group 1 of the Class trial affixed a marker whose visual representation included the word 'individual' to a letter box which was part of a group of letterboxes hanging in a public walkway of Îlot 13. They used the mobile phone to take a photo of the letterboxes with the marker on them, shown in Figure 7.9, they associated the photo with the marker and added the caption "Postbox". Later, using the web interface, they added a text item with title "comment" saying:
" individual? repetition...
a long row of identical post boxes... each one individual? "
In this case the digital content refers to the object on which the marker is affixed and its surroundings.

In another instance, the same group affixed a marker displaying the word 'communication' over the mouth of the subject of a piece of street art. They took a photo of it, shown in Figure 7.10, and associated with the marker, with the caption "Communication graffiti". They later added the textual comment *"something to say?"* from the web. Seeing the word "communication" over the mouth of a figure on the side of a building is in itself a message, and indeed the participants who placed it told us during the interview that in this case for them *"the marker becomes like a sign"*. Yet, the textual content attached says something additional to the uWiki user. This example shows how a marker allows, then, for a compound form of communication: messages are carried by the locational context in which the marker is placed, the visual representation of the marker itself, and the digital content to which that marker provides access.

Group 1 from the Class trial, again, attached a marker displaying the text 'open' to a plant in the inner courtyard of Îlot 13. They associated with it two photos of the garden, one featuring the marker and one without it, with captions "open garden" and "open garden 2", both shown in Figure 7.11. Group 1 later added from the web a text note with title "comment" and the following text:

Figure 7.11: Details of the courtyards of Îlot 13. Both photographs were taken by Group 1 and associated with a marker displaying the text 'open', visible in the left photograph. The marker is attached in an unstable position, and it is portrayed in one of the photos associated with it.

the garden is a consecution of open, semi-open and rather closed spaces. a very interesting kind of 'parkour' is established by the possibilities of view into the open sky, followed by semi-open spaces, means interaction between a structure (natural or built), and then sheltered places which are probably used by the hole community living in illot 13. for this topic, see also our marker 'closed'.

In this case the digital content refers to the large surrounding of the marker, and not the specific object it was attached to.

In a few instances the position of the markers defined a point of view on a physical object of interest. Team A of the contest trial affixed a marker to a balcony from which there is a view in the distance of the " Learning Centre" building on EPFL campus, as it is visible in Figure 7.12. The text content associated with this marker referred precisely to this distant building, it has title "Learning Centre" and it reads:

Here you can have a little snapshot of how the Learning Centre looks like.
The building's surface is about two football fields. Its role is to host the new book library and to create 500 new working places for the students. There will also be a restaurant and a Bank !
Unfortunately, you won't be able to acquire a watch in there, as Rolex didn't open a shop inside.

Figure 7.12: A marker defining a point of view. The visual aspect of the marker as well as its digital content refer to the 'Rolex Learning Centre', the building visible in the distance. Jean from Team A of the Contest trial tests the marker as he had just affixed.

The placement of markers can then be seen not only as an act of associating digital content to the physical place, but also as a way to influence how users see and experience that place. Moreover, as discussed below, in this case the visual aspect of the marker was also designed to refer to the Learning Centre, as it represented the building iconic plan.

Participants of the Contest trial described the ways in which they physically positioned markers in order to make them easier, or harder, for people to access. Unlike, for instance, the 2-dimensionality of a GPS hotspot, markers are inherently in a fixed position in a 3-dimensional spot. A set of markers attached to a metal pole would all fall on the same location on a standard map, however, the height at which they were affixed would affect the user's experience. Positioning a marker is a nuanced task that goes beyond geo-localisation coordinates.

On the opposite end, most of the content created by Team B of the contest trial does not have a direct relationship with the position of the associated markers. For example, one of the four markers displaying the text "the end" was affixed to a pillar opposite a shop on EPFL campus – as shown in Figure 7.13. Team B later associated with this marker, using the uWiki web interface, an audio clip containing a dialogue which is part of their narrative entry. The dialogue is set at the home of one of the characters, and there is no reference in it about the physical location of the marker. During the post-task interview, when asked about this marker Frank explained: *" it's just that that shop is easy to miss, but it's really nice shop, so I like the idea that someone may go around looking for this marker and then find it and then go 'ah there's a shop here...' "* In this case the digital content is used to attract visitors to a specific location, indeed as Frank further explained: *" all of my content was essentially rewards for finding the markers, and the game was to find the markers, and then when you find the markers you got a piece of content... "*

The different relationships between content and marker locations may sometimes reflect different practices that were adopted for the activity of placing markers and relating digital content to them. As it is visible in the interaction logs, in the vast majority of cases in the class trial, participants affixed markers first, then scanned them and finally created content. As reported in the previous

Figure 7.13: The marker placed by Frank from Team B of the Context trial in front of a shop on EPFL campus. Neither the visual aspect of the marker nor the digital content associated with it are related to its physical location.

section, in the majority of the cases, the photographs associated with the markers portrayed the markers themselves in context, as can be seen in Figure 7.9. This action, then, effectively serves to render the physical marker and its specific position in space as content itself. Yet this strategy was not *always* used in the class trial, as Marie from Group 6 of the class trial described during the interview: *"we were taking a picture of the marker... then a photo.. and then we were placing the marker.."* Here *'Taking a picture of the marker'* refers to scanning it, while holding the marker in the hands, as it was revealed by the interaction logs, while the *'photo'* refers to one that was associated with the marker and stored on the server. In this case, then, the digital content is produced before the marker is affixed: in fact, all of the photos produced by group 6 do not portray the marker.

In the Class trial digital content was created and associated with the marker *in situ* by taking a picture with the camera phone and adding a short caption for it. This mobile-generated content was often commented, explained or expanded later with additional digital content, generally text, added through the uWiki web interface. In this context the practice of physical tagging can be seen as a way of *recording* and *gathering* information which is meaningful to the creator. This interpretation is supported by the observation that that often the content created included a record of how the marker was attached in context. Much like web-based tagging, the act of placing a marker in this way creates, in itself, new information. Contrastingly, in the Contest trial participants placed and scanned their markers as they distributed them around the campus and only later on they attached audio files and other content to the markers through the uWiki web interface. This style of adding content served to position the marker as a link or pointer to another piece of content, rather than to refer back directly to its moment of placement. Attaching content to markers for the participants in the Contest trial, then, can be seen as an act of content *sharing* and *dissemination*. These two extremes demonstrate the ways in which physical tagging can be used to both *gather* and *distribute* content. It is the subtle differences, though, that allow us to see that there is no *one* way of physical tagging, rather, a variety of distinct styles, which span a continuum, can emerge through, and are rooted in, the use of such a technology.

Figure 7.14: On the left a marker unstably attached by participants of the Class trial to a bicycle at Îlot 13. On the right the marker on the floor, a few minutes after it was placed.

7.5.2 Physicality and Fragility of Markers

The examples reported in the previous subsection show that markers generally communicate both digitally and physically. When observing the participants of our study, it became clear that physical nature of markers played a strong role in the practice of physical tagging, influencing also its digital aspects. The physicality of the markers, and the fact that they are placed in the real world, allows them to be manipulated beyond the control of their creators. This is particularly evident for those markers that were coded as 'unstable' or 'fragile', such as in the example of the 'open' marker affixed by Group 1 to a plant, described above (Figure 7.11 Left). In these cases markers were placed in *unstable* ways, from which they could easily fall – and indeed in some cases they did, as illustrated in Figure 7.14 – thus rupturing the link between digital and physical. Likewise, other people cohabiting the space where the markers are placed can move or remove them. In fact, when we returned to the site of the Class trial two days after the trial, most of the markers which the students placed had been removed[2]. The virtual world of content in the uWiki system, is also, then, very much *of* the physical world, as the pathways to that content can be altered by physical actions. This physicality and "being in the real world" resonate with Dourish definition of *embodied interaction*: *"Embodied interaction is the creation, manipulation, and sharing of meaning through engaged interaction with artefacts.(Dourish 2001, p. 126)"*

[2]we were told that a homeless person who is often in Îlot 13 was seen collecting the markers from most of the area, it still remains a mystery what may have driven them to do so.

Figure 7.15: Two markers used by Group 2 (Class trial) to reference spaces behind glass windows. In one case (left) the marker is on the window's glass, while in the other it's affixed next to it. Both photos were taken by Group 2 and associated with the corresponding markers.

Participants of the Class trial were quite aware of the volatility of markers. As Marie of the Class trial said, *"[The markers being removed felt] predictable, normal.. if I find a paper on my letterbox I would remove it too."* In some instances it could be seen that it is the act of physical tagging, rather than the physical marker itself, which was deemed important. As Louis told us, *"I think the most important thing is the movement of sticking, and not... you stick it and so the action is down, and that's it. And if it stays or not it's not for me so important, it's the action of doing it."* This statement resonates with some of the findings of a study of graffiti artists Schacter (2008):

> *For the graffiti-artists, however, destruction was* not *in fact seen as a negative act, rather, ephemerality was seen to be part of the very* process *of street-art. Many of the artists argued that the life-span of their images was genuinely irrelevant, the act of* production *counting as the vital part of the process; 'I just write for that moment of writing and the feeling I get from it. I really have no interest in what happens to my work once I've finished creating it' (WK, personal communication).*

The statement from Luis might suggest that the importance of markers is only given by the act of scanning them, as a way to access the digital layer of the uWiki system. Yet, when during the interview participants from the Class trial were asked whether they had ever avoided placing a marker once they had scanned it (i.e., if they just threw it away), they unanimously answered *no* because they considered it important to *"leave a trace."*

7.5.3 Negotiation of Space & Legitimacy of Marker Placement

The potential of markers to both affect, and be affected by, other people in the space where they are placed points towards the fact that markers are not only embedded physically into our world, but also socially. The social impact of markers was evident in a variety of ways, for example when the act of physical tagging involved a negotiation of boundaries. In 2 instances Group 2 (Class trial) placed a marker to refer to a room with people, from the exterior, behind a window. In the first case, illustrated in Figure 7.15 Left, when the room contained a group of people singing, the marker was placed on the glass, in the centre of the window; in the second, Figure 7.15 Right, when the occupant was a lone man at work, the marker was placed less obtrusively next to the window, on the wall. When asked about the differences between these 2 instances, Carlos of Group 2 explained: *"It's also like a psychological aspect. Here he was working and imagine a guy who is working and he turns around and you are like putting a thing against the glass. And here it was like they are all singing; they are like really far. I mean they were here just next to us but however far in a psychological way. They were occupied naturally. Here we have kind of a permission to do this [place the marker] because we weren't noticed some how."*

In the Class trial, however, the awareness of the temporary nature of the markers, due to the semi-permanent adhesive, somehow reduced the concerns about legitimacy. Louis, from Group 1 of the Class trial, placed a marker on a Vespa parked near Îlot 13, as shown in Figure 7.17; when asked if he took into account the possibility of the owner of the scooter turning up while he was attaching the marker and being upset about it, Louis answered that he did not consider it a problem: *"we don't destroy anything.. so I think it's not an act of vandalism."*

The participants' perception of the marker placement was quite different in the Contest trial: while Frank from Team B was placing his markers he found himself in a situation where more explicit negotiation was required. While at the campus information point he had he following conversation:

> *Frank:* Is it ok to place something here?
> *Receptionist:* No. If you ask the lady there, outside, and see what she says that would be good.
> *Frank:* We are doing an experiment around campus.
> *Supervisor:* Where do you come from?
> *Researcher:* I&C, computer science.
> *Supervisor:* You can put on the table.
> *Frank:* No, not on the table...
> *Supervisor:* On the window? Not possible sorry.
> *Frank:* No, not on the window, on the blue [pillar].
> *Supervisor:* Ah yes, ok.

This interaction demonstrates that the process of legitimising markers is a complex one which can rely on a variety of factors and requires situated negotiation.

Finally, it is not only the position of markers which can be negotiated, their appearance can be another means of tacit negotiation. Jean of Team A described how he designed the visual appearance of a marker: *"We thought if we have the Satellite [the campus bar] logo they may let it [stay]."* Jean designed some markers to blend with the visual landscape in order that they might be seen as legitimate and accepted by the people cohabiting, and controlling, the space.

It was also observed that participants used markers to demarcate boundaries found in the environment and give a physical representation to the tensions they represent. For example, some of the markers from the Class trial included the word "private" in their visual representation; Carlos of Group 2 described placing one of these markers (see Figure 7.16), saying:

Figure 7.16: A marker placed and photographed by Group 2 of the Class trial, showing a transition from a shared space into a private one.

Carlos: ..you felt however a kind of privacy [issue] here because you didn't want to be next to a bed of someone you don't know.. and also you could.. it was like a boundary not visible.. not really a physical one but you felt it and so we didn't.. I mean we wanted to respect this and we didn't want to enter here to put it here [beyond the pillar].. you know.. and also the way we wanted to show the picture was like.. really.. from.. going from the shared space into the private.. so you took sort of distance to.. actually take the picture afterwards..
Researcher: ..because you wanted the marker to be in the picture?
Carlos: ..yeah..
Researcher: ..you wanted to basically have this as a sort of a physical..
Carlos: yeah.. like a 'stop' you know.. it was also like this: 'private!'.. 'keep private!'.. 'don't go here!'.. it was like a sign for somebody else also.. here.

Group 2 placed this marker in such a way as to reinforce, and help others recognise, this boundary between public and private space which they perceived.

In summary, because markers are physical objects affixed in real places which may be inhabited by other people, their placement raises questions of legitimacy that have no counterpart in most other forms of digital communication. Participants of our trials addressed these questions in different ways: sometimes explicitly asking for permission, other times using their own judgement or feelings to decide what could be an acceptable place to place a marker. It was also interesting to note that in some situations participants' decisions were consciously influenced by their awareness of the markers' adhesive being only semi-permanent, and therefore not causing real disturbance.

7.5.4 Marker-mediated Interaction with Others

At a general level, the uWiki system is designed as communication technology: it enables users to post and retrieve content using the mobile phones or the web interface. However, the markers *per-se* can also support communication even without the technology mediation, just as graphic signs,

Figure 7.17: Marker on a Vespa scooter, attached and photographed by Group 1 of the Class trial.

and participants demonstrated a strong awareness of this potential, as illustrated in the previous example. As another example, Louis (Group 1 of the Class trial) during the interview commented about a marker that his group placed on the Vespa scooter (Figure 7.17):

> *I was also thinking about how people.. if you saw this somewhere and you are not the one who put and you see it.. it's also a marker that takes your interest.. I was also thinking about the people who would see it.. for example we put this one [on the Vespa] it was more for the guy who drives this Vespa because it was kind of funny.*

The architecture students valued very much this potential interaction with the inhabitants of the site through the visual aspect and position of the markers. They pointed out multiple times during the interviews that they considered it important for the visual content of the markers to be *'interesting'* or *'funny'*, for those who may find the markers on their property and are not equipped with the uWiki technology. They went as far as imagining that inhabitants may displace the markers, thanks to the semi-permanent adhesive: *" ..since we have something like glue that we can take out and glue again in another place, there is also this idea of mobility and we can easily imagine that our marker, when we come back to Îlot 13, it is displaced somewhere else and there is this journey that happens through the îlot. "*

The visual affordance of markers was by itself a way of communicating even within participants of the Class trial. Marie from Group 2 explained:

> *I was in the second shift, so there were already markers placed, which sometimes provided good ideas.. and I found it interesting, but we would not put new markers at the same place...for example there was a marker on the letter boxes and this gave us the idea [to affix a different marker there]... without it we wouldn't have tilted on the place... we wouldn't have noticed it.*

Figure 7.18: Alex from Team C of the Context trial affixing 4 markers near the EPFL Language Centre.

Interaction through the digital content of uWiki did not occur frequently. However, when it happened it was perceived as positive. For example Group 9 added some text content to one of the markers affixed by Group 2; when asked how he felt about this, Carlos from Group 2 said:

> *I mean in a way you are like 'oh somebody uses what we have done' and it's like.. yeah you enjoy it a little bit.. I mean nothing extreme 'wow! yes!' but just a little.. yes, someone used it and it made sense.. so you get a certain confirmation from someone who is outside, and that was good I think.*

In the Contest trial each group was less aware of other groups markers. However, because markers could be left on for a longer period of time, they were noticed by other members of EPFL, who also noticed my name on them (the name was reported in the short piece of text that was added to each marker asking not to remove them). So within 10 days from the end of the Contest trial, I received 2 emails. These unexpectedly provided more interesting data to analyse the trial. The first one was from the director of the language centre, and it referred to the four markers that Team C placed near it, as illustrated in Figure 7.18. In the text printed on the markers, the marker itself were referred to as an 'affiche', which is here translated back as 'poster.' The email (originally in French) read:

> *Subject: posters*
>
> *Body: Hello,*
> *I really appreciated the posters that you have created and displayed at the entrance of the Language Centre.*
> *Could you possibly also make a poster commissioned for the open house for new students?*
> *If yes, thank you for contacting me. I could not reach you by phone.*
>
> *With my best regards*
> ...

The markers were not only noticed, but their graphic design was so much appreciated to receive a request for the design of more 'posters.' Another email did not get as far as requesting more markers, but it clearly shows that the artefacts attracted the attention of a colleague, who was curious about them, and took the chance of a "fallen marker" (the one in Figure 7.12) to contact us and let us know:

> *Subject: Scientific experiment poster found*
>
> *Body: Hello,*
> *This is not the 1st time I am puzzled in front of a small sign "Rolex" attached to the wall in the stairwell of [building X] near the shop. I found it odd "to advertise" for Rolex on the EPFL site. Finally ...*
>
> *This morning, passing by the walkway outside (2nd floor) between [building X] and [building Y], I found another poster (white spots on a black background with some beige-brown points) which was lying on the floor. Guessing that should not be its place, I took it with me and I have it in my office.*
>
> *If you want to get it back, I can either send it by internal mail or of you want to stop by please give me a call to make sure I am in my office.*
>
> *Good day and best regards.*
> ...

Markers, then, can communicate not only to d-touch users who are able to scan them with their mobile phones, but also to anyone who perceives simply their visual nature and sees them as signs or posters. Because of this support of multiple viewpoints designable markers may constitute an example of *boundary objects* as defined by Star and Griesemer (1989): objects that carry different significance for different groups, and yet allow communication between them.

7.5.5 Visual Design of Markers

Participants in the Class trial and in the Contest trial experienced the visual design of d-touch markers in different ways: while in the Contest trial participants were able to design their own markers, the Class trial participants were provided with ready-made markers already printed. Yet, across the two trials, the visual aspect of the markers was always used to convey meaning, rather than, for example, in a decorative or generic way.

As described above, Architecture students were aware of the potential of the visual aspect of markers and they actively took advantage of it, in combination with the markers' position, for communicating with others. As an attempt to relax the limitation caused by having only 10 marker-words, students were invited to use the markers beyond their visual meaning, yet, they generally used them adhering literally to the words included in them. Moreover, as already discussed, the markers' visual design was very often featured in the photographs that students captured with the camera phones on site, and it was in this way rendered as digital content. When asked about this action students from the Class trial commented: *"We took a photo of the marker to localise the theme and make it more easy to understand.. ..we did it in a very mechanical way. "*

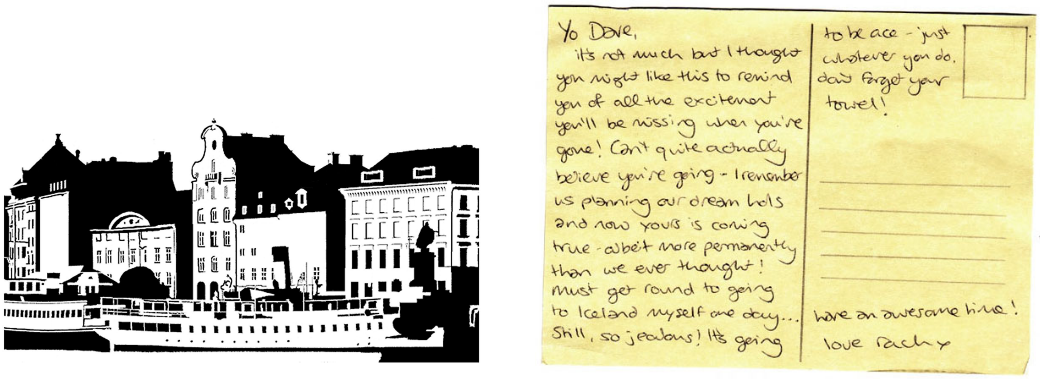


Figure 7.19: The “postcard” marker by Team B of the Contest trial. On the left the marker visual design, representing the front of the postcard – note that not the entire symbol is a visual marker, but only one part of it, on the top-right. On the right the digital content associated with it, representing the back of the postcard.

Participants of the Contest trial used the visual design of the markers to convey meaning in a very direct way. This is most evident in the “Postcard” (Figure 7.19) and “Answering Machine” (Figure 7.8) markers created by Team B. The Postcard marker, illustrated in Figure 7.19 left, represents the front of a postcard, a black and white photograph of Stockholm. The back of the postcard, a hand-written text message, exists in the form of a digital image associated with the marker (Figure 7.19 right); this is an actual scan of a real postcard received by the participant. The Answering Machine marker embeds actual functionality in its visual design – by moving black magnets in or out of the marker users can change its topology, and therefore its ID. The design of the marker also includes visual cues to the functionality (the arrow indicating where to add the magnets) and the digital content (the phone icon and the text).

All of the markers from Teams A and C are iconic, for both teams the visual design of the marker was representative of the digital content associated with them. Markers from Team A mostly show logos of student associations, events or businesses (e.g. restaurants and bars) present on campus and their digital content provides information or reviews about the same as well as clues to solve the puzzle invented by Team A. Logos, which are often used on existing signs and plaques on campus, were modified to fulfil the d-touch marker design constraints. Team C similarly provided digital content about campus businesses and university services, but used more abstract visual symbols, such as the icon of an hamburger to indicate campus restaurants (Figure 7.20).

Rather than suggesting to replace the existing signs with their interactive ones, Teams A and C created a separate layer of signage, which is interactive and independent from the other one. From the interview it is clear that these participants were well aware of the visual distortion that the d-touch constraints induce in existing logos, and they considered an element to play with: *“ I was even thinking it would be fun [to have] something you don’t even know it’s a marker and suddenly you try to scan it and ‘oh it’s a marker’ so.. obviously most of the time if you know what a d-touch marker is, you can see well that picture looks a little bit weird you somehow kind of fun...”*

Other interesting markers were produced by Team A, beside the ones based on logos. The marker shown in Figure 7.12 represents the floor plan of the Learning Centre building on EPFL campus, to which the content associated with the marker refers to. In this case the visual design of the

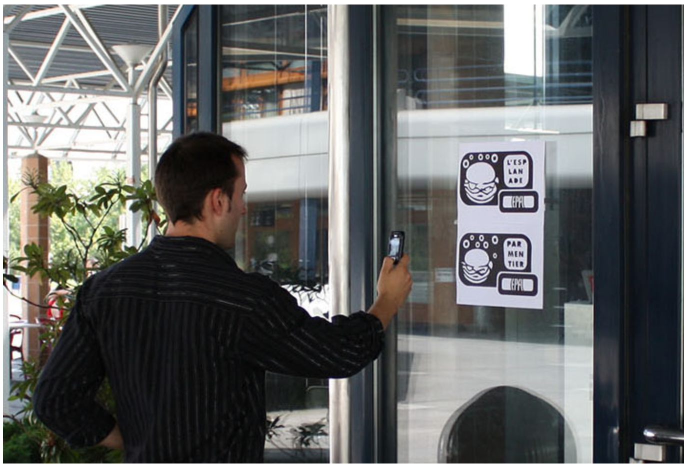


Figure 7.20: Two markers affixed by Team C referring to restaurants on EPFL campus.

marker represents the actual object it refers to. Similarly the same team produced a marker based on a black and white photo of the president of EPFL, which they affixed near his office.

Despite the functional use of the visual design, participants were also aware of the aesthetic appeal of the markers: Frank of Team B considered the postcard marker as *" the best looking marker that I had so it would draw people's eyes."* Therefore, he positioned it in front of a location that considered important on campus, the main office of the students' association, to attract visitors attention there.

Participants were very aware of the visual potential, and in some cases wanted to demonstrate it through their contest entry, as Alex from Team C told us referring to one the language centre markers (Figure 7.18): *" We wanted to make the Chinese [marker] because it gives the idea that with these markers there may be something that does not make sense to you, it's just visual but with the system you get the content."*

In summary, several observations show that participants were very aware of the visual affordance of markers. Both when they directly determined their design and when they worked with pre-designed markers, the visual aspect of markers had a central role. The visual design was actively used to convey meaning, rather than just as decoration. Often visual aspect and digital content were deeply connected in ways that changed from instance to instance.

7.5.6 Marker Interrelations

Sometimes, in the Contest trial, they were based on the markers' visual design, for example all markers from Team C maintain a clear visual identity that allows them to be recognised as parts of the same group. Similarly, the "beginning", "middle" and "end" markers from Team B are visually coherent, but they are also set in a sequential relation. Sequential relationships within markers were other times created through their digital content: Team A created this relationship in several ways. For some markers there are very clear instructions within the content about where to find the next marker (e.g., *"You should go along the wall you have on your right, and then turn*

Figure 7.21: Two “alternative” markers affixed by Team B as the beginning of their campus tour. They allow users to choose English or French as language for the digital content.

right. But only do this if you trust what a phone tells you...”), while others included less explicit clues (e.g., a marker whose content included *“If you’re lost, you should look around for a map.”* which lead users to the subsequent marker placed on a public informational map nearby).

The relations observed between markers were not only sequential. Team A placed two markers next to each other as to create an alternative, to let users select the language for the digital content to be either English or French (Figure 7.21). Other times markers were simply grouped: the ‘Language Centre’ markers (Figure 7.18) from Team C representing and providing information about different languages, or the two restaurant markers from the same team. Markers were grouped to provide convenient access to different pieces of content, as Alex from Team C explained: *“the idea is that the user can see all menus for EPFL in the same place”* – yet, rather than condensing all content to one maker, different visual designs form a physical selection interface for the mobile phone. The dynamic ‘answering machine’ marker from Team B (Figure 7.8) is also in effect a group of markers combined together, which also form with a physical interface.

Position was also used during the Class trial to establish relationships between markers from different groups. For example having noticed a marker visually representing “individual” on one of the letter boxes, Group 6 placed a marker with the word “community” on another box – Marie explained, *“ we would have not placed the same marker because it’s a waste to use [the same marker] twice on the same thing.. so.. we wanted to think about the letterbox in a different way.”* Group 2 placed a pair of markers which one member, Carlos, expressed as tightly related to one another by saying, *"..this one [private] and this one [shared].. it is a specific one [case] because they were linked.."* Group 1 enacted a similar relationship between two markers by placing in the content of the marker with the word 'open' in its visual representation the textual comment *"for this topic, see also our marker 'closed.' "* These examples highlight the way in which people work markers into a sort of dialogue with one another, a bi-directional bond. It is clear, then, that while uWiki supports this sort of 'traditional' sequential ordering of markers often seen in location-based tour applications, it is flexible enough to allow users to enact this relationship in a variety of ways.

7.5.7 Technological Ecology

Markers must be situated within a broader social and physical landscape, where other forms of technology also play a rle. This is evidenced by the fact that the uWiki system itself stands as an ecology of technologies which are worked together by the users. As mentioned earlier, at the most basic level, the visual and physical representations of the markers must be designed and then manufactured; there is a mobile phone application which is used to scan markers; the same application allows the user to utilise the camera and keyboard of the mobile phone in order to create content to add to the marker; the web interface allows for further addition and manipulation of the content; and, finally, both the mobile and web interfaces also allow users to view content from the system. The important point here is not only that uWiki is comprised of a variety of technologies but that these aspects of the system, and their interrelations, are at the very core of what a physical tagging is–a multilayer techno-social system. And this system itself is embedded in the even broader ecology of technologies encountered in the everyday world. For instance, participants in the Class trial brought with them other technologies, such as cameras, to gather data in parallel, or to collect additional information at higher resolution, during the class. Carlos told us: *We wanted a high resolution picture. So there the problem was the technology actually, because we wanted to do a photo montage afterwards so we took our own cameras.*

Further, during the Contest trial the participants used a variety of external editing software to generate both their markers and the content which they attached to them. Here then the limitations of the uWiki system become evident, but this is not to say that the solution is better mobile phone cameras or more built-in content editing. Rather, it helps us to recognise that for any system one may build there will always be other devices and applications which users carry with them or encounter in the spaces they move through. It is important, then, to acknowledge the ways in which physical tagging systems can be come a part of, rather than replace, a larger technological ecology.

One of the ways in which I have made efforts to that effect was prompted by participants of the trials. During the course of the study I was asked to expand the uWiki system to include user-selected URLs in addition to the media items stored on the server. This was indeed implemented, as mentioned earlier in the system description. Here I would like to draw attention to the fact that this decision was made to enhance the relevance and the interconnectedness of the uWiki system with users' everyday technological practices. Physical tagging systems can indeed benefit, then, from creating new relationships with other large-scale systems, like the Internet, rather than trying to emulate or replace them. And in fact the uWiki pages themselves were intentionally designed to exist alongside the many sites already found online, and as webpages they allow users to link to

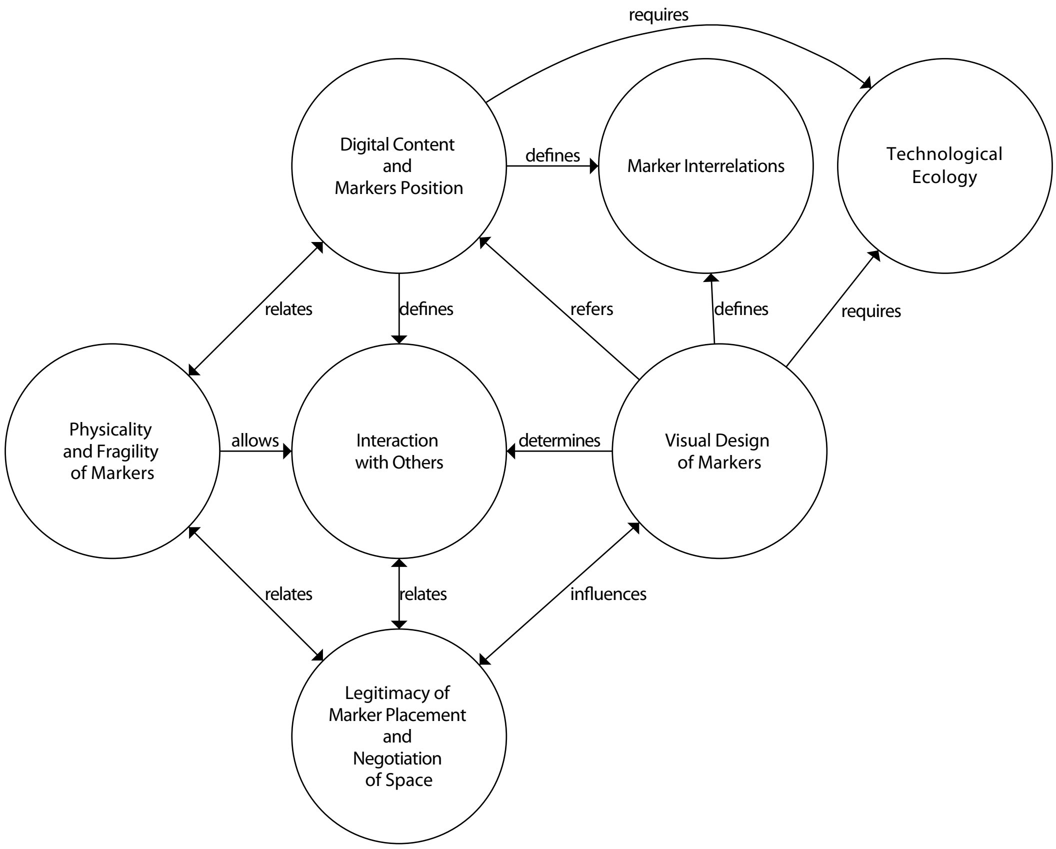


Figure 7.22: Diagram showing the relations between codes extracted from the trials data.

them from other sources besides markers and to export their content for reuse elsewhere. Indeed, as mentioned earlier 4 out of the 5 reports created by the architecture students for their class used content from uWiki. Not only the content itself, but also the experience of physical tagging, of using the system, influenced the users; when asking Louis if he used uWiki content for the final report he told us: *We used the ideas. The photo quality is not good, so we took the photo again, the very same one.*

Clearly, then, uWiki, and other systems like it, do not exist in isolation, in a vacuum. Rather they exist in a world characterised by its diverse technologies. Any physical tagging system, including this one, is part of a broader set of practices and cannot be thought of as existing in a world void of other technological interventions. The uWiki application sits alongside other mobile phone applications on the same phone; the mobile phone itself is one of many other devices people carry with them and encounter. This conception of physical tagging systems then moves beyond the way in which the classic museum guide application. Rather than being the sole technological intervention, the only way of digitally interacting with the museum, these findings urge us to consider the ways in which one can both design, and conceive of, physical tagging systems that are figured into, and make use of, a broader technological ecology.

7.6 Relations between Categories & Implications for Future Design

The 7 categories described in the previous subsections are far from being isolated. Often they overlap and influence each other. These relationships are summarised in the diagram in Figure 7.22.

As it is shown in the diagram, at the centre of these relationships is the category *interaction with others* – this is not surprising as physical tagging is a form of communication. Three different factors have an effect on *Interaction with others*: the *visual design* of the markers, their *position & digital content* in the physical world and their physicality or fragility. The *visual design* as well as the *position & digital content* convey meaning to others. Their *physicality* allows *others* to remove them, which indeed is also a form of interaction. The category *legitimacy of placement and negotiation of space* has bidirectional relationships with 3 other categories: the *physicality and fragility of markers* creates the question of their legitimacy, as they physically occupy a space and can be removed from it; this is also why the category is also related to *interaction with others*; finally the *legitimacy of placement* is influenced and influences the visual design of the marker, as observed for example in the case of the Satellite campus bar.

Marker interrelations are defined through their *positions & content* or through their *visual design*. Both the visual design of the markers and the creation of the content associated with them require the use of other technological tools – this is expressed in Figure 7.22 with the arrows connecting these two categories with the category *technological ecology*.

Even though the results of this study are preliminary, they serve to demonstrate the range of potential for the future of physical tagging systems. In summary, physical tagging is a practice which is rooted in both the physical and digital aspects of physical markers themselves; physical tags can be used as a way to negotiate the tensions found in the everyday world, but at the same time, the act of physical tagging itself involves a negotiation of those same tensions.

These observations suggest that systems supporting user generated physical tagging have potential to attract the interest of real users. Future designs should carefully define the physical nature of tags, including how firmly or loosely they can be attached to the environment and how visible or hidden they can be made, as these factors may strongly influence the tagging behaviour of end-users. The relationship between content and markers should be made more flexible, allowing users to define conditional branching based on which markers have already been scanned. The participants of the uWiki trials did take advantage of the openness of uWiki for importing and exporting digital content from and to other applications, in particular through the web interface. This observation suggests that the design of future physical tagging systems should take into account how they fit into broader technological ecologies of use.

7.7 Limitations

The study reported in this chapter represents just an initial attempt to understand the phenomena of location-based content creation and in particular physical tagging, which are rather understudied, as discussed in Chapter 2. Of course, this work being early and exploratory is only scratching the surface. Participants in the first trial were almost exclusively Architecture students, who may have a particular sensitivity for social use of space and for design. This might have been particularly beneficial in facilitating the emergence of interesting physical tagging practices. Moreover, the duration of both trials was quite short. More studies are necessary to understand how different user groups approach physical tagging and how the practices observed in this study might be supported in longer term engagements and in a variety of contexts and applications.

7.8 Conclusion

The qualitative analysis of the field trials presented in this chapter answers the fourth and last of the research questions formulated in the introduction of this dissertation. The rich variety of content and behaviours observed confirm that visual markers, indeed, support physical tagging practices. The qualitative and quantitative results of the study demonstrate that people can use physical tagging systems to express, distribute and create content. Moreover, the designable feature of d-touch markers was also observed to support specific practices: participants of the field trials were very aware and in several instances actively used the visual affordance of markers. Future designs should carefully define the physical nature of tags, including how firmly or loosely they can be attached to the environment and how visible or hidden they can be made, as these factors may strongly influence the tagging behaviour of end-users. The participants of the trials did take advantage of the openness of uWiki for importing and exporting digital content from and to other applications, in particular through the web interface. This observation suggests that the design of future physical tagging systems should take into account how they fit into broader technological ecologies of use.

8. Limitations and Future Work

This dissertation presented the concept of designable visual markers and an exploration of how they support interaction with mobile devices. The subject was explored through a wide interdisciplinary perspective and different research tools were employed to observe how users design markers and how they interact with them through mobile devices, both in a laboratory and in the field. The main limitation of this approach was that it favoured a macroscopic perspective, but it did not drill down into the details of specific issues. This chapter outlines how each section of the dissertation could be extended through further research.

8.1 Marker Recognition

D-touch, the specific implementation of designable markers recognition based on image topology was introduced in Chapter 3 as a working example, and it enabled the experiments and field trials reported in the rest of the dissertation. The d-touch system was developed only up to the point of being usable, without the claim of being an optimal solution to the engineering problem of designable marker recognition: there is therefore an opportunity for future research to further investigate technical solutions to this problem.

The work could start by probing further the limits of the d-touch system itself, including the marker definition. The consequences of allowing deeper nesting of the marker regions, i.e. having marker threes of height larger than 3, should be explored in terms of computational efficiency as well as resilience to false positives. Future user studies of marker design should asses how these different constraints may impact human expression ability. Further user studies on marker design should also look at different user groups, such as graphic design students and professionals – having received specific training, these users may have a different approach and different needs from the application.

Systematic testing of the d-touch operational limits and comparison with other marker recognition technologies can be a starting point to improve its performance. Performance for marker recognition can be defined in terms of: detection rate, misclassification rate, information density (bits per unit area at a fixed resolution), positional accuracy, robustness to deformation and lighting and processing time. In particular, possible improvements of the thresholding scheme could take into account scale variations and application of non-linear front-end filters. These could improve detection rate, misclassification rate, positional accuracy, and robustness to deformation and lighting.

Further investigation should asses the opportunity of complementing the topology-based recognition with geometrical processing, for example taking into account the shape of the markers regions and their relative positions. Geometrical information could be used to increase the bandwidth of the system, allowing to store more information in the markers, or provide error detection potential. A mixed topology-geometry approach may also enable pose estimation, which would in turn allow the visualisation and manipulation of three-dimensional virtual objects through the markers. Error-detection could also be introduced through additional feedback on the interface: for example, application may display a visual representation of the marker detected by the system, so that users could easily verify that this corresponds to the one they meant to scan, and try again if necessary.

The d-touch analyser application should be extended to include additional models of distortion, beside blur, in order to provide a more accurate "robustness" analysis. In particular, it was noticed that inner areas of the markers in proximity of contours with high curvature are sometimes erroneously handled by the current robustness analysis. Digitally rotating the candidate marker images by incremental steps, in addition to scaling them, could probably improve the quality of the results. Moreover, the application should be extended to allow users to analyse not only individual markers, but also collections of markers, in order to assess and limit risks of misclassification arising from markers in the same collection having similar IDs.

At a more general level, further research about designable marker recognition should investigate implementation strategies alternative to the topology-based approach of d-touch. Attractive candidates for this are the natural features tracking methods discussed in Chapter 2, such as SIFT-based methods. Still it is unclear what level of real-time performance can be achieved through this methods on devices with limited computing power an memory such as phones, and what the trade-off between processing speed and number of symbols or templates that can be recognised. In any case, both the designs reported in Chapter 6 and the observations discussed in Chapter 7 suggest that the physicality of printed markers and their visual affordance making them recognisable by users as *interactive* or *scannable objects* is important. As a consequence future work on designable markers should give high importance to these features, rather than merging back into the field of generic object recognition and tracking.

8.2 Usability of Marker-based Mobile Interfaces

Chapter 5 presented a usability study of marker-based mobile interfaces. This work could be extended in various directions. Future usability experiments could be carried out to compare the performance of designable markers to 2D barcodes. The results reported by Toye et al. (2007) and Rohs and Oulasvirta (2008) suggest that the split of visual attention between the phone display and the physical environment (to get context about what marker the phone should be pointed at) hinders performance during the pointing task. Experiments could, therefore, be designed to test the hypothesis that designable markers reduce the need of such visual attention split, as the information is all visually contained in the marker itself, which can be observed through the viewfinder.

Another direction of investigation could aim at assessing the effects of marker recognition processing speed on usability. As discussed in Chapter3, the real-time performance of the d-touch implementations enables rich visual feedback through an interactive viewfinder, where markers are highlighted as soon as they are inside the finder and it is possible to create hovering effects: displaying information on a marker before the user selects it, similarly to hovering effects with the mouse pointer. Other marker recognition systems, or phones with more limited processing resources, may require longer processing time to analyse an image captured through the camera, making it impossible to implement an interactive viewfinder. A usability experiment could be designed to evaluate marker selection through an interactive or non-interactive viewfinder, measuring performance differences, if any. An additional dimension of comparison could be defined on the resolution at which images are processed: processing higher resolution images should allow the recognition of smaller markers (or equivalently markers that are further away from the phone), but it would result in slower frame-rates. Controlled experiments could be designed to asses the effects of resolution on marker selection tasks, trying to identify possible flooring and ceiling effects: systems may be completely unusable below a certain resolution/frame-rate, while differences may not be noticeable above a certain other resolution/frame-rate.

8.3 Marker-based Mobile Applications

In order to gain more insight on the integration of markers in mobile interfaces, several of the applications outlined in Chapter 6 could be further developed and deployed with real users. Mobile application stores (e.g. the Apple App Store, Nokia Ovi and the Android Market) provide an opportunity for more easily deploy applications and reach a wide audience of potential users. Remotely logging the applications' usage may provide a low-cost strategy to learn more about users' interaction with them, to be complemented by observation sessions and interviews. Yet, on mobile platforms remote logging may involve technical challenges as phones are not likely to be connected to the internet all the time, while running the applications.

In particular, the initial work, development and evaluation, briefly reported for the Mobile Audio Cards application in Chapter 6 shows that such "mobile tangible interface" has the potential to attract children's attention. More work is required to overcome the interface design issues and evaluate whether the system actually enriches children storytelling, similarly to the work of Shankar (2004), by which it was inspired. Future field trial could aim at comparing different conditions where the use of cards with pictograms is augmented through the phone or not, to understand if audio augmentation can enhance the story telling activity.

8.4 Physical Tagging

Chapter 7 reported two trials of physical tagging and their analysis. Further observations are required to generalise the understanding of this kind of systems and create models that can predict users' behaviours. It would be interesting to observe interaction in other social contexts and application areas (museums, narrative applications), to evaluate whether this technology may have positive effects for example for learning.

The field trials of uWiki reported in this dissertation were limited to a small number of participants, and to prevent shortcomings due to a network effect, external conditions were arranged to motivate the use the system. In this way it was possible to observe *how* participants used uWiki, showing that it is generally usable. However, the actual usefulness of physical tagging is still to be determined, and future observations should address it by creating usage conditions that are more loose and through longer term trials.

The uWiki system should be extended based on the results of the trials reported in Chapter 7. For example, the system may let users express how specific of a reference they want to cast through the markers, this could be defined in categories such as: *specific*, *immediate surrounding*, *vast surrounding*, *point of view* and *generic*. Given that very often an image of the marker is associated with the marker itself, uWiki could automatically capture a photograph when users can the marker for the first time and offer it to be associated with the marker.

Similar to the way standard websites take advantage of users' profiles to filter and recommend content that may better match their interest, uWiki could arrange and filter content based on what markers users scanned or what content they posted. Such additions would make it easier to manage large amounts of content, especially on the mobile interface. Filtering and recommendation techniques related to user profiles may realise more of the potential of mobile technology – mobile phones are truly personal devices, with us most of the time, which may be able, in principle, to sense much contextual information that could be integrated with physical tagging systems.

Future studies should attempt to compare designable visual markers with alternative physical tagging technologies. How do 2D barcodes, RFID, GPS or Wifi positioning affect the practices of physical tagging? Observing and contrasting different implementation techniques may reveal interesting patterns and help in the design of multi-modal systems.

From the technical point of view, deploying the d-touch and uWiki technology on recent mobile platforms, such as the Apple iPhone and Google Android, may make it easier to arrange future studies. While Symbian S60 was considered the most convenient mobile platform at the beginning of the project, the introduction of the Apple iPhone and Google Android SDKs require reconsidering the situation. Not only these devices are very popular, but they support easy installation of 3rd party applications through centralised application repositories, making it easier to distribute marker-based prototypes to a larger audience. It must be noted, however, that porting d-touch to Android would require a radical amount of re-writing, as this platform only supports the Java programming language and not C/C++. On the other hand, the availability of the Qt library (in C++) on Nokia devices, as well as Windows Mobile ones, make this platform also attractive, even though the lack of a central applications repository is indeed a strong deterrent. On the server side, to facilitate further development and deployment the project should be migrated from TikiWiki to a more manageable and modular platform, such as the Django web framework in Python.

9. Conclusion

The general contribution of this dissertation is the introduction of the idea of *designable markers*, visual markers that are both machine-readable and visually communicative to humans, and the results of an investigation of the ways in which they can support mobile human-computer interaction. This investigation was conducted through a variety of methods: through formal usability experiments, through the creation and analysis of example designs, as well as through the qualitative analysis of two field trials. All 3 approaches were enabled by the engineering and development of d-touch, an actual recognition system that supports designable visual markers and by its integration in a variety of applications and experimental probes. D-touch was described in Chapter 3.

As listed in the introduction, specific contributions included:

- an experimental confirmation that topology-based recognition allows for designable visual markers, and that markers that are both functional and visually expressive can be easily produced by a wide spectrum of users, without much training;
- the design, implementation and evaluation of a software tool that supports users in creating valid markers;
- a study-based comparison of a small variety of marker-based mobile interfaces, illustrating how interface variations affect usability;
- an overview and discussion of mobile application designs demonstrating the ways in which designable markers can be adopted for mobile interaction, including the design and implementation of prototypes enabling study of practices emerging around their use and the assessment of the interfaces' usability;
- the analysis of different physical tagging practices supported by visual markers, as observed in two field trials.

D-touch is based on image topology, and its markers are defined in terms of constraints on the nesting of dark and light regions. First of all, this dissertation assessed the question of *whether such topology-based approach could actually support designable visual markers*. The answer was provided in Chapter 4 through a controlled user study, where novice users were introduced to the d-touch systems and asked to draw as many markers as they could within a fixed period of time. These results form the first specific contribution of the dissertation.

The study reported in Chapter 4 also verified the usefulness of the *d-touch analyser*, demonstrating how *software tools can be designed to support users in the creation of valid d-touch markers*. This is the answer to the second research question formulated in the introduction and the second specific contribution.

To understand *in which new ways designable visual markers can be integrated in mobile interfaces* and *how different marker-based interfaces affect users' performance and preferences*, 5 interface prototypes were created and tested through the lab-based usability experiments reported in Chapter 5, and 8 mobile application designs were presented and discussed in Chapter 6. The study results show that, under the specific experimental conditions, all marker-based mapping strategies produced correct selection rates above 94.5%, and task completion time ranging, in average, from 2.28s (st. dev. 0.91) to 3.59s (st. dev. 1.23). The results provide a characterization of the different interfaces, the third contribution of the dissertation. Participants reported a strong preference for interfaces that involved only simple, coarse, marker scanning, compared to those involving a combination of marker scanning and key-presses or touch selections. The mobile application examples illustrated that the visual design of markers can take different roles. These range from instances where the markers' design is left open for end-users to define, to situations where the visual aspect of the markers conveys cues about their interactive role, to instances where iconic markers are blended into the physical world, while still being clearly recognizable as interactive elements, to markers that are "invisible" because they are hidden into existing graphic elements. The examples were also varied with respect to the functional role of markers: from being complementary to existing printed information, to be central or primary elements of the interaction design.

Finally, to study *what physical tagging practices can emerge from, and be supported by, the use of markers* two field trials were reported and analysed in Chapter 7. The trials were centred around the use of uWiki, a physical tagging functional prototype that allows users to associate digital content to d-touch markers. The rich variety of content and behaviours observed show different physical tagging practices supported by that visual markers, in the context of the trials. The qualitative and quantitative results of the study demonstrate that people can use physical tagging systems to express, distribute and create content. Moreover, the designable feature of d-touch markers was also observed to support specific practices: participants of the field trials were very aware of the visual affordance of markers and in several instances explicitly took advantage of it. The observations suggest that systems supporting user generated physical tagging have potential to attract the interest of real users. Though the results of this work are preliminary, they serve to demonstrate the range of potential for the future of physical tagging systems.

The research questions formulated in the introduction were answered in this dissertation using the d-touch topology-based recognition. Yet, the same questions could be addressed through alternative technologies for the implementation of designable markers prototypes, leading perhaps to different answers. As mentioned in Chapter 8, comparing the effects of different technical solutions for designable markers and physical tagging may be interesting in itself and should be considered as future work.

I hope that the work presented in this dissertation, besides addressing the research questions listed in the introduction, may be helpful to other HCI researchers as an example of the combination of different research methods. More in general, I hope that the material presented here will open up new opportunities for the design of mobile interfaces that relate digital information to the physical world, and that this may, in turn, allow people to better communicate through mobile devices, being more aware of their immediate surroundings.

References

Gregory D. Abowd, Christopher G. Atkeson, Jason Hong, Sue Long, Rob Kooper, and Mike Pinkerton. Cyberguide: a mobile context-aware tour guide. *Wirel. Netw.*, 3(5):421–433, 1997. ISSN 1022-0038. doi: http://dx.doi.org/10.1023/A:1019194325861.

Morgan Ames and Mor Naaman. Why we tag: motivations for annotation in mobile and online media. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: http://doi.acm.org/10.1145/1240624.1240772.

Mattias Müller Andreas Ullrich, Matthias Marx, editor. *Stickers. From the 1st International Sticker Award.* Die Gestalten Verlag, 2006.

Alice Angus, Dikaios Papadogkonas, George Papamarkos, George Roussos, Giles Lane, Karen Martin, Nick West, Sarah Thelwall, Zoetanya Sujon, and Roger Silverstone. Urban social tapestries. *IEEE Pervasive Computing*, 7(4):44–51, 2008. ISSN 1536-1268. doi: http://dx.doi.org/10.1109/MPRV.2008.84.

Paul M. Aoki, Rebecca E. Grinter, Amy Hurst, Margaret H. Szymanski, James D. Thornton, and Allison Woodruff. Sotto voce: exploring the interplay of conversation and mobile audio spaces. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 431–438, New York, NY, USA, 2002. ACM. ISBN 1-58113-453-3. doi: http://doi.acm.org/10.1145/503376.503454.

Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21 (6):34–47, 2001.

Ross Bencina, Martin Kaltenbrunner, and Sergi Jordà. Improved topological fiducial tracking in the reactivision system. In *Proc. IEEE International Workshop on Projector-Camera Systems (Procams 2005)*, San Diego, USA, 2005.

Steve Benford, Gabriella Giannachi, Boriana Koleva, and Tom Rodden. From interaction to trajectories: designing coherent journeys through user experiences. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 709–718, New York, NY, USA, 2009a. ACM. ISBN 978-1-60558-246-7. doi: http://doi.acm.org/10.1145/1518701.1518812.

Steve Benford, Gabriella Giannachi, Boriana Koleva, and Tom Rodden. From interaction to trajectories: designing coherent journeys through user experiences. In *CHI '09: Proceedings of the*

27th international conference on Human factors in computing systems, pages 709–718, New York, NY, USA, 2009b. ACM. ISBN 978-1-60558-246-7. doi: http://doi.acm.org/10.1145/1518701.1518812.

J. Bernsen. Dynamic thresholding of grey-level images. In *Proc. Eighth Intl Conf. on Pattern Recognition*, pages 1251–1255, Paris, France, Oct. 1986.

Barry Brown, Matthew Chalmers, Marek Bell, Malcolm Hall, Ian MacColl, and Paul Rudman. Sharing the square: collaborative leisure in the city streets. In *Proc. of European Conf. on Computer Supported Cooperative Work ECSCW '05*, pages 427–447, New York, NY, USA, 2005. Springer-Verlag New York, Inc. ISBN 978-1402040221.

A. J. Bernheim Brush, Tammara Combs Turner, Marc A. Smith, and Neeti Gupta. Scanning objects in the wild: Assessing an object triggered information system. In *Ubicomp*, pages 305–322, 2005.

Scott Burnham, editor. *Droog Event 2:Urban Play (catalogue)*. Droog Publisher, 2008.

Jenna Burrell and Geri Gay. E-graffiti: evaluating real-world use of a context-aware system. *Interacting with Computers*, 14(4):301–312, 2002.

M. Cherubini, F. Hong, P. Dillenbourg, and F. Girardin. Ubiquitous collaborative annotations of mobile maps: how and why people might want ot share geographical notes. In *Proceedings of the 9th International Workshop on Collaborative Editing Systems (IWCES'07)*, 2007.

Mauro Cherubini. *Annotations of maps in collaborative work at a distance*. PhD thesis, EPFL, Lausanne, Switzerland, 2008.

Keith Cheverst, Nigel Davies, Keith Mitchell, and Adrian Friday. Experiences of developing and deploying a context-aware tourist guide: the guide project. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 20–31, New York, NY, USA, 2000a. ACM. ISBN 1-58113-197-6. doi: http://doi.acm.org/10.1145/345910.345916.

Keith Cheverst, Nigel Davies, Keith Mitchell, Adrian Friday, and Christos Efstratiou. Developing a context-aware electronic tourist guide: some issues and experiences. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 17–24, New York, NY, USA, 2000b. ACM. ISBN 1-58113-216-6. doi: http://doi.acm.org/10.1145/332040.332047.

Mobile Codes Consortium. Website. http://www.mobilecodes.org/, February 2007. URL `http://www.mobilecodes.org/`.

Juliet Corbin and Anselm Strauss. *Basics of Qualitative Research*. Sage, 3rd edition, 2008.

Enrico Costanza and Mirja Leinss. Telling a story on a tag: The importance of markers' visual design for real world applications (paper presentation). In *Proceedings of the Mobile Interaction with the Real World Workshop at Mobile HCI2006*, Espoo, Finland, sep 2006.

Enrico Costanza and John Robinson. A region adjacency tree approach to the detection and design of fiducials. In *VVG*, pages 63–69, 2003.

Enrico Costanza, Simon B Shelley, and John Robinson. D-touch: A consumer-grade tangible interface module and musical applications. In *Proceedings of Conference on Human-Computer Interaction (HCI03)*, 2003a.

Enrico Costanza, Simon B Shelley, and John Robinson. Introducing audio d-touch: A tangible user interface for music composition and performance. In *Proceedings of the 2003 International Conference on Digital Audio EffectsDAFx03*, 2003b.

Enrico Costanza, Andreas Kunz, and Morten Fjeld. Mixed reality: A survey. *Lecture Notes in Computer Science*, 5440:47–68, 2009a.

Enrico Costanza, Andreas M. Kunz, and Morten Fjeld. Mixed reality: A survey. *Lecture Notes in Computer Science*, 5440:47–68, 2009b.

Andy Crabtree, Steve Benford, Tom Rodden, Chris Greenhalgh, Martin Flintham, Rob Anastasi, Adam Drozd, Matt Adams, Ju Row-Farr, Nick Tandavanitj, and Anthony Steed. Orchestrating a mixed reality game 'on the ground'. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 391–398, New York, NY, USA, 2004. ACM. ISBN 1-58113-702-8. doi: http://doi.acm.org/10.1145/985692.985742.

D Crow, P Pan, and G Davenport. System architecture for developing mobile cinema. In *ACM Multimedia Conference*, November 2003.

Diego Lopez de Ipina, Paulo R. S. Mendonca, and Andy Hopper. Trip: A low-cost vision-based location system for ubiquitous computing. *Personal Ubiquitous Comput.*, 6(3):206–219, 2002. ISSN 1617-4909. doi: http://dx.doi.org/10.1007/s007790200020.

David Dearman, Kirstie Hawkey, and Kori M. Inkpen. Effect of location-awareness on rendezvous behaviour. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1929–1932, New York, NY, USA, 2005. ACM. ISBN 1-59593-002-7. doi: http://doi.acm.org/10.1145/1056808.1057059.

Paul Dourish. *Where the action is: the foundations of embodied interaction*. MIT Press, Cambridge, MA, USA, 2001. ISBN 0-262-04196-0.

DROPDROP Agency, editor. *The Book of Tags*. Kitchen 93, 2006.

Mark Fiala. Artag, a fiducial marker system using digital techniques. In *CVPR (2)*, pages 590–596, 2005.

Margaret Fleck, Marcos Frid, Tim Kindberg, Eamonn O'Brien-Strain, Rakhi Rajani, and Mirjana Spasojevic. From informing to remembering: Ubiquitous systems in interactive museums. *IEEE Pervasive Computing*, 1(2):13–21, 2002. ISSN 1536-1268. doi: http://dx.doi.org/10.1109/MPRV.2002.1012333.

G Fowler. Qr codes: In japan, billboards take code-crazy ads to new heights. Wall Street Journal, Oct, 10 2005.

Arjan Geven, Peter Strassl, Bernhard Ferro, Manfred Tscheligi, and Harald Schwab. Experiencing real-world interaction: results from a nfc user experience field trial. In *MobileHCI '07:*

Proceedings of the 9th international conference on Human computer interaction with mobile devices and services, pages 234–237, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-862-6. doi: http://doi.acm.org/10.1145/1377999.1378012.

Morten Gisselbaek, Peter Haefeli, and Pierre Hollmuller. Participative retrofit of the ilôt 13 area in geneva. In *Proc. of Conf. Passive and Low Energy Architecture PLEA '06*, 2006.

Rebecca E. Grinter, Paul M. Aoki, Margaret H. Szymanski, James D. Thornton, Allison Woodruff, and Amy Hurst. Revisiting the visit: understanding how technology can shape the museum visit. In *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 146–155, New York, NY, USA, 2002. ACM. ISBN 1-58113-560-2. doi: http://doi.acm.org/10.1145/587078.587100.

William G. Griswold, Patricia Shanahan, Steven W. Brown, Robert Boyer, Matt Ratto, R. Benjamin Shapiro, and Tan Minh Truong. Activecampus: Experiments in community-oriented ubiquitous computing. *Computer*, 37(10):73–81, 2004. ISSN 0018-9162. doi: http://dx.doi.org/10.1109/MC.2004.149.

Juha Häikiö, Arto Wallin, Minna Isomursu, Heikki Ailisto, Tapio Matinmikko, and Tua Huomo. Touch-based user interface for elderly users. In *MobileHCI '07: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, pages 289–296, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-862-6. doi: http://doi.acm.org/10.1145/1377999.1378021.

Frank Allan Hansen and Kaj Grønbæk. Social web applications in the city: a lightweight infrastructure for urban computing. In *Hypertext*, pages 175–180, 2008.

Frank Allan Hansen, Karen Johanne Kortbek, and Kaj Grønbæk. Mobile urban drama — setting the stage with location based technologies. In *ICIDS '08: Proceedings of the 1st Joint International Conference on Interactive Digital Storytelling*, pages 20–31, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-89424-7.

Thomas T Hewett, Ronald Baecker, Stuart Card, Tom Carey, Jean Gasen, Marilyn Mantei, Gary Perlman, Gary Strong, and William Verplank. Curricula for human-computer interaction. http://old.sigchi.org/cdg/webnotes.html, 1996.

ISO. International standard, iso/iec16022international symbology specification, data matrix. ISO standard, 2006.

ISO. International standard, iso/iec18004. ISO International Standard, Jun 2000.

M. Isomursu. Tags and the city. *PsychNology Journal*, 6(2):131–156, 2008.

Minna Isomursu and Mari Ervasti. Touch-based access to mobile internet: User experience findings. *International Journal of Mobile Computer-Human Interaction (IJMHCI)*, 1(4):58–79, 2009.

Minna Isomursu, Pekka Isomursu, and Mervi Komulainen-Horneman. Touch to access the mobile internet. In *OZCHI '08: Proceedings of the 20th Australasian Conference on Computer-Human Interaction*, pages 17–24, New York, NY, USA, 2008. ACM. ISBN 0-9803063-4-5. doi: http://doi.acm.org/10.1145/1517744.1517751.

Woodland Norman J. and Bernard Silver. Classifying apparatus and method, October 1952. URL `http://www.freepatentsonline.com/2612994.html`.

David J. Johnston and Adrian F. Clark. A vision-based location system using fiducials. In *VVG*, pages 159–166, 2003.

Younghee Jung, Per Persson, and Jan Blom. Dede: design and evaluation of a context-enhanced mobile messaging system. In *Proc. of Conf. Human Factors in Computing Systems CHI'05*, pages 351–360, New York, NY, USA, 2005. ACM. ISBN 1-58113-998-5. doi: http://doi.acm.org/10.1145/1054972.1055021.

H. Kato and Mark Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99)*, San Francisco, USA, October 1999.

David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004. ISSN 0920-5691. doi: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94.

Kaj Mäkelä, Sara Belt, Dan Greenblatt, and Jonna Häkkilä. Mobile interaction with visual and rfid tags: a field study on user perceptions. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 991–994, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: http://doi.acm.org/10.1145/1240624.1240774.

Tristan Manco. *Street Logos*. Thames & Hudson, 2004.

Kurt Mehlhorn and Stefan Näher. *LEDA: a platform for combinatorial and geometric computing*. Cambridge University Press, Cambridge, November 1999. ISBN 0-521-56329-1.

M.B. Miles and A.M. Huberman. *Qualitative Data Analysis : An Expanded Sourcebook*. Sage, Thousand Oaks, 2nd edition, 1994.

Natasa Milic-Frayling, Martin Hicks, Rachel Jones, and Jamie Costello. On the design and evaluation of web augmented mobile applications. In *Proc. of Conf. Human computer interaction with mobile devices and services MobileHCI '07*, pages 226–233, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-862-6. doi: http://doi.acm.org/10.1145/1377999.1378011.

MIT Mobile Experience Lab. The electronic lens project. http://mobile.mit.edu/p/elens/index.html (last retrieved Jan 2010), June 2006.

Bill Moggridge. *Designing Interactions*. MIT Press, 2007.

Mathias Mohring, Christian Lessig, and Oliver Bimber. Video see-through ar on consumer cellphones. In *ISMAR '04: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04)*, pages 252–253, Washington, DC, USA, 2004. IEEE Computer Society. ISBN 0-7695-2191-6. doi: http://dx.doi.org/10.1109/ISMAR.2004.63.

Takao Nakamura, Atsushi Katayama, Masashi Yamamuro, and Noboru Sonehara. Fast watermark detection scheme from camera-captured images on mobile phones. *IJPRAI*, 20(4):543–564, 2006.

Valentina Nisi, Ian Oakley, and Mads Haahr. Inner city locative media: Design and experience of a location-aware mobile narrative for the dublin liberties neighborhood. *intelligent agent*, 6(2), 2007.

Nicolas Nova, Fabien Girardin, Gaëlle Molinari, and Pierre Dillenbourg. The underwhelming effects of location-awareness of others on collaboration in a pervasive game. In *Proceeding of the 2006 conference on Cooperative Systems Design*, pages 224–238, Amsterdam, The Netherlands, The Netherlands, 2006. IOS Press. ISBN 1-58603-604-1.

Kenton O'Hara, Tim Kindberg, Maxine Glancy, Luciana Baptista, Byju Sukumaran, Gil Kahana, and Julie Rowbotham. Social practices in location-based collecting. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1225–1234, New York, NY, USA, 2007a. ACM. ISBN 978-1-59593-593-9. doi: http://doi.acm.org/10.1145/1240624.1240810.

Kenton O'Hara, Tim Kindberg, Maxine Glancy, Luciana Baptista, Byju Sukumaran, Gil Kahana, and Julie Rowbotham. Collecting and sharing location-based content on mobile phones in a zoo visitor experience. *Comput. Supported Coop. Work*, 16(1-2):11–44, 2007b. ISSN 0925-9724. doi: http://dx.doi.org/10.1007/s10606-007-9039-2.

Eamonn O'Neill, Peter Thompson, Stavros Garzonis, and Andrew Warr. Reach out and touch: Using nfc and 2d barcodes for service discovery and interaction with mobile devices. In Anthony LaMarca, Marc Langheinrich, and Khai N. Truong, editors, *Pervasive*, volume 4480 of *Lecture Notes in Computer Science*, pages 19–36. Springer, 2007. ISBN 978-3-540-72036-2.

N. Otsu. A threshold selection method from gray levelhistograms. *IEEE Trans. Systems, Man and Cybernetics*, 9:62–66, mar 1979. minimize inter class variance.

M Ozuysal, P Fua, and Lepetit V. Fast keypoint recognition in ten lines of code. In *Proceedings of Conference on Computer Vision and Pattern Recognition.*, 2007.

Tapan S. Parikh, Paul Javid, Sasikumar K., Kaushik Ghosh, and Kentaro Toyama. Mobile phones and paper documents: evaluating a new approach for capturing microfinance data in rural india. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 551–560, New York, NY, USA, 2006. ACM. ISBN 1-59593-372-7. doi: http://doi.acm.org/10.1145/1124772.1124857.

James Patten. Corporate fallout detector. website http://www.jamespatten.com/cfd/, 2005. Last visited 25 March 2010.

Per Persson and Petra Fagerberg. Geonotes: a real-use study of a public location-aware community system. Technical report, SICS, Sweden, 2002. http://www.sics.se/ petra/techReport.pdf (last accessed Nov 2009).

Per Persson, Fredrik Espinoza, and Elenor Cacciatore. Geonotes: social enhancement of physical space. In *Proc. of Conf. Human Factors in Computing Systems CHI '01*, pages 43–44, New York, NY, USA, 2001. ACM. ISBN 1-58113-340-5. doi: http://doi.acm.org/10.1145/634067.634097.

Salil Pradhan, Cyril Brignone, Jun-Hong Cui, Alan McReynolds, and Mark T. Smith. Websigns: Hyperlinking physical locations to the web. *Computer*, 34(8):42–48, 2001. ISSN 0018-9162. doi: http://dx.doi.org/10.1109/2.940012.

Jenny Preece, Yvonne Rogers, and Helen Sharp. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, Inc., New York, NY, USA, 2nd ed. edition, 2007. ISBN 0471492787.

Ali Rahimi. Fast connected components on images. website: http://people.csail.mit.edu/rahimi/connected/, jul 2001. URL `http://people.csail.mit.edu/rahimi/connected/`.

Jun Rekimoto. Navicam: A magnifying glass approach to augmented reality systems. *Teleoperators and Virtual Environ.*, 6(4):399–412, 1997.

Jun Rekimoto and Yuji Ayatsuka. Cybercode: designing augmented reality environments with visual tags. In *DARE '00: Proceedings of DARE 2000 on Designing augmented reality environments*, pages 1–10, New York, NY, USA, 2000a. ACM Press. doi: http://doi.acm.org/10.1145/354666.354667.

Jun Rekimoto and Yuji Ayatsuka. Cybercode: designing augmented reality environments with visual tags. In *DARE '00: Proceedings of DARE 2000 on Designing augmented reality environments*, pages 1–10, New York, NY, USA, 2000b. ACM. doi: http://doi.acm.org/10.1145/354666.354667.

Jun Rekimoto and Katashi Nagao. The world through the computer: computer augmented interaction with real world environments. In *UIST '95: Proceedings of the 8th annual ACM symposium on User interface and software technology*, pages 29–36, New York, NY, USA, 1995. ACM. ISBN 0-89791-709-X. doi: http://doi.acm.org/10.1145/215585.215639.

Michael Rohs. *Linking Physical and Virtual Worlds with Visual Markers and Handheld Devices*. PhD thesis, ETH Zurich, Zurich, Switzerland, August 2005.

Michael Rohs and Antti Oulasvirta. Target acquisition with camera phones when used as magic lenses. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1409–1418, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1. doi: http://doi.acm.org/10.1145/1357054.1357275.

Michael Rohs and Philipp Zweifel. A conceptual framework for camera phone-based interaction techniques. In *Pervasive Computing: Third International Conference, PERVASIVE 2005*, number 3468 in Lecture Notes in Computer Science (LNCS), Munich, Germany, May 2005. Springer-Verlag.

Azriel Rosenfeld. Adjacency in digital pictures. *Information and Control*, 26(1):24–33, September 1974.

Jaakko J. Sauvola and Matti Pietikäinen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, 2000.

R Schacter. An ethnography of iconoclash: An investigation into the production, consumption and destruction of street-art in london. *Journal of material culture*, 13(1)(13):35–61, 2008.

Bill N. Schilit, Norman Adams, Rich Gold, Michael Tso, and Roy Want. The parctab mobile computing system. In *Proc. of Workshop on Workstation Operating Systems*, pages 34–39. IEEE, 1993.

Richard P. Schuler, Nathaniel Laws, Sameer Bajaj, Sukeshini A. Grandhi, and Quentin Jones. Finding your way with campuswiki: a location-aware wiki. In *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*, pages 2639–2644, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-642-4. doi: http://doi.acm.org/10.1145/1240866.1241055.

Semapedia. http://en.semapedia.org/community/learnmore, 2005.

Tara Michelle Rosenberger Shankar. *Speaking on the Record*. PhD thesis, MIT Media Lab, 2004.

Shepard Fairey & The Institute of Contemporary Art / Boston, editor. *Obey: Supply & Demand : The Art of Shepard Fairey*. Gingko Press, 20th anniversary edition edition, 2009.

Ben Shneiderman and Catherine Plaisant. *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Addison Wesley, 4 edition, April 2004. ISBN 0321197860.

Erkki Siira, Tuomo Tuikka, and Vili T´orm´anen. Location-based mobile wiki using nfc tag infrastructure. In *Proceedings of the First International Workshop on Near Field Communication*, 2009.

David Silverman. *Interpreting Qualitative Data*. Sage Publications Ltd, 3rd edition, 2006.

Steven S. Skiena. *The Algorithm Design Manual*. Springer-Verlag, 1998. URL http://www2.toki.or.id/book/AlgDesignManual/INDEX.HTM.

Susan Leigh Star and James R. Griesemer. Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in berkeley's museum of vertebrate zoology, 1907-39. *Social Studies of Science*, 19(3):387–420, 1989.

Bruce Sterling. *Shaping Things (Mediaworks Pamphlets)*. The MIT Press, September 2005. ISBN 0262693267. URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0262693267.

Atau Tanaka and Petra Gemeinboeck. A framework for spatial interaction in locative media. In *NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression*, pages 26–30, Paris, France, France, 2006. IRCAM — Centre Pompidou. ISBN 2-84426-314-3.

G. Thomas, J. Jin, T. Niblett, and C. Urquhart. A versatile camera position measurement system for virtual reality tv production. In *Proc. Internat. Broadcasting Convention (IBC'97)*, pages 284–289, 1997.

Eleanor Toye, Richard Sharp, Anil Madhavapeddy, David Scott, Eben Upton, and Alan Blackwell. Interacting with mobile services: an evaluation of camera-phones and visual tags. *Personal Ubiquitous Comput.*, 11(2):97–106, 2007. ISSN 1617-4909. doi: http://dx.doi.org/10.1007/s00779-006-0064-9.

Yanna Vogiazou, Bas Raijmakers, Erik Geelhoed, Josephine Reid, and Marc Eisenstadt. Design for emergence: experiments with a mixed reality urban playground game. *Personal Ubiquitous Comput.*, 11(1):45–58, 2006. ISSN 1617-4909. doi: http://dx.doi.org/10.1007/s00779-006-0068-5.

Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. Pose tracking from natural features on mobile phones. In *ISMAR '08: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 125–134, Washington, DC, USA, 2008. IEEE Computer Society. ISBN 978-1-4244-2840-3. doi: http://dx.doi.org/10.1109/ISMAR.2008.4637338.

Roy Want, Andy Hopper, Veronica Falc ao, and Jonathan Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10(1):91–102, 1992. ISSN 1046-8188. doi: http://doi.acm.org/10.1145/128756.128759.

Peter Wildbur and Michael Burke. *Information Graphics*. Thames & Hudson, 1998.

Appendices

Appendix A. Marker Design Study Instructions

On the following pages the instructions for the marker design user study (Chapter 4) are reported, as they were presented to participants. The version reported here is for the second experiment, where participants were asked to draw markers that were not only valid but also resilient to a given amount of distortion. Notes indicate the sentences or paragraphs specific to the second experiment.

Subject Number: ......
Date: ... ...
Time: ... ...

d-touch Marker Design Instructions

Introduction

d-touch is a system that allows to create graphic symbols, or *visual markers*, that can be read automatically by a computer or a mobile phone, and at the same time can be visually meaningful for people. This is because the algorithm used in d-touch is quite flexible in reading shapes, as long as a number of rules are respected.

In this experiment we are trying to assess how easy or difficult it is to draw following these rules, and to test a software application, the DTAnalyser, designed to help in this task. Therefore, we ask you to collaboratively try to draw a number of different markers. The reason we ask you to perform this task in collaboration with another person is to facilitate the analysis and facilitate the finding of possible problems in the process or in the application.

In principle the markers can be drawn using any tool you like, however, for the purpose of this experiment we ask you to use a dry-erase white board. A camera connected to a computer will be used to take pictures of your drawings and check them in the DTAnalyser application.

This document will first introduce the rules that have to be followed to draw functional d-touch visual markers, then it will describe the functioning of the DTAnalyser application which should help you in drawing functional markers , and finally it will present you a task to complete. Please read the document carefully and do not hesitate to ask if you have any questions or doubts. However, please ask questions only after having read the entire document.

You have one hour for the entire experiment, this includes reading these instructions and then draw as many valid markers as you can, as detailed later.

Essential Constraints: Nesting of Black and White Regions

For simplicity this experiment we will consider only black and white markers. The essential rule to create a valid d-touch visual marker is in the nesting of black and white *areas* or *regions*. To illustrate what we mean by "nesting" and "regions" let's take into account a few examples. The following is just a black region:

The shape of the region does not matter, so the following 2 are also just black regions:

Next is the first example of *nesting*, 1 black region containing 1 white region:

And now more *nesting*, 1 black region containing 2 white regions (note that it's just 1 black region):

Let's look at more *nesting*, 1 black region containing 1 white region which in turn contains 1 black region:

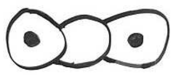
Finally even more *nesting*, a black region containing 2 white regions, the one on the left containing 1 black region and the one on the right containing 2 black regions:

Definition of d-touch Markers

Now that the concept of nesting black and white regions should be clear, let's introduce the basic rule that defines a d-touch marker:

> a valid marker can be composed of a black region containing 3 or more white regions, and the majority (i.e. more than half) of these white regions must contain one or more black regions. This makes exactly 3 levels of nesting – it must be no more and no less. However, there is no limit in the number and shape of the regions.

The following is the "minimal" valid marker:



It is a black region containing 3 white regions (the minimum number of regions allowed); 2 of the white regions contain 1 black region (the majority 2 is larger than half of 3), one contains none.

Let's now look at some other valid markers.

A black region containing 5 white regions, 4 of them containing 2 black regions and 1 of them empty:

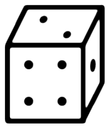
A black region containing 7 white regions, 6 of them containing 1 black region and 1 of them containing 3 black regions:

A black region containing 4 white regions, one white region contains 5 black regions (the "hello" balloon), one white region contains 3 black regions (the face with eyes and mouth), and two white regions contain 1 black region (the two ears).

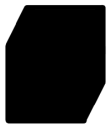
Below are 3 additional examples:

When you draw a marker, you can draw it any order you prefer, starting from a detail and then going out, or starting from the outline and then adding details inside. In fact, you can draw the markers however you like using the tools provided, and to erase and modify them as much as you want.

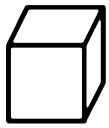
Here follows some other examples, drawn on the computer:



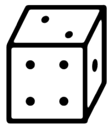
this marker reads as:



a black region



that contains 3 white regions



each of the white regions contains 1, 2 and 4 black regions

this marker reads as:

a black region

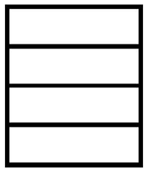
that contains 3 white regions

one of the white regions contains 3 black regions, and the other two white regions contain 4 black regions

---

this marker reads as:

a black region



that contains 4 white regions

each of the white regions contains 5, 9, 14 and 15 black regions

---

this marker reads as:

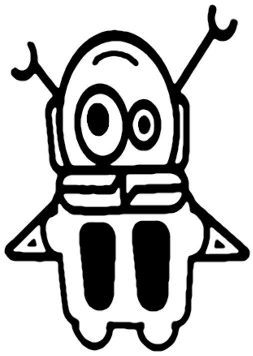
a black region

that contains 7 white regions

one of the white regions contains 0 black regions, four of them contain 1 black region, and the last one contains 3 black regions

---

Two more examples:



Marker ID

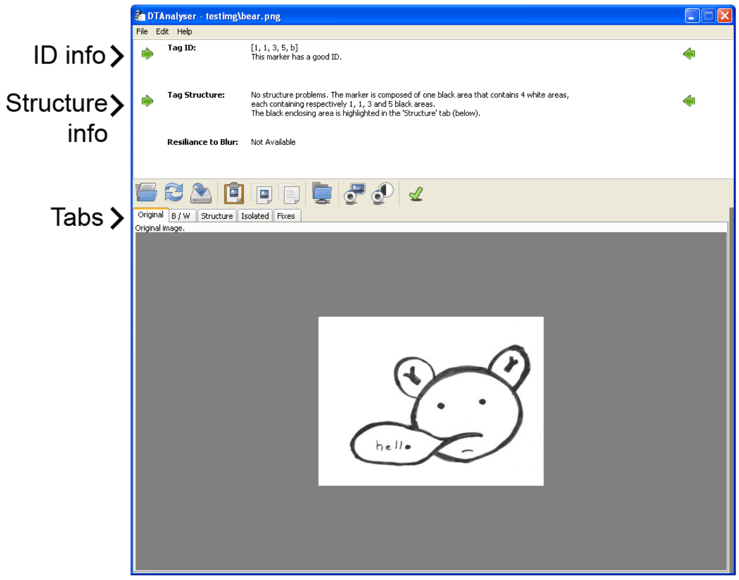
Each valid marker is associated to a numerical identifier, an *ID*, which can be used to link digital information to it. The ID is defined by the structure of each marker, in the sense of how black and white regions are nested. Markers that have the same structure will be associated to the same ID, even if they have different shapes.

Robust Markers

If the rules mentioned and exemplified above are followed, any drawing can be recognised by the d-touch algorithm, at least if a clear, steady picture is taken from a *reasonable* distance. However, often pictures can be blurred, perhaps because shot while moving, and the definition of "reasonable distance" depends on the resolution of the camera used – mobile phone cameras can have very low resolution. d-touch markers can be more or less *robust* to blur and low resolution issues – at a general level, small details are the first to suffer from blur and low resolution, at least compared to more bold parts of the same drawing. The DTAnalyser application, introduced in the following session, provides some information about the *robustness* of the analysed markers.

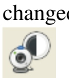
The DTAnalyser

The DTAnalyser application was developed to help drawing valid d-touch markers. Please use it and try to get as much information as possible from it during the experiment.



Please draw on the white board provided, a camera (webcam) will import the images in the DTAnalyser application when you press the following button on the DTAnalyser interface:

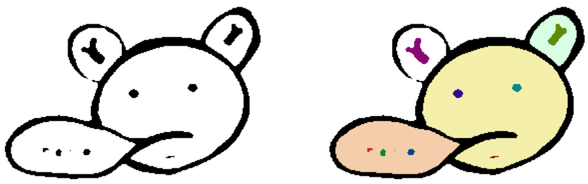
A separate small window (normally on the right of the computer screen) shows what the video feed from the camera. This window can be in "natural" or "high contrast" mode – "high contrast" shows how the marker is "seen" by the recognition algorithm. It is strongly suggested to keep this window always in high contrast mode, the mode can be changed using the following button:



Once a drawing has been imported in DTAnalyser, the application analyses it (please be patient this can take up to some minutes) and displays the results. The first thing to look at is the message under the toolbar, this shows either "Valid marker" and the marker ID or "Not a valid marker". Please ignore the trailing letter in the marker ID (it will normally be 'b' or sometimes 'w'). Under this message one or more other lines of text provides more information from the analysis. Below this text are 3 to 7 small images: you can select each of them by clicking on it and it gets displayed at full size at the bottom of the application. You can enlarge or zoom out by using the following two buttons in the toolbar:

If the drawing is not a valid marker, the DTAnalyser tries to guess which specific aspect of the rules is violated, and this information is displayed in the second line of text (above the row of icons). Examples include regions being nested too deeply, too many white regions being empty or having less than 3 white regions. Please note that this is a guess. If the drawing does not follow at all the d-touch rules, the application will suggest to check the instructions: please do! Sometimes problems can be caused by strokes being too thin or by gaps inside areas which are supposed to be filled , this is particularly common using the white board . The different views offered by the DTAnalyser can help spotting and fixing these problems.

The first view, named "Original" is simply the image acquired by the webcam. Next to it is the "B / W" view which shows how the image is converted to pure black and white. In this conversion sometimes there may be unexpected results, in particular, thin lines or small regions can be read as white rather than black. It is possible to examine the B / W view to try and spot this kind of problems, however the next view "Regions" should make it a lot easier. The "Regions" view shows how the computer divides the drawing into separate regions by colouring them with different colours: white regions are coloured in light colours and black regions in dark ones. Different regions have different colours, so for example if two white regions are separated by a black line they should show in two different colours. However, if the black line is not complete, the supposedly-two regions will in reality be only one region, all filled in the same colour. For example consider the following drawing and its "Regions" decomposition:

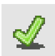


You can notice that both the background and the left ear are white. This means that the two are in fact the same region. A closer inspection (in the application you can zoom in) reveals that the black line enclosing the left ear, towards the bottom near the head, is actually broken. (please note that the same may happen with different colours, not just white)

[The following paragraph was included only for Experiment 2]
If the drawing is a valid marker, the application will determine how easily it can be

read by a low resolution camera (e.g. from a mobile phone). This information is displayed in the text line under the marker ID, and in the 4 views on the right: "Weak", "Distorted", "Broken B/W" and "Broken Regs.". The "Weak" view highlights in red the parts of the marker that are least robust, typically this are small details. If the marker contains a lot of red highlights, it is actually a good sign, as it means that it is uniform. The "Distorted" view shows the amount of distortion (due for example to blur or low resolution scanning) that causes the problems highlighted in the "Weak" view. The "Broken B/W" and "Broken Regs." views have the same function as the "B/W" and "Regions" views described above, except that they are relative to the distorted marker. In fact, these last four views can be used to understand how the distortion modifies the structure of the marker, and therefore how to modify the marker to make it more robust, if necessary. Please note that any marker, for how robust it is, will always break at some low resolution. The goal is normally (but not necessarily always) to produce markers that are good to read from mobile phones.

Once you are happy with your design, please try to register it by pressing the following button in the tool bar:



Please note that this function is available only for valid markers. The application will then check if the ID of the current marker clashes with any existing marker. If there is no conflict a dialogue window will appear asking you to add a brief text description for the marker. Please do so and upload the data. In case of a clash, please try to modify the marker and solve the conflict.

Hints and Common Problems

One common problem in the creation of markers is the occurrence of too many levels of nesting. For example in the following case there is a black region containing 3 white regions, and each of the white regions contain a number of letters (5, 5 and 6 from top to bottom), but the letters 'e', 'o', 'd' and 'a' contain inside them a white region, which is not allowed by the system.

Please note:

- one way to draw robust markers is to draw them small – if they are too small, however, they may not be recognised correctly, you may need to find a reasonable balance;
- you have about one hour for the entire experiment, a time-out counter on the computer will help you keep track of time;
- thin lines/strokes are a common source of problems, please press the pen hard when you draw and use the "Regions" and "B/W" views to detect mistakes;
- when your hands or their shadow are in the view of the camera, or when there are no drawings the system may see a lot of noise – please ignore this, but make sure that your hands are not in the way when you import your drawing in DTAnalyser;
- the system is quite sensitive to the dust created by erasing the marker, please make sure there isn't any;
- even though the system has been tested before the experiment, there may still be bugs – if you notice something unexpected or not working please let us know;
- please talk in English, even with the other subject.

To familiarise yourself with the software, please try to reproduce one of the example markers depicted above. Try to reproduce it exactly as it is and then try to introduce some errors and see how the application reports them. At the beginning please ignore the robustness and resolution issues, just make valid markers. At a later stage, please try to make markers that are "Good for a mobile phone".

As soon as you feel comfortable please move on to the task on the next page. Remember you have just 1 hour for the entire experiment.

Task

In collaboration with the other participant, please try to draw as many markers as you can related to any of the scenarios described below. The goal is to draw each marker so that someone else could guess which of the scenarios it was associated to, and to draw as many valid markers as possible within the allowed time (so to complete each marker as quickly as you can). [The following sentence was included only for Experiment 2] The markers should be “Good for mobile phones” or at least “Not easy for a mobile phone” but not “High resolution only”.

Please remember that the total time for the experiment is about 1 hour. A countdown timer on the computer will help you to keep track of how long you have left. Within the available time you are free to do all the attempts you want, use the DTAnalyser as much or as little as you want and to record the designs that you are happy with. Please name your designs according to the theme (e.g. Music, Pollution, etc..).

Please swap role and swap chairs after drawing each marker, so that each of you will draw half of the markers.

1. **Music**. Imagine that you want to advertise a concert through a d-touch marker, so that people would notice the marker, and by scanning it with their mobile phone they would get information about when and where the concert is.
2. **Animals**. Imagine you want to raise a concern about animals in your neighbourhood. Please design tags that would attract people’s attention on this topic, so that when they scan they with their phones they would be able to get your message and react to it.
3. **Pollution and energy consumption**. Imagine you want to find people in your neighbourhood who are concerned about reducing pollution and energy consumption. Please design tags that would attract people’s attention on this topic, so that when they scan they with their phones they would be able to get your message and react to it.
4. **Children**. Imagine you want to raise a concern about children in your neighbourhood, for example in relation to their playing or their learning. Please design tags that would attract people’s attention on this topic, so that when they scan they with their phones they would be able to get your message and react to it.

Please fill the questionnaire on the following page only at the end of the experiment

Background Information

Gender: ❐ M ❐ F

Age: ____

Mother tongue: _____________

	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
I am familiar with computers	❍	❍	❍	❍	❍
I am familiar with algorithms (I can write them)	❍	❍	❍	❍	❍
I often draw freehand	❍	❍	❍	❍	❍
I can draw well	❍	❍	❍	❍	❍
I am familiar with geometry	❍	❍	❍	❍	❍
I am familiar with mathematical topology	❍	❍	❍	❍	❍
I know what a connected component is (in mathematical terms)	❍	❍	❍	❍	❍

In which field do you currently work?

____

____

What is your background? (e.g. what did you study?)

____

____

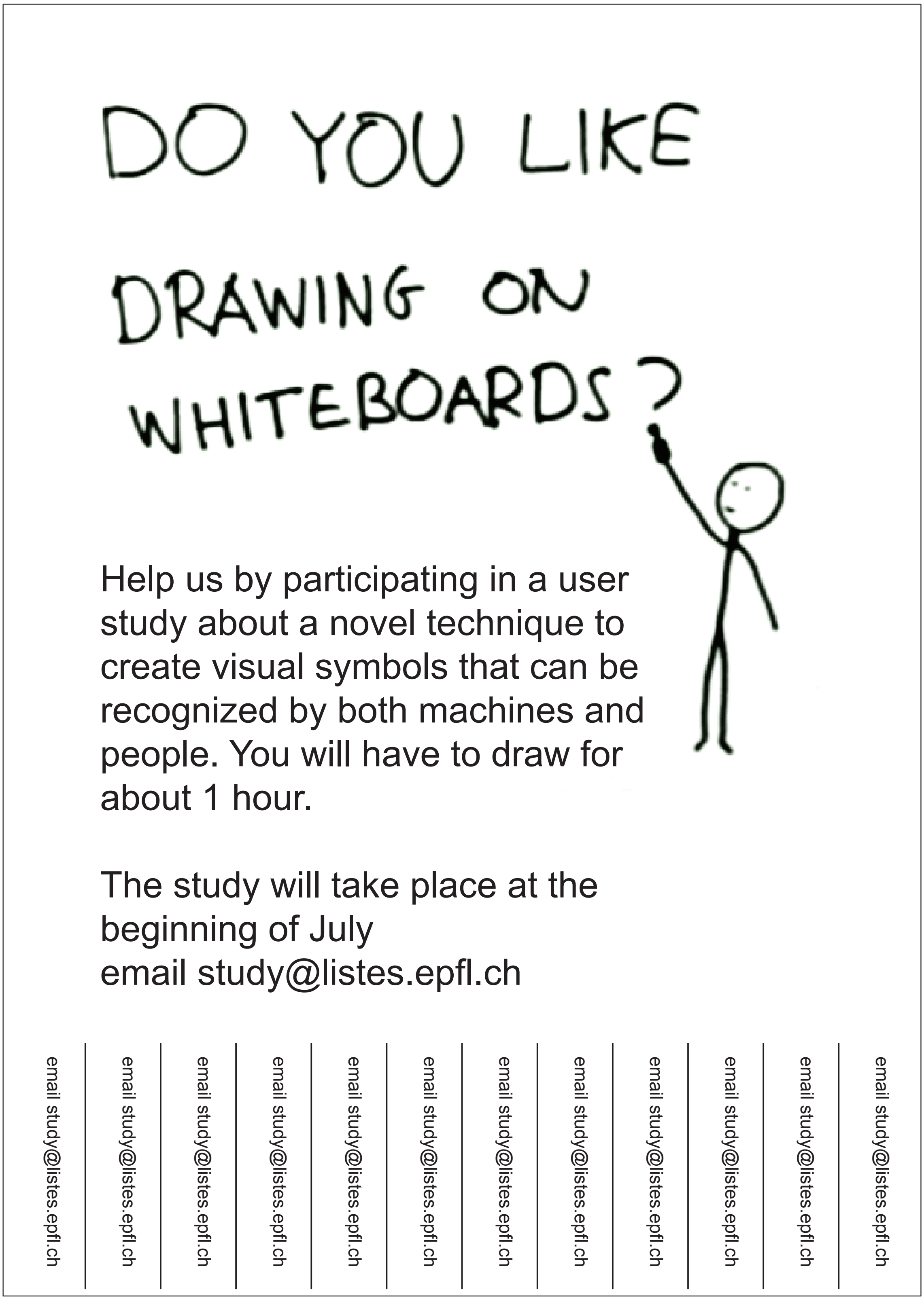
Did you know of the d-touch system before the experiment? What did you know?

____

____

Appendix B. Marker Design Study Advertising Posters

The posters on the following two pages were affixed around EPFL campus to recruit subjects for the marker design user study (Chapter 4).


DO YOU LIKE
DRAWING ON
WHITEBOARDS?
Help us by participating in a user study about a novel technique to create visual symbols that can be recognized by both machines and people. You will have to draw for about 1 hour.
The study will take place at the beginning of July
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch

CAN YOU EXPRESS PEACE & LOVE WITH BARCODES?

Help us by participating in a user study about a novel technique to create visual symbols that can be recognized by both machines and people. You will have to draw for about 1 hour.

The study will take place at the beginning of July
email study@listes.epfl.ch

email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch
email study@listes.epfl.ch

Appendix C. Marker Design Study Advertising Email

The following email was sent to several mailing lists at EPFL to recruit participants for the marker design user study (Chapter 4).

—

Subject: User study on graphic symbols meaningful to machines and people (20CHF)

Can you draw symbols that are meaningful to people as well as to machines?

We are working on a recognition system that works with symbols that can be easily read by a machine (computer or phone) but can also be visually meaningful to people. We are now running a user study to understand how easy or difficult it is for people to draw this type of symbols, and we are looking for subjects.
The experiment lasts for about 1 hour, and it involves drawing in collaboration with another person using a small white board connected to an interactive validation tool. It takes place on EPFL campus and participants will receive CHF 20 for their time.

If you are interested please contact us on study@listes.epfl.ch Thanks!

- -
Enrico Costanza
Assistant-doctorant
Media and Design Laboratory
Ecole Polytechnique Fdrale de Lausanne

http://web.media.mit.edu/ enrico

Appendix D. Marker Design Study Consent Form

On the following pages the consent form for the marker design user study (Chapter 4) is reported, as it was presented to participants.

INFORMED CONSENT

Visual Marker Design
Principal Investigator: Enrico Costanza

You are asked to participate in a research study conducted by Enrico Costanza (MEng), from the Informatique et Communication Department at the Ecole Polytechnique Fédérale de Lausanne (EPFL). You were selected as a possible participant in this study because you expressed interest in participating. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

- **PARTICIPATION AND WITHDRAWAL**

Your participation in this study is completely voluntary. If you choose to be in this study, you may withdraw from it at any time without penalty or consequences of any kind and you may request that any data collected be destroyed. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

- **PURPOSE OF THE STUDY**

This study evaluates a system to recognize visual markers that can be hand-drawn and a software application to help in the design of recognizable markers. The experiments are aimed at measuring the qualities of the system and its interface, NOT the abilities of the subjects.

- **PROCEDURES**

If you volunteer to participate in this study, we would ask you to perform a task using our system in collaboration with another volunteer participant. In particular you will be asked to draw a number of graphic symbols, or “markers”, to visually represent some concepts. The markers should be drawn following a number of rules that allow them to be recognized by our system. A software application should help you to understand whether or not your drawings are readable by our system.

The total estimated time for the completion of this experiment session is less than 1 hour. The entire experiment will be held in the room where you are currently sitting.

- **CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law.

The information collected during the experiment will never be associated with your personal identity. All information will be coded to remove identifying information; the information will be stored on secured EPFL computers. Any data published from this experiment will be done in aggregate form.

The pictures that you will draw may be included as examples in scientific publications and on our website or used in other research projects in which our team is involved. In all cases the pictures will be credited to subjects of the user study without specifying names, unless you explicitly ask us to do so.

During the experiment, audio and video recordings may be made. All transcripts, and questionnaire data will be stored securely. The only individuals with access to this data are the principal investigators of this study. If any content from these recordings is published, it will be in aggregate form.

If you decide you do not want these recordings to be made, or if you decide at a later date that you do not want these recordings to be kept in the form specified above, you may request that we delete these recordings from our dataset.

- **RISKS**

The experiment does not involve any risk other than those related to using standard computer and office equipment.

- **BENEFITS**

You will get to try out new communication technology and you will participate in a fun game.

- **IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please contact:

Enrico Costanza (principal investigator)
office: BC 121
email: enrico.costanza@epfl.ch
phone: 021 693 1299

- **RIGHTS OF RESEARCH SUBJECTS**

You are not waiving any legal claims, rights or remedies because of your participation in this research study.

- **FURTHER DATA USAGE**

If the usage of the data collected will be used for any other purposes than the ones outlined here, the principal investigators will contact you at the email address you provide below.

Your email address: ______________________________________ (optional)

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

____
Name of Subject

____ ______________
Signature of Subject Date

SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

____ ______________
Signature of Investigator Date

PERMISSION FOR VIDEORECORDINGS TO BE USED IN PRESENTATION AND COMMUNICATION MATERIALS RELATED TO EXPERIMENT

Visual Marker Design

During this experiment, you may be videotaped as part of our data-collection procedure. We the experimenters will keep this videotape in a secure place after it has been collected, and only we will have access to it. We do find it useful sometimes to show video of our work and experiments to sponsors of the media lab and to the outside world to communicate our procedures and findings. This form gives you an opportunity to accept or decline the possibility of us using your footage in documentation that may be shown to other people.

This is completely voluntary and you are under no obligation, express or implied, to allow the footage taken during your experimental trial to be shown to people other than the experimenters.

☐ I give permission for video footage taken during my experimental trial to be used in video documentation that may be shown to other people.

☐ I **do not** give permission for video footage taken during my experimental trial to be used in video documentation that may be shown to other people.

Signed: ______________________________________ Date: ___________________

Appendix E. Usability Experiment of Marker-based Interfaces Instructions

On the following pages the instructions for the usability study of marker-based mobile interfaces (Chapter 5) are reported, as they were presented to participants. The version reported here is for the first experiment, the only differences in the second experiment were the description of the interfaces and the data collection form (data was collected through a screen-based form).

Mobile Audio Cards Experiment

We are experimenting with 3 new methods to interact with mobile phones, and in this experiment we are asking you to test them. In all cases you will work with an application to play and record sounds using a mobile phone and a set of physical cards. Please note that the aim of this experiment is to compare the performance of the 3 interfaces, and not your ability.

You will be introduced to each of the 3 different interfaces and you will be given 2 minutes to become familiar with it. After this familiarization phase, you will be asked to complete 20 tasks with the interface. For each task the phone will display what to perform. Each task is always composed of an action ("play", "record" or "append") and an object ("girl", "dog" or "factory").
Please try to complete each task **as quickly and as accurately as possible**. So don't forget to take time to read what's on the screen.

For each task, we will be measuring the accuracy of your selection and the time it takes to perform it. The time is measured starting from the moment that you lift up the phone from the table (using a sensor) to the moment you trigger the action. Please note that you only have one attempt for each task. After each task you will need to place the mobile phone back on the table, in order to proceed to the next task.

Interface A

This interface includes a set of cards representing actions and a set of cards representing objects. To perform a task you need to first point the phone to the card corresponding to the action that you want ("play", "record" or "append") and then point the phone to the card corresponding to the object that you want ("girl", "dog" or "factory").

Interface B

This interface includes a set of cards representing different objects ("girl", "dog" or "factory"), while actions are selected through the arrow keys of the phone. As shown on the phone display, "up" is "record", "left" is "play" and "right" is "append".
To perform a task with this interface, you need to first point the phone to the card corresponding to the object that you want and then press the key corresponding to the action that you want.

Interface C

This interface includes a set of cards representing different objects ("girl", "dog" or "factory"), and actions are selected through the relative position of the phone and the object-card. As shown on the phone display, for "Play" you need to place the symbol on the card in the left of the display, for "Record" you need to place the symbol on the card in the top of the display and for "Append" you need to place the symbol on the card in the right of the display.

Gender: M F

Age: _____

Mother tongue: ____________

What are your current occupation and your background?
(e.g. what are you studying or what did you study)

____

____

____

(Please turn the page)

Which interface did you find easier to use?

A B C

Why?

____

____

Which interface did you find most accurate?

A B C

Why?

____

____

Which interface did you find more enjoyable or fun to use?

A B C

Why?

____

____

Which interface did you prefer overall?

A B C

Why?

____

____

Appendix F. Usability Experiment of Marker-based Interfaces Advertising Email

The following email was sent to several mailing lists at EPFL to recruit participants for the usability study of marker-based mobile interfaces (Chapter 5).

—

Try out some new methods to interact with mobile phones, and have a chance to win CHF 100!

We are running an experiment to evaluate some new playful ways to interact with mobile phones. The experiment lasts 20-30 minutes and it takes place on EPFL campus.

If you are interested please send an email *as soon as possible* to study@listes.epfl.ch

Thank you in advance!

- -
Enrico Costanza
Assistant-doctorant
Media and Design Laboratory
Ecole Polytechnique Fdrale de Lausanne

http://web.media.mit.edu/ enrico

Appendix G. Usability Experiment of Marker-based Interfaces Consent Form

On the following pages the consent form for the usability study of marker-based mobile interfaces (Chapter 5) is reported, as it was presented to participants.

INFORMED CONSENT

Mobile Audio Cards Experiment
Investigators: Enrico Costanza and Francois Suter, EPFL Media and Design Lab

You are asked to participate in a research study conducted by Enrico Costanza (Meng, MS) and Francois Suter, from the Informatique et Communication Department at the Ecole Polytechnique Federeale the Lausanne (EPFL). You were selected as a possible participant in this study because you expressed an interest in participating. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

- **PARTICIPATION AND WITHDRAWAL**

Your participation in this study is completely voluntary. If you choose to be in this study, you may withdraw from it at any time without penalty or consequences of any kind and you may request that any data collected be destroyed. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

- **PURPOSE OF THE STUDY**

This study evaluates a number of different interfaces to interact with mobile phones through printed cards. The experiments are aimed at measuring the qualities of the interfaces, NOT the abilities of the subjects.

- **PROCEDURES**

If you volunteer to participate in this study, we would ask you to repeat a number of tasks, using different variations of a mobile phone interface. In each case the aim is to select one action and one item on the phone interface. More details are provided in a separate document.

The total estimated time for the completion of this experiment session is 20 to 30 minutes. The entire experiment will be held in the room where you currently are.

- **CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law.

The information collected during the experiment will never be associated with your personal identity. All information will be coded to remove identifying information; the information will be securely stored on EPFL computers. Any information published from this experiment will be done in aggregate form.

During the experiment, audio and video recordings may be made. All videotapes, transcripts, and questionnaire data will be stored in a securely locked cabinet. The only individuals with access to this data are the principal investigators of this study. If any content from these recordings is published, it will be in aggregate form.

If you decide you do not want these recordings to be made, or if you decide at a later date that you do not want these recordings to be kept in the form specified above, you may request that we delete these recordings from our data set.

- **RISKS**

The experiment does not involve any risk other than those related to using standard electronic equipment.

- **BENEFITS**

You will get to try out new communication technology and you will participate in a fun game.

- **IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please contact:

Enrico Costanza
office: BC 121
email: enrico.costanza@epfl.ch
phone: 021 693 1299

- **RIGHTS OF RESEARCH SUBJECTS**

You are not waiving any legal claims, rights or remedies because of your participation in this research study.

- **CONFIDENTIALITY**

If the usage of the data collected will be used for any other purposes than the ones outlined here, the principal investigators will contact you at the email address you provide below.

Your email address: ______________________________________ (optional)

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

____
Name of Subject

____ ______________
Signature of Subject Date

SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

____ ______________
Signature of Investigator Date

PERMISSION FOR VIDEORECORDINGS TO BE USED IN PRESENTATION AND COMMUNICATION MATERIALS RELATED TO EXPERIMENT

Mobile Audio Cards Experiment

During this experiment, you may be videotaped as part of our data-collection procedure. We the experimenters will keep this videotape in a secure place after it has been collected, and only we will have access to it. We do find it useful sometimes to show video of our work and experiments to sponsors of the media lab and to the outside world to communicate our procedures and findings. This form gives you an opportunity to accept or decline the possibility of us using your footage in documentation that may be shown to other people.

This is completely voluntary and you are under no obligation, express or implied, to allow the footage taken during your experimental trial to be shown to people other than the experimenters.

☐ I give permission for video footage taken during my experimental trial to be used in video documentation that may be shown to other people.

☐ I **do not** give permission for video footage taken during my experimental trial to be used in video documentation that may be shown to other people.

Signed: ______________________________________ Date: ___________________

Appendix H. Contest Field Trial Advertising Email

The following email was sent to several mailing lists at EPFL to recruit participants for the Contest trial (Chapter 7).

—

Subject: Contest / Experiment on mobile-media stories – 1st prize 300 CHF

Version franaise ci-dessous

Can you think of a story set on or around EPFL campus? Mystery, science fiction, horror, or anything else that comes to your mind!

The EPFL Media and Design Lab is deploying a novel mobile phone technology that opens the doors to a new type of mobile media. You can think of it as multimedia/ meets/ treasure hunt/ meets/ street art. Using our system you can express your story on mobile phones through any combination of video, audio, images or text.

We are organising a creative contest for teams or individuals. Spend about one day after you are done with the exams producing a story set on or around EPFL campus using our system (we will provide all you need to use it).

The contest entries will be reviewed by a panel of mobile media, media art and design experts from Switzerland and abroad and the best entry will be awarded CHF 300. A second prize of CHF 150 will be also assigned.

Interested? Send an email to mobilemedia@listes.epfl.ch to get more information! Register your interest by June 26. The contest will take place in the first half of July.

—

Pouvez-vous imaginer une histoire qui se droule dans les environs du campus de l'EPFL? Mystre, science-fiction, horreur, ou n'importe quoi d'autre qui vous passe par l'esprit!

Le Media and Design Lab de l'EPFL met en place une technologie innovante sur tlphone portable qui ouvre la porte un nouveau type de mdia mobile. C'est une sorte de croisement entre du contenu multimdia, de l'art urbain et une chasse au trsor. Notre systme permet de produire des histoires en combinant de la vido, du son, des images et du texte.

Nous organisons un concours cratif par quipe ou individuel. Passez environ une journe aprs la session d'examens pour raliser une histoire dans les environs de l'EPFL en utlisant notre systme (nous vous fournirons tout le ncessaire pour l'utiliser).
Les histoires ainsi produites seront examines par un panel d'experts en mobile media, media art & design de Suisse et de l'tranger. La meilleure contribution recevra une rcompense de 300 CHF. Un deuxime prix de 150 CHF sera aussi remis.

Intress? Envoyez un email mobilemedia@listes.epfl.ch pour plus d'informations!
Enregistrez-vous avant le 26 Juin. Le concours aura lieu dans la premire moiti du mois de Juillet.

- -

Enrico Costanza
Assistant-doctorant
Media and Design Laboratory
Ecole Polytechnique Fdrale de Lausanne

http://web.media.mit.edu/ enrico

Enrico Costanza

Education, Research and Teaching Experience

July 2006 – January 2010
Swiss Federal Institute of Technology, Media & Design Lab Lausanne, CH
"Assistant-doctorant": Research & Teaching Assistant and PhD Student
PhD research on *designable visual markers* and their application to HCI and Mobile HCI. My duties included writing of grant proposals for public and industrial funding, supervision of student projects and the design, coordination, teaching and grading of a master-level course in computer science on mobile HCI and mobile phone programming (2008, 2009). Academic advisor: Prof Jeffrey Huang.

September 2004 – June 2006
Massachusetts Institute of Technology, Media Lab Cambridge, MA, USA
Research Assistant and MSc in "Media Art and Sciences" GPA: 5.0
Course based on individual research and taught classes. Topics include location based applications and systems, affective computing & bio-signals, sensor technologies for interactive environments. Master thesis on intimate interfaces for mobile human-computer interaction. Duties as RA included the frequent production and presentation of research demonstrators to sponsors and potential sponsors of the lab, and the supervision of undergraduate student projects. Academic advisor: Prof Pattie Maes.

September 2003 – January 2005
MIT Media Lab Europe Dublin, Ireland
Research Assistant
Media Lab Europe was the European Research Partner of MIT Media Lab. I worked in the Liminal Devices research group under the guidance of Prof Rebecca Allen. My focus within the group was on human interaction with mobile devices, with emphasis on unobtrusive interfaces, location based interaction and mixed reality (selected results were published at international conferences). Duties as RA included the frequent production and presentation of research demonstrators to sponsors and potential sponsors of the lab. Working with Prof Allen I also had a chance to be exposed to development and production of art installations and demos.

June – August 2003
The University of York and HP Research Laboratories York U.K.
Research and Development on Audiophoto Desk (consultancy)
I developed the image processing and sound reproduction algorithms for the Audiophoto Desk. It is "an inclusive way of reviewing audiophotos from physical prints. An overhead camera and hidden computer are used to recognize a printed photograph and play its associated sound automatically". Designed by David Frohlich (HP Laboratories), it was implemented in C++ on Linux and works in real time. The prototype was successfully exhibited at the Helen Hamlyn show 2003 at the RCA in London and at the Second International Conference on Appliance Design.

October 2000 – June 2003
The University of York York, U.K.
MEng in "Electronic and Communication Engineering" 1. Class Hons.

The course gives a strong electronics background with an emphasis on application to communication technologies. I chose the Electronics Department at the University of York because a research group in visual information engineering and augmented reality had recently been set up there. This allowed me to broaden the course content, studying visual perception, computer vision, and video production (both from a *technical* and a *creative* point of view) and work on research into augmented reality and tangible user interfaces.
I received the **Texas Instrument Prize for Best MEng Final Project**: "A Tangible User Interface for Display and Manipulation of Multimedia Information".

September 1996 – June 2000
University of Palermo Palermo, Italy
Degree course in Electronic Engineering (spec. Telecommunications)
I transferred to York University having completed half of the course, with average result of 90% (27/30). The subjects covered in depth mathematics (algebra, advanced calculus, numerical methods, statistics, geometry, mathematical physics) physics (mechanics, EM, advanced EM, thermodynamics), chemistry foundations, computer science, signal processing, circuit theory, and solid state devices. The course had a strong theoretical approach.

September 1991 – June 1996
Liceo Classico "G. Garibaldi" Palermo, ITALY
A-level equivalent (cumulative grade: 56/60)
Mathematics, Physics, Inorganic and Organic Chem., Philosophy, Latin, Ancient Greek, Italian, English, History, Art History.

Peer-reviewed Publications

Costanza E., Panchard J., Zufferey G., Nembrini J, Freudiger J., Huang J., Hubaux JP. "SensorTune: a Mobile Auditory Interface for DIY Wireless Sensor Networks." To appear in *Proc. ACM CHI2010*, April 2010, Atlanta, GA, USA.

Costanza E., Huang J. "Designable Visual Markers." Full paper in *Proc. ACM CHI2009*, April 2009, Boston, CA, USA. (acceptance rate: 24.5%) Nominated for Best Paper Award (14% of accepted p.)

Costanza E., Inverso S. A., Allen R., Maes P., "EMG For Subtle, Intimate Interfaces," chapter in Lumsden J. (Ed.), "Handbook of Research on User Interface Design and Evaluation for Mobile Technology", 2007 Idea Group Reference (acceptance rate: n.a.)

Costanza E., Inverso S. A., Allen R., Maes P. " Intimate Interfaces in Action: Assessing Usability and Subtlety of EMG-based Motionless Gestures." Full paper in *Proc. ACM CHI2007*, April 2007, San Jose, CA, USA (acceptance rate: 25%, 4 citations)

Costanza E., Inverso S. A., Pavlov E., Allen R., Maes P., "eye-q: Eyeglass Peripheral Display for Subtle Intimate Notifications." Full paper in *Proc. of MobileHCI 2006*, September 2006, Espoo, Finland. (acceptance rate: 25%. 1 citation)

Costanza E., Inverso S. A., Allen R. "Toward Subtle Intimate Interfaces for Mobile Devices Using an EMG Controller." Full paper in *Proc. ACM CHI2005*, April 2005, Portland, OR, USA (acceptance rate: 25%, 9 citations)

Costanza E., Perdomo A., Inverso S.A., Allen R., "EMG as a Subtle Input Interface for Mobile Computing," in *Proc of MOBILE HCI 04*, Springer 2004 (acceptance rate: n.a., 3 citation)

Costanza, E., Shelley, S.B., Robinson, J., "Introducing Audio d-touch: a Novel Tangible User Interface for Music Composition and Performance," in *Proc 6th int. conf on digital audio effects (DAFx-03)*, London September 2003. (acceptance rate: n.a., 8 citations)

Costanza, E., Shelley, S. B., Robinson, J., "d-touch: a Consumer-Grade Tangible Interface Module and Musical Applications," in *Proceedings of Conference on Human-Computer Interaction (HCI03)*, Bath September 2003. (acceptance rate: n.a., 2 citations)

Costanza, E., Robinson, J., "A Region Adjacency Tree Approach to the Detection and Design of Fiducials," in *Proceedings of Vision, Video and Graphics*, 2003, Bath, UK, July 2003. (acceptance rate: n.a., 8 citations)

Other Publications and Professional Activities

Costanza, E., Kunz, A., Fjeld, M., "Mixed Reality: a Survey ", invited book chapter in "Human machine interaction," LNCS 5440, pp. 47-68, Springer, 2009

Invited lecturer at CMU-Portugal, Madeira for 5-days HCI workshop on mobile applications and service design (http://design.epfl.ch/nf). July 2008.

Short Term Scientific Mission within the COST 287 ConGAS EU project at Universitat Pompeu Fabra, Barcelona, Spain, 28.10.2004 – 8.3.2004

Short Term Scientific Mission within the COST IC0601 Action on Sonic Interaction Design EU project at Holon Institute of Technology, Holon, Israel, 25.10.2009 – 9.11.2009

Reviewer for ACM CHI conference since 2006.

Editorial Board for the Int. J. of Mobile Human Computer Interaction (IJMHCI), May 2008 - Apr. 2010.

Reviewer for IEEE & ACM *International Symposium on Mixed and Augmented Reality* (*ISMAR*) 2009.

Reviewer for Interacting with Computers journal, 2009.

I am member of the IET, IEEE (both since 2000) and ACM (since 2007).

Artistic Collaborations and Interest in Visual Media

From September 2006 composer and cello player Giovanni Sollima has been using my d-touch sequencer, including in 4 public concerts, documentation available on http://d-touch.org/audio/concerts/

In July 2004 I collaborated with composer and cello player Giovanni Sollima on his "Songs from the Divine Comedy" show. The work involved the design and implementation of an interface for Sollima to "play images" through a midi keyboard, allowing him to write the visual sequences using the language that is most natural for him: the standard score notation. The system was used in a number of live performances in Italy in 2004 and 2005.

Experience in photography and darkroom techniques (personal exhibition in Palermo in December 2001, which I also set up and arranged), and in video making; one of my videos entered the final selection in an Italian national contest in June 2000 (all films in this contest had to be exactly 60 seconds long); I have completed video assignments for the courses that I have taken at York. From 2002 I produce videos documenting my research work.

Languages

Italian (native), English (fluent), Spanish (conversational), basic German (ZD), basic French.

Computer Skills

Programming: C and C++ - *including: embedded programming; Symbian OS programming; Bluetooth; sockets; use and creation of a variety of libraries and toolkits for real time image/audio/signal acquisition and processing; libraries for GUI and computer graphics*. Python, desktop and S60 mobile phones. Java - *including J2ME, SWING*. PHP. MySQL. HTML. CSS. XML. PureData. VHDL and Assembly to a basic level.
Operating Systems: Linux desktop & server (LAMP), OS X, Windows.
Engineering Applications: MATLAB (including statistical analysis), Eagle (PCB schematic and layout) MapleV, Mathematica, CVS, Subversion.
Other Applications: MS Office & Open Office, graphics & multimedia authoring applications (PhotoShop, Premiere, Illustrator, Flash, Avid Cinema, wave editors, ...). TikiWiki.

Other Skills and Interests

Clean full driving licence. Biking, Cinema, Concerts, Music, Art Exhibits.