Content and Context-Aware Interfaces for Smarter Media Control

THÈSE N^{O} 5348 (2012)

PRÉSENTÉE LE 22 JUIN 2012

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE DE RÉALITÉ VIRTUELLE
PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Mathieu HOPMANN

acceptée sur proposition du jury:

Prof. A. Schiper, président du jury
Prof. D. Thalmann, Dr F. Vexo, directeurs de thèse
Dr T. Conde, rapporteur
Prof. J. Huang, rapporteur
Prof. C. Joslin, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2012

Abstract

Since the arrival of the computer at home during the '80s, the importance of the digital world in our daily lives has hugely increased. First centralized on a unique computer, the digital world spread over several devices that surround us: smartphones, tablets, cameras, music players, etc. However, despite the increasing number of devices around our digital lives, the way we interact with them has barely evolved. Digital devices are not aware of their surrounding environment, and they often use interfaces and mechanisms which are computer-oriented. The next step in the evolution of our digital life will be centered on a better integration of the digital world with the ecosystem of devices that surrounds it. In this work, we illustrate this upcoming evolution through the development and evaluation of original interfaces dedicated to the control of digital media. Based on our research, we organize these interfaces in three areas of improvement: unification, automation and familiarity.

This thesis is divided in four main chapters. The first chapter presents the evolution of our digital world over the last 30 years. The next three chapters are centered on the three areas of improvement that we defined: unification, automation and familiarity. Unification corresponds to the fusion of digital worlds as a unique and coherent whole that devices should interact with. Automation consists in using information obtained from the user and her environment in order to simplify or automate the interaction with the digital world. Finally, familiarity allows benefiting of users' past experiences in order to create friendly and intuitive interfaces.

Keywords: Human computer interfaces, interaction, digital content, media control, natural interfaces

Résumé

Depuis l'entrée en scène de l'ordinateur auprès du grand public dans les années 80, le monde numérique a pris une place importante dans notre quotidien. D'abord centralisé sur un unique ordinateur, le monde numérique s'est répandu sur un ensemble de périphériques qui nous entourent: téléphones, tablettes, appareils photos, lecteurs musicaux, et bien d'autres encore. Toutefois, malgré l'explosion du nombre d'appareils entourant notre vie numérique, la manière dont nous interagissons avec n'a que peu évolué. Les appareils électroniques n'ont pas conscience de leur environnement proche, et utilisent le plus souvent des interfaces et mécanismes issus directement des ordinateurs. Dans ce contexte, la prochaine évolution de notre vie numérique sera axée sur une meilleure intégration du monde numérique avec l'écosystème qui l'entoure. Dans ce travail, nous illustrons cette évolution via la présentation d'applications originales dédiées au contrôle de nos media numériques. Nous avons classé ces applications en trois axes d'amélioration: **unification**, **automatisation** et **familiarité**.

Cette thèse est divisée en quatre chapitres principaux. Le premier chapitre présente l'évolution de notre monde numérique sur les 30 dernières années. Les trois chapitres suivants se concentrent sur nos trois axes soit l'unification, l'automatisation et la familiarité. L'unification correspond à la fusion des mondes numériques en un ensemble unique et cohérent avec lequel les périphériques interagissent. L'automatisation consiste à tirer parti d'informations concernant l'utilisateur et son environnement dans le but de simplifier ou d'automatiser l'interaction avec le monde numérique. Finalement, la familiarité permet à l'utilisateur de profiter de ses expériences passées afin de rapidement rendre l'interaction plus agréable et intuitive.

Mots-clés: Interfaces homme-machine, interaction, contenu digital, gestion des média, interfaces naturelles

Remerciements

Tout d'abord, j'aimerai remercier Daniel Thalmann et Frédéric Vexo pour m'avoir offert l'opportunité de faire ce doctorat, ainsi que pour leur aide et support. Merci aussi à Mario et à sa patience sans limite durant l'écriture et relecture de cette thèse, ainsi qu'à Ronan et au IIG pour m'avoir accueilli durant les derniers mois. Enfin, je n'oublie pas Josiane qui a été d'une aide précieuse durant toutes les années au VRLab.

Je tiens aussi à remercier l'ensemble de mes colocataires de bureau : Patrick, le plus endurant avec presque 5 ans, Achille, ainsi que Patricia, Jiri, Mario, Anthony et Nicolas à l'incubateur Logitech. Ces années auraient été bien plus fades sans les nombreuses pauses café prises à vos côtés !

Merci aussi à tous ceux qui étaient là le jour au laboratoire mais aussi le soir pour les soirées jeux et soirées tout court d'ailleurs : Renaud, Sylvain, Patrick, Xavier, Alejandra, Benoit, Helena, Quentin, Stéphane, Nicolas Pronost, AJ, Eray.

Un gros merci à Quentin et Marjory, mes coachs brico et cuisine (mais il reste encore du boulot !), à Stéphane et Sylvie, mes partenaires tarot et JDR, et à Nicolas, mon coach Tennis.

Un gros bisou d'amour à Luce pour avoir rendu bien plus jolies ces dernières années de thèse et pour m'avoir supporté durant la rédaction malgré le stress des révisions.

Enfin, merci à ceux qui m'ont encouragé de loin. Ma famille à Bordeaux bien sûr, ainsi que mes amis : Aurélia depuis Bordeaux puis depuis le Québec, Nada depuis Paris, et Damien depuis divers endroits.

Et merci à tous ceux que j'ai oublié et à qui j'offrirai volontiers un verre pour me faire pardonner !

Contents

List of Figures

Chapter 1

Introduction

Contents

Since the arrival of personal computers at home, many faces of our everyday life have changed. Technology has drastically influenced the way we are communicating, consuming, or accessing the information. This revolution has also shaped the way we are consuming multimedia and entertainment. With the explosion of digital formats and the Internet, our “digital life" has hugely gained in importance. Pictures, songs, movies, almost all of our multimedia content could be dematerialized.

Many of the electronic devices dedicated to manage our digital data are now strongly connected to our lifestyle, and are part of how we live. The diversity of devices related to our digital life has also drastically increased: laptops, digital cameras, mp3 players, digital photo frames, smartphones... In a USA household, there is now an average of 20 digital devices.[1] But despite this evolution in terms of hardware, the evolution in terms of interaction has been less pronounced.

The interaction mechanisms to manage our digital content have poorly changed : it is still based on computer-oriented interfaces. These interfaces are often designed to interact with a single device, and are not adapted anymore to the digital ecosystem which is now surrounding our everyday life. Future interfaces should unify this ecosystem, and should diminish the gap between our physical and digital worlds.

[1] http://www.rd.com/family/take-control-of-family-tech-time/

1.1 Motivation

“We live between two realms: our physical environment and cyberspace. Despite our dual citizenship, the absence of seamless couplings between these parallel existences leaves a great divide between the worlds of bits and atoms. At the present, we are torn between these parallel but disjoint spaces.” Hiroshi Ishii used these worlds to introduce his paper Tangible Bits [Ishii 1997]. It perfectly expresses the issue we are exploring in this work: current interfaces represent an obstacle between users and their digital world. They are clumsy windows to the digital universe.

Different reasons could explain that they are still not the perfect bridge between our physical and digital worlds.

First, there is an important lack of inter-device communication. Many of our digital devices are not aware at all of their surroundings. They are just independent devices, even if they are connected to the rest of the world via their Internet capabilities.

Second, current interfaces are often computer-oriented. Despite the hardware evolution, we are still browsing files and folders in graphical interfaces. It is a convenient representation to interact efficiently using a mouse and a keyboard, but not the most natural way to interact with digital multimedia content in other contexts. People that are not confident with computers cannot fully enjoy their digital content on many other devices because of unfamiliar interfaces.

Finally, modern life demands access to large digital data collections. New paradigms are mandatory to navigate efficiently through this content.

The next evolution of our digital world should tackle these issues and be focused on the creation of smarter interfaces that diminish the gap between the user and the digital world. In the next chapters, we will present content and context-aware interfaces for smarter control of our digital media. We classified these interfaces into three areas of improvement. These three areas could be summarized in three words: **unification**, **automation** and **familiarity**.

- **Unification** corresponds to the fusion of all the sets of digital data spread over digital devices into a unique and coherent whole. Then, the interaction should be seamless between the user, the digital world and the ecosystem of digital devices. New interaction paradigms are necessary in order to benefit from this unification.
- **Automation** is a mean to enhance interface’s efficiency by anticipating user’s desires and by adapting the interface to the current environment.

But automation also gives the opportunity to the user to sense the presence of the digital world, to feel that actions performed in one world have repercussions in the other one.

- **Familiarity** is a key notion to propose intuitive interfaces. Intuitive interfaces are nothing more than interfaces which remind us past experiences. Taking advantage of familiarity greatly accelerates user's performance and acceptance.

1.2 Outline

This document is structured as follows.

Chapter 2 gives an overview of how our perception of the digital world has changed over the last 30 years. We describe this change in parallel with the evolution of computer hardware that shaped this digital world, and the lacks in how we are currently interacting with this digital world.

Chapter 2 is followed by the description of our prototypes dedicated to enhance the control of our digital media. We categorized these prototypes in three chapters, one for each area of improvement.

Chapter 3 relates to the unification of the digital world. After related works on this topic, we present tangible drag-and-drop, an adaptation of the well-known graphical interface paradigm to the physical world. We present two applications based on this paradigm, an interface to easily transfer digital pictures and a second one to manage a multi-room audio system.

Chapter 4 is dedicated to user awareness, and how to automate part of the interaction by taking advantage of user's position or actions. We give an overview of ambient intelligence, and describe 2 prototypes: "Sound will follow" - a multi-room audio system which locates the user (via Wi-Fi networks) and plays music in the room where she is situated - and "Natural Activation" - a prototype that replaces the need of an unnatural gesture to trigger a gesture recognition system by an automatic system based on gaze estimation.

Chapter 5 refers to familiarity and to advantages that come with this notion. We describe two prototypes which are dedicated to music browsing: the "virtual shelf" and the "vintage radio" interfaces.

Chapter 6 summarizes obtained results and opens up possibilities of future works.

1.3 Notes for readers

References to scientific articles (paper, journal, proceeding, etc) are cited with the name of the first author and the article's publication year (e.g. [Hopmann 2012]). They are detailed in the bibliography at the end of this dissertation. If you are using the PDF file, you can access directly to the corresponding reference by clicking on it.

References to websites are directly inserted in the text as footnotes. URLs have been checked for validity in January 2012.

CHAPTER 2

Digital Life Evolution

Contents

2.1 The digital world

2.1.1 "The computer moves in"

Before starting with the birth of our digital life, we need to go back to the first steps of computers at home. Indeed, the evolution of our digital world has been mainly driven by the progress in computer hardware.

It would take several years before the introduction of personal computers in the mass market. In 1977 appeared the "Trinity" of personal computers: the PET 2001, from Commodore; the Apple II; and the TRS-80, by Tandy Corporation. Hundreds of thousands of these computers were sold, and each of these machines was declined in several models during the following years. Many companies joined afterwards this market contributing to the fall of the component prices.

In August 1981, IBM joined the market with its IBM PC, and quickly became one of the main challengers, next to the Commodore 64 and the Apple II. Millions of these computers were sold: it was the beginning of the PC era. The impact of the computer on the society was clear, and the Time Magazine decided in January 1983 to name the computer "Machine of the

Year" [Friedrich 1983]. It was the first and only occasion that a nonhuman has won this award so far.

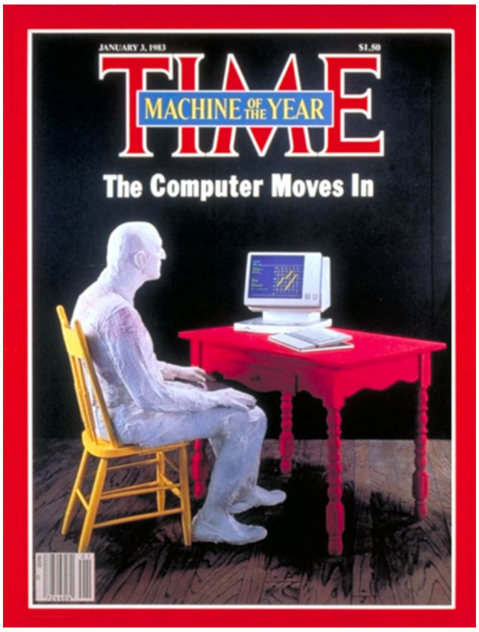


Figure 2.1: Time Magazine - January 1983

We can perceive in the Time's article that excitement and expectations were high: "The information revolution that futurists have long predicted have finally arrived, bringing with it the promise of dramatic changes in the way people live and work, perhaps even in the way they think." But with this excitation came uncertainty. Uncertainty about what this digital revolution would offer. The article ended with this open question: "God knows what will happen now".

At this time, main applications were limited to word processing, spreadsheet and some early attempt at drawing and image processing. Home computer games started to gain commercial success, assisted by the golden age of video arcade games. Computer games largely contributed as a motor for development of better hardware (graphics, sound, interfaces, etc), and played an important role in the success of the most popular computers. They also contributed to let people dream about what would be the future of PCs: Would the future be the immersion in digital worlds ?

2.1.2 Immersion in the digital world

Science fiction fed on early uncertainty and novel capacities of the PC, and created a new genre, the cyberpunk, based on the doubt and fears offered by this new computer era and more generally by the invasion of the information technologies. Novels and movies based on this new kind of science fiction usually took place in a dark future where computers were omnipresent and controlled everything. An interesting point of the cyberpunk movement is that it gives a good idea of how people of the '80s imagined the digital world. The digital world was usually represented as a world completely separated from the real world, accessible by connecting our nervous system to the computer.

William Gibson, considered as the creator of the cyberpunk genre, published in 1984 the novel "Neuromancer", in which he described the "Matrix", a global computer network represented as a virtual reality dataspace. In this world, the technology invades the human life and eventually, the digital world replaces the physical world. Several movies have been inspired by this novel (Matrix, The Thirteenth Floor, etc), and each time the virtual world, or more generally technologies, are represented as negative for human beings. The main reason of this negative picturing is generally the isolating aspect of the technology, which separates humans from each one and from the real world.

Figure 2.2: Cyberpunk presented a dark vision of the technology (Illustration from http://pyxelated.deviantart.com)

The cyberpunk genre inspired and motivated the creation of virtual worlds, leading to the birth of the Virtual Reality (VR) as a research and business field. The systems described in the cyberpunk literature (figure 2.2) could be characterized as full immersive virtual reality systems In the virtual reality field, full immersion devices describe systems that completely isolate the user from the real world, in order to enhance the efficiency of the simulation. Full immersion was one of the trending topics during the early years of virtual reality field. It was thought as the best way to achieve one of the goals of VR: "create imaginary worlds that would be indistinguishable from the real world" [Gutierrez 2008]. VR gains in importance during the '80s in the academic domain. Virtual humans appeared to populate these virtual worlds [Magnenat-Thalmann 1987], and with them came great expectations: being able to soon replace flesh and blood actors by virtual actors. VR popularity was also visible in the industry: Jaron Lanier, pioneer in the VR field, founded VPL Research in 1985, the first company to sell VR goggles and VR gloves.

Unfortunately, at this point, the technology was not able to fulfill all the expectations: the hardware was too expensive and too slow to provide the so waited experience. Soon, technology and science showed their limitations. The lack of progress in the Artificial Intelligence research field greatly limited the progress of virtual humans. VR devices experienced issues such as Cybersickness [LaViola 2000], which particularly affects fully immersive systems. The desire of full immersion in realistic virtual worlds has also been confronted to the uncanny valley hypothesis [Mori 1970]: a simulation that is almost, but not perfectly realistic, will cause a revulsion response to the observer.

Slowly, the immersion trend fizzled out, replaced by none or semi immersive systems. In 1992, a semi immersive environment was presented: the CAVE [Cruz-Neira 1992]. The CAVE had several advantages compared to the previous full immersive technologies: it had better resolution, was less invasive, provided a one-to-many visualization system, and it was not isolating users from the real world. Non immersive systems also gained popularity: they were cheaper, less invasive, and easier to use and deploy.

Finally, if the full immersion principle is the first concept that comes into our mind in order to interact naturally with virtual worlds, it is not the only aspect to considerate. Many works are focused on the concept of presence, which has been defined by Slater et al as "a state of consciousness, the (psychological) sense of being in the virtual environment" [Slater 1995]. This topic is popular as many researchers believe that presence could be a general measure of the efficiency of VR systems [Zeltzer 1992] [Sheridan 1992]. This notion of presence will have an important role in the management of the digital world as we know it today, the management of our digital data.

VR contributed to the creation of our digital world by accelerating advances in visualization, interaction mechanisms and paradigms. However it was another invention which opened the door to a truly digital world.

2.1.3 Birth of the digital data world

The emergence of a "digital world" as we conceive it today started in the middle of the '90s, with the popularization of the World Wide Web. Created at CERN by Tim Berners-Lee [Berners-Lee 1993], this global information medium drastically gains in popularity with the release in 1993 of the Mosaic web browser. Mosaic was the first browser to display text and images in the same window, in a clean and easily understood user interface. It was the first step in the digital world: the possibility of accessing a huge quantity of information from a personal computer. This digital world was not as immersive as expected, but made up of text and images, gathered in web pages. People started to share information, which later turned also into sharing thoughts, feelings, and emotions. Little by little, the means to enrich this content increased, with the use of digital multimedia data for example.

Besides this revolution, another part of our "digital world" emerged: the digitizing of our multimedia content. It first started with the music collection: the Fraunhofer Society released during the second half of 1994 I3enc, the first software MP3 encoder. The small size of MP3 files was adapted to low Internet speed connections, and people started to share music ripped from CD albums through the network. From this point, MP3 files spread rapidly through the Internet, helped by the Winamp audio player, released in 1997, and by the Napster filesharing network in 1999. At this time, the legal offer was nonexistent, and the only way to legally obtain digital music files was to buy and rip CDs. The ease of obtaining MP3 via Napster resulted on lawsuits between them and the major record companies, and the beginning of a long copyright battle for the industry. Indeed, major companies accused file sharing to have a significant impact on the music purchases. Some studies acknowledged this fact [Peitz 2004], some others did not [Oberholzer-Gee 2007], but the Naspter network had got at least one positive aspect: popularizing the usage of digital music and forcing the majors to adapt their offer to the digital world. Sony Music Entertainment was the first major to open an online music store, in April 2000. But the real boom in the digital music industry was caused by the release of the iTunes Music Store in 2003.

After the digitizing of the music, the same pattern happened during the late '90s and the early 2000s with the video industry and the DivX codec. The DivX codec quickly became popular for its capacity to compress video segments of large sizes into smaller bits, without compromising too much the quality of the resulting output. Similarly to the MP3 case, the compression ability of the DivX was adapted to low Internet speed connections. People started to share movies via peer-to-peer networks, and DivX files spread rapidly over the Internet, resulting again on copyright issues and the emergence of legal offers.

The transition to the digital world has been less chaotic for the photography medium. The creation of the JPEG [Wallace 1991] and MPEG [Le Gall 1991] standards was an important step in this transition. These algorithms allow the compression of the data before storing it. Again, the small size of the output files was crucial in the adoption of these formats. JPEG and MPEG were adapted to the Internet speed for displaying images and videos in web pages, but also to store images in the limited memory of digital cameras.

The first completely digital consumer camera, the Dycam Model 1 [Said 1990] (also known as Logitech Photoman) was released in 1990. From this point, the hardware evolution has been incremental: Casio released the first camera with a LCD display in 1995; Kodak the first camera that used CompactFlash in 1996; Nikon released the first digital reflex in 1999. In the early 2000s, the digital camera rapidly took the advantage over the film camera in terms of sales.

To sum up, the late '90s have seen the emergence of the digital world: the digital information medium - the web - and the digital multimedia files. At this time, all this content was often reachable from one point in the house: the desktop computer. Our content was stored on this computer, and the interface to access this world was mainly limited to a keyboard and a mouse. The following decade has seen the spread of these digital files from the desktop computer to many other devices.

2.1.4 Digital data are leaving the computer

After the arrival of this digital world at home, the devices around our digital life increased drastically. The desktop computer did not turn out into a full immersion system, as described in 2.1.2. Instead of being accessible from one point, the digital world moved from the desktop computer to a "wide variety of platforms" [Choi 2003] [Norman 1999].

The computer has changed, it does not only have the shape of a desktop PC, but tends to become smaller and less isolated. Laptop computers gained in popularity with the improvement of the technology: performance, battery, power consumption, LCD, etc. Another important evolution that helped the development of laptops was the new connectivity possibilities: the USB in 1997 made easier the connection with external devices; the Wi-Fi in 1999 eased the connection to the Internet. Little by little, laptops took the place of the desktop PC at home, and the laptop sales outpaced the desktop in 2008. From that time, the laptop shape evolved in turn: the netbook market, which proposed lightweight and inexpensive laptops, exploded in 2008. Two years after, the iPad was released, and the tablet PC market appeared as a challenger for the laptop market. Analysts see the tablet PC market outpacing the netbook market in 2012, and the desktop market in 2013 [1]. It is still unclear whether tablet PCs will replace or complement laptop PCs, but it is clear that the way we access our digital world has dramatically evolved over the last few years.

Along this change, many devices related to our digital life have been widely adopted. Indeed, even if the usage of digital files gained in popularity during the '90s, it was one decade later when devices dedicated to this content appeared. Digital photo frames have been released to enjoy pictures taken by the digital camera. Portable media players and network music players have been released to enjoy the music files (and later video files) previously stuck on the desktop PC. PDA, or Personal Digital Assistant, appeared to access easily the Internet and the web using a pocket device, but have been quickly replaced by smartphones.

Smartphones have been a revolution in the usage of our digital world. If the PC could be considered as a window on our digital world, we could qualify the smartphone as a porthole: it gives an overview of our entire digital world, and gives access to the web and all our multimedia files. Smartphones are not replacing the desktop or laptop computer, but they are a big step in the merging of our digital life and our physical world. Instead of being fully immersed in the digital world using a device available in a specific room, it is the digital world itself that is surrounding us via the multiplication of the connected devices: : we have reached the post-PC era [Hennessy 1999] [Bergman 2000].

In conclusion, the last decade has seen the multiplication of devices around our digital world. These devices have a common point: they are more and more connected to the Internet. This connection is a key point in the evolution of our digital world during the next decade.

[1] http://www.inquisitr.com/76157/tablets-to-overtake-desktop-sales-by-2015-laptops-will-still-reign/

2.1.5 Welcome to the Cloud

The last decade has seen an explosion of devices dedicated to our digital world, the 2010-2020 decade will probably see all these devices connected to the Internet. This evolution has already started: 68% of all consumer electronic devices sold in 2011 are connectable to the Internet [2]. The connection of physical things to the Internet creates new possibilities: remotely controlling the physical devices from a distance; creating devices that communicate with each other; or creating devices that communicate with the outside (such as the smart refrigerator posting an online order when you are low on milk). These possibilities are part of the "Internet of Things" paradigm [Atzori 2010], and will probably changed the way we live in the near future, as the Internet changed it few years ago. It will drastically reinforce the presence of the digital world in our everyday life, through the connection of all our physical devices and home appliances to the Internet.

Another major evolution that comes with the Internet capabilities of electronic devices is cloud computing: providing resources and data access as a service, over the Internet. The challenge of providing computing resources as a service is an old dream [Parkhill 1966]. In the late '90s, Oracle Corporation designed the "network computer", a diskless computer on which data would be stored in Oracle databases. This network computer was a failure: it was too expensive, the software was not mature and the home Internet connection speed was insufficient at this time. But this failure led to the development of cloud computing [Roth 2010], and infrastructures are now ready to make it real.

Over the last few years, cloud computing took an important place in our Internet use. It includes web-based email (Gmail, Hotmail, etc), social networks (Facebook, Twitter, etc), multimedia sharing platforms (Flickr, Youtube, etc) and streaming platforms (Spotify, Google Music, etc). Most of the "Web 2.0" websites are platforms that offer cloud services. There are several advantages to the cloud computing approach: for example, deporting the need of computing resources from devices to the cloud, or centralizing the digital data. It is essential in our ecosystem of devices: without cloud computing, mobile devices such as smartphones and tablets would not be able to propose powerful applications such as Google Maps on Google's Android or Siri on Apple's iOS.

Centralizing the storage of our digital data is also an important contribution of the cloud computing. Indeed, with a computer, a laptop, a MP3

[2] http://http://mobility.cbronline.com/news/connected-devices-to-contribute-70-revenues-to-ce-industry-in-2011-study-190811

player, a smartphone, a tablet, etc, we usually tend to spread our digital files all over these devices. By consequence, instead of having one collection of music, video and photo, we have several little collections spread over all our devices. Using the cloud as a data storage location could be the solution to easily access our complete digital media collections.

But cloud computing will not solve all our digital life issues. Having all our data accessible from everywhere is an important step, but other problems in the interaction process will still be present. One of these issues is clearly the lack of interaction between the devices: all our electronic devices tend to be connected to the Internet, but they are still acting as stand-alone devices, and often ignore surrounding devices. Sharing a music album with a friend, sending a picture on a photo frame, playing a movie on the TV in front of us, selecting the albums of a specific singer, all these actions are still painful and generally not facilitated by current technologies. Instead of accessing to a digital media, we are accessing to this media through a particular device, and many simple actions require many unnatural steps (e.g. the number of actions to move a picture from a camera to a digital photo frame).

Thus, current interfaces fail to provide the perfect bridge between users and their digital content. New paradigms are needed in order to interact with this content in a more natural way, and to minimize the gap between the user and her digital content.

2.2 Towards a smarter control of our digital media

As described above, multimedia content has become more and more dematerialized over last years, and the number of devices around it has drastically increased. Dematerialization brings advantages: you can carry all your music, video and photos in a single device. It could be your MP3 player, your external hard drive, or a place accessible from the web, allowing you to reach your content from everywhere. But dematerialization also brings an inconvenient: the lack of easiness and tangibility in the manipulation of our media. With a physical medium, the possible actions are concrete: selecting an album on the compact disk (CD) shelf, putting the CD in the Hi-Fi system and pressing the play button. Using digital files is not that concrete, especially for people who are not familiar with computers. But if dematerialized content is not concrete and tangible, it doesn't mean that the interaction process should be abstract as well.

Thus, we target the problem of the interface, and how to bring back easiness in the control of digital content. This research aims at designing and implementing new interfaces dedicated to enhance the manipulation of our digital media. We created these new interfaces with two goals: bringing easiness to allow people who are not familiar with computers to interact with digital media, and increasing the efficiency of digital file manipulation. We wanted to explore different ways of merging digital media with the set of devices that surrounds us. People should sense the digital world via their interactions with the technology. This is where current interfaces are failing: they target a good usability in terms of device, and not in terms of global experience. With our prototypes, we wanted to explore how to enhance the usability of the digital world itself.

Usability, as a fundamental determinant of user acceptance, is central in the interaction process. Usability is composed of three parts: effectiveness, efficiency, and satisfaction in a specified context of use [ISO 1998] [Frøkjær 2000]. Effectiveness means that the user is able to achieve the intended task, efficiency is related to the time and effort she used to perform the task, and satisfaction is how pleasant is the system to use [Nielsen 1997].

Efficiency includes task completion time and learning time. In our case, we are not looking for the user to be efficient, but for the system. The goal is not to achieve a task as fast as possible, but to achieve it easily and naturally, and avoid the feeling of being unable to perform an action. As defined by [Hallnäs 2002], “the use of new computational things in everyday life implies a shift from efficient use to meaningful presence”. Some of our devices are a part of our lifestyle, they are not only something we use, but also a “part of who we are, how we live, and how we express ourselves”. This integration of computational devices into our lifestyle is an important step towards an improved digital experience. The more devices and interfaces will be integrated to our lifestyle, the less we will be limited in terms of interaction possibilities and sensing.

This integration is only possible with devices that fit together, which can communicate and interact together. Devices are currently too separated to offer this experience: there is no unified digital world, but an addition of small digital sets which are poorly communicating among them. Consequently, it is hardly possible to provide a seamless experience between the user and their digital data. The Cloud (cf 2.1.5) is a step towards this digital world unification: placing online our digital content in order to access it from everywhere. This unification should come with new interaction paradigms to take advantage of this fusion of our digital worlds. In [Winograd 1997], the author argues that the design role is "the construction of the "interspace" in which people live, rather than an "interface" with which they interact". This "interspace" is populated by "multiple people, workstations, servers, and other devices in a complex web of interactions".

Efficiency could also be greatly improved through the automation of many of our interactions. Our devices are more and more equipped with sensors: we can find accelerometers, cameras and GPS in many smartphones, tablets or MP3 players. At home, new televisions[3] are equipped with cameras for video chat and gesture recognition. New interfaces should take advantage of these sensors in order to increase user's efficiency. For example by automating some tasks thanks to user's location or behavior.

Concerning **effectiveness**, it is important to maintain user's attention. When the user is not able to produce the desired result, she ends up being frustrated and focuses on the device instead of the activity. It is strongly connected to the notion of "break in presence" (BIP). In virtual reality, a participant in a virtual environment receives two streams of sensory data, from the real world and from the virtual world. According to [Slater 2003], "a break in presence (BIP) occurs when the participant stops responding to the virtual stream and instead responds to the real sensory stream". These breaks of presence are generally dysfunctions or unexpected events that occur during the navigation in a virtual environment.

Finally, **user satisfaction** is the attitude of the user towards the interface she is manipulating. User satisfaction is also a critical point of interface's success. Despite the quality in terms efficiency and effectiveness of an interface, the system may not be used because of user's dissatisfaction [Chin 1988]. According to Raskin, "user satisfaction and early productivity are strongly dependent on familiarity" [Raskin 1994]. Indeed, a user will be more receptive to a system which reminds her past experiences compared to a completely novel interface. Familiarity is a key point in this acceptance: a familiar interface benefits from the past experiences of the user, and accelerates user's performance and satisfaction.

Based on these assumptions, we developed several user interfaces to enhance the manipulation of our digital media. We classified them into three areas of improvement that could be summarized in three words: unification, automation, and familiarity.

[3] for example the Samsung UE55ES8000, released in march 2012

- **Unification**, because currently our digital world is not a single set, but the addition of multiple sets spread over all our digital devices. These sets should be unified into a coherent whole, and all our devices should interact with this unique set.
- **Automation**, because many of our actions should be performed transparently in order to spare unnecessary actions to the user. It is an important part in the process of making technology more invisible and more integrated to our home. This axis could be related to ubiquitous computing. In the reference article of this domain, Mark Weiser described the computer of the 21st century as elements of hardware and software "so ubiquitous that no one will notice their presence" [Weiser 1991].
- **Familiarity**, because it is a natural process to quickly integer new things in our lifestyle. In [Van de Walle 2003], familiarity is described as very important for HCI, but also neglected. Researchers have come to the conclusion that familiarity is related to understanding, and to "doing without thinking".

In the next chapters, we will detail these three axes, and we will present concrete examples of prototype implementations.

Chapter 3

Unification

Contents

Many interaction paradigms have been developed to interact with digital files. These techniques are adapted to graphical interfaces, and are efficient to manage digital files. But almost all these paradigms are dedicated to computer graphical user interfaces (GUI), using a keyboard and/or a mouse. The user has to be in front of a computer or screen to efficiently interact with digital media.

We are reaching the post-PC era, and this usage is no longer adapted : our digital environment has changed, the desktop PC is not the center of our digital life anymore [Rekimoto 2003b]. This new era implies that "accessing information no longer means sitting in front of a computer screen, but carrying out this process at any time in any place through the use of a wide variety of platforms" [Choi 2003][Norman 1999]. New interaction paradigms have to support this evolution, and have to take advantage of this variety of platforms.

Creating new interfaces that ease the interaction between all the digital devices will help users to conceive the digital world as a seamless experience, and not as digital files spread among different platforms.

3.1 Towards a unified digital world

At the end of the '90s, digital content at home was limited and mainly stored in one place: the desktop PC. But with the explosion of digital formats described earlier in 2.1, our "digital life" has hugely gained in importance. Pictures, songs, movies, almost all of our multimedia content is dematerialized. The diversity of devices to manage this digital life has also drastically increased: laptops, digital cameras, MP3 players, digital photo frames, Smartphones ...

A problem resulting from this device multiplication is the content management. It is difficult to manage digital information on a single computer: organizing files or emails for example is a very demanding task [Barreau 1995][Whittaker 1996]. This issue is exacerbated with our variety of devices [Dearman 2008]: having many devices often results on having different content on each device, and synchronizing everything is a baffling problem. Thus, our digital world is not a single set, but the addition of multiple sets spread over our digital devices. It results in several issues in terms of usability: digital collections are not synchronized, files are present on some devices but not others, etc. Unifying these sets in order to obtain a coherent whole, with all our devices interacting with this set, would be an important improvement on how we interact with our digital media.

The cloud (cf 2.1.5) is an step towards a solution. With the multiplication of connected devices, the Internet becomes more and more the "aggregator" of our digital data. We store our emails on a webmail provider, we share our pictures on social networks, and listen to music via music recommender systems or more recently store and sync through the Cloud. In terms of storage, the cloud is a good way to obtain a unified cyberspace and to avoid the loss and dispersion of digital files. But it is not sufficient to obtain the coherent whole we are looking for.

If the cloud is the solution to gather digital files in a unified cyberspace, we need to perform the same work in the physical world. Even if many of our digital devices are connected to the Internet, interaction between these devices is limited or inexistent: they are isolated. Transferring files between different digital devices is often painful, therefore digital content is often confined on the device we are currently interacting with. New interfaces are needed to address this problem, and to make our digital ecosystem interact as a single space.

3.2 Related work

In our ubiquitous computing environment, sharing digital content between several devices is often painful, not to say impossible.

The main problem addressed so far is operating multiple devices. Many remote controls have been commercialized in order to interact with several devices[1][2], and the idea of the universal remote control has more than 20 years [Rumbolt 1988]. However, the goal of these universal remote controllers remains basic: to replace several remotes by a single one. There is no possibility of interaction between the different paired devices, the goal is only to limit the number of remote control or the number of steps you need to perform an action.

In the academic field, several studies have been tackling issues of connecting and interacting with multiple devices. The majority of these researches are mainly dedicated to meetings and collaborative environments. Indeed, it is a common situation where people need to easily share information between several platforms, such as whiteboards, table displays or wall displays.

Rekimoto et al. [Rekimoto 1997] proposed with Pick-and-Drop one of the first systems to manage digital objects in a multiple-computer environment. They tackled the problem by using a pen: inspired by the drag-and-drop paradigm from graphical user interfaces, the user had to pick up a computer object by tapping it on the screen, and then she could release it by tapping the destination screen. They conclude by explaining the importance of balancing "the virtuality and physicalness of the target area" in the user interface design. A similar idea to Pick-and-Drop is described in [Zigelbaum 2008] with Slurp, an eyedropper with haptic and visual feedback which extracted digital media from physical objects in order to inject it into other devices.

Another concept dedicated to collaborative environment is presented in Augmented Surfaces [Rekimoto 1999]. Inspired by how easy it is to share and reorganize physical documents, they wanted to make it easy for digital files as well. Meeting participants had to connect their portable computers to the computerized table and wall, and then take advantage of an augmented environment composed of laptops, a table display and a wall display. Users just had to use normal mouse operations to move files among different displays.

In a different context, Rekimoto proposed SyncTap, an easy way to create device associations [Rekimoto 2003a]. The concept is to simplify the connection between two network devices by simultaneously pressing the SyncTap

[1] http://www.logitech.com/index.cfm/remotes/universal_remotes
[2] http://www.remotecontrol.philips.com/

button of each device. They illustrated this concept via several examples, e.g. to transfer pictures from a digital camera to a PC: when the user pressed the SyncTap button of both device (it could be the shutter button of the camera and the escape button on the keyboard for example), then a wireless communication was established and a window appeared on the PC in order to drag pictures from the camera to the computer.

A comparison between these techniques and other techniques dedicated to multiple device interaction was performed in [Nacenta 2005]. After the evaluation of the different techniques, they summarized the results via several lessons learned. For example, they advise designers to consider Radar Views in spatially-aware environments. A radar view is a world in miniature [Stoakley 1995] where the surrounding environment is represented as a map, and manipulations within the map are translated into actions in the physical environment. In other environment, they consider that the Pick-and-Drop technique is probably the best option in order to interact with multiple devices, as long as they are within hand's reach. They insisted on the importance to provide feedback locally, and also to avoid discontinuities in feedback.

Recently, Lee et al. [Lee 2008][Lee 2009] suggested to replace the cumbersome process of connecting and interacting with multiple devices by using simple selecting and pointing hand gestures. They implemented the interface in a meeting room populated by a tabletop, wall displays, laptops and mobile devices.

In [Dickie 2006], authors directly used where the user is looking at for switching input between multiple devices. They evaluated their system, LookPoint, in the context of a typing task: they automatically routed the keyboard input on the computer connected to the display the user is currently looking at.

At home, the concept of enhancing our digital life using multi-device interaction paradigms has also been addressed in research works.

In [Seifried 2009a][Seifried 2009b], the CRISTAL system simplifies "the control of our digital devices in and around the living room". In these papers, they described a system based on an interactive multi-touch surface and a camera. Users can interact with the different devices in the living room by manipulating their representation on the screen. It is an interesting solution for mixing digital files and physical devices on a single surface, but cumbersome and invasive: having a live camera in the living room may be unacceptable for some users.

In this chapter, we will describe tangible drag-and-drop, a spatial adaptation of the drag-and-drop technique used in graphical user interfaces. Tangible drag-and-drop is a multiple-device user interface, designed to bring easiness and efficiency in the manipulation of digital multimedia content. Unlike the Pick-and-Drop technique described earlier, we are using a remote control as input device, and actions could be performed at a distance. Thus, it is more adapted to situations where devices are not within hand's reach, for example in the living room, when the user is sitting on her couch. We will now describe a concrete scenario involving the transfer of digital pictures and how the drag-and-drop paradigm could be used to perform that.

3.3 Tangible drag-and-drop

3.3.1 Concept

Mouse devices, which have accompanied computers for more than 20 years, are still the main device to select, move and interact with digital content. The reason is probably because it combines efficiency, easiness, and low physical effort. Standard interaction tasks offered by the mouse are selection, pointing, drag-and-drop, copy/cut and paste... These tasks are now massively adopted by computer users, and we use them everyday in a natural way. A question we asked ourselves was: among these interaction tasks, which are the ones more closely related to everyday life gestures.

Pointing is one of them. Every child points her finger to pick up something, and everyday we are pointing devices with remote controls to interact with them. It is natural, but also limited: pointing allows to select, but generally not to perform an action. On the other hand, copy and paste allows to transfer text, files or objects from a source to a destination, but is hardly comparable to an everyday life gesture: we are still not able to clone things right away ! On the contrary, drag-and-drop is an action performed continually in everyday life: when we take an object to put it in a different location, it is kind of a drag-and-drop action.

In human-computer interaction, drag-and-drop is the action of clicking on a virtual object, and dragging it to a different location. In graphical user interfaces (GUI), it is intended to drag virtual objects from a location to another in a natural way. However, drag-and-drop is not limited to the action of moving files to a new location. In GUI, you can drag a window border to increase its size and drop it to validate your choice. You can drag a file on a program icon to execute an action on this file. You can drag a picture from a web browser to the desktop in order to create a file which contains this picture. Besides being a fast interaction task to adopt, drag-and-drop is flexible and allows for diverse functionality.

Our idea is to take advantage of this efficient and easy-to-learn interaction task and adapt it to the physical world. Using a “drag-and-drop remote control”, we want to be able to select digital content, drag it physically with the remote and drop it on another physical device.

In this chapter, we present an interaction technique, called tangible drag-and-drop, and two prototypes that have been developed in order to evaluate our concept. The first one is intended to ease the interaction with digital pictures, whereas the second one aims at managing a multi-room audio system. We describe for both of them the implementation and the feedback we got from users.

3.3.2 Prototype #1: The Picture Manager

3.3.2.1 Concept

We present here an everyday life situation, how we perform it now and how we propose to perform it using the drag-and-drop paradigm.

Scenario

You are looking at your last holiday digital photos on your TV screen, comfortably seated in your sofa, and want to select this beautiful picture to add it on your photo frame.

Solution using current hardware

After having plugged the laptop on your TV screen with a VGA/DVI/HDMI cable, you look at your pictures using the classic keyboard/mouse combo or a more multimedia oriented device. Then, there is this picture, and you want to add it on your photo frame. To perform that, you have to take the photo frame memory card, plug it in your laptop, close the slideshow mode, find the picture in your hard drive, and copy it on the memory card.

Solution using tangible drag-and-drop

You point your TV screen with your remote control and click on the drag-and-drop button, move your arms to point to the picture frame with the remote and release the button. Then the picture is automatically transferred in your photo frame memory and appears on its screen.

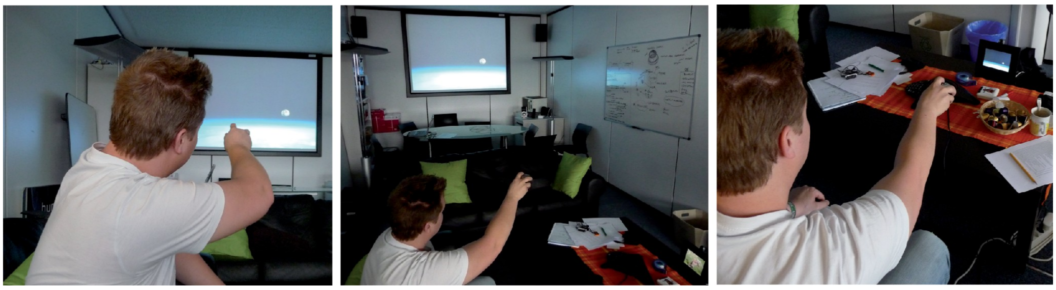


Figure 3.1: Drag-and-dropping pictures from the TV to the photo frame. The user presses the mouse button with the mouse facing the television, and releases it when the mouse is facing the photo frame: the picture is then automatically transferred.

3.3.2.2 Implementation

We describe here the hardware and software needed to build our proof of concept.

Hardware description

To provide our tangible drag-and-drop experience, we need:

- A big screen to display photos as a slideshow (the TV screen)
- A digital photo frame with network capabilities to send pictures on it using our interface
- A remote control to select the picture on the TV and to perform the drag-and-drop action

To display the pictures, we have chosen a laptop connected to a video projector. The video projector allows to comfortably display the photos on a big screen, and the laptop gives us the freedom to run our application.

The digital photo frame is the Parrot photo viewer[3]. This photo frame is wireless oriented and allows to send pictures on it using the Bluetooth technology. Designed to send pictures from mobile phones, we are using this ability to send pictures from our application.

Finally, the remote control was the most problematic hardware to choose. It has to manage the photos (at least allowing to go to the next picture on the TV), but most of all it has to manage the drag-and-drop action. We had two

[3] http://www.parrot.com/

possibilities at this point: we could use a complex remote control with all the hardware embedded in it (gyroscopic motion capture system, or a webcam for example) to detect where the user is pointing; or we could use a simple device and perform the pointing direction detection by an external hardware. We chose the second approach, in order to have a smaller remote control: we took a computer mouse. The mouse buttons are working as follows: left button to go to the previous image on the TV screen, right button to go to the next image, and middle button to perform the drag-and-drop action. The user has to push the button in front of the TV screen to select the picture, and release the drag-and-drop button in front of the photo frame to launch the transfer.

One problem is still remaining: how to detect where the mouse is pointing? To perform that, we have decided to use webcams (one under the TV, the other on the photo frame) and we fixed a LED under the mouse. Each webcam is connected to a computer which processes the video signal. The idea is to detect if the LED is visible to know if a device is pointed or not.

Figure 3.2 summarizes our installation in a schema.

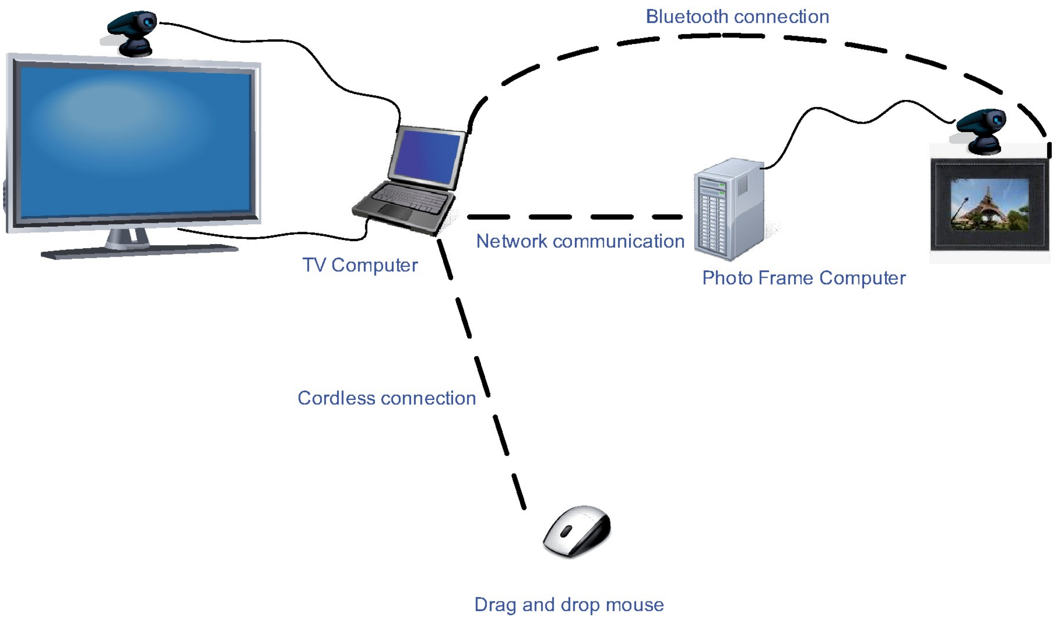


Figure 3.2: Installation of the first drag-and-drop prototype

Software description

Our software is composed of a server and two clients which communicate using TCP.

The client has only one purpose: analyzing the video stream to detect if the remote control is pointing in our direction or not, and sending the information to the server. One client is based on the TV screen computer, the other on the photo frame computer.

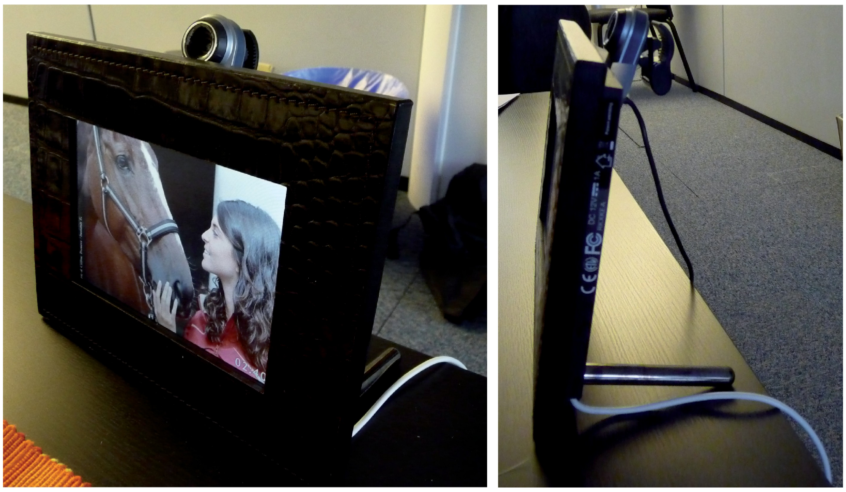


Figure 3.3: The photo frame with the webcam stuck on it. The video signal given by the webcam is analyzed in order to detect if the remote control is pointing it or not

The server is the part of the application that manages the slideshow, the remote control (the mouse), and the drag-and-drop part. For the slideshow, the server just has to display pictures in full screen, go to the previous picture on a left click, and to the next picture on a right click. For the drag-and-drop action, the server is continuously informed by the two clients whether the TV or the photo frame is pointed or not by the remote control. When the user clicks on the drag-and-drop button, the server checks whether the TV is pointed or not. If yes, the current picture is memorized and we are considered in drag mode. When the user releases the button, it checks if the photo frame is pointed. If yes, and if we are in drag mode, the memorized picture is transferred to the photo frame. In other cases, nothing happens.

The detection of the remote control is done by the two clients. Each client reads the webcam video stream and analyzes it to detect whether the remote LED is in range or not. To perform that, we have used a home-made framework previously designed for a multi-touch table. The purpose was the same: detecting fingers enlightened by LEDs, and returning their positions. The only difference is that we are directly detecting one LED, and we don't need any information about the LED's position . OpenCV[4] (Open Source Computer Vision) could have been an alternative way to obtain a similar result.

Figure 3.4: The mouse with its LED

The picture transfer is launched when the drag-and-drop process has been successful, as mentioned in the general description. To perform the Bluetooth transfer, we have chosen the easiest way. Because we are running a Linux-based operating system, we take advantage of the available applications. bluetooth-sendto is one of these applications, specially designed to transfer files over Bluetooth. It only takes two parameters: the MAC address of the destination device (the photo frame), and the path of the file to send, which is the path of the photo currently displayed on the TV. An advantage of using bluetooth-sendto is that the application opens a progress bar window to inform the user about the transfer position. It gives a feedback which could be interesting for the user after the drag-and-drop action (for example if the picture is heavy and doesn't appear immediately on the photo frame screen).

3.3.2.3 Experiment

In order to evaluate our proof of concept, we asked twelve users to test our application and rate it. For the test assessment, we designed a SUMI[5]-like

[4]http://sourceforge.net/projects/opencvlibrary/
[5]http://sumi.ucc.ie/

questionnaire to measure the software quality from the user's point of view. It consists in a list of 35 simple questions. The user only has to reply "yes", "no", or "I don't know". The criteria evaluated by these questions are mental and physical effort, efficiency and how much our interface is natural.

Regarding the test, we asked users to follow the scenario described in section 3.3.2.1. They had to look at the pictures on the TV screen, navigate using the remote control, and select their three favorite photos. We explained the purpose of the prototype (adapting the drag-and-drop paradigm to the physical world). Except that, we did not give any additional instructions about how to use the drag-and-drop remote.

After the tests, users' feedback was really enthusiastic. All the users answered positively to the questions relative to the usability and the interest using our application. They all have thought that using our tangible drag-and-drop allows to be more efficient than using existing solution, with an intuitive manner. Physical and mental effort has been qualified as normal and not excessive.

Concerning the easiness of using our proof of concept, all the users took less than one minute to perform the first drag-and-drop, without giving any instruction about how to use it. We can deduce that our application requires a low learning curve in order to be fully usable. If this result is really satisfying about the easiness and intuitive aspect of our concept, we have to take account that users were people from our lab and school. They use computers almost every day and are already familiar with the classical drag-and-drop concept. The good point is that the transfer from the graphical interface to the physical world seemed very natural for people who are familiar with computers.

Concerning the inconvenient aspects of our application, some users experienced problems with the LED detection. During our tests, four people missed one time the drag-and-drop process because they were not precise enough. The field of view of the webcam is indeed intentionally limited to avoid the false positive (when the picture transfer is started whereas the user did not point the photo frame or TV). Thus, users have to precisely point a device when they want to select it. If missing the drag-and-drop process seems frustrating, precision is inevitable to detect which device is pointed. Adapting precisely the field of view of each webcam (maximizing detections and at the same time avoiding false positive) is a possibility to solve this issue.

Another inconvenient pointed by the users is the lack of output given by the application. For example, if you point the TV with your remote and click on the drag-and-drop button, nothing indicates that the TV has been successfully selected. Thus, when nothing happens after a bad selection, you have no feedback to help you understand why the drag-and-drop process failed: TV was not selected, photo frame was not pointed during the drop, the application was not working, etc. Different solutions could be implemented to solve this issue. A simple sound feedback for example, to indicate that the user successfully selected the devices. Another solution could be to give a visual feedback: when you select the picture on the TV, the picture could move according to your movement, and disappear when you leave the visual field of the webcam TV.

Finally, one user mentioned the impossibility of cancelling the drag-and-drop action (during the transfer of the picture).

This first prototype of our tangible drag-and-drop gave interesting results, and we decided to conduct a second study of the drag-and-drop remote control. For this second prototype, we based our study on a different medium: digital music.

3.3.3 Prototype #2: The Multi-Room Audio Manager

3.3.3.1 Concept

The dematerialization of the music content brings advantages: you can carry all your music collection on a single device, which could be your mp3 player or your cell phone for example. Digital music allows you to have access to your entire music collection everywhere. Paradoxically, playing music at home has not changed that much. People who want to play their digital music are often using computers to do that, and so they are limited to one room.

Different commercial audio systems are now widely available to enjoy a digital music collection in different rooms. For example, Philips with the Streamium wireless music center[6] and Logitech with the Squeezebox audio system[7]. Technically, these solutions are fast and easy to deploy, they can play a lot of different digital formats. On the other hand, navigating in these devices is not so comfortable. You have to navigate through a classical linear menu to select artists, albums or playlists, using the buttons on the device or a basic remote control.

Our idea is to use the tangible drag-and-drop technique in order to control a multi-room audio system. Our system is based on the Logitech Squeezebox server for the multi-room audio system, and on the Nintendo Wii Remote (also nicknamed Wiimote) for the drag-and-drop remote control. We want to fully control the audio player next to us, but also the players situated in all

[6] www.streamium.com

[7] http://www.logitechsqueezebox.com/

the other rooms of the house, while sitting in our living room. Thus, we need to be able to:

Browse and select content The music album collection has to be accessible and browsable. We need to be able to select an album, a web radio or a playlist with our remote control, and to create playlists with it.

Play selected content After selecting a music content, we could choose to play it on the device in the same room, but we have also the possibility to play it on a further device.

Control the devices Our remote control has to support all the basic features of a sound system, such as play, pause, next or previous song, etc.

Transfer content The devices have to communicate with each other. For example, we want to be able to transfer the playlist currently played on a device to another device.

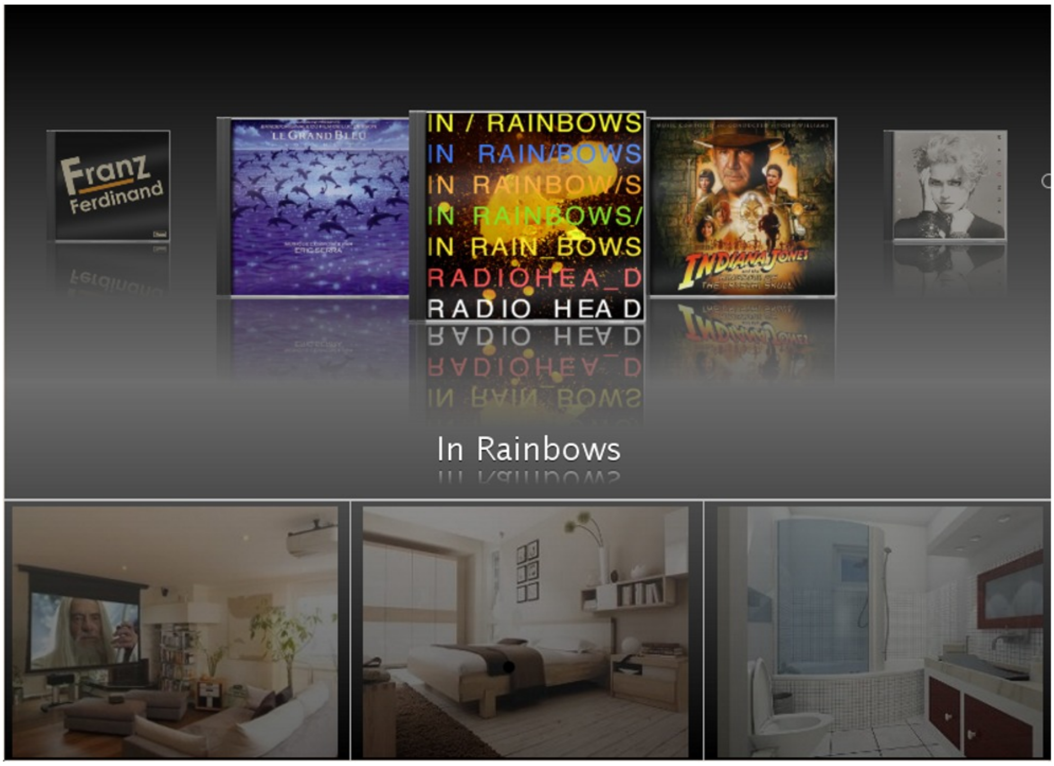


Figure 3.5: Graphical user interface of the second prototype

A possible use case could be the following: You are in your living room, sitting in your sofa, and wish to listen to some music. Your entire music collection is displayed on your TV. You select the album you want to play, and you drag-and-drop it from the television to the music player. After some tracks, you want to go to the bathroom to take a shower. Thus, you target your music player, and drag-and-drop its content to the representation of the bathroom displayed on the screen. The album (or the playlist if you choose several albums) are now playing in the bathroom.

3.3.3.2 Implementation

This second tangible drag-and-drop prototype is not based on the first prototype. For this experiment, we have selected products which are already available on the market, and we have designed the software to make them communicate with each other.

This approach has the following advantages: First, we did not have to develop a complete multi-room audio system. As explained before, commercial solutions already exist to perform that and they work well. Secondly, users are already familiar with these devices, they wont be facing an unknown environment.

On the other hand, we need the devices to be fully customizable. We choose the Squeezebox of Logitech to be the multi-room audio system, and the Nintendo's Wiimote[8] to be our tangible drag-and-drop remote control. Both devices can be controlled and customized with open source software libraries.

Audio system

Our audio installation is composed of a computer running the Squeezebox Server, and three Squeezebox Radios.

The server provides the access to the music library, and can control the entire system. From the server, we can create a playlist, choose an album/radio/playlist and on which player we want to listen to it, or control the players (play, pause, stop, increase volume, etc.). This is exactly what we want. To communicate with the server, the Squeezebox Server provides a command-line interface via TCP/IP. Commands and queries could be sent using TCP/IP in order to fully control the server. This way, we are able to ask programmatically the album list, the covers, and interact with all the players.

[8] http://www.nintendo.com/wii/what/controllers

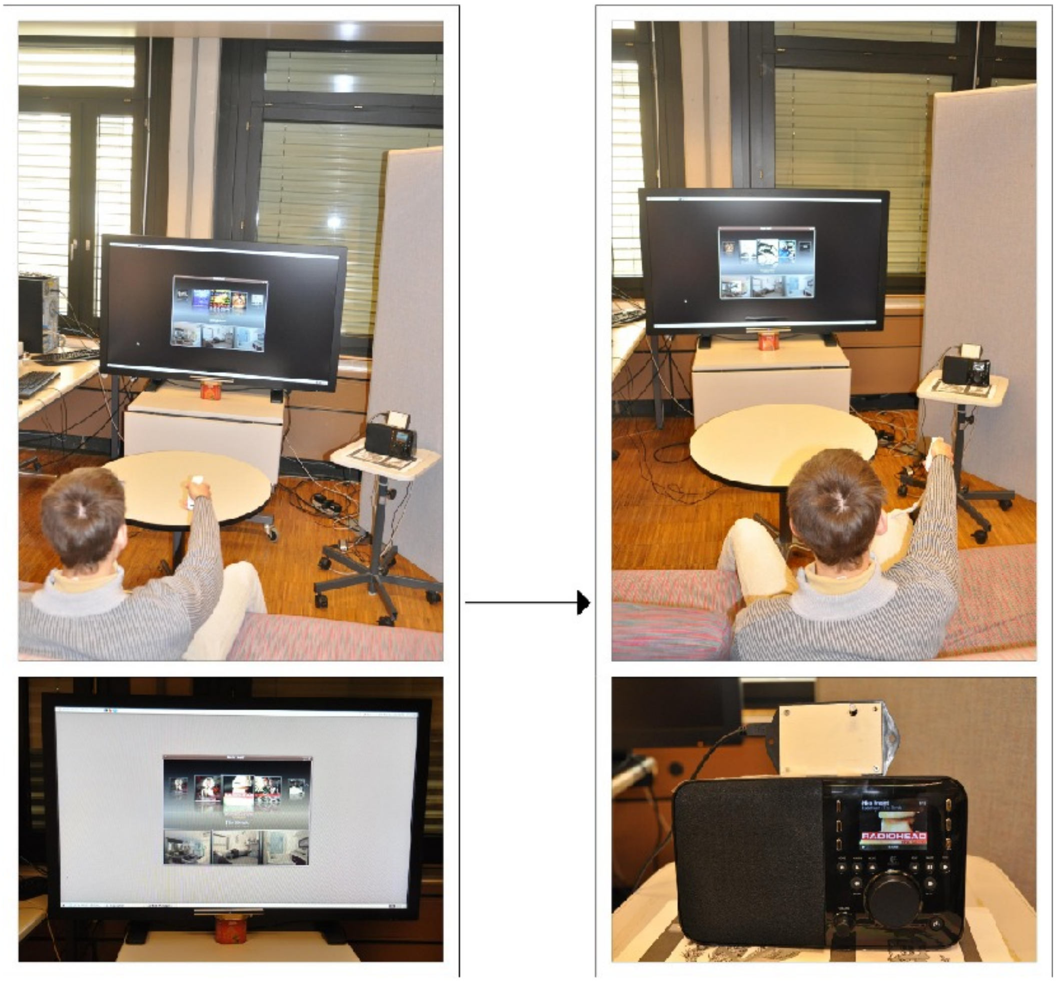


Figure 3.6: Concept and hardware of the second prototype

We choose to use javaSlimServer[9], a Java API used to control the Squeezebox Server via the command-line interface, to easily communicate with the server.

Remote control

As tangible drag-and-drop remote control, we choose this time the Nintendo Wii Remote. The reasons are the following: Firstly, the Wiimote integrates a 1024 x 768 100Hz infrared camera. This is a good solution to detect which device the user is currently pointing at. It favorably replaces the multiple webcams of our first prototype. Secondly, the Wiimote has some output possibilities: it can rumble for a few milliseconds or seconds, there is a speaker so it can play sounds, and there is four LEDs. We take advantage of these Wiimote capacities to give a feedback when the pointed device has changed. Now, when the user moves the remote control in order to point another device, the Wiimote rumbles 100 milliseconds to indicate that the new device has been detected. Finally, another advantage of the Wiimote is that several API already exist to access to the Wiimote abilities. As we are using Java to control the Squeezebox Server, we choose a Java library to communicate with the Wiimote: motej[10].

A challenge of this second prototype was to detect where the user is pointing the remote control. Our first idea was to put a different number of infrared LED on the devices. The Wiimote camera can detect four different LEDs at the same time, so we wanted to take advantage of this ability. For example, putting one LED on the television, two on the first Squeezebox Radio, etc., and then identifying which device is targeted depending on the number of LEDs detected. But this solution did not work. At a distance of two meters and more, the Wiimote mixes the different LEDs into a single one. In order to avoid this problem, the different LEDs have to be separated by at least 20 cm. (which is actually the size of the Wii™Sensor Bar). It is acceptable for one or two LEDs, but too invasive for three or four.

Our second idea was to put a single infrared LED on each device, and make each LED blink at a different rate. This was the way we implemented the system. In order to make our LED blink, we use Arduino Duemilanove boards[11]. Arduino is a platform intended to easily create electronic prototypes. We just have to plug our infrared LED on the board and program the

[9] http://code.google.com/p/javaslimserver/
[10] http://motej.sourceforge.net/
[11] http://arduino.cc/

blink rate. We equipped each Squeezebox Radio with an Arduino board and a LED blinking at a particular rate.

The television does not have a blinking infrared LED. Indeed, the user needs to know where she is pointing on the graphical user interface, and we do not want to have a blinking pointer on the screen. For this reason, the screen is equipped with a standard Wii Sensor Bar. This way, there is no device identification problem (the sound systems have LEDs with different blink rates while the screen has a standard infrared LED) and our application can precisely detect where the user is pointing the remote on the screen.

Graphical user interface

The graphical user interface is displayed on the television. This is the central part of our application, as it allows to control all the devices.

Concerning the visualization of the media, we choose to display the albums using the cover flow interface. Cover flow is an interface created by Andrew Coulter Enright [Enright 2004], but is mainly known for its integration in iTunes 7.0. We select this interface because it is adapted to our prototype environment (television + Wiimote), and also because it is near to the physical representation of a music album. In order to go to the next or previous album, the user just has to target the television and press the left and right button of the Wiimote.

The first version of the main screen of the graphical user interface was limited to the cover flow. The idea was to display the sound devices only when the drag-and-drop button was pressed. The user had to point at the television, select the album through the cover flow interface, press the drag-and-drop button on the Wiimote (at this time, the sound devices appear on the visual interface) and release it when she points at the desire device on the screen. But after some early tests, we noticed that this interface had a major drawback: it did not allow to directly interact with the distant devices (for example, to stop a device that you forget to turn off).

The second version tackles this drawback. We add the devices directly on the main screen of the graphical interface, under the cover flow. In order to represent the sound devices on the screen, we choose to display a representation of the room where they are located. Indeed, it is easier to identify the destination of the selection (bathroom, living room, bedroom) rather than three identical pictures of the Squeezebox Radio with different names. So, when the user wants to interact with the device of a distant room, she just has to target the room on the T.V. and press the corresponding Wiimote button.

Controlling the tangible drag-and-drop audio system

To give an idea of how to control the system, the following list partially presents a possible action sequence of the application.

Browse the album collection Point at the screen and press Wiimote's left or right button

Play an album Use the drag-and-drop button (Wiimote's trigger) to transfer the album from the collection to the physical device or its representation on the screen

Change volume Point at a playing device (physical or on the screen) and press the plus or minus button of the Wiimote

Transfer content from a player to another Point at the playing device, press the drag-and-drop button and release it on the destination

Stop playing/clear playlist Point at the playing device, press the drag-and-drop button and release it on the album collection

3.3.3.3 Experiment

In order to get a feedback and evaluate our second prototype, fifteen participants (5 women, 10 men) aged between 25 and 45 (mean 32.5) participated in our experiment. Two-thirds are working in the computer science field, and the others frequently use computers at work or at home (all of them were aware of the drag-and-drop concept in graphical user interfaces). Six of the participants do not use frequently the Wiimote, but they all had at least a little experience with it.

After explaining the tangible drag-and-drop concept and how to control the system, the user had some minutes to play with it and be familiar with the prototype. After this time, we asked them to perform a list of actions:

- Select an album in the GUI and play it on the player in this room (i.e. the living room)
- Change the volume of the device
- Transfer the content from the current player to the bedroom player

- Transfer the content from the bedroom to the living room

During the test, we observed where the user was pointing during the drag-and-drop operations (on the device or on the representation of the device on the screen), and after we asked them to fill up a form. The form was divided in three parts: one about the prototype, one about the tangible drag-and-drop concept in general, and one with questions. The first and second part was questions with answers from 1 to 5, 1 for “not at all” and 5 for “definitely yes”.

About the prototype questions, it was to ensure that the quality of the experience was high enough to not interfere with the opinion about tangible drag-and-drop. Indeed, if a person did not like the experience, we want to know if the reason was the prototype or the concept. We asked if the remote control and the graphical user interface were easy to use, reactive and precise enough. All the answers are between 4 to 5, which is very good, except for the pointing precision of the Wiimote at the screen (some 3 and one 2, but an average of 4). Thus, the experience was good enough to fairly evaluate the tangible drag-and-drop concept.

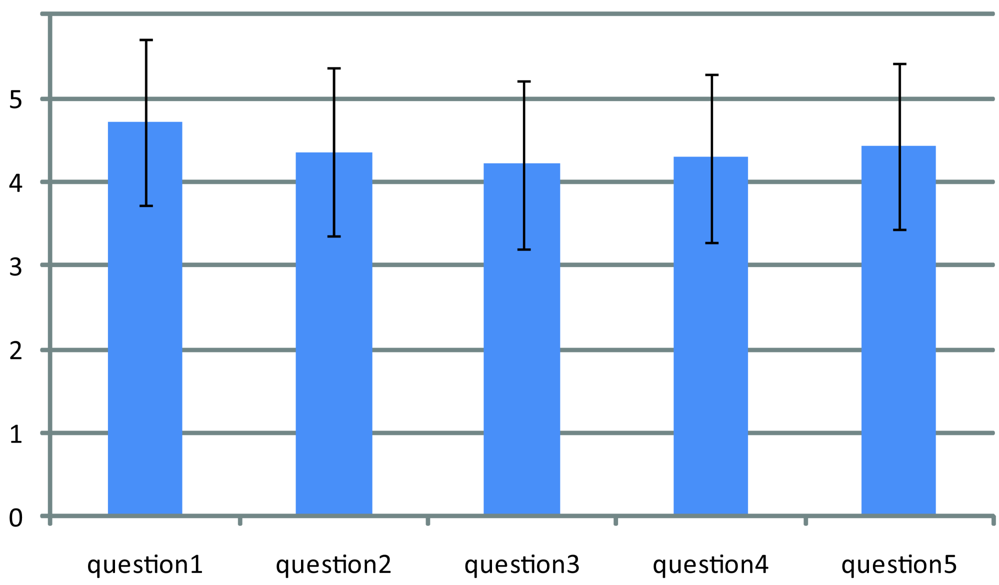


Figure 3.7: User answers for the five questions about the concept. Columns represent the means of the 15 answers, the Y error bars are the standard deviations

The user evaluation of tangible drag-and-drop was encouraging as well. We asked five questions (check-boxes from 1 to 5) to users about the concept.

The questions were, in this order:

1. Is this concept an interesting idea?
2. Are you convinced by this concept?
3. Is this concept natural?
4. Is this concept efficient?
5. Is this concept not demanding/tiring?

Results are presented in figure 3.7. As we can see, users were satisfied about all these points of the application.

Paradoxically, users pointed at the physical device only 57% of the time. This value moderates the natural aspect of the concept. After informally asking some questions to the users, several reasons can explain this result.

Firstly, it is important to consider the difficulty of fully breaking the border between digital and physical worlds. Indeed, all the participants were familiar with graphical user interface, and kept their focus on this interface in order to perform all possible actions.

Secondly, the feedback of the music device does not require to look at it. In the first prototype, users looked at the photo frame to check if the picture transfer were done. In this audio prototype, users can "hear" if she successfully transfer the album, and for this reason can stay focused on the graphical interface.

Finally, the Fitt's law [MacKenzie 1992] index of performance is inferior for tangible drag-and-drop than for the graphical user interface. Indeed, this value directly depends on the distance from the starting point to the center of the target, and the distance from the album selection interface to the digital representation of the device was smaller.

Despite that, it is important to take into account that the graphical interface was really uncluttered, with only three rooms, one device per room, and few options. With five devices or more on the GUI, or more than one device per room, the graphical interface becomes more complicated and the tangible drag-and-drop gains in interest. Some preliminary tests with two squeezebox in the same room are going in favor of this assumption.

In order to better assess the usability and applicability of the drag-and-drop interaction technique we performed a third experiment focused on evaluating the two variations of the technique: on-screen and spatial. Next subsection presents our results.

3.3.4 Evaluation of the tangible drag-and-drop technique

With this evaluation, we wanted to compare separately what we call *on-screen* drag-and-drop technique -inside a graphical interface, and the *spatial* tangible drag-and-drop technique.

We created two applications out of prototype #2 (cf section 3.3.3). In both applications, the concept was to play music, while sitting in a couch in the living room in front of the TV. The first application implemented an *on-screen* drag-and-drop interface (GUI-based). We kept the coverflow in order to display the albums, but we displayed only a picture of a Squeezebox device instead of the pictures of the rooms. If the user wanted to play an album, she had to drag-and-drop it from the coverflow to the device icon, and all the interaction regarding to the device (volume up/down, previous/next track, etc) was performed by pointing at this icon. The second interface was based on the tangible -*spatial*- drag-and-drop technique. The graphical interface on the TV was limited to the coverflow, and nothing else. Compared to prototype #2, we removed all the pictures of the rooms. If the user wanted to play an album, she had to drag-and-drop it from the TV to the device using the remote control.

For this study, we asked 10 participants (four women, six men) to test our two interfaces, and to evaluate it with an AttrakDiff [Hassenzahl 2003] questionnaire. All the participants were different from the previous tests. After a short training session, we asked them to play at least four albums with each system. After the test, we asked them their opinion concerning the two interfaces, and then to fill in the questionnaire. Figure 3.8 presents the results of the questionnaire.

From the AttrakDiff questionnaire, we can see that the tangible drag-and-drop technique performs better than the *on-screen* drag-and-drop technique. The difference in terms of pragmatic quality is not statistically significant, but the difference in terms of hedonic quality is important. The tangible drag-and-drop technique is perceived as much more human, innovative and attractive, but less predictable. The two interfaces are perceived as equally efficient, which explained the close result concerning the pragmatic quality.

During this evaluation, we learned that both interfaces are not mutually exclusive. Even if the interaction task - drag-and-drop - is the same, the two techniques are quite different.

Tangible drag-and-drop was appreciated because of its spatial aspect: you don't need to be precise, you don't even need to look at the device when you

Figure 3.8: Results from the AttrakDiff evaluation for the technique evaluation

transfer an album on it. It is a natural way to interact when involved devices are within the field of view.

The graphical interface is more demanding in terms of precision: you have to be focused on the screen to interact with the system. The *on-screen* drag-and-drop was appreciated because of its iconic representation. You can interact with devices which are not in the field of view, like we did in the section 3.3.3. Another advantage is that you can create symbolic icons: for example, one user suggested implementing a Wikipedia icon that could be used to display the information of the album dropped on it. This kind of interaction is not possible using tangible drag-and-drop, as Wikipedia is not a physical interface.

To conclude this evaluation, we can say that the two interfaces are more complementary than in competition: one is using an iconic representation of the devices, the other one is more suitable for actions to be performed within a well-delimited area.

3.4 Discussion

Breaking the border between the digital and the physical world is difficult, especially for computer users who are used to stay focused on the graphical user interface in order to manipulate their digital content. Physical devices have to manifest their presence. In the context of our prototypes, this means: letting the user know that the tangible drag-and-drop action is possible when a particular content is selected. For example, a solution could be to illuminate the devices when the user is pressing the button. This way, it will attract her attention and the user will know that the drag-and-drop is possible on these devices.

Another solution could be to add additional information directly on the remote control itself. Indeed, as users of the first prototype were complaining about the lack of feedback, we tried to add a picoprojector on the remote control. The idea is to display the selected content during the drag-and-drop action. There are two benefits of such an approach: besides providing visual feedback, it also shows digital media moving in the physical world. It is an interesting solution to "bridge" the digital and physical world. Figure 3.9 is a photo of this prototype.

The action performed after a tangible drag-and-drop movement has to be clear and consistent. For example, if a user wants to visualize photos located on a digital camera, she will drag-and-drop pictures from the camera to the television. But what will happen to the pictures located on the digital camera? Will they be erased or just transferred? Generally, it is better to divide the action as much as possible. If the user wants to erase the pictures, she will perform another action (for example, drag-and-drop the pictures from the digital camera to a trash can). Another solution to inform the user about what will happen after a tangible drag-and-drop action could be to use a remote control equipped with a screen.

Tangible -spatial- drag-and-drop is interesting when all the devices are in range. It is interesting for example in the living room where there is often a large number of devices (television, VCR, media center, etc.). To transfer content physically from one room to another, a different paradigm such as on-screen drag-and-drop could fit better. Finally, one interesting point we have learned is that the on-screen drag-and-drop and the spatial drag-and-drop are not mutually exclusive, and could be used together to create a good user experience.

Concerning the unification of our digital world, it should be eased by the use of the Cloud and the popularization of the Internet connection over all

Figure 3.9: Drag-and-drop remote control with pico projector

our different devices. But this unification will only be possible through the creation and adoption of open standards in order to help computing devices to communicate between them. The Digital Living Network Alliance[12] (DLNA) is a step towards this unification. Established by Sony in 2003, DLNA is responsible of the definition of "interoperability guidelines" and is gathering more than 200 companies. But despite the 440 million of DLNA devices installed in user's homes in 2011[13], additional work is still necessary to enhance the way we share media from device to device. Moreover, standard organizations (ISO, IEEE-SA, ITU-T,...) should involve the HCI community in the creation of new standards [Lund 2012].

3.5 Conclusion

We detailed in this chapter the need of unification of our digital world. This unification should bring new possibilities of interaction, more consistent and natural. Our spatial technique tangible drag-and-drop - a graphical user interface paradigm adapted to the physical world - is an illustration of these possibilities. This technique brings efficiency in the manipulation of digital files in our ubiquitous environment populated by more and more computing devices. But this technique will only be possible with a central access on our digital data and with the creation of standards that should make the communication from device to device much easier.

[12] http://www.dlna.org/
[13] http://www.eeherald.com/section/news/nws201101245.html

Chapter 4

Automation

Contents

In the previous chapter, we detailed the importance of the digital world unification: when the user is interacting with a device, she wants to interact with her digital world in a global way, and not only with the files stored in one specific device. This way, interactions from the user in the physical world have repercussions in the digital world. Conversely, our digital devices should not have to wait for user's input and should anticipate user's desires without conscious mediation [Zelkha 1998]. To estimate what the user wants, devices need to know who the user is, where the user is and what is she doing. Digital devices should sense their surroundings in order to provide this information and to react according to it.

In this chapter, we illustrate the integration of automation in the interaction with digital content via the presentation of two prototypes. "Sound will follow" is an example of the integration of user awareness - location awareness in this case - in an application dedicated to music listening. The second prototype - natural activation of gesture recognition systems - illustrates a second aspect of user awareness. It infers where the user is looking at in order to activate or deactivate the recognition system.

4.1 Taking advantage of automation

A domain dedicated to the integration of automation at home is Ambient Intelligence. The concept of Ambient Intelligence [Aarts 2002][Ducatel 2001] emerged at the end of the '90s. Firstly developed by Philips[1], this concept involves the presence of surrounding electronic devices in everyday life which react to the user behavior and habits in the most natural way. One of the most innovative concept in ambient intelligence is user involvement: how to bring ambient intelligence to people and their homes. It is not simply to bring new devices and sensors into their homes, but to develop new paradigms and strong metaphors that will allow natural interactions within ambient intelligent environments. In [Augusto 2007], the author describes an Ambient Intelligent system as "a digital environment that proactively, but sensibly, supports people in their daily lives".

To refine the notion of Ambient Intelligence, Aarts and Marzano [Aarts 2003][Aarts 2004] define the key technology features as:

- Embedded - devices are integrated into the environment
- Context aware - the system senses users and the current context
- Personalized - the system can change depending on user needs
- Adaptive - the system adapts itself in response to the user
- Anticipatory - the system anticipates user's desires

One of the characteristics highlighted by Ambient Intelligence is that human-centric computer interfaces should be context aware and natural [Cook 2009].

About the context, it can be any information that can be used to characterize the situation of a person, place, or object [Abowd 1999]. Taking advantage of context is an entire topic, and hundreds of applications and dedicated frameworks have been developed in order to take advantage of this information [yi Hong 2009]. In our case, we are using context information in order to increase the usability of our digital devices at home, and to increase the presence of the digital world. Indeed, making our devices and environment reacting to our moves and actions is an opportunity to increase this feeling: our digital world can be responsive to our physical actions.

To illustrate this aspect, we designed a prototype that reacts to user's location at home. Based on a signal map of surrounding Wi-Fi networks, the

[1] http://www.research.philips.com/technologies/ambintel.html

system detects user's current room and adapts where the music should be played.

4.2 Sound will follow

Music has become more and more dematerialized over last years due to the development of the Internet and the MP3 format. Thanks to digital music devices, the music follows the user wherever they are : they can access to their entire musical content in the bus, in the street, or in her car by plugging her MP3 player on the head unit. At home, sound systems in different rooms are often completely independent. Each system needs its own remote control, and each system could only play the music directly available on it. The music collection is often split between CDs, PCs and MP3 players in different rooms.

With the development of the Internet, Wi-Fi networks at home have become widely available. If the user can access to streaming music over the Internet, why can't they access to their entire music collection everywhere in their house ?

In this section we propose a multi-room sound system which takes advantage of indoor remote control localization in order to provide smart "sound follow" experience. The location of the user is delivered by the analysis of Wi-Fi signal strength values and a graph describing the environment topology. The music collection is centralized on one device, the main sound player, and is completely accessible from everywhere in the house thanks to the remote control.

In the next section, we present a short overview of the existing multi-room audio systems and existing indoor location systems. We continue with a description of our prototype, our experiments, and the validation of our proof of concept.

4.2.1 Related work

Several multi-room music systems are available on the market, such as the Philips Streamium wireless music center[2] or the Logitech SqueezeBox[3]. These systems are often divided in several components : the main unit with the music collection, and smaller units connected wirelessly to the main one and able to play all the music contained on it using the UPnP (Universal Plug and Play) network protocol. This way, the user could have access to her entire musical content wherever she is.

[2] http://www.streamium.com/

[3] http://www.logitech.com/en-ch/speakers-audio/wireless-music-systems

Figure 4.1: Example of multi-room sound system: the Logitech Squeezebox

These multi-room music systems provide a good integration of digital music at home. The music is embedded in the environment, not on a physical format such as CDs or DVDs, allowing the user to access all of her music collection everywhere and without constraints. The personalized and adaptive features are given by the remote control : the user chooses her music and the remote control will play the music on the pointed device. One of the Philip's Streamium advanced mode is the "Music follows me" system. To achieve that, the user pushes the "Music follows me" button one time in the current room, and one time in the destination room, and the music is transferred from the first room to the second room. By this way, this system partially benefits of the context awareness, but requires a user interaction to be possible. And despite the "Music follows me" slogan, their concept is based on the music transfer and not on music that completely follows the user. This interaction process could strongly take advantage of an invisible indoor geolocation system, allowing a fully automated process and a real follow of the user. It refers to the context aware [Abowd 1999] and anticipatory features : the music follows the user with the location system, and the system anticipates user desires without conscious mediation by modifying dynamically the music volume. That's why we will now study indoor location system alternative.

Classically, the most commonly used geolocation system is the Global Positioning System (GPS). However, this system is designed to be functional only outdoor, whereas Ambient Intelligence is mostly oriented in indoor use. Indoor localization and tracking of persons begin many years ago with specialized hardware requirements (infrared, ultrasound, electromagnetism, etc), which are precise but also really expensive and often limited to one room or less. Recent systems tried to avoid these cost and deployment problems by using technologies already deployed such as GSM, Wi-Fi or Bluetooth. One of the first works in this direction was RADAR [Bahl 2000], a tracking system based on signal strength information given by radio frequency devices. RADAR system worked in two phases: the off-line phase which consisted in collecting signal strength data in order to build a signal map, and the real-time phase, corresponding to the localization phase and based on the previously acquired signal map. This system provided a median resolution of nearly 3 meters. This two-phase model has inspired several papers, which tried to enhance this system in different manners and for different purposes: enhance accuracy and precision [Astrain 2006][Raman Kumar 2006], localize in high signal fluctuation areas [Ho 2006], use only one access point [ZÃ ruba 2007]. Most of these papers are based on the Wi-Fi technology, and shows that localization using this technology is conceivable. Our idea is to implement our Wi-Fi localization system, and use it to provide context awareness to applications which could take advantage of this information.

4.2.2 Concept

Our multi-room sound system is based on several music players distributed among the rooms of the house and a remote control which is able to localize itself with the help of available wireless networks.

Our application scenario : the user selects through the remote control a music to be played or a playlist. Then, when she is moving in the house, she carries the remote with her and the remote localizes itself and transmits the position to the player to start or mute them. A player which is too far away from the user is mute (cf figure 4.2). When the user is in a room with a sound player, it starts playing music. Finally, when the user is in a room or a corridor without a sound player, closest sound systems play music with a volume which depends on the distance between the user and the sound players.

4.2.2.1 Hardware

To perform our experiments, we needed two kind of hardware responding to these points:

Remote control The remote control must be pocket size, with a screen big enough to display clearly the user interface. It must have Wi-Fi capabilities to communicate with the sound devices and to be localized.

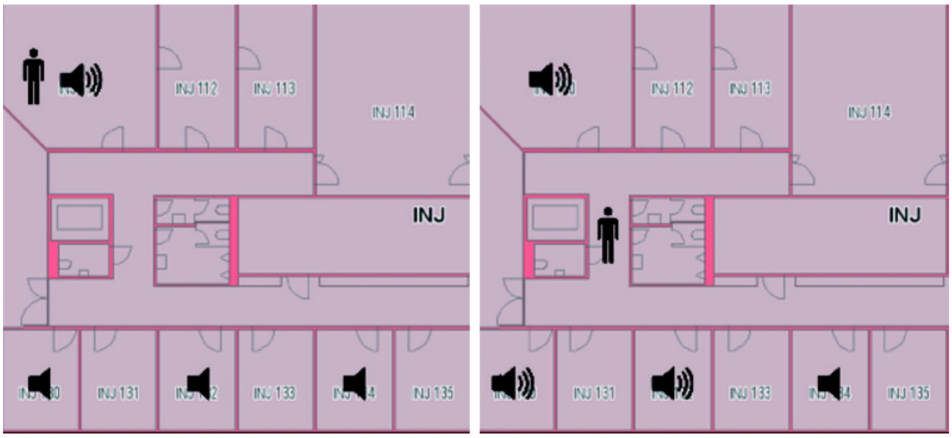


Figure 4.2: Influence of the user position on the sound system activation

Sound player The sound players must be connected to the network in order to communicate with the remote control.

As proof of concept, we chose the Ultra-Mobile PC (UMPC) Sony VAIO UX90 (cf picture 4.3) for the remote control. This device offers more hardware and software freedom compared to other handheld devices such as smartphones or PDAs : power (CPU, memory), USB connectivity, choice of operating system, choice of development tools. The main drawback is its size, which is not really adapted for the pocket, but this is not an issue for our tests.

Figure 4.3: UMPC Sony VAIO UX90 and Linksys WUSB200. Smartphones are another option for the remote control and could replace the UMPC.

The Wi-Fi chipset embedded on the Ultra Mobile PC was not satisfying to perform localization: the signal strength values range was too narrow (between -50 and -90dB) compared to the chipset embedded on an Intel Centrino laptop for example (between -30 and -90db). Thus, it was difficult to differentiate the signal values between two neighbor rooms and then to correctly estimate user's location. For this reason, we looked for a Wi-Fi USB adapter, and chose after some tests the Linksys WUSB200, which provides a better value range (equivalent to the Centrino). Another advantage of having a second Wi-Fi chipset is the network connection possibility. Indeed, when you are associated with an access point, the values are updated less quickly or only for the linked access point. Finally, our sound players are PCs connected on our network. The advantage is the ability to easily communicate with the remote. Our application doesn't do complex computation on these PCs, the tasks are limited to play, mute, or change the volume of the music.

4.2.2.2 Software

User positioning The first software we have to integrate in the remote control is the localization algorithm. To begin, we need to collect signal strength values. In order to do this, we have chosen Windows Vista or Windows 7: Microsoft provides with these operating systems a network API, the Native Wi-Fi API[4]. This interface offers functions to update and access Wi-Fi information. The only drawback is a slower Wi-Fi refresh rate when the adapter is connected to an access point. This point is not really annoying for us because of our two Wi-Fi chipsets : the WUSB200 to read WiFi information, and the UMPC chipset for the network connection. Concretely, with the Windows Native Wi-Fi API and our Linksys WUSB200, the Wi-Fi information (BSSID and RSSI) are updated one time per second.

About our localization software, our needs are a program which is :

Light Our program is intended to run on a handheld device

Fast The music has to follow the user, so we need to update user position as fast as possible.

Accurate Precision is not our priority, but we want a reliable position to always play music in the user current room.

As other radio frequency positioning systems, our system works in two phases: the off-line phase which consists in collecting signal strength data in

[4]http://msdn2.microsoft.com/en-us/library/ms706556(VS.85).aspx

order to build a signal map, and the real-time phase, corresponding to the localization phase and based on the previously acquired signal map. During the off-line phase, the user creates a navigation graph: each room (or part of corridor) is represented as a vertex of the graph, then the user has to link neighbor vertex between them and record signal data (access point identifiers and signal strengths) for each vertex.

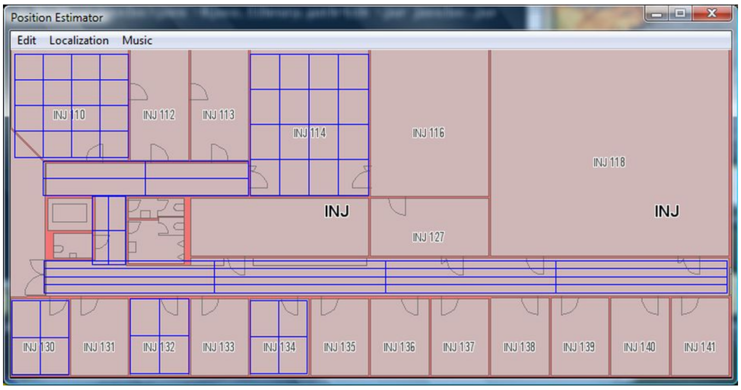


Figure 4.4: The editor allows the calibration of Wi-Fi information and sound players position

After this quick training phase (15 seconds per vertex to capture signal data), the localization system is ready to use: the application takes Wi-Fi samples every second during three seconds, smoothes the received signal strengths by removing the values too far from the mean if the standard deviation is too high, locates the user (by a simple distance calculation) according to the signal map acquired during the off-line phase. The system uses the navigation graph to ensure that the new position is realistic (connected to the previous one), and then avoid important errors due to signal noise.

Device communication We will now describe the communication between the remote and sound players. The first communication is the interaction between the remote control and the main sound player, which contains the music file, in order to select the current music to play, and diffused it on the network. The second interaction is between the remote control and all the sound players : the remote control decides which player is on and the volume of the current players. Figure 4.5 shows the interaction between each device.

To facilitate the communication between all the devices, we have chosen Java RMI (for Remote Method Invocation[5]) : this programming interface

[5]http://java.sun.com/javase/technologies/core/basic/rmi/index.jsp

allows applications to easily access distant objects. In our case, it is helpful to dynamically turn on/off a music player or change its current volume from the remote control.

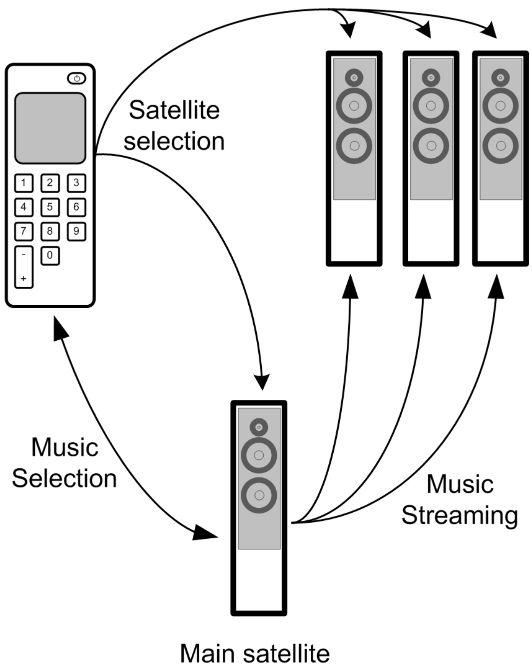


Figure 4.5: Illustration of our architecture architecture

Music diffusion For the music diffusion, we have one RMI client and one RMI server.

The RMI server is the main music player, which contains all the music files, and offers three services : a function to list the music files, a function to start the music diffusion and a function to stop the music diffusion. When the server has to start the music diffusion, the selected file is diffused on the network using the multicast Real-time Transport Protocol (RTP): this way, all the devices connected on the same network are able to play the music diffused by the main music player. To diffuse the music using multicast RTP protocol, we have chosen Java Media Framework (JMF[6]), which is not supported anymore but still working.

The RMI client is the remote control : when the user wants to access the music database, the remote control asks the server to list all the files. Then, the user chooses his music, and could start the music diffusion. At this

[6]http://java.sun.com/products/java-media/jmf/index.jsp

moment, the main music player begins to start the diffusion over the network, and the remote control starts to localize the user in order to activate adapted players.

Sound players activation For the sound player activation, we have one RMI client for the remote control and one RMI server for each music player.

Each server allows the client only to start or stop the music. The start function takes as input the volume given by the client. When the client asks to the server to start the music, the server just read the multicast RTP flow (sending by the main sound player) using JMF.

The client, based on the remote control, reacts each time a new position is evaluated (i.e. every three seconds as described in subsection 4.2.2.2). If the new position is in a room containing a music player, the client turns only this music player on. On the contrary, if the user is in a room without a music player (in a corridor for example), each close player (for our experiments, we choose a distance of less than 20 meters) is turned on, with an increased volume. To calculate the new volume, we are using the sound attenuation formula: this formula is normally reserved to environment without obstacles, but it gives a good idea on how we have to increase the volume to compensate the distance. The formula is

$$V2 = V1 - 20log\frac{D2}{D1}$$

with V1 (dB) and D1 the original volume and distance, and V2 (dB) and D2 the new volume and distance. When the user modifies the volume of the music, she modifies the V1 parameter.

4.2.3 Preliminary tests and results

The system has been tested in our laboratory. The surface is 817 square meters (43m x 19m), representing 18 offices of variable size. The floor is directly covered by 4 access points, but we can acquire up to 7 different signals in some rooms (from access points from nearby floors or buildings).

In order to evaluate our system and have a feedback on it, we asked 12 persons to use our music application. The purpose of the experiment was to test the accuracy of our localization system and to see how immersive the system was.

Users were asked to check the position shown in the UMPC and qualifying the systems accuracy as wrong, near or good. The good qualification was applied if the localization system shown the exact room or corridor part, the

Table 4.1: Localization experiment result

Good	Near	Wrong
81%	15%	4%

near qualification when the system indicated not the exact position but a juxtaposed position, and the wrong was for all other cases.

Table 4.1 shows the result of our localization experiment. During these tests, we have noticed that 15% near and 4% wrong were mostly obtained in the corridor, because of the lack of important changes in the signal compared to pass through a door for example. Mistakes in corridor are not really annoying because they only influence the sound volume of the music players. But mistakes in rooms are a real issue: even if it happens rarely, it is really frustrating for the user to have the music playing in the wrong room. Unfortunately, these mistakes can't be completely removed with a localization system only based on Wi-Fi, because noise will always occurs. Adding a second localization system could be helpful : a short-range localization system (based on Bluetooth or RFID tags for example) to ensure user position in room with music players, and Wi-Fi for corridor or rooms without sound players.

To conclude, we presented in this section a multi-room music system which takes advantage of an indoor localization system based on Wi-Fi to allow sound follow. Our application takes advantage of location information by using existing Wi-Fi access points in order to localize user's remote control in the house. The user position allows the program to select on which 5 sound systems the music will be played, and to modify the sound volume if necessary. Our experiment proves that using indoor localization to enhance the user experience is feasible. Our localization system works well in an urban environment where we are able to acquire numerous Wi-Fi networks, but the precision and accuracy drop down dramatically in an individual house where only one access point is visible at a time. Using short-range radio frequency modules, such as Bluetooth or RFID, could be a solution to this issue: it would give reliable information when the user is closed to a device. Unfortunately, the system would not have any clue on user's location if the system was only based on short range RF, but a combination of short range RF and Wi-Fi could give interesting results. Finally, our system is based on a UMPC, but a smartphone should be more adapted to our current lifestyle. However, wireless information such as Wi-Fi signal strength values was not accessible on smartphones when the prototype was developed.

Next section presents another form of context-awareness. Instead of taking advantage of user's location, we are adapting the system to user's behavior by analyzing where she is looking at. This second prototype is presented in the following section.

4.3 Natural Activation for Gesture Recognition Systems

For several years, interacting using hand gestures has been one of the hottest areas of human-computer interfaces. Even though hundreds of new developments have been made within this research field, only very few products have hit the mass market. Different reasons could explain this: the need for specialized and expensive hardware to obtain an acceptable robustness; the social acceptance of gesturing in front of a computer or television; or the problem of the system activation - triggering gesture -. The last issue is the one we want to address: when to turn on or off the gesture recognition system to avoid false positive detection due to natural movements such as grasping one's cup of tea in front of us?

Indeed, one of the main problems in hand gesture interfaces is how to differentiate when the user is actually interacting with the system and when the user is simply moving in front of the camera. In the context of gesture interaction in video games, this difficulty is reduced: when someone wants to play video games, she turns on the games console and starts playing by doing expressive, unnatural, and recognizable gestures. In this case, the user intention is clear. On the contrary, it is much harder to differentiate a gesture interaction from an insignificant movement while watching TV. Effectively, in front of the television, people could grab something on the coffee table, discuss with other persons, stand up to bring something in the kitchen, etc. These insignificant gestures bring ambiguity in the recognition system and could activate unwanted functionalities. This uncertainty could be sidestepped by using complicated hand signals to avoid false positive and be sure that the user really wants to interact with the system. Obviously, this solution is not satisfying, because it is not natural and generally tiring for the user to perform such gestures. On the other hand, using natural gestures often brings false positive detections. In the context of watching television in the living room, this is a major issue as the scene is "complicated, unpredictable, and the hand is not a dominant part of the image" [Freeman 1995].

In this section, we target this system activation issue by exploring two solutions that aim at helping the gesture recognition system to decide whether the user wants to interact or not with the television. These solutions are designed to enhance the interaction flow in hand gesture-based interfaces in general. This research does not directly targets videogame interfaces since the gesture activation problem is not that relevant in this context. The first solution is based on the point that we naturally look at the objects and people we want to interact with. Based on a face tracking system and a common webcam, we are able to evaluate where user's attention is focused and then decide whether consider a gesture or not. The second solution requires a remote control as people are used to grab one to start interacting with a television, e.g. channel swapping. The user just has to grab the remote control to "wake up" the gesture recognition system. In order to test these two concepts, we integrated them in a simple cover flow music system, and we asked some participants to test this application displayed on a TV in a living room.

This section is organized as follows: first we overview some works related to gesture interaction. We continue with a description of our concept and proposed solutions. Last section describes our evaluation and results.

4.3.1 Related work

Using a vision-based gesture interface could be a powerful tool to control the living room devices from the couch. Even though the video games industry has just hit the market with products such as Sony's EyeToy or recently Microsoft's Kinect system, the idea of controlling devices in the living room with hand gestures has been present since a long time ago in the research field. Freeman et al. presented a television controlled by hand gestures in [Freeman 1995]. Users had to move the palm of their hands if front of the TV in order to interact with it. Although this prototype appeared exciting for the participants, they complained about the tiring aspect of keeping this position to interact. For this reason, authors suggested to maintain the open hand exclusively as a trigger gesture to start the recognition system.

Even if hardware, techniques, and precision of gesture recognition systems have considerably evolved since [Freeman 1995], trigger gestures as a starting point for the complete gesture recognition are still used nowadays. It is still used in the academic field [Fang 2007], but also in the industry. For example, Hitachi presented a gesture-controlled television[7] based on GestureTek software. In this video, we could see that the system is waiting for a trigger gesture to start the interaction, and this trigger gesture looks quite unnatural.

In [Ashbrook 2010], Ashbrook et al. presented a different solution to avoid unwanted gestures. They conceived the EGL, for Everyday Gesture Library,

[7] http://www.youtube.com/watch?v=O21SYHDEPOs

a data set which aims to detect if the recognition system will confuse a created gesture with an end-users' everyday movement. The drawback of this approach is that EGL could eliminate gestures which are considered as easy and natural by the end-user.

The concept of using user's attention or gaze direction in human-computer interfaces has been applied several times. In [Stiefelhagen 2002], Stiefelhagen et al. estimated the focus of attention of meeting participants in order to build gaze-aware social robots. In [Nickel 2003], Nickel et al. used the head orientation as well as the forearm of the user to estimate a 3D pointing direction. They indicated the usefulness of head orientation to determine the target of the user. In this section, our objective is not to determine which device is targeted as in the work of Nawaz et al [Nawaz 2008], but to know whether the user is focused on one device in order to interact with it. To the best of our knowledge, it is the first time that head orientation is used as a "trigger gesture" in the context of gesture interaction.

4.3.2 Concept

Interpreting human gestures in order to interact with a device is often perceived as the future of human-computer interaction. But many challenges are still remaining in gesture recognition. If most of the researches are focused on the accuracy of the systems, or how to make the interaction more natural, an important problem is to understand when the user is interacting with the system and when they are not. We describe in this section three approaches: the trigger gesture activation, the gaze estimation activation and the remote control activation. The two last approaches are the solutions we propose in this section.

Trigger gesture The trigger gesture is nowadays the most common way to start interacting with a gesture recognition system. It consists in performing an unnatural gesture (i.e. a gesture you would normally not do in this context) to inform the system that we are ready to interact with it. This gesture could be waving your hands, like in the Kinect menu for example. We include this solution to have a point of comparison with our two proposed activation methods: gaze estimation and remote control detection.

Gaze estimation We naturally look at the objects or people when we want to interact with them. The eye-gaze often expresses where user's attention is focused onto, and several research works [Matsumoto 2000][Nickel 2003][Stiefelhagen 2002] already use this fact for several purposes. Also, according to [Glenstrup 1995], "communication through the direction of the eyes is faster than any other mode of human communication". We want to take advantage of these assumptions to know when the gesture system has to be activated. If the user is not looking at the screen, we assume that they do not want to interact. In this case, gestures are then ignored. There are at least three parameters to determine where a person is looking at: head location, head orientation and the gaze direction. In our application, we consider the head orientation as a sufficient parameter to determine where user's attention is as shown on figure 4.6.

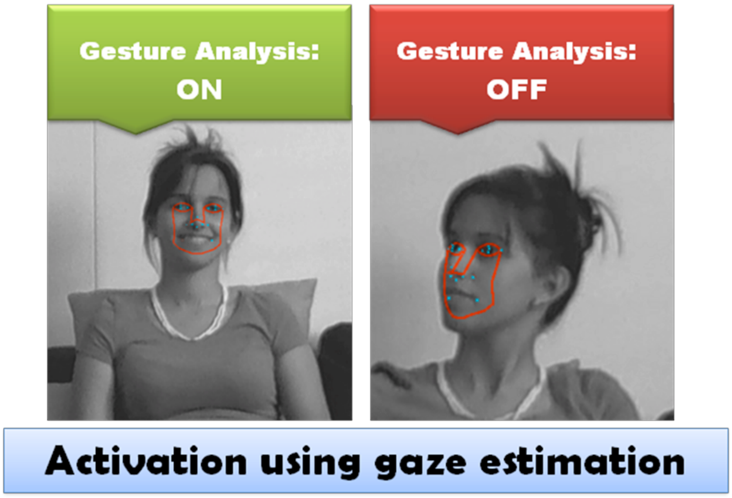


Figure 4.6: Gaze estimation activation

[Stiefelhagen 2002] empirically demonstrated the feasibility of such an approach, which does not need specific hardware compared to eye tracking systems. Our face tracking uses a standard Logitech webcam, the HD Pro Webcam C910[8]. In terms of software, we are using faceAPI[9], a robust and real time 3D face tracking API developed by Seeing Machines. Based on this API, and with the webcam on top of the television, we are able to track if user's attention is on the screen or not. We need two angles to evaluate that: the vertical angle and the horizontal angle.

Figure 4.7 illustrates the vertical angle: if the vertical face angle is between zero and alpha degrees, the user is considered as focused. Alpha is calculated using the screen height and the distance of the user from the screen. The same measure is applied for the horizontal angle of the face, using the screen width instead of the screen height.

[8]http://www.logitech.com/en-us/webcam-communications/webcams/devices/6816
[9]http://www.seeingmachines.com/product/faceapi

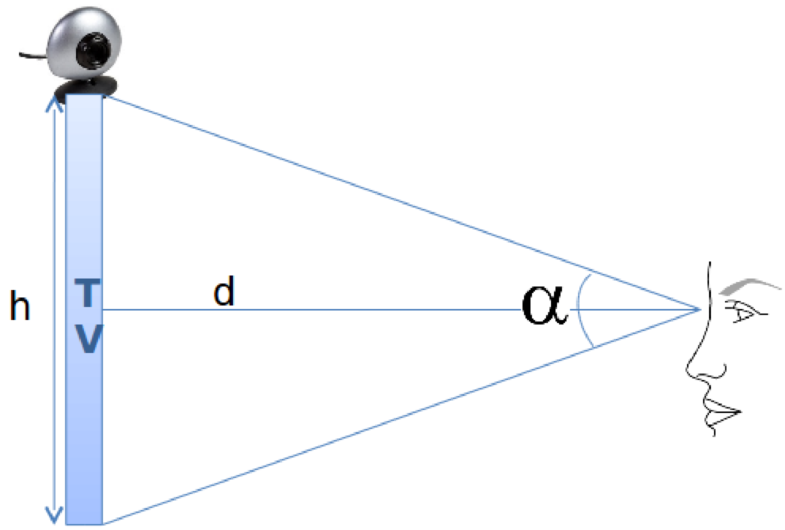


Figure 4.7: Left Side View of the User and TV

Remote control detection Based on the French expression “prendre la parole” (literally “take the words”), we use the paradigm of grasping an object to interact with the system.

Moreover, it has been decades that each television is sold with a remote control for interaction. According to this, we decided to take advantage of a remote control to trigger the gesture recognition system: the gesture detection is then activated while and as long as the user grasps the remote control. This solution can be considered as very robust. Nevertheless, even if it does not prevent any hand gesture, one of the user hands must grasp the remote control, which may be disturbing.

In order to know if the user is holding the remote control (RC), our gesture detection system checks the activation of the RC accelerometers. Indeed, accelerometers are sensitive enough to provide a robust “holding detection” mechanism. When the remote control is on the coffee table, acceleration values are close or equal to zero, and are different from zero when the remote control is in user’s hand. In order to validate both concepts, we made some experimentations that we will now be described more in detail.

4.3.3 Evaluation

We conducted a user study with twelve participants (six males and six females with different professions and school qualifications) aged from 23 to 39.

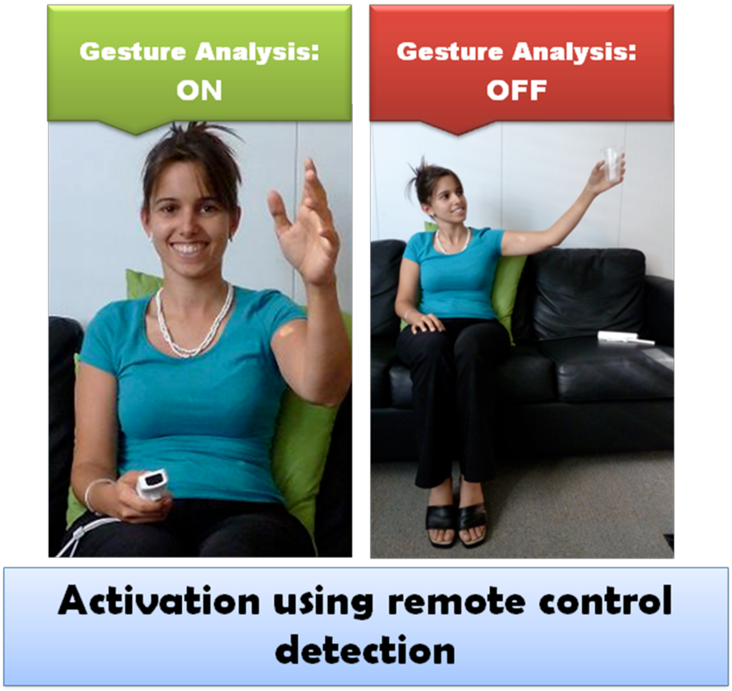


Figure 4.8: Remote control detection activation

No specific selection criteria were applied except for balancing the number of male and female. None of them had experience with gesture interface, except few hours on Sony's EyeToy for two persons. We started the experimentation by giving them a short explanation on the goal of the experiment (interacting with TV) and how to do it (description of the gestures and activation methods). They were then asked to sit down in front of the TV, and then to start performing a set of predefined tasks written in the protocol we gave them (described later in the section). Finally, they were asked to answer an AttrakDiff [Hassenzahl 2003] questionnaire, an tool designed to evaluate and compare interactive products.

Test Applications

For this test case, we developed three versions of a basic cover flow[10] application. The only difference between each version was the activation method of the gesture recognition system: using a trigger gesture - holding up the arms, showing palms of the hands to the TV; holding the remote control; or using the gaze estimation.

[10] http://en.wikipedia.org/wiki/Cover_Flow

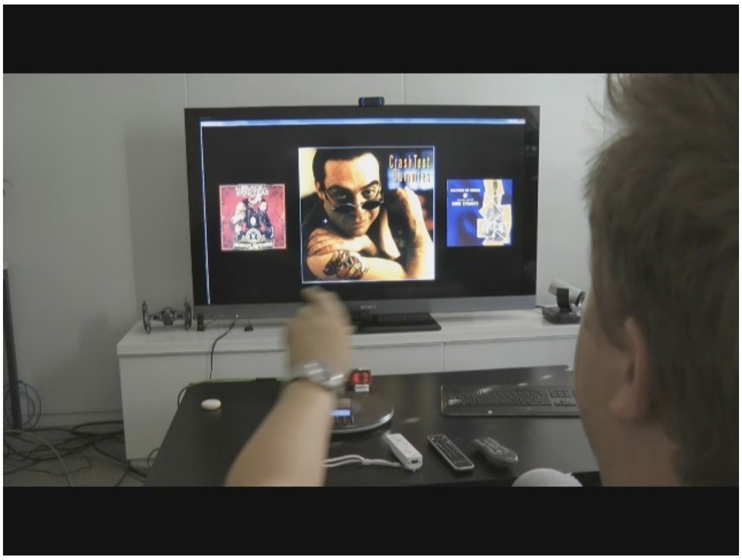


Figure 4.9: Graphical interface of our test applications

The application displayed three CD covers: the currently playing one, the previous and next ones. The user can "browse" the available CDs from left to right (and vice versa) with the help of hand gestures in the desired direction. Once she found the CD she wants to listen, she simply selects it by pointing at it.

We should notice that the gesture recognition of the application was performed by a Wizard of Oz[11] setup. Indeed, the goal was to evaluate the techniques to activate the gesture recognition system, and not the recognition system itself. The Wizard of Oz used the information from the gaze estimation or the remote control detection to decide when to convert a gesture or not.

Protocol

The whole experimentation consisted of three phases for a total duration of 30 minutes. Firstly, we gave an explanation of the problematic and why we were performing these tests. We described the three methods to activate the gesture system and the different gestures available in the application (previous album, next album, play/pause). After that, we chose randomly one of the three applications, launched it and let the participant using the system in order to be familiar with the gestures.

Secondly, the user was considered to feel comfortable with our system and

[11] http://en.wikipedia.org/wiki/Wizard_of_Oz_experiment

was asked to perform two actions with the current application: a “quick” action and a “long” action. The quick action consisted in starting the gesture system, selecting the previous or next CD, playing this song and stopping the gesture system. The long action consisted in exactly the same steps, but the selecting task was longer (e.g. selecting the n^{th} CD on the left, with n randomly chose between 8 and 12).

We repeated these steps for the two remaining activation methods. Finally, once the experiment performed, we asked the user to order the activation system by preference, and then to fill in an AttrakDiff questionnaire for each activation method.

Results

For each questionnaire, a 2D graphic is generated, which places our system in terms of pragmatic quality (i.e. usability of the system) and hedonic quality (i.e. novelty, interest, identification, etc.).

Activation with trigger gesture The activation system using the trigger gesture has been qualified as neutral. It means that the system is average both in terms of pragmatic and hedonic qualities. The confidence rectangle is large for the two dimensions, which means that participant opinions were mixed. In terms of attractiveness, the overall impression is that this system is moderately attractive. According to our users, the system is really “manageable”, but also “unpleasant”. Figure 4.10 shows the result of the AttrakDiff questionnaire for the trigger gesture activation.

Figure 4.10: Result for trigger gesture activation

Activation with remote control The activation system using the remote control has also been qualified as neutral. Compared to the trigger gesture, the pragmatic quality is equivalent but the hedonic quality is lower. Again, the confidence rectangle is large for both HQ and PQ. In terms of attractiveness, the overall impression is that this system is moderately attractive. The good points of the system are that it is "manageable" and "simple", and the bad points are that it is "cautious" (maybe because remote controls are already used to control the television?) and "cheap" (we used a Nintendo's Wiimote for our tests). Figure 4.11 shows the result of the AttrakDiff questionnaire for the remote control activation.

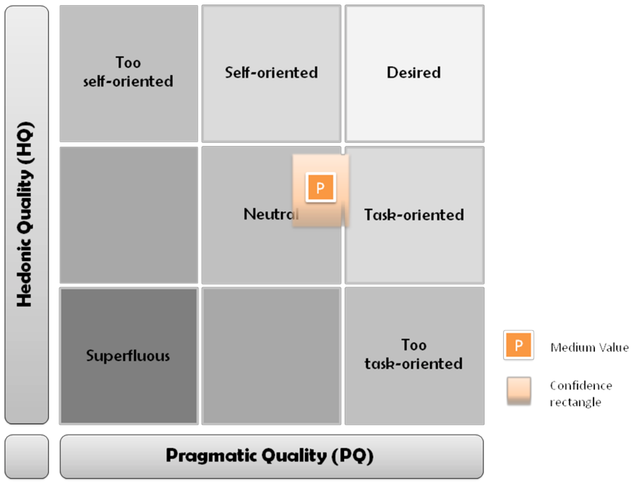


Figure 4.11: Result for remote control activation

Activation with gaze estimation The activation system using the gaze estimation has been qualified as desired. It means that the system assists optimally the user, and that users are motivated and stimulated by it. The confidence rectangle is small, particularly in terms of hedonic quality. In terms of attractiveness, it has been qualified as very attractive. The good points of the system are that it is simple, innovative and pleasant, and the bad point is that it is not challenging for the user. Figure 4.12 shows the result of the AttrakDiff questionnaire for the gaze estimation activation.

Activation systems ranking In addition to the questionnaire, we asked the participants to rank the different activation system. For 11 users, the best activation system was the one using the gaze estimation (cf figure 4.13). Nobody selected the remote control activation.

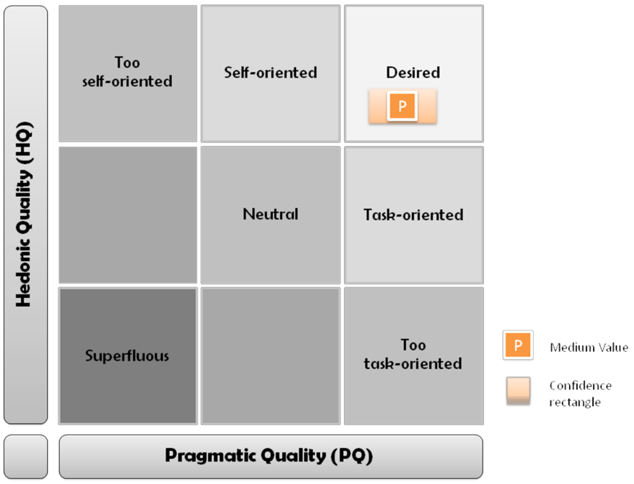


Figure 4.12: Result for gaze estimation activation

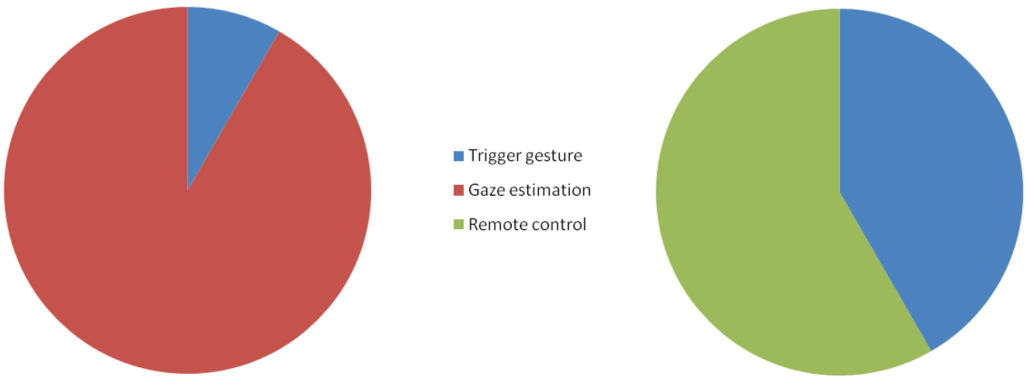


Figure 4.13: On the left: Favorite activation system. On the right: Worse activation system

For seven participants, the worse activation system was the one with the remote control. The five remaining participants selected the trigger gesture.

Discussion After this preliminary study, the results are clear: all the users enjoyed the gaze estimation activation method, nevertheless the benefit of using the remote control activation system is much more questionable. Indeed, seven participants preferred the trigger gesture to this system. The main reason mentioned by the users is that it is unnatural to grab an object and not interact with it. If you grab something, you want to use the buttons on it, or to move it in front of the TV. Thus, if the remote control system was considered as a better system in terms of activation compared to the trigger gesture, it was also considered as worse in terms of interaction. Indeed, it left the user with only one hand available to perform the gestures.

The main goal of testing the remote control activation system was to evaluate whether grabbing a physical object was reassuring for the user. Indeed, with such a system, you can fully control when the camera is watching you or not. With the two other systems, the camera is always on to detect the trigger gesture or the user's face. But our participants were more focused on the cumbersome aspect of having a physical object in their hands.

On the other hand, results concerning the gaze estimation activation system were fully positive. Among our 12 participants, only one put this system at the second position. She preferred the trigger gesture because she had more control, compared to the gaze estimation, and the remote control system was not natural enough. The other participants enjoyed to have no action to perform in order to start the gesture recognition system, and found it very natural.

Another satisfying point concerns the gaze estimation method and its robustness. Indeed, even if it is a first prototype, the error rate was less than 3%. Some improvement could be performed concerning the position of the user on the couch: during our tests, users were normally sitting in front of the TV. In their living room, people tend to be more "relaxed" (e.g. lying down on the couch), and the application has to consider this parameter.

Currently, one of the main limitations of the gaze estimation activation system is its sensitivity to illumination conditions. In the dark, the face tracking system would probably not be able to detect user's face accurately. This could be a problem in a scenario where the user is watching a movie in the dark and wants to interact with the system. A solution to this issue could be to use a 3D IR camera (like the Kinect) instead of a standard webcam in order to determine where the user is looking at, regardless of the illumination conditions.

To conclude, we have presented in this chapter two new solutions dedicated to the activation of gesture recognition systems in the living room: the remote control activation and the gaze estimation activation methods. Even though the first one did not convinced our panel of users, they were really enthusiast about the gaze estimation method. Our first prototype also showed that this method is feasible and robust enough. Future researches should focus on how to manage the gaze estimation activation system when several people are seated in front of the television. Indeed, with a remote control, only one person at a time could interact. Thus, how should our system react when several persons are looking at it? Another improvement could be to implement the gaze estimation activation with a 3D IR camera such as the Kinect to resolve the illumination conditions problem.

4.4 General discussion

Sensing the world and learning about where the user is and what she is doing is important in order to increase usability of computing devices. It is a means to increase efficiency of interfaces, for example by offering a limited number of options in a GUI depending on user's location, activity, and environment, or simply by doing automatically redundant tasks.

Despite these advantages, there are several difficulties and criticisms concerning user awareness and Ambient Intelligence.

One of the most important issues is about the loss of privacy. To offer user awareness, there is a need of intrusive systems that detect where you are, what you are doing, with whom, etc. This has raised some outspoken critics: in [Araya 1995], the author described Ambient Intelligence, or more generally Ubiquitous Computing, as "an attempt at a violent technological penetration of everyday life", and "the substitution of real world situations by digital surrogates". Other works - [Talbott 2000][Adamowsky 2000] - also addressed this issue. But recent developments have shown that people are ready to sacrifice a part of their private life - in the case of Facebook[12] for example. Would they accept an additional intrusion in order to gain more comfort ? The question is far from easy, and the difficulty of entering people's house in order to add more sensors and computing devices is one of the main challenges of Ubiquitous Computing[Edwards 2001].

[12]www.facebook.com

Close to this privacy problem, acceptance is also an issue. Indeed, as described in 2.1, computing devices changed our life, but as a side effect. Ambient Intelligence "expressly propose to transform society by fully computerizing it" [Weber 2005]. Entering people's house and everyday life is not a simple task, and making them accept to be surrounded by computing devices and sensors can only be possible little by little. Computers, Internet, Wi-Fi, smartphones entered in our life one by one, and it will probably be the same for future computing devices. For this reason, Ambient Intelligence will probably not penetrate our everyday life in one shot, but little by little, taking advantage of computing devices that are already common at home. Our first prototype - Sound will follow, is an example of this approach.

Finally, an issue which is more related to the design process is how much a system or activity should be automated. Automation is a good way to relieve the user of unnecessary or redundant tasks, but she probably still wants to feel that she is controlling the system, and not the opposite. In the case of our second prototype, the benefit is clear - removing an unnecessary step of the interaction process - and positive. After discussions with engineers specialized in gesture recognition, they told us that people usually forgot to perform the "trigger gesture" in order to activate the gesture recognition system. They directly tried to perform gestures to interact with the system, and then finished frustrated by the non-response of this system. Thus, the second prototype is an illustration of an efficient use of Ubiquitous Computing. On the other hand, in the case of the first prototype, the use of people's location in order to change the current music player is more questionable. If the user really wants the music to follow her, it relieves her of unnecessary tasks, but if she doesn't want to, or if there is other people in the room, then she would not appreciate and would feel that she is not in control of the system. In this case, it could be better to add an additional step, for example by displaying a message on her smartphone asking her if she wants to play music on the sound system close to her. But this problem illustrates the difficulty of interacting with sensor-based interfaces, or more globally with "invisible" devices.

The philosophical aspect of this same question - "what should be automated?" - has been presented in [Tedre 2008] as a fundamental question in human-centered computing. The author describes that the question initially started in the '60s by "What can be automated?" [Forsythe 1968], and then shifted to "What can be efficiently automated?" [Comer 1989]. Figure 4.14 illustrates the fact that the answers of these different questions are targeting different sets: it is possible that a process should not be automated despite the technical possibility to do it. Furthermore, these questions are not targeting the same disciplines, and the question "what should be automated?"

needs knowledge in sociocultural, ethical, economic and other topics outside computing science. Automating tasks is not only a matter of efficiency, but a problem that is also including social and cultural dimensions.

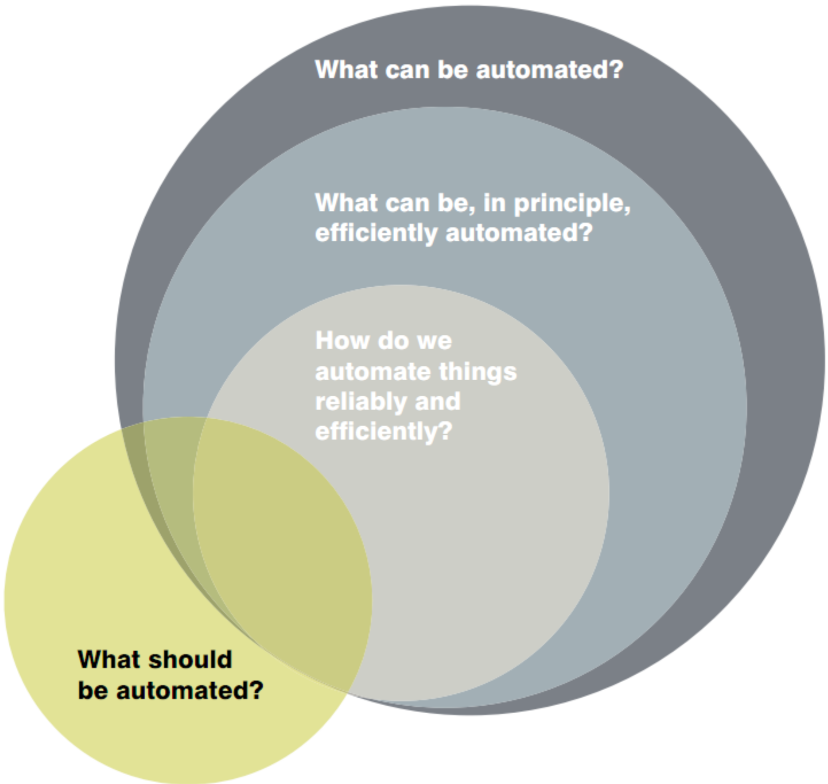


Figure 4.14: Automation sets. Image from [Tedre 2008]

4.5 Conclusion

We presented in this chapter two prototypes which illustrate automation through user awareness. The first prototype, Sound will follow, takes advantage of user location in order to automatically switch sound players depending on user's position in the house. The second prototype uses the face position in order to know whether the user wants to interact with a gesture interface or not. Both of them illustrates the possibilities offered by user awareness: performing automatically tasks that the user would do - like shutting down a sound player and starting another one - but also tasks that she would prefer not to do, or she would forget - like doing unnatural gestures to trigger a gesture system-. Automating these tasks are naturally increasing usability, efficiency in particular, and increase the sense of presence of the digital world by showing to the user that her actions in the real world are affecting digital devices and their content.

Chapter 5

Familiarity

Contents

We have presented the importance of unifying the digital world and making it react to our moves and actions. These notions are central to efficiently increase our interaction with digital media. In this chapter, we will focus on the relevance of using familiarity in the design of interfaces. The notion of familiarity is more axed on the users' satisfaction factor, and on the creation of interfaces which accelerate user's performance and acceptance.

In order to illustrate this principle, we developed two prototypes dedicated to the navigation and interaction with digital music collections. The "virtual shelf" is a graphical user interface based on a CD shelf metaphor, and is enriched with additional capacities such as tangible drag-and-drop and user's identification. The second prototype, the "vintage radio" interface, proposes a representation of digital music collections as radio frequency bars commonly displayed on analog radios. The presentation of these prototypes is preceded by the results of a survey about how people enjoy their digital music collection nowadays.

5.1 Taking advantage of familiarity

One of the most expected features in every interactive interface is that it should be "intuitive". An intuitive interface is an interface that users can manipulate without the need of a learning period, or at least after a short one. This notion has been discussed several times, claiming that intuitive interaction involves using existing skills and knowledge gained through previous experiences and products. Thus, an interface is intuitive if "it resembles or is identical to something the user has already learned" [Raskin 1994], and intuitive products are those with features that users already encountered in the past. In this context, "intuitive" is synonym to "familiar". Minimizing human learning can greatly accelerate user's performance and acceptance [Galitz 1997], and for that familiarity is a key notion.

The importance of familiarity was discussed by Van de Walle et al. in [Van de Walle 2003]. They noticed that even if familiarity has been found very important in HCI, it also has been neglected. In their work, they empirically studied the impact of familiarity over 50 elderly persons. They interpreted the results using the philosophical writings of Heidegger about familiarity, and interrelated familiarity with concepts such as understanding, unconsciousness, "unthinking", involvement, easiness, comfort and friendliness.

There have been many practical approaches and even the creation of fields of study that use familiarity in computing interfaces.

An example is the creation of Tangible User Interfaces (TUI) [Sharlin 2004]. The concept of TUI is to take advantage of a sense which is generally unused when we are interacting with digital data: our sense of touch. As we are manipulating physical objects in our everyday life, TUI proposes to manipulate digital data through the use of graspable objects. This concept has been introduced in the middle of the 90's by Fitzmaurice et al. with Bricks [Fitzmaurice 1995]. With Bricks, they presented a "Graspable User Interface" to control virtual objects through physical handles. Ishii et al. continued in this way with Tangible Bits [Ishii 1997], where they suggested to bridge the gap between bits and atoms by coupling digital data with everyday physical objects.

Using metaphors is another form of familiarity in digital interfaces. A metaphor consists in using an image to represent something intangible by something more intelligible. It is considered as a central element of design practice [Blackwell 2006]. An example is the "desktop metaphor", which is often associated to the success of graphical user interfaces. Metaphors are a good mean to increase the initial familiarity between the user and the application [Dix 1997]. In [Galitz 1997], the author argued that real world metaphors are most often the best choice. Jacob et al. proposed the notion of Reality-Based Interaction (RBI) to describe Post-WIMP (Windows, Icon, Menu, Pointing device) interfaces that use "actions that correspond to daily practices within the non-digital world" [Jacob 2008]. They also presented in this work a framework to classify these new interaction styles.

We previously cited two interfaces based on metaphors: the Slurp [Zigelbaum 2008], an eyedropper which "extracts" digital media from physical objects in order to "inject" it into other devices, and Augmented Surface [Rekimoto 1999], inspired by how easy it is to share and reorganize physical documents.

Another domain inspired by familiarity is the Natural User Interfaces (NUI) domain. NUI are interfaces that "enable users to interact with computers in the way we interact with the world" [Jain 2011]. It often refers to interfaces that use speech recognition [Cooke 2001][De La Torre 2005], gesture recognition [LaViola 1999][Moeslund 2003] or even whistle recognition [Esnaola 2006]. Our prototype presented in section 4.3 can be considered as a NUI: the gaze estimation part is inspired by the fact that we tend to look at objects or people we want to interact with.

In this chapter, we present two prototypes that take advantage of familiarity. They are dedicated to digital music collections, and are mainly based on real-life metaphors: the "virtual shelf" reminds the organization of our CD albums in a shelf, whereas the approach of the "vintage radio" prototype is to remind analog radios. Before describing these two prototypes, we present in the following section the results of a survey dedicated to digital music collection. We conducted this survey in order to have a better understanding of people's habits concerning their music acquisition and interaction.

5.2 Listening to music nowadays

Since the end of the '90s, our music habits have considerably evolved. Music sales continue to shift to digital distribution, and CDs slowly become obsolete.

We wanted to know how people are listening to music nowadays, and we conducted a survey dedicated to music listening during the summer 2011. We asked people about the size of their digital collection, the music devices they have, the importance of physical medium according to them, or how they are navigating in their musical collections.

This survey was sent to students and workers at EPFL[1]. We received 986 answers, from persons from different ages and disciplines. Detailed results of this survey are presented in appendix A.

We categorized the answers in three parts: music storage, music selection and interaction, and music discovery.

5.2.1 Music storage

The dematerialization of the music content brings advantages: we can carry all our music collection on a single device, which could be our laptop, our mp3 player, our cell phone, etc. It allows an access on our entire music collection everywhere. But a digital album doesn't have the "tangibility" of a vinyl or CD album. For 63%, it is not important to have albums on physical medium. And in the future, this number will increase over and over: the importance of having physical medium is twice inferior for people aged less than 20 years compared to people aged more than 45 years.

On the other hand, organizing a digital music collection is often a real challenge. Organizing digital files on a single device is a demanding task [Barreau 1995], doing it for all computing devices is even more complicated. The size of our music collection make it even harder: we often have thousands of digital songs. The possibility of playing music on several devices often results in having different songs on different devices. This pain point was selected as one of the most annoying: it takes too much time to perfectly organize a music collection over all our computing devices, with the full tags and album covers. Moreover, the multiple audio formats make this task even harder.

Ideally, a single location accessible from everywhere could be a solution to store our digital music collection. Brown et al. [Brown 2001] already suggested in 2001 the "benefits from having digital versions of one's collection being available in a central place, with various networked, digital appliances

[1] www.epfl.ch

able to access the collection". The central location solution is nowadays technically possible using online services such as iTunes Match[2] or Spotify[3]. But with this solution comes the fear of unavailability: no network means no music.

5.2.2 Music interaction

Building up a music collection is one thing, interacting with this collection is another.

Listening to a music collection can be done in an active manner or passively. When we are active, we browse the collection until we find what we want to listen to. Making playlists is an alternative to select actively digital files. Some tools, such as iTunes' smart playlists, give the user the possibility to select automatically songs depending on some basic criteria (artist name, genre, etc) or more advanced criteria (last played date, date added, personal rating, play count). On the other hand, when we passively listen to music, we let the player choose for us: it could be random mode on a genre, artist, or on the complete collection.

Our interviewees are generally more passive than active when they are listening to music. We also observed that the more the digital collection is large, the more listeners are actively interacting with the collection. The more people are accumulating music, the more they need to actively browse their music collection in order to find a song they want to listen to. We noticed that even though people often used manual selection and random mode, smart playlists were not very popular.

CDs are not popular anymore: Hi-Fi is rarely used and portable CD players are almost completely abandoned. The most popular device to listen to music is clearly the computer. It illustrates the problem of unification described in chapter 3: a computer is not the best device to play music, but people are using it because data are stored on this device. After the computer comes the MP3 player, which is much more used to play music than smartphones. Finally, streaming audio devices - Logitech Squeezebox[4] or Sonos products[5] for example - are still not very popular.

5.2.3 Music discovery and acquisition

The last point is about how people discover and obtain new songs.

[2] http://www.apple.com/itunes/itunes-match/
[3] www.spotify.com
[4] http://www.logitech.com/en-us/speakers-audio/wireless-music-systems
[5] http://www.sonos.com/

The main way to acquire new tracks and albums is clearly the download, legal or not Buying CDs come at the second place. The use of free streaming - for example youtube - or dedicated streaming services - like Spotify - is limited. Physical medium are by far the main way of acquisition for people aged from 35 and more, whereas younger participants preferred to download music.

Concerning the main way to discover new songs and bands, the most common way taken by our participants is clearly recommendations from friends. We can also find this observation in [Brown 2001] and [Brown 2002]: friends had an important influence on what people later bought or listened to. After recommendation via friends came listening radio, streaming and then recommendation via social networks. We noticed that these answers are slightly related to the size of the music collection: participants with large music collection tend to be more active in the research of new songs - via the purchase of music magazines for example - compared to people with small collections which are much more influenced in a passive manner - via the radio or advertisement for example. Sease et al. also noticed in [Sease 2009] that participants with the largest collections commonly researched new media through external sources or strangers.

Thus, the digitization of our music collection brings a lot of new possibilities, but the interfaces to manage and interact with this content are still limited nowadays. In the following we provide additional details about existing user interfaces for content discovery and retrieval. This research provided us with enough elements to design new interfaces based on familiar concepts which still have their advantages.

5.3 Content discovery and retrieval

Our prototypes are focused on discovery and retrieval of digital music. We can distinguish a few different ways of approaching the problem: recommendation systems, content-based retrieval, visual representations and tangible interfaces.

Discovery through recommendation, a social network approach which takes into account the preferences of my group of contacts to offer me new songs to try out. Similarity measures and collaborative filtering are some of the main techniques involved. Examples of this approach are described in [Chedrawy 2009], [Nakahara 2009]. This approach is suitable for discovering new music and is mostly used in the context of on-line stores, as a marketing tool. We are interested in discovering music within a local collection. Our system should be suitable for re-visiting tasks. We wanted to design an interface that could help finding songs that we may have forgotten that we had. Another important goal was for the interface to ease the selection of songs that fit within a given genre or mood.

A second approach to music discovery is based on the actual content. Query-by-example systems such as the popular Shazam [Wang 2006] and other pattern recognition approaches such as [Blackburn 1998] provide content-based navigation systems. This kind of interface can be used for "re-visiting" purposes within a local collection but still require the user to have a clear idea of what to look for.

Data visualization approaches are a third alternative, they focus on providing the user with visual representations of the music, which can help to distinguish between the available genres/artists. This approach allows for a user to perform a free exploration of the data repository. A representative example of novel visual representations of digital music is presented in [Bossard 2009]. In this system, a music collection is represented and explored through an abstract representation of songs visualized as clusters in a multidimensional space. Another visualization system which implements a graph-like representation and helps with playlists creation is presented in [Crampes 2007]. In [Dias 2010], the authors describe a visualization and exploration tool for large music collections. The visualization is based on a 2D area representation of songs according to different filters and similarity measures. The system is mainly used for automatic creation of playlists based on a seed set, using similarity measures. Other researchers have focused on 3D visualizations of music collections, using similarity measures to create clusters which can be represented as landscapes, such as the work described in [Knees 2006]. This interface is enriched with semantic information and a surround sound model. This can create a compelling experience. However the system may not be adapted for daily use, e.g. the system requires a joypad for navigating the 3D musical landscape, which may not be the most adapted interface for listening to music.

Finally, tangible interfaces are another alternative that move the interface and/or content to external objects. With iCandy [Graham 2008] for example, Graham et al. created a tangible interface for organizing and sharing multimedia content in an iTunes library. They chose a paper-based interface: song, albums and playlists are represented as cards with bar code. Users can select a song to play by browsing the card album and scanning the bar code of the corresponding card. They also created a sharing system based on these cards. A similar system is described in [Brown 2002]. Instead of using cards, the concept is based on books: each book represents one album, and each book contains a small RF tag to identify it and connect it to the online copy of the album. When the book is close enough to a suitable player, the album is downloaded from the internet and played. This is an example of a centralized music system, as addressed in the previous section.

Alonso and Keyson [Alonso 2005] designed the MusicCube, another solution to browse music and physically interact with it Users interacted with the MusicCube, a wireless cube-like object, using gestures to shuffle music and a rotary dial for song navigation and volume control. Pot a Musique [Strachan 2010] is a similar concept that implements a tangible music controller. This device takes a form of a pot - an everyday object that may stay on a desk - and is enriched with inertial sensors and vibrotactile feedback in order to control the music player via gestures.

Next section describes the virtual shelf, a system which combines two aspects: the easiness of interacting with a familiar "classic" shelf and the efficiency of using a remote control to interact with the digital content and with the surrounding audio devices.

5.4 The virtual shelf

Digital lifestyle [Hofmann 2008] is a life "in which we can create, share, and enjoy multimedia information in a personal and mobile environment". Three points are essential to enjoy this digital life:

- an easily accessible multimedia content
- a communication infrastructure for this content
- intelligent devices to deliver it in a user-friendly manner

With virtual shelf, our goal is to offer an application that fits these points. Virtual shelf allows the user to browse her personal music collection in a familiar way, to easily share music with friends, and to transfer music from the shelf to other sound devices using our drag-and-drop remote control described in the previous chapter.

This section is organized as follows: first we present an overview of how people listen to music nowadays. We continue with related works about digital music interaction followed by a description of our virtual shelf application and its implementation.

5.4.1 Concept

Despite the fact that digital media has already invade our homes, interaction models have not changed: it is still hard to interact with our music collection without a computer. Our main goal is to make this interaction more simple and efficient, even for people with few computer skills. We focus our prototype on three points:

- Browsing our music collection in a natural way
- Sharing music between friends
- Sharing music between devices

In order to browse our music collection, we choose to display it in a familiar context: as we had CDs at home stored in a shelf, our application displays albums on a screen inside a virtual shelf. The user doesn't need to enter a login or a password. Indeed, the application detects the user's cell phone, scans the music database and displays his/her collection inside the virtual shelf on the TV screen. With the remote control, she can choose the album she wants to play, and drag-and-drop it on the sound player close to her and listen to it. If a friend enjoys the music, she can use the application too: a second virtual shelf appears on the TV with her music collection, she just has to select the album in the other shelf and drops it in her collection.

To put it briefly, our application adapts the drag-and-drop paradigm to transfer digital music files from virtual to physical world (from the user's music collection to the selected sound player) and from a virtual world to another (to allow music sharing between friends).

5.4.1.1 Music centralization

The storage of digital music is a real challenge. Keeping our music collection ordered in a single location requires too much time. With this in mind, we decided to use a single place to store user's music collection. For our prototype, we have stored all the music files on a web server, to make it accessible from everywhere.

We have created two MySQL databases to manage the music collections: the first one stores all the basic information about songs (band, album, cover, song name, song number) and their location (a hyperlink to the song path on the server); the second one stores information about users (first name, last name, bluetooth MAC address of the cellphone) and the songs they have in their music collection (a reference to the first database). This is a technical choice above all: sharing music with a friend just consists in adding an entry

Figure 5.1: From top left to bottom right: (I) User's cell phone is detected, a shelf appears with his music collection. (II) The user is browsing his music collection using the MX Air mouse. (III) The user drops an album on the sound player after selected it in the virtual shelf. (IV) Two users are sharing albums from their music collection.

in the database. In the discussion, we get back to that point with the point of view of the testers.

5.4.1.2 Virtual Shelf

The virtual shelf is a simple and familiar application to browse, share and play our music collection.

Hardware description The user interacts with our application with a remote control, which is a custom Logitech MX Air. This mouse is a cordless air mouse which transforms hand motions into a cursor control: the mouse follows the user's hand, using a gyroscopic sensor and an accelerometer. This ability is interesting to control the mouse pointer on a TV for example, from the sofa.

We have stuck an infrared LED under the mouse in order to detect where the user is pointing the mouse. As we did for the prototype presented in section 3.3.2.1, a webcam is connected to a computer near the sound player in order to detect if the user is pointing the device or not. Technically, the sound player is connected to this computer, but for the user experience we hid the computer. The user must have the sensation of pointing the shelf and dropping the digital CD in the sound player directly.

Finally, the virtual shelf computer has bluetooth capability in order to detect user's bluetooth cell phone.

User identification To access her music collection, the user doesn't need to enter a login or password: she is identified with her bluetooth cell phone. As almost everybody has a cell phone, it seems to be the best way to identify transparently the user. To allow this possibility, a Java software is running on the virtual shelf computer. It permanently calls a linux application (hcitools) in order to scan the bluetooth devices in the room, and if a new one is detected, the Java application sends the bluetooth MAC address to the flash-based virtual shelf application using a TCP connection. The virtual shelf application checks if the user is present or not inside the centralized music database. If yes, a new shelf appears with the user's music collection. We limited the application to two shelves (thus, two users): firstly for a size reason, and secondly because it seems useless to allow more users on a single screen (sharing music is often between two persons at the same time).

Browsing music We have designed the virtual shelf to be simple and familiar, with albums stored inside the shelf Browsing a musical collection is really easy, even for people without computer skills. At this point, we have mainly worked on the easiness aspect of the browsing and we used MX Air's capabilities: putting the mouse cursor on a CD displays information about it (album and artist name, songs in this album, CD cover), whereas clicking on the CD animates it as if you remove a CD from a real shelf. Then you can share your CD with your friends, listen to it, or release it and an animation replaces the CD at its initial position.

As we explained before, the music information is stored on a distant database, and not directly on the computer. And because the virtual shelf has been developed in flash, it could be run on other computers with a web browser and flash capabilities (the only additional part to install is the bluetooth java module to detect the user's cell phone). Later, we could imagine that an application such as the virtual shelf would directly run on a Smart TV[6] without the need of an additional computer.

Sharing music As we saw in the previous section, sharing our music with our friends is not an easy task nowadays. You need to parse the music files as any other file type, and sometimes it is even not possible (with ipod or iphone for example if you are not on the computer synchronized with the device). With our application, sharing music just consists on drag-and-drop music from your friend's music collection to your own music collection: you pick up a CD from her shelf and release it on your shelf, and the copy is done. Visually, it appears as a cloning: when you release the mouse button, a new CD appears under the mouse cursor, and goes next to your other CDs, whereas the original CD goes to its initial position. Technically, the virtual shelf application adds new entries in the user's music database, which are the same entries already present inside the other user's music database. This way, the copy is done instantaneous: the music files are not duplicated, only the database entries are.

Playing music To play the CD selected on the shelf, the user has two possibilities:

- Dropping the CD on the jukebox displayed between the shelves.
- Drag-and-dropping the CD from the flash application to the physical sound player

For the first point, the virtual shelf application just plays the CD on the TV speakers. There is no technical challenge with this solution: Flash has

[6] http://en.wikipedia.org/wiki/Smart_TV

sound capabilities and could play distant MP3s without any problem. So when a user drops a CD in the jukebox, we just launch the album reading by giving to Flash the MP3 hyperlink of the first track. The mouse wheel allows to change the album tracks, and the mouse click on the jukebox will turn off the current music.

The second point was more challenging to implement. In the previous chapter, we already developed an application to transfer files from TV to a digital photo frame. We have adapted our software for the virtual shelf application. An additional application dedicated to the webcam image analysis is running on the sound player computer, and informs the virtual shelf application about it. If the user selects a CD and releases the button when the mouse is pointing to the sound player device, then the sound player computer plays the corresponding album (thanks to Java Media Framework). When the sound player is on, the user can stop the music by pointing the device with the mouse, clicking on the left button and releasing the button outside of the webcam area: this way, the user drag-and-drops the sound outside of the music player. Technically, in order to perform the mouse LED detection, we have used cvBlobsLib, a blob detection API based on the OpenCV API. Coupled with an infrared LED and a webcam that filters visible light, detecting if the user is pointing or not the mouse in front of the sound player is functional.

Next part is a discussion about the opinion of the participants who tested our application.

5.4.2 Discussion

Twelve participants tested our application, and gave their feedback about it. They particularly enjoyed the efficiency of drag-and-dropping albums from a collection to another and from the collection to the sound device. Different remarks and concerns have been made about some points of our application after a little interview post-test.

Concerning the music centralization, the place where the music has been stored seems very important for the user. On one hand, they really enjoyed the idea of having a place accessible from everywhere (at home or on a distant server) to store their music collection, for the convenient aspect of having only one collection. On the other hand, users are still suspicious with online storage instead of local storage, for connectivity reasons. For instance, one participant pointed out that music will not be accessible during a flight the music is online. Ideally, online and local storage (for an MP3 player for example) should be synchronized in both ways, to be accessible even without any connection.

About the identification process, users find it too invasive. Indeed, our system automatically detects user's phone and immediately displays her music collection. An additional step is necessary to give more control to the user. With the recent smartphones and their available SDKs, it could be easy to create an application which asks the user if she wants to display her music collection on the virtual shelf or not.

Browsing on the virtual shelf has been described as easy and natural. But the possibilities were limited, and not adapted for browsing big music collections. Albums are classified by names, and other criteria should be added. For example, one shelf level could be dedicated to one genre. Creating and managing playlists is another option to add, without reducing the natural aspect of the application.

Sharing music between two collections has been described as really easy and interesting, but raised some legal questions. We mainly focus on how to share media with our friends in the Web 2.0 era, not on how resolve these legal questions. Furthermore, our centralized system for music collection could be adapted to deal with these problems: for example, instead of copying the album from one shelf to another, the system could ask to the user if she wants to buy this album on a digital music store. Another example : on a music store with a flat monthly fee (such as the Zune Pass or Spotify), subscribers can download unlimited music. Thus, copying a track from one shelf to another could also inform this music store about the transfer, and then adapt the remuneration of each artist depending on the number of downloads.

Finally, our testers have really enjoyed the music transfer from the shelf to the music player using the drag-and-drop paradigm. They described it as convenient and efficient. The only problem with this system is that it only works for devices closed to the virtual shelf. For a multi-room usage, we could visually display the other sound players on the screen where the virtual shelf is present. Then, with just one virtual shelf, rooms would be displayed as in our prototype presented in section 3.3.3, and with two virtual shelves - when two persons want to share music - the rooms would be hidden by the second shelf. Indeed, when two persons are sharing music, they probably don't need to listen to music in a distant room. They would most likely listen to music in the current room in order to "try" it before putting it in their collections.

Next section presents our second prototype, which is also dedicated to digital music collection. Based on an "analog radio" metaphor, digital tracks are visualized as radio frequency bars. The goal is to provide a familiar and intuitive interface to browse digital music collection, but also to propose an efficient way to browse this content by flattening it as a one-dimensional space.

5.5 Vintage Radio Interface: Analog Control for Digital Collections Navigation

In this research, we address the problem of content discovery and retrieval in personal digital collections. In particular, we target the browsing -discovery- and selection -retrieval- of music within a personal repository. Personal digital collections have reached sizes which are difficult to manage with "conventional" interfaces based on scrolling lists and hierarchical menus.

Scrolling lists can be too slow to locate a song within a set of thousands of tracks. Having a personal collection of more than five thousand songs is quite common nowadays [Brinegar 2010]. Hierarchical menus may require too many "clicks" to reach to the desired album and then we are still confronted to a scrolling list.

Direct search by typing a song's title can be the fastest solution, provided that we know what we are looking for. However, when listening to music this is not always the case. Depending on a number of factors, people like to freely browse their music collection in order to (re)discover tracks, jump from one genre to another, and so on.

Common interfaces seem cumbersome when it comes to casual content discovery and retrieval. This work is focused on free -casual- browsing of large music collections, and on offering a compelling way to freely explore a given collection and retrieve content with minimal effort from the user.

5.5.1 Concept

We have implemented a system based on a compelling visual representation of a music collection. Our main goal was to simplify both the data organization and the interaction mechanisms required for browsing and selecting content. Instead of designing a novel abstract visualization scheme, we turned to a more familiar representation which has proofed its efficacy since well before the digital age: the analog radio tuner. We based our data visualization on the old analog radio tuners. Analog radio tuners represented the receivable frequencies on a number line. We observe that the old radio tuner interface is an intuitive mechanism for discovery and retrieval. It provides a simple interaction mechanism and a clear visualization of the "data collection". Navigation is done by turning a knob. A marker shows the frequency being tuned and provides feedback so the user can easily remember the part of the collection that has been already visited. To the best of our knowledge, a radio tuner metaphor has been applied to digital music only once before: the MIT Media

Lab website[7] contains a brief description of a "Radio-ish" Media Player, developed by Barry Vercoe and Wu-Hsi Li. We were unable to find more details about this or similar designs.

Our prototype presents two main aspects :

- The one-dimensional navigation in the music collection
- The vintage radio user interface

5.5.1.1 One-dimensional navigation

We designed our interface with one key idea in mind: getting over with the hierarchical organization (genre/artist/album) which is the prevailing paradigm in current music players. Even though hierarchical organization is efficient to find a specific track, it is not adapted for free navigation in a complete music collection : users need to perform several selection actions -clicks- before they reach the track level. It is painful for browsing a music collection, especially when you don't know what you want to listen to.

To avoid this hierarchical organization, we decided to display the music collection as a one-dimensional space. Two incremental versions of a prototype were developed. In the first version, we designed a simple visual interface (see figure 5.2), with a text list of all the songs. Tracks were ordered by album, artist and genre, and the first song of each album had a cover art on the left.

To control this interface, we use a Griffin PowerMate, a USB control knob[8]. This kind of controllers are mainly used in video editing: they are precise and fast to navigate in one-dimensional content. Control knobs are also familiar in the music environment since they are often used as volume/frequency controllers on analog radios and on Hi-Fi systems.

We used the first prototype to obtain early feedback about how the radio graphical interface should look like. We realized the interface should be more visual, with less text. With too much text, the user doesn't know where to focus her attention. Secondly, it is important to give dominant place to the important information. During the browsing phase, it is not crucial to have several track names displayed at the same time for example. Finally, it is preferable to have a more global view over the collection rather than only two or three albums displayed at a time. Based on these assumptions, we developed the second version our vintage radio concept.

[7] http://www.media.mit.edu/research/groups/music-mind-and-machine
[8] http://store.griffintechnology.com/powermate-622

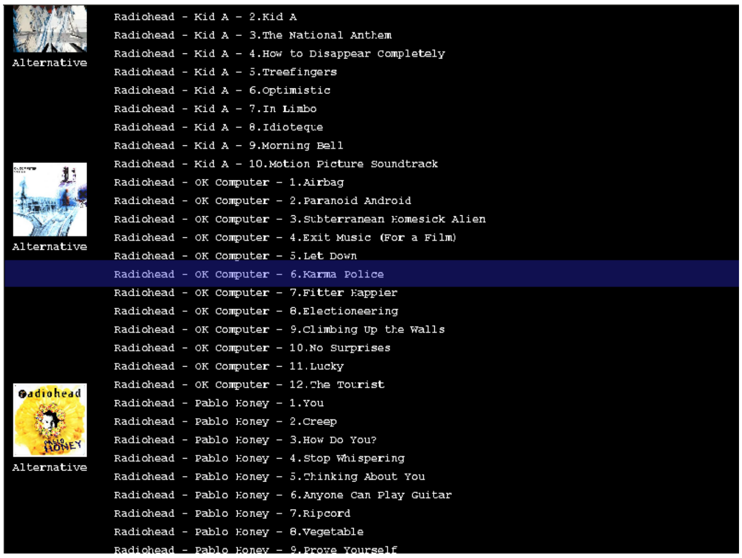


Figure 5.2: First version of the graphical interface

5.5.1.2 The vintage radio

In order to go a step further in the radio paradigm, we drew inspiration from the old analog radios to conceive our graphical interface. Our goal was to have on one hand a familiar and pleasant device, and on the other hand a practical interface to explore a digital music collection.

Hardware In terms of hardware, we conceived a prototype based on a 8" LCD screen and the USB control knob as shown in figure 5.3. We integrated both components in a wood frame: in order to complete the radio metaphor, we wanted the user to feel that it was self-contained device.

The screen and the USB knob are connected to a computer on which the vintage radio software is running. The 8" screen is simply configured as a second screen on this computer.

Software On the software side, the graphical user interface has been developed using Processing, a open source API dedicated to fast prototyping. A C program is running at the same time to communicate the information from the USB knob to the GUI using UDP message passing. The Processing program parses the music directory given as parameter, and reads the artist, album, genre and title ID3 tags from each mp3 file. The songs are then classified and

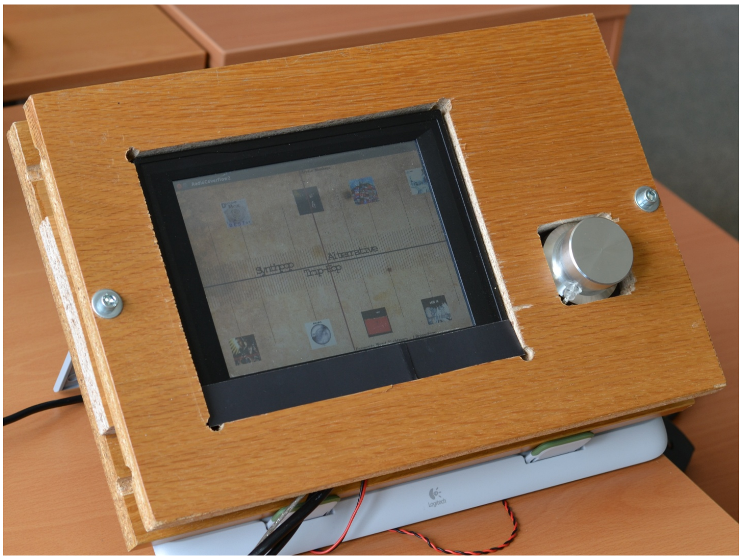


Figure 5.3: Our vintage radio prototype

displayed on the screen using lines to remind the analog radio systems. The music is visually organized as follows:

- The red line on the middle is the current playing song
- Each major line represents the first song of an album
- Each minor line represents a song of the current album

Albums are classified by genre and by artists. Each genre is displayed when it begins, on the "frequency line", and is displayed again every five albums. Figure 5.4 gives an overview of this interface. A video showing the system in action can be watched on-line[9].

The movement and acceleration of the frequency line is mapped on the knob rotation. Acceleration is linear according to the knob rotation speed, and is limited by a threshold. Deceleration starts when the knob returns to its inactive state and is mapped to the current speed (fast at the beginning, slow at the end). During the rotation of the knob, a "radio noise" (the familiar mix of white noise and disturbing frequencies) is emitted to remind the sound of analog radio frequency changes.

[9] http://www.youtube.com/watch?v=oxTWrBi94hM

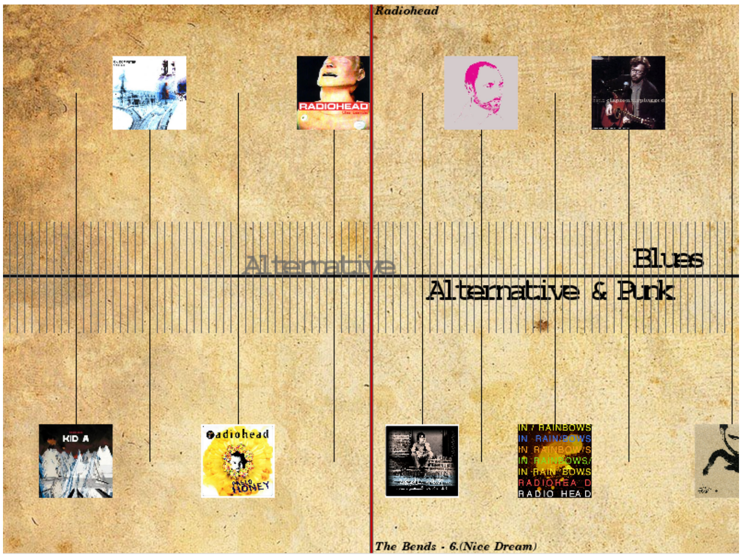


Figure 5.4: Graphical user interface of the vintage radio prototype

5.5.2 Evaluation

We conducted a user study with 24 participants. The main idea was to get feedback concerning the usability of the interface, its efficiency and general acceptance.

Experiment The participants were asked whether they were familiar with the iPod music player. We consider that this is a fairly standard and representative interface for navigating music collections and we use it as the reference model. All the participants were familiar with this kind of device, with two exceptions who did not have a digitized music collection. When participants didn't have an iPod at home, we gave them one to use during five minutes in order to freely navigate in the music collection, and then find a particular track and artist. We decided to compare our interface with the iPod interface for several reasons. Firstly because it is the most popular MP3 player. Secondly, because it is adapted to play music in the living room, with a docking station. Finally, because almost all music interfaces are using this same hierarchical organization.

After the iPod testing, we explained the purpose of our vintage radio prototype, and let the participants play with it during five to ten minutes. Participants first had to navigate freely inside the music collection, and then were asked to find a particular track and artist.

The music collection size of our prototype was 100 albums, 1296 songs in total. Since the participants were not completely familiar with the music collection, we did not evaluate the performance time to find the track and the artist. The purpose of these tasks were to give to face our participants with the two phases of navigation: browsing freely inside a collection, and looking for a precise song or album.

After the test, we asked participants to fill in a usability survey. The results are described in section 5.5.3.

User group The participants to our tests were 11 males and 13 females, aged from 24 to 51, with an average age of 33. Professions and school qualifications were diverse. No specific selection criteria were applied except for balancing the number of males and females.

Only two participants didn't have a local digital collection. One participant declared she didn't use digital music because she didn't have the patience to digitize her CD collection. The second one is using an online music streaming service (Spotify).

Our prototype contains 100 albums. It is slightly less than our users mean:

- 11 participants have between 50 and 149 albums.
- 7 participants have between 150 and 500 albums.
- 6 participants have more than 500 albums (with a maximum close to 4000 albums)

We were also interested in how often they are listening to music at home. Ten of them are listening to music everyday, eleven are listening some hours per week, and the last three are rarely listening to music at home.

We wanted to know which device they are using the most at home for listening to music. Six are using laptops with computer-based audio player. Most of them (five) are active when they are listening to music (i.e. they are choosing the songs they want to play). Ten others are using MP3 players in combination with a docking station. Most of them (seven) are passive when they are listening to music (i.e. they just play songs on shuffle mode). The other participants are using network music players: music collections are on a computer but they are using a distant device to access it (such as Logitech Squeezebox or Apple TV). Two of them are mainly using these devices to listen to web radios, whereas the others are listening to their personal music collections. They are mainly passive when they are listening to music with network music players.

5.5.3 Results and discussion

After the test, our participants had to complete two AttrakDiff[Hassenzahl 2003] usability questionnaires: one for the iPod interface, and a second for our vintage radio prototype. For each questionnaire, a 2D graphic is generated, which situates our system in terms of pragmatic quality (i.e. usability of the system) and hedonic quality (i.e. novelty, interest, identification, etc.).

- The iPod interface has been qualified as neutral/task-oriented. Results are shown in figure 5.5. It is good in terms of pragmatic quality, and average in terms of hedonic quality. The confidence rectangle is large for the two dimensions, which means that participant opinions were mixed. In terms of attractiveness, the overall impression is that this interface is moderately attractive. According to our users, the good points of the ipod interface are that it is “practical” and “clearly structured”.

- The vintage radio interface has been qualified as desired. Results are shown in figure 5.6. It is good in terms of pragmatic and hedonic quality, but room for improvement exists. The confidence rectangle is small on both dimensions, meaning most of the participants shared the same opinion. In terms of attractiveness, the overall impression is that this interface is very attractive. According to our users, the good points of the vintage radio interface are that it is “simple”, “stylish” and “good”.

The average values of both questionnaires gave us some additional information. The pragmatic quality and the identification of the user with the system have been described as slightly better with the vintage radio interface compared to the iPod interface. The attractiveness and stimulation have been qualified as much better on the vintage radio interface. This last point is encouraging because one of the main purposes of this interface was to propose something more stimulating than the hierarchical menu.

During the experiment, participants raised several comments. We have classified these into three categories: Organization, Visual aspect, and Interaction.

Figure 5.5: Result of the ipod interface

Figure 5.6: Result of the radio prototype interface

Organization The main general concern from our participants was the organization of the collection. Many of them are not using the genre in order to classify their music collection. Hence, they often don't know the genre of a particular album and then can't find quickly the album they are looking for with this classification. They would prefer an organization by artist or album name. On the other hand, organization by genre is better when you don't know precisely what you want to listen to. Indeed, people usually choose a song according to their current mood, and don't want drastic changes. Thus, classifying songs by genre is also a desired feature.

Another question directly related to this issue is "what will be the next song?". We asked this question to all our participants, and most of them suggested a shuffle in the current genre. A minority prefered to listen tracks in the album order.

Therefore, to satisfy every user, our vintage radio should propose these options:

- Album classification: Genre/Artist name/ Album name/ Album year
- Next song: Next in Album or Shuffle on Artist/Genre/All.

Depending on which classification has been chosen, the text displayed on the center of the vintage radio application would be adapted.

Visual aspect Comments about the graphical interface were very good. All our test users enjoyed the analog radio visual aspect. The main complaint about the graphical interface was about track names: many users would have preferred to see all the track names of the current album at a glance. The reason is that they preferred to know the track names rather than the position of the track in the album. A possible solution to this issue could be to have a zoom effect on the center of the graphical application: the cover art would be bigger, and the name of each track would be vertically displayed after each "small line". The album on the left and right of the current album would keep the current visual aspect.

Another point raised by our participants was the lack of fixed mark during the music browsing. For example, it would be better to fix the current genre in the center of the graphical interface, next to the red bar. This way, it would be easier to read the current genre.

Interaction One of the questions we asked to our users was about the importance of the knob for interacting with the prototype, instead of interacting directly with a touch screen. According to them, it is essential for the vintage radio metaphor: if this device is supposed to remind old analog radios, then the knob is important. For interaction purposes, answers were mixed: the USB knob is an advantage compared to a touch screen (more precise), but not crucial. Ideally, the application should take advantage of both technologies: the knob to navigate inside the collection, and the touch screen to set parameters (such as the album classification discussed earlier). Three of our participants would prefer a stand alone application, to use with their smart phones or tablets.

Another relevant point about the interaction concerns the lack of a remote control. Indeed, as this device is mainly dedicated to take place in a living room, it would be interesting to control it directly from the couch. A simple wireless knob would be a perfect remote control to satisfy this point.

In general terms, we believe our prototype succeeded in providing an enjoyable, yet efficient interface for music discovery and retrieval. From the user's feedback we can conclude that our interface is more appealing than conventional user interfaces such as the iPod's. Besides the suggested modifications aimed at enhancing the functionality and flexibility, the test users liked the concept and were willing to adopt it. This might be a new trend in UI design: creating interfaces for free exploration and discovery. We have revisited an old interaction paradigm that was created to overcome technical constraints: tuning different frequencies and adapted for the digital age. The current trend is to accumulate music, and interfaces dedicated to free exploration instead of searching could be essential in the near future.

Finally, about possible extensions, future researches could focus on how to extend the concept to other domains. We already think about applying our "knob navigation" to web radios: a small rotation would select a radio close to the previous one in terms of genre/style, whereas a fast rotation would select a completely different kind of radio.

5.6 General discussion

Designing familiar interfaces is important in human-computer interaction. As described before, familiarity is closely connected to the notion of intuitiveness, but is also interrelated with unconsciousness, involvement, easiness, comfort and friendliness. These notions are essential for devices that should integrate people's homes.

But whether a familiar interface is intuitive, it doesn't mean it is efficient. Familiarity could not compensate inefficiency, and a user confronted to an inefficient interface will probably stop using it after a short while.

In the virtual shelf prototype, our participants pointed up the benefits of the features around the virtual shelf - the drag-and-drop capabilities for example - but were not so enthusiastic about the graphical interface. They appreciated how easy it was to interact with it, but they never mentioned it when we asked them what they liked the most. Browsing one by one CD albums was adapted when we had small music collections, but it is not adapted anymore with hundreds of digital albums or tracks to browse. It does not mean that the virtual shelf is not adapted at all to the digital universe, but it requires improvements in terms of efficiency to fully benefit of the familiarity of the concept. The same remark is valid for the second prototype: new classification possibilities (by album name, by artists, etc) have to be offered in order to not frustrate people who are not interested in the genre when they are browsing music.

To put it briefly, familiarity does not suppress the need of designing efficient interfaces. It is an "ingredient" to increase user's satisfaction during the first usages of an interface, whereas efficiency will preserve user's satisfaction on the long term.

Several points have to be taken account when dealing with familiarity.

Familiar elements are not the same for everybody. It is often relative to many factors, such as age or origin. For example, the vintage radio paradigm will not be familiar for children who never saw any analog radio. It is the same with gestures: despite a common base, there are sometimes disparities between cultures - with simple head-shake for example .

Concerning the use of metaphors, some people have been concerned that using them too much could be a burden. According to them, always using familiar features could lead to a loss of innovation in interface design. In [Blackler 2005], the authors disagreed as user's knowledge can be applied or transferred in an innovative way, and not always in the same manner.

The popularity of natural user interfaces has also raised several interrogations. In [Norman 2010a], the author expressed that "natural user interfaces are not natural". He described issues relative to current gesture interfaces: these interfaces were released fast on the market, without a real effort of standardization. Thus, even if gesture interaction appears natural, "most gestures are neither natural nor easy to learn or remember. Few are innate or readily predisposed to rapid and easy learning."

Norman also pointed out the lack of visibility in current gesture interfaces. In graphical user interfaces, all possible actions can be discovered through the exploration of the menus. This "discoverable" aspect is missing from current gesture interfaces, and without concrete standards and conventions, it is not easy to identify the possible interactions to perform via a gesture interface.

Similar issues - visibility, feedback, consistency issues, etc - were presented in [Norman 2010b] about touch interfaces.

Natural user interfaces come with a great potential, just as GUI came with a great potential few decades ago. And in the same manner, NUI will have to deal with problems and challenges. A lot of rethinking will be required to provide solutions to current difficulties, such as how to provide feedback, how to give hints to the user concerning possible actions, etc.

It also leads us to really think about what is really familiar and what is not. Talking to another person to express our desires and feelings is familiar and natural. Thus, it could be argued that talking to machines is an obvious choice to design better interfaces. Despite that, it is not that obvious, and even worse, familiarity works against us in this case as people try to communicate with machines in the same way we communicate with persons[Noyes 2001].

To conclude, familiarity is a great tool and source of knowledge in order to create intuitive interfaces, but interface designers have to be careful about what is really natural and familiar and what is not. Tools and evaluation procedures can be developed to evaluate that [Gamberini 2011], but many social factors have to be taken into account when familiarity and naturalness are entering in the design process.

5.7 Conclusion

We presented in this chapter the importance of familiarity in the design of new interfaces. Familiarity, as a fundamental component of intuitiveness, is a central ingredient for computing devices to penetrate people's home. It is a mean to greatly accelerate user's performance and acceptance, but also to propose more friendly representations of our digital data. Tangible and natural user interfaces or the use of metaphors are different possibilities to achieve this goal. In this chapter, we presented two interfaces mainly based on metaphors: the virtual shelf and the vintage radio interface. The first one displayed user's collection as CD albums in a virtual shelf, and permitted an easy transfer of albums between two collections, but also between two devices.

The second one, the “vintage radio” interface, was focused on the representation of the tracks as analog frequencies. These prototypes were examples of the integration of familiarity in interfaces dedicated to the browsing of digital collections. Finally, having familiar interfaces does not per se guarantee having efficient interfaces. Intuitiveness should not be considered as a replacement or compensation of efficiency, and these two factors should be conceived as complementary.

CHAPTER 6

Conclusion

Contents

6.1 Summary of contributions

From the arrival of personal computers at home to nowadays, our digital life has drastically changed. Digital data are now surrounding our everyday life, accessible from a wide variety of computing objects. Indeed, firstly located on a single device, our digital world has migrated to several devices: laptops, digital cameras, mp3 players, digital photo frames, smartphones, etc. But the interaction between these devices and the digital world is neither smooth nor natural. The interaction mechanisms have poorly changed, and are not adapted anymore to our digital ecosystem.

In this thesis, we wanted to explore what could compose the upcoming evolution of our digital world. This evolution should contribute to enhance the way we are interacting with the digital world, and to integrate more naturally this digital world in our daily life. We presented several user interfaces that aim at enhancing the manipulation of our digital media. We classified them into three areas of improvement: unification, automation, and familiarity.

Unification is central in the manipulation of our media. Currently, the digital world is not a coherent whole, but many sets spread over several devices. The Cloud will probably be an important step towards this unification, but new interaction paradigms are mandatory in order to take advantage of this unification. To illustrate this point, we presented tangible drag-and-drop, a spatial technique conceived to easily interact with multiple devices. We described two applications built upon this technique: the Picture Manager (section 3.3.2) and the Multi-room Audio Manager (section 3.3.3). We also compared tangible drag-and-drop with the standard drag-and-drop technique from graphical user interfaces, and concluded that these two techniques are more complementary than rival. But the introduction of new paradigms in devices at home will strongly depend on the good will of industrial actors and on the implication of the HCI community in the creation of new standards.

Automation is important to give coherency to digital media in the physical world. Users need to feel that digital and physical worlds are interconnected, and that actions performed in one world have repercussions in the other one. Automation is also a precious tool to enhance interface's efficiency by anticipating user's desires and by adapting the interface to the current environment. We illustrated the integration of automation at home with two prototypes: Sound Will Follow(section 4.2) and Natural Activation (section 4.3). The first application was designed to benefit from location awareness at home, and to demonstrate the feasibility of such a concept without cumbersome installations. The second application demonstrated the advantage of estimating where user's attention is directed instead of using an unnatural movement in order to start a gesture recognition system. These prototypes demonstrated the possibilities offered by automation through user awareness: performing automatically tasks that the user would do - like shutting down a sound player and starting another one - but also tasks that she would prefer not to do, or she would forget - like doing unnatural movements to trigger a gesture system-. Automating these tasks are naturally increasing interface's efficiency, and give consistency to the digital world by showing to the user that real world actions are affecting digital devices and their content.

Familiarity is an effective factor to increase user satisfaction. Future interfaces to interact with digital content should look like real world objects in order to ease the first uses of a device. A familiar interface benefits from the past experiences of the user, and greatly accelerates user's performance and acceptance. There is several ways to put into practice familiarity. Using metaphors is one of them. We illustrated the use of metaphors with two applications: the Virtual Shelf (section 5.4) and the Vintage Radio Interface (section 5.5). The first prototype was designed as a graphical user interface based on a CD shelf metaphor. It aimed at reminding the way we were classifying our music collection when albums were tangible. The second prototype proposed a representation of digital music collections that reminded frequency bars commonly displayed on analog radios. These prototypes were examples of the integration of familiarity in interfaces dedicated to the browsing of digital collections.

6.2 Limitations and future work

As described above, we wanted in this thesis to explore what could compose the upcoming evolution of our digital world. We worked for that on the creation and evaluation of several prototypes that aimed at improving the control of digital media, and we identified three areas of improvement. The next step would be to go a step further by identifying what are all the key points that would be necessary to build the perfect "bridge" between the physical and digital worlds.

This bridge is probably not universal, and could be different for all users. It depends on the age, the culture, the experience with computers, the way we use and manipulate digital data, etc. This is probably the main limitation of this work: we did not have the chance to evaluate our prototypes with completely different categories of people. The "perfect bridge" is probably not the same for a teenager and a 70 years old man. In a future work, it would be interesting to identify these different parameters and evaluate new prototypes depending on these factors.

Another interesting future work would be to create new interfaces for other aspects of our digital world. We mainly focused this work on the interaction with digital media. Interaction with other components of our digital world should also be improved in the future. There are for example emails and other forms of direct communication via digital means, but also all the social network side which is more and more present in our digital life. The way we are accessing to the richness of information present on the World Wide Web could also be more integrated and familiar than today. An helpful progress to make that possible will be the development of the Semantic Web[1]. With the semantic web, users should be able to find, share, and combine information more easily, and new interfaces could be built upon this progress.

We also focused this thesis on the use of digital media at home. As a next step, it would be interesting to find new interaction possibilities in a different context, such as outside or at work. Outside, the number of devices and screens is much more limited, and the challenge of designing smarter interfaces in these conditions is different. The smartphone has become the main device to access our digital data outside home. Smartphones have the benefit of being small and can be easily transported everywhere. Smartphones are adapted devices to provide unification (thanks to the cloud) and user awareness (thanks to all the sensors), and new interfaces should be developed to increase the

[1] http://www.w3.org/2001/sw/

presence of the digital world. Augmented reality became popular over the last years, and could be a good opportunity for that.

Appendix A

Music Habits of 2011/2012

Since the end of the '90s, our music habits have considerably evolved. Music sales continue to shift to digital distribution, and CDs become slowly obsolete.

We wanted to know how people are listening to music nowadays, and we conducted a survey dedicated to music listening during the summer 2011. We asked people about the size of their digital collection, the music devices they have, the importance of having a physical medium, or how they are navigating in their musical collections.

This survey was sent to students and workers at EPFL[1]. We received 986 answers, from persons from different ages and disciplines. Figure A.1 shows the repartition of survey participant by gender, age, occupation and school.

We categorized the answers in three parts: music storage, music selection and interaction, and music discovery.

A.1 Music storage

The dematerialization of the music content brings advantages: we can carry all our music collection on a single device, which could be our laptop, our mp3 player, our cell phone, etc. It allows an access on our entire music collection everywhere. But a digital album doesn't have the "tangibility" of a vinyl or CD album. We asked our participants if it is still important for them to have albums on physical media. For 63%, it is not important. And in the future, this number will increase over and over: indeed, as we can see on figure A.2, the importance of having a physical medium is twice inferior for people aged less than 20 years compared to people aged more than 45 years.

On the other hand, organizing a digital music collection is often a real challenge. Organizing digital files on a single device is a demanding task [Barreau 1995], doing it for all computing devices is even more complicated. The size of our music collection make it even harder: we often have thousands of digital songs (figure A.4). The possibility of playing music on several devices often results on having different songs on different devices. This pain point

[1] www.epfl.ch

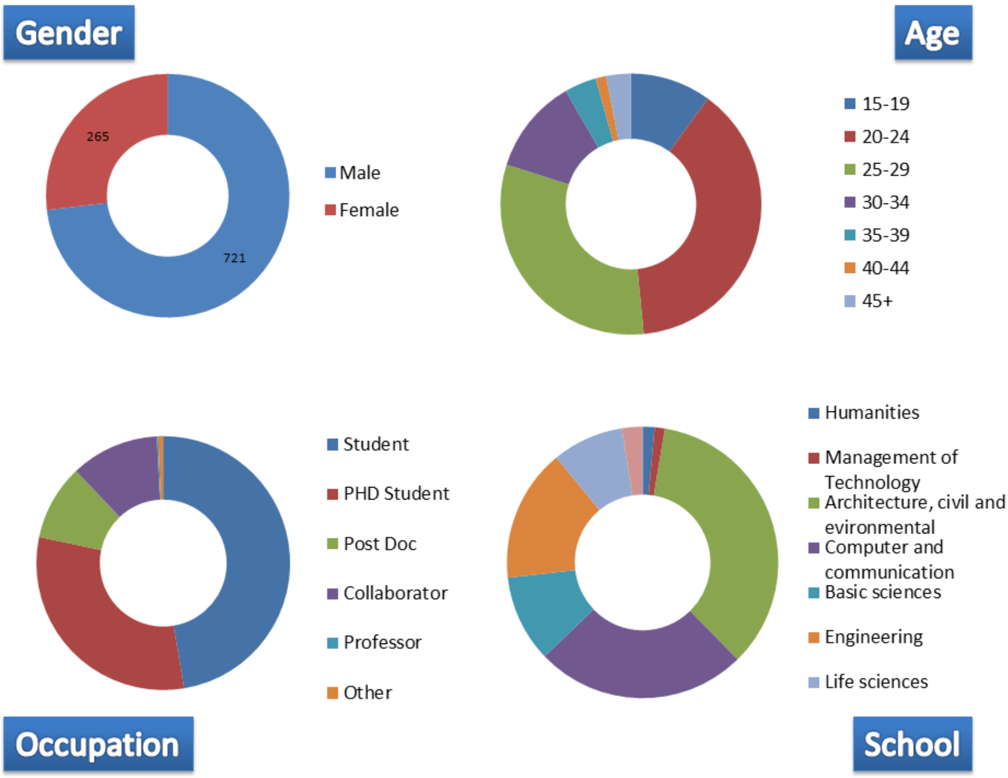


Figure A.1: Gender, Age, Occupation and School of survey participants

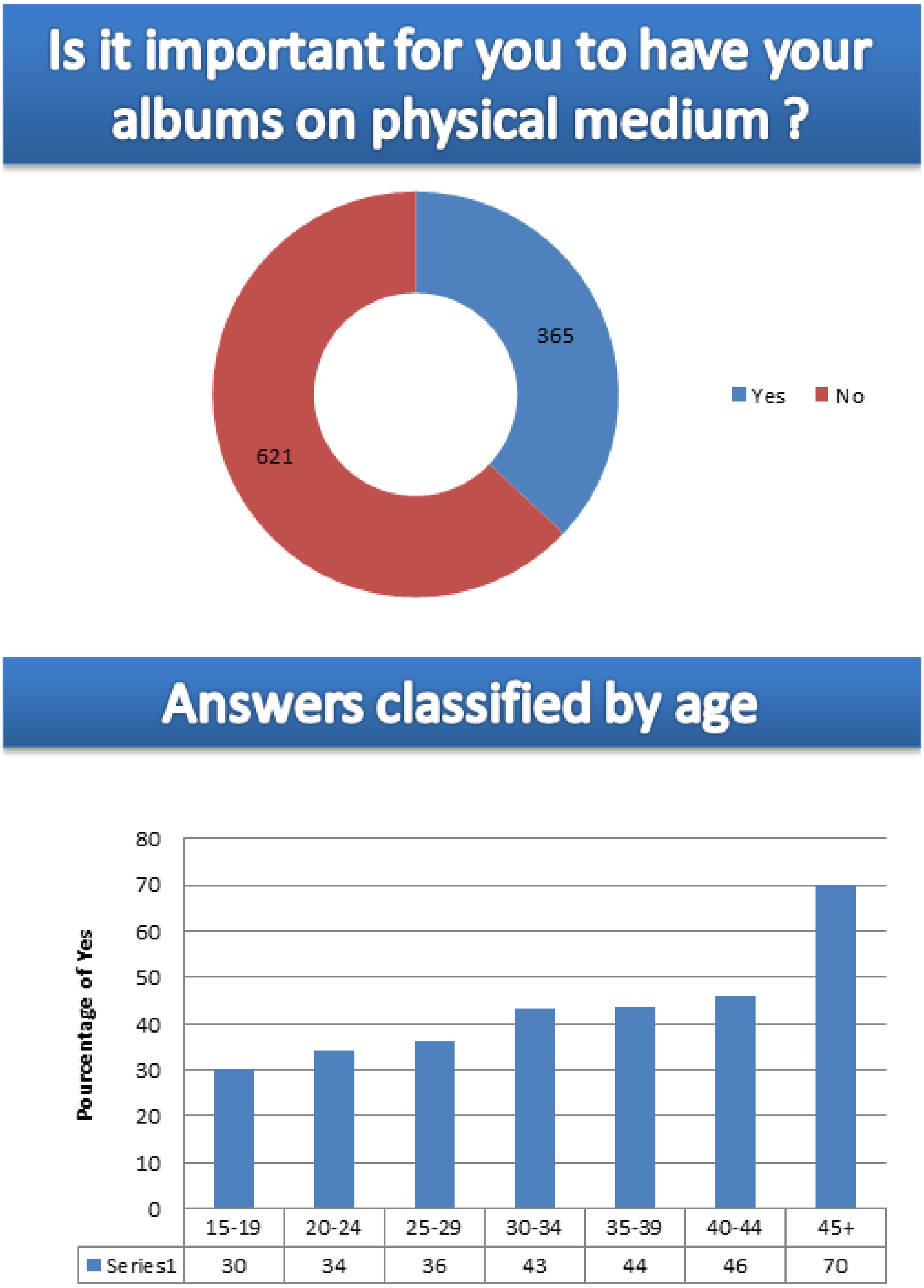


Figure A.2: Importance of having a physical medium was selected as one of the most annoying: it takes too much time to perfectly organize a music collection over all our computing devices, with the full tags and album covers. Moreover, the multiple audio formats make this task even harder.

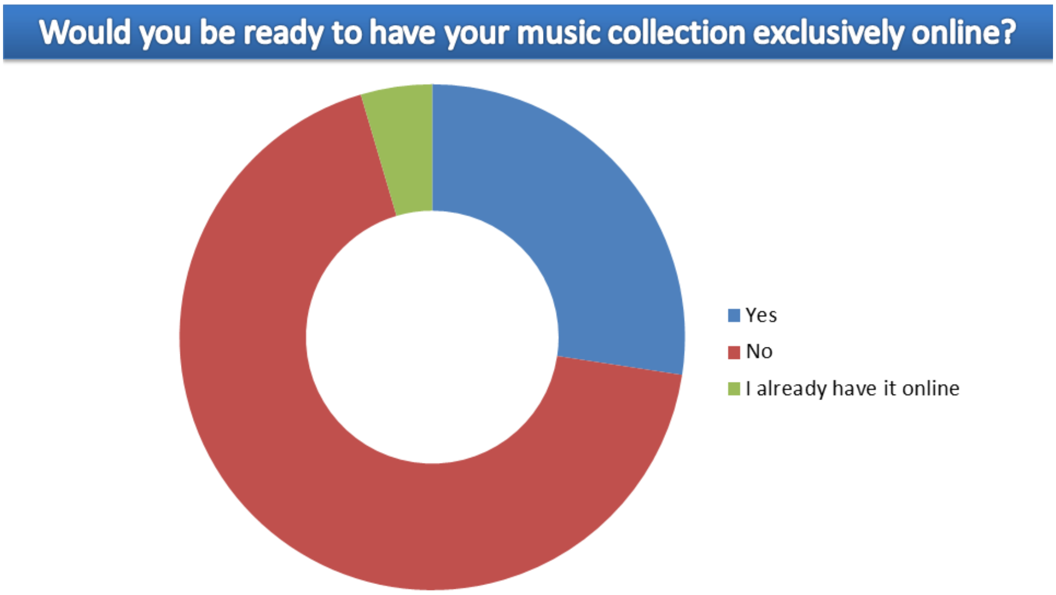


Figure A.3: Results about online music digital collections

Ideally, a single location accessible from everywhere could be a solution to store our digital music collection. Brown et al. [Brown 2001] already suggested in 2001 the "benefits from having digital versions of one's collection being available in a central place, with various networked, digital appliances able to access the collection". The central location solution is nowadays technically possible using online services such as iTunes Match[2] or Spotify[3]. We asked to our participants if they would be ready to have their music collection exclusively online. Figure A.3 shows the results: 68% answered "No" to this question, with an equal distribution among the age (15-19 answered "No" at 66%).

The main reason mentioned to explain this lack of interest is the fear of unavailability: no network means no music. 83% of our interviewees agreed or strongly agreed with this fact. Participants were mixed about the other

[2] http://www.apple.com/itunes/itunes-match/
[3] www.spotify.com

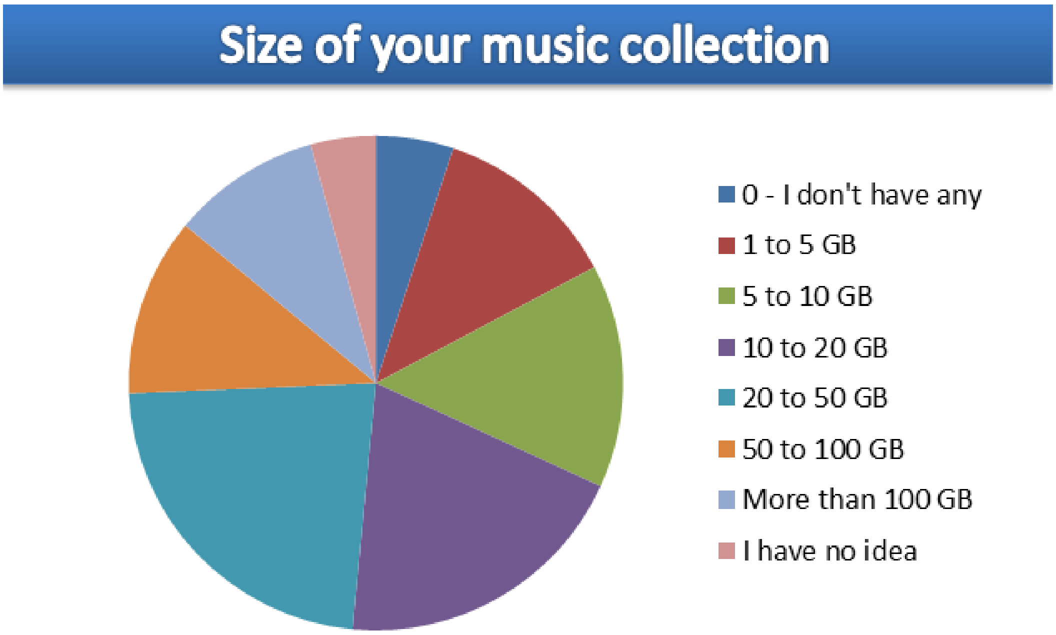


Figure A.4: Size of the music collection of our interviewees

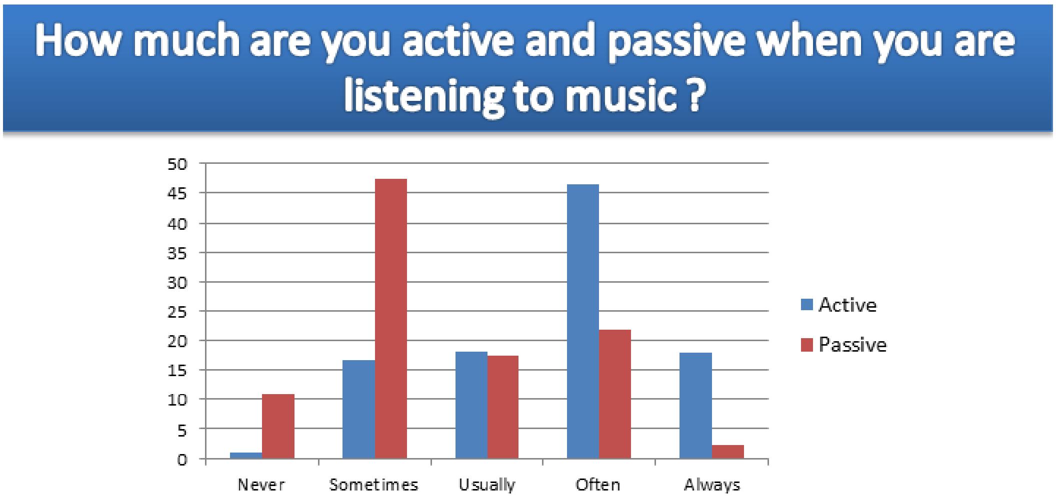
choices we gave. For example, 48% of our participants agreed or strongly agreed that the quality of the compression would be an issue (therefore, this criterion is not important for 52% of the pool's population).

A.2 Music interaction

Building up a music collection is one thing, interacting with this collection is another.

Listening to a music collection can be done in an active manner or passively. When we are active, we browse the collection until we find what we want to listen to. Making playlists is an alternative to select actively digital files. Some tools, such as iTunes' smart playlists, give the user the possibility to select automatically songs depending on some basic criteria (artist name, genre, etc) or more advanced criteria (last played date, date added, personal rating, play count). On the other hand, when we passively listen to music, we let the player choose for us: it could be random mode on a genre, artist, or on the complete collection.

Figure A.5 shows that our interviewees are generally more passive than active when they are listening to music. On the second figure, we categorize the answers for the active part according to the collection size: we can observe that the more the digital collection is large, the more listeners are actively interacting with the collection. The more people are accumulating music, the more they need to actively browse their music collection in order to find a song they want to listen to.



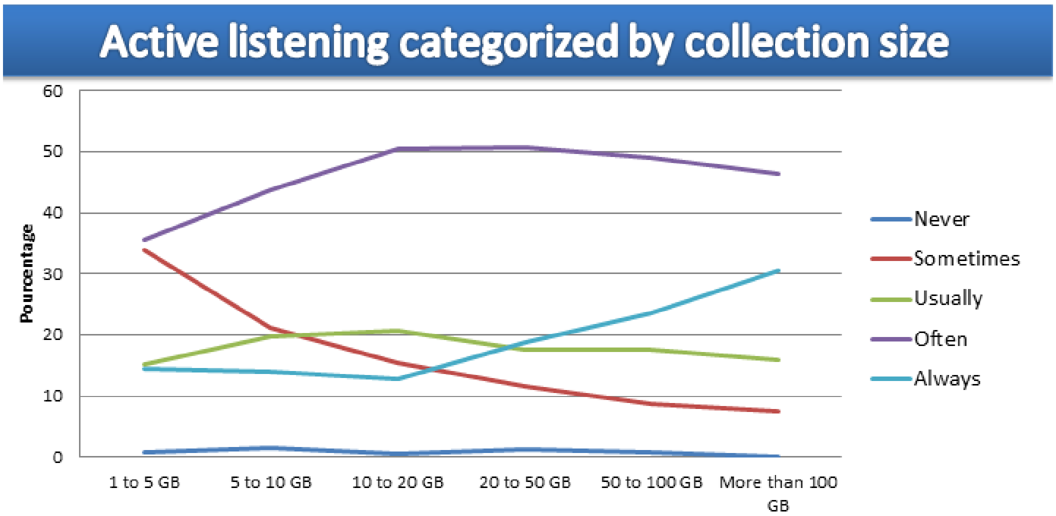


Figure A.5: Passive and active listening of music collection

We noticed that even though people often used manual selection and random mode, smart playlists were not very popular (56% of our pool's population never used it).

We also asked on which devices they were generally listening to music. Figure A.6 gives an overview of the answers. We can observe that CDs are not popular anymore: Hi-Fi is rarely used and portable CD players are almost completely abandoned. The most popular device to listen to music is clearly the computer: 66% of our participants are often or always using it in order to play music. It illustrates the problem of unification described in chapter 3: a computer is not the best device to play music, but people are using it to do so as data are stored on this device.

After the computer comes the MP3 player, which is much more used to play music than smartphones. Finally, streaming audio devices - Logitech Squeezebox[4] or Sonos products[5] for example - are still not very popular.

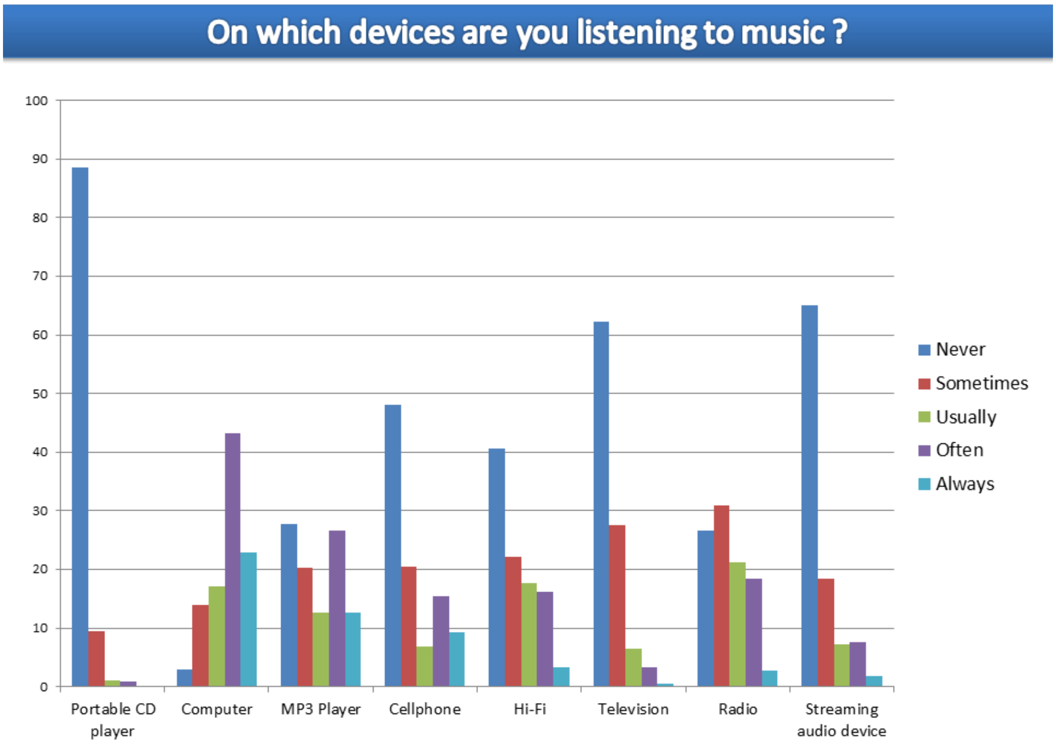


Figure A.6: Question about the devices they are using

[4]http://www.logitech.com/en-us/speakers-audio/wireless-music-systems
[5]http://www.sonos.com/

A.3 Music discovery and acquisition

The last point is about how people discover and obtain new songs.

The main way to acquire new tracks and albums is clearly the download, legal or not: 66% of our participants are using download as the main source of music acquisition. Buying CDs come at the second place, with 25%. The use of free streaming - for example youtube - or dedicated streaming services - like Spotify - is limited (8% for the first, less than 1% for the second). But as we can see on figure A.7, these results are very different according to the age of the participant. Physical media are by far the main way of acquisition for people aged from 35 and more (65% vs 25% for download), whereas younger participants preferred the download way (76% of the 15-19 use download as the main source of acquisition).

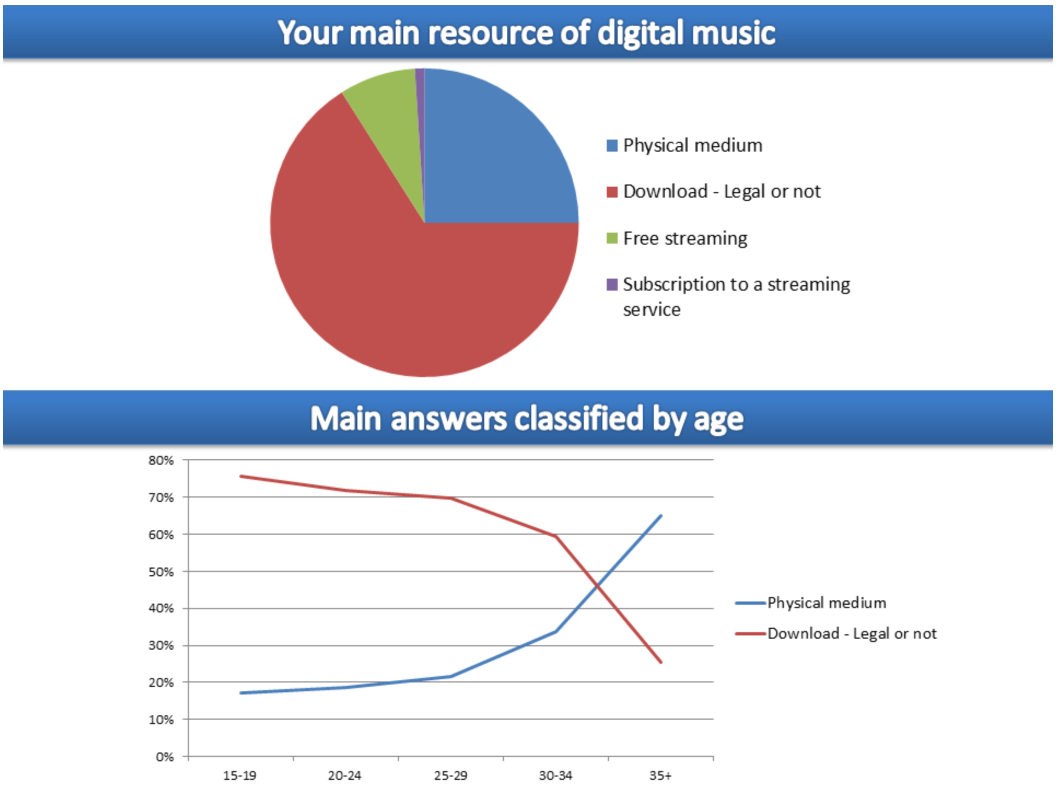


Figure A.7: Acquisition of digital music

Figure A.8 shows how our participants find new songs and bands, The most common way is clearly recommendations from friends. We can also find this observation in [Brown 2001] and [Brown 2002]: friends had an important influence on what people later bought or listened to. 85% of our participants confirmed that, followed by listening radio (65%), recommendation via streaming (32%) and recommendation via social networks (27%). We noticed that these answers are slightly related to the size of the music collection: participants with large music collection tend to be more active in the research of new songs - via the purchase of music magazines for example - compared to people with small collections which are much more influenced in a passive manner - via the radio or advertisement for example. Sease et al. also noticed in [Sease 2009] that participants with the largest collections commonly researched new media through external sources or strangers.

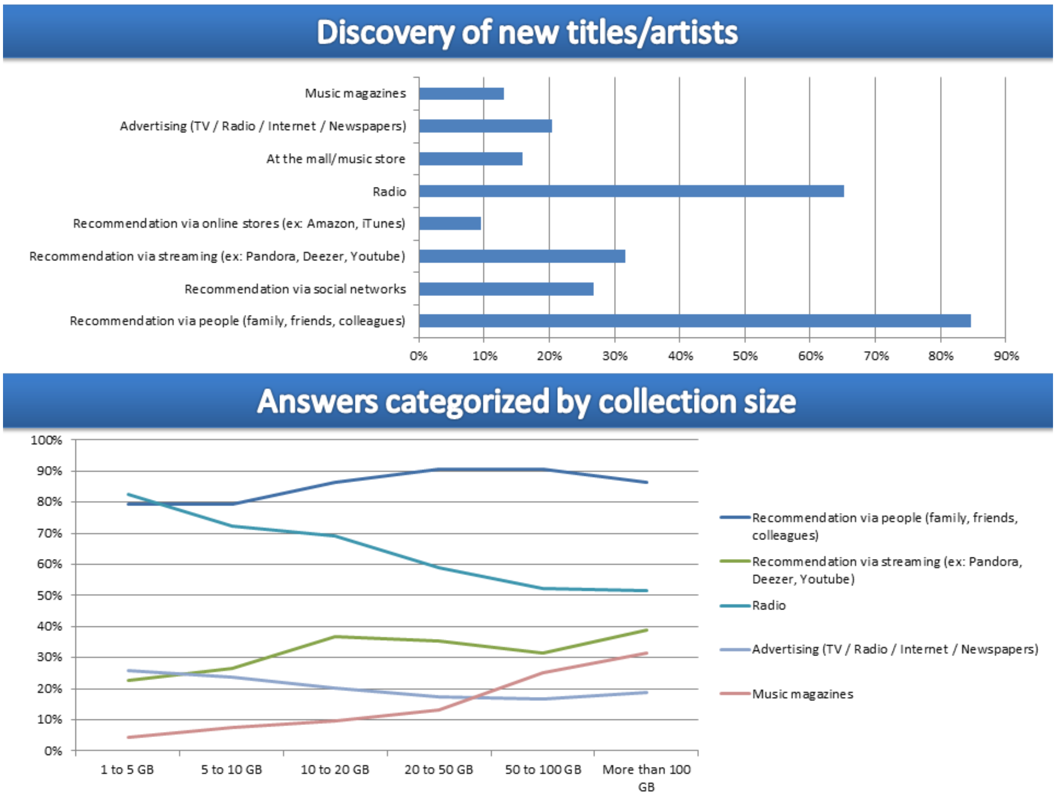


Figure A.8: Discovery of new songs and artists

Bibliography

[Aarts 2002] Emile Aarts, Rick Harwig and Martin Schuurmans. *Ambient intelligence.* pages 235–250, 2002. (Cited on page 44.)

[Aarts 2003] E.H.L. Aarts and S. Marzano. *The New Everyday: Views on Ambient Intelligence.* 2003. (Cited on page 44.)

[Aarts 2004] E. Aarts, P. Res and N. Eindhoven. *Ambient intelligence: a multimedia perspective.* Multimedia, IEEE, vol. 11, no. 1, pages 12–19, 2004. (Cited on page 44.)

[Abowd 1999] Gregory Abowd, Anind Dey, Peter Brown, Nigel Davies, Mark Smith and Pete Steggles. *Towards a Better Understanding of Context and Context-Awareness.* In Hans-W. Gellersen, editeur, Handheld and Ubiquitous Computing, volume 1707 of *Lecture Notes in Computer Science*, pages 304–307. Springer Berlin / Heidelberg, 1999. (Cited on pages 44 and 46.)

[Adamowsky 2000] N. Adamowsky. *Kulturelle Relevanz.* Ladenburger Diskurs 'Ubiquitous Computing', vol. 10, page 2000, 2000. (Cited on page 65.)

[Alonso 2005] Miguel Bruns Alonso and David V. Keyson. *MusicCube: making digital music tangible.* In CHI '05 extended abstracts on Human factors in computing systems, CHI EA '05, pages 1176–1179, New York, NY, USA, 2005. ACM. (Cited on page 76.)

[Araya 1995] Agustin A. Araya. *Questioning ubiquitous computing.* In Proceedings of the 1995 ACM 23rd annual conference on Computer science, CSC '95, pages 230–237, New York, NY, USA, 1995. ACM. (Cited on page 65.)

[Ashbrook 2010] Daniel Ashbrook and Thad Starner. *MAGIC: a motion gesture design tool.* In Proceedings of the 28th international conference on Human factors in computing systems, CHI '10, pages 2159–2168, New York, NY, USA, 2010. ACM. (Cited on page 55.)

[Astrain 2006] J. J. Astrain, J. Villadangos, J. R. Garitagoitia, J. R. González de Mendívil and V. Cholvi. *Fuzzy location and tracking on wireless networks.* In Proceedings of the 4th ACM international workshop on Mobility management and wireless access, MobiWac '06, pages 84–91, New York, NY, USA, 2006. ACM. (Cited on page 47.)

[Atzori 2010] Luigi Atzori, Antonio Iera and Giacomo Morabito. *The Internet of Things: A survey.* Computer Networks, vol. 54, no. 15, pages 2787 – 2805, 2010. (Cited on page 12.)

[Augusto 2007] J.C. Augusto. *Ambient intelligence: the confluence of ubiquitous/pervasive computing and artificial intelligence.* Intelligent Computing Everywhere, pages 213–234, 2007. (Cited on page 44.)

[Bahl 2000] P. Bahl and V.N. Padmanabhan. *RADAR: an in-building RF-based user location and tracking system.* In INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, volume 2, pages 775 –784 vol.2, 2000. (Cited on page 47.)

[Barreau 1995] Deborah Barreau and Bonnie A. Nardi. *Finding and reminding: file organization from the desktop.* SIGCHI Bull., vol. 27, pages 39–43, July 1995. (Cited on pages 18, 72 and 101.)

[Bergman 2000] E. Bergman. Information appliances and beyond: interaction design for consumer products. Morgan Kaufmann, 2000. (Cited on page 11.)

[Berners-Lee 1993] T. Berners-Lee and D. Connolly. *Hypertext markup language.* Internet Working Draft, vol. 13, 1993. (Cited on page 9.)

[Blackburn 1998] Steven Blackburn and David DeRoure. *A tool for content based navigation of music.* In Proceedings of the sixth ACM international conference on Multimedia, MULTIMEDIA '98, pages 361–368, New York, NY, USA, 1998. ACM. (Cited on page 75.)

[Blackler 2005] Alethea L. Blackler, Vesna Popovic and Douglas P. Mahar. *Intuitive Interaction Applied to Interface Design.* In International Design Congress - IASDR 2005, Douliou, Taiwan, 2005. (Cited on page 93.)

[Blackwell 2006] Alan F. Blackwell. *The reification of metaphor as a design tool.* ACM Trans. Comput.-Hum. Interact., vol. 13, pages 490–530, December 2006. (Cited on page 71.)

[Bossard 2009] Lukas Bossard, Michael Kuhn and Roger Wattenhofer. *Visually and Acoustically Exploring the High-Dimensional Space of Music.* In Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 04, pages 58–65, Washington, DC, USA, 2009. IEEE Computer Society. (Cited on page 75.)

[Brinegar 2010] Justin Brinegar and Robert Capra. *Understanding personal digital music collections.* In Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47, ASIS&T '10, pages 136:1–136:2, Silver Springs, MD, USA, 2010. American Society for Information Science. (Cited on page 83.)

[Brown 2001] B.A.T. Brown, E. Geelhoed and A.J. Sellen. *The use of conventional and new music media: Implications for future technologies.* In Human-computer interaction: INTERACT'01: IFIP TC. 13 International Conference on Human-Comupter Interaction, 9th-13th July 2001, Tokyo, Japan, page 67. IOS Press, 2001. (Cited on pages 72, 74, 104 and 108.)

[Brown 2002] Barry Brown, Abigail J. Sellen and Erik Geelhoed. *Music Sharing as a Computer Supported Collaborative Application.* In Wolfgang Prinz, Matthias Jarke, Yvonne Rogers, Kjeld Schmidt and Volker Wulf, editeurs, ECSCW 2001, pages 179–198. Springer Netherlands, 2002. 10.1007/0-306-48019-0_10. (Cited on pages 74, 75 and 108.)

[Chedrawy 2009] Zeina Chedrawy and Syed Sibte Abidi. *A Web Recommender System for Recommending, Predicting and Personalizing Music Playlists.* In Proceedings of the 10th International Conference on Web Information Systems Engineering, WISE '09, pages 335–342, Berlin, Heidelberg, 2009. Springer-Verlag. (Cited on page 74.)

[Chin 1988] John P. Chin, Virginia A. Diehl and Kent L. Norman. *Development of an instrument measuring user satisfaction of the human-computer interface.* In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '88, pages 213–218, New York, NY, USA, 1988. ACM. (Cited on page 15.)

[Choi 2003] Hun Choi, Minseok Choi, Jinwoo Kim and Hyoshik Yu. *An empirical study on the adoption of information appliances with a focus on interactive TV.* Telematics and Informatics, vol. 20, no. 2, pages 161 – 183, 2003. (Cited on pages 10 and 17.)

[Comer 1989] D. E. Comer, David Gries, Michael C. Mulder, Allen Tucker, A. Joe Turner and Paul R. Young. *Computing as a discipline.* Commun. ACM, vol. 32, pages 9–23, January 1989. (Cited on page 66.)

[Cook 2009] Diane J. Cook, Juan C. Augusto and Vikramaditya R. Jakkula. *Ambient intelligence: Technologies, applications, and opportunities.* Pervasive and Mobile Computing, vol. 5, no. 4, pages 277 – 298, 2009. (Cited on page 44.)

[Cooke 2001] M. Cooke, P. Green, L. Josifovski and A. Vizinho. *Robust automatic speech recognition with missing and unreliable acoustic data.* Speech communication, vol. 34, no. 3, pages 267–285, 2001. (Cited on page 71.)

[Crampes 2007] Michel Crampes, Jean Villerd, Andrew Emery and Sylvie Ranwez. *Automatic playlist composition in a dynamic music landscape.* In Proceedings of the 2007 international workshop on Semantically aware document processing and indexing, SADPI '07, pages 15–20, New York, NY, USA, 2007. ACM. (Cited on page 75.)

[Cruz-Neira 1992] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Robert V. Kenyon and John C. Hart. *The CAVE: audio visual experience automatic virtual environment.* Commun. ACM, vol. 35, pages 64–72, June 1992. (Cited on page 8.)

[De La Torre 2005] A. De La Torre, A.M. Peinado, J.C. Segura, J.L. Pérez-Córdoba, M.C. Benítez and A.J. Rubio. *Histogram equalization of speech representation for robust speech recognition.* Speech and Audio Processing, IEEE Transactions on, vol. 13, no. 3, pages 355–366, 2005. (Cited on page 71.)

[Dearman 2008] David Dearman and Jeffery S. Pierce. *It's on my other computer!: computing with multiple devices.* In Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, CHI '08, pages 767–776, New York, NY, USA, 2008. ACM. (Cited on page 18.)

[Dias 2010] Ricardo Dias and Manuel J. Fonseca. *MuVis: an application for interactive exploration of large music collections.* In Proceedings of the international conference on Multimedia, MM '10, pages 1043–1046, New York, NY, USA, 2010. ACM. (Cited on page 75.)

[Dickie 2006] Connor Dickie, Jamie Hart, Roel Vertegaal and Alex Eiser. *LookPoint: an evaluation of eye input for hands-free switching of input devices between multiple computers.* In Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments, OZCHI '06, pages 119–126, New York, NY, USA, 2006. ACM. (Cited on page 20.)

[Dix 1997] A. Dix, J. Finlay, G. Abowd and R. Beale. *Human Computer Interaction (second edition).* Prentice Hall, UK, vol. 1, page 149, 1997. (Cited on page 71.)

[Ducatel 2001] K. Ducatel, M. BOGDANOWICZ, F. SCAPOLO, J. LEIJTEN and JC BURGELMAN. *ISTAG Scenarios for Ambient Intelligence in 2010. Final Report.* European Commission, Feb, 2001. (Cited on page 44.)

[Edwards 2001] W. Edwards and Rebecca Grinter. *At Home with Ubiquitous Computing: Seven Challenges.* In Gregory Abowd, Barry Brumitt and Steven Shafer, editeurs, Ubicomp 2001: Ubiquitous Computing, volume 2201 of *Lecture Notes in Computer Science*, pages 256–272. Springer Berlin / Heidelberg, 2001. 10.1007/3-540-45427-6_22. (Cited on page 65.)

[Enright 2004] Andrew Coulter Enright. *Dissatisfaction Sows Innovation. http://tinyurl.com/2zr7bw*, 2004. (Cited on page 33.)

[Esnaola 2006] Urko Esnaola and Tim Smithers. *Whistling to Machines.* In Yang Cai and Julio Abascal, editeurs, Ambient Intelligence in Everyday Life, volume 3864 of *Lecture Notes in Computer Science*, pages 198–226. Springer Berlin / Heidelberg, 2006. (Cited on page 71.)

[Fang 2007] Yikai Fang, Kongqiao Wang, Jian Cheng and Hanqing Lu. *A Real-Time Hand Gesture Recognition Method.* In Multimedia and Expo, 2007 IEEE International Conference on, pages 995 –998, july 2007. (Cited on page 55.)

[Fitzmaurice 1995] George W. Fitzmaurice, Hiroshi Ishii and William A. S. Buxton. *Bricks: laying the foundations for graspable user interfaces*. In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '95, pages 442–449, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co. (Cited on page 70.)

[Forsythe 1968] George E. Forsythe. *Computer science and education*. In IFIP Congress (2), pages 1025–1039, 1968. (Cited on page 66.)

[Freeman 1995] W.T. Freeman and C. Weissman. *Television control by hand gestures*. In Proc. of Intl. Workshop on Automatic Face and Gesture Recognition, pages 179–183. Citeseer, 1995. (Cited on pages 54 and 55.)

[Friedrich 1983] O. Friedrich. *Machine of the year–The computer moves in*. Retrieved September, vol. 9, page 2002, 1983. (Cited on page 6.)

[Frøkjær 2000] E. Frøkjær, M. Hertzum and K. Hornbæk. *Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?* In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 345–352. ACM, 2000. (Cited on page 14.)

[Galitz 1997] W.O. Galitz. The essential guide to user interface design: an introduction to gui design principles and techniques. Wiley Computer Pub., 1997. (Cited on pages 70 and 71.)

[Gamberini 2011] Luciano Gamberini, Anna Spagnolli, Lisa Prontu, Sarah Furlan, Francesco Martino, Beatriz Solaz, Mariano Alcañiz and Josè Lozano. *How natural is a natural interface? An evaluation procedure based on action breakdowns*. Personal and Ubiquitous Computing, pages 1–11, 2011. 10.1007/s00779-011-0476-z. (Cited on page 94.)

[Glenstrup 1995] Arne John Glenstrup and Theo Engell-nielsen. *Eye Controlled Media: Present and Future State*. Rapport technique, 1995. (Cited on page 56.)

[Graham 2008] Jamey Graham and Jonathan J. Hull. *Icandy: a tangible user interface for itunes*. In CHI '08 extended abstracts on Human factors in computing systems, CHI EA '08, pages 2343–2348, New York, NY, USA, 2008. ACM. (Cited on page 75.)

[Gutierrez 2008] Mario Gutierrez, F Vexo and Daniel Thalmann. Stepping into virtual reality. Springer-Verlag TELOS, Santa Clara, CA, USA, 1 édition, 2008. (Cited on page 8.)

[Hallnäs 2002] Lars Hallnäs and Johan Redström. *From use to presence: on the expressions and aesthetics of everyday computational things*. ACM Trans. Comput.-Hum. Interact., vol. 9, pages 106–124, June 2002. (Cited on page 14.)

[Hassenzahl 2003] M. Hassenzahl, M. Burmester and F. Koller. *AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität*. In Mensch & computer, pages 187–196, 2003. (Cited on pages 37, 59 and 89.)

[Hennessy 1999] J. Hennessy. *The future of systems research*. Computer, vol. 32, no. 8, pages 27 –33, aug 1999. (Cited on page 11.)

[Ho 2006] Wenyao Ho, Asim Smailagic, Daniel P. Siewiorek and Christos Faloutsos. *An Adaptive Two-Phase Approach to WiFi Location Sensing*. Pervasive Computing and Communications Workshops, IEEE International Conference on, vol. 0, pages 452–456, 2006. (Cited on page 47.)

[Hofmann 2008] Gunter Hofmann and Gerhard Thomas. *Digital Lifestyle 2020*. IEEE Multimedia, vol. 15, pages 4–7, 2008. (Cited on page 76.)

[Ishii 1997] Hiroshi Ishii and Brygg Ullmer. *Tangible bits: towards seamless interfaces between people, bits and atoms*. In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '97, pages 234–241, New York, NY, USA, 1997. ACM. (Cited on pages 2 and 70.)

[ISO 1998] ISO. *ISO 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability*, 1998. (Cited on page 14.)

[Jacob 2008] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey and Jamie Zigelbaum. *Reality-based interaction: a framework for post-WIMP interfaces*. In Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, CHI '08, pages 201–210, New York, NY, USA, 2008. ACM. (Cited on page 71.)

[Jain 2011] Jhilmil Jain, Arnold Lund and Dennis Wixon. *The future of natural user interfaces*. In Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, CHI EA '11, pages 211–214, New York, NY, USA, 2011. ACM. (Cited on page 71.)

[Knees 2006] Peter Knees, Markus Schedl, Tim Pohle and Gerhard Widmer. *An innovative three-dimensional user interface for exploring music collections enriched*. In Proceedings of the 14th annual ACM international conference on Multimedia, MULTIMEDIA '06, pages 17–24, New York, NY, USA, 2006. ACM. (Cited on page 75.)

[LaViola 1999] Joseph J. LaViola Jr. *A Survey of Hand Posture and Gesture Recognition Techniques and Technology*. Rapport technique, Providence, RI, USA, 1999. (Cited on page 71.)

[LaViola 2000] Joseph J. LaViola Jr. *A discussion of cybersickness in virtual environments*. SIGCHI Bull., vol. 32, pages 47–56, January 2000. (Cited on page 8.)

[Le Gall 1991] D. Le Gall. *MPEG: A video compression standard for multimedia applications*. Communications of the ACM, vol. 34, no. 4, pages 46–58, 1991. (Cited on page 10.)

[Lee 2008] Hyunglae Lee, Heeseok Jeong, JoongHo Lee, Ki-Won Yeom, Hyun-Jin Shin and Ji-Hyung Park. *Select-and-point: a novel interface for multi-device connection and control based on simple hand gestures*. In CHI '08 extended abstracts on Human factors in computing systems, CHI EA '08, pages 3357–3362, New York, NY, USA, 2008. ACM. (Cited on page 20.)

[Lee 2009] Hyunglae Lee, Heeseok Jeong, Joongho Lee, Ki-Won Yeom and Ji-Hyung Park. *Gesture-Based Interface for Connection and Control of Multi-device in a Tabletop Display Environment*. In Julie Jacko, editeur, Human-Computer Interaction. Novel Interaction Methods and Techniques, volume 5611 of *Lecture Notes in Computer Science*, pages 216–225. Springer Berlin / Heidelberg, 2009. 10.1007/978-3-642-02577-8_24. (Cited on page 20.)

[Lund 2012] Arnie Lund, Jean Scholtz and Nigel Bevan. *Why the CHI community should be involved in standards: stories from three CHI participants*. interactions, vol. 19, pages 70–74, January 2012. (Cited on page 41.)

[MacKenzie 1992] I. Scott MacKenzie. *Fitts' law as a research and design tool in human-computer interaction*. Hum.-Comput. Interact., vol. 7, pages 91–139, March 1992. (Cited on page 36.)

[Magnenat-Thalmann 1987] N. Magnenat-Thalmann and D. Thalmann. *The Direction of Synthetic Actors in the Film Rendez-Vous a Montreal*. Computer Graphics and Applications, IEEE, vol. 7, no. 12, pages 9 –19, dec. 1987. (Cited on page 8.)

[Matsumoto 2000] Y. Matsumoto, T. Ogasawara and A. Zelinsky. *Behavior recognition based on head pose and gaze direction measurement*. In Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on, volume 3, pages 2127 –2132 vol.3, 2000. (Cited on page 56.)

[Moeslund 2003] T.B. Moeslund and L. Norgard. *A brief overview of hand gestures used in wearable human computer interfaces*. Computer Vision and Media Technology Lab., Aalborg University, DK, Tech. Rep, 2003. (Cited on page 71.)

[Mori 1970] M. Mori. *The uncanny valley*. Energy, vol. 7, no. 4, pages 33–35, 1970. (Cited on page 8.)

[Nacenta 2005] Miguel A. Nacenta, Dzmitry Aliakseyeu, Sriram Subramanian and Carl Gutwin. *A comparison of techniques for multi-display reaching*. In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '05, pages 371–380, New York, NY, USA, 2005. ACM. (Cited on page 20.)

[Nakahara 2009] Takanobu Nakahara and Hiroyuki Morita. *Recommender System for Music CDs Using a Graph Partitioning Method*. In Proceedings of the 13th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems: Part II, KES '09, pages 259–269, Berlin, Heidelberg, 2009. Springer-Verlag. (Cited on page 74.)

[Nawaz 2008] T. Nawaz, M. Mian and H. Habib. *Infotainment devices control by eye gaze and gesture recognition fusion*. Consumer Electronics, IEEE Transactions on, vol. 54, no. 2, pages 277 –282, may 2008. (Cited on page 56.)

[Nickel 2003] Kai Nickel and Rainer Stiefelhagen. *Pointing gesture recognition based on 3D-tracking of face, hands and head orientation*. In Proceedings of the 5th international conference on Multimodal interfaces, ICMI '03, pages 140–146, New York, NY, USA, 2003. ACM. (Cited on page 56.)

[Nielsen 1997] Jakob Nielsen. *Usability Engineering*. In The Computer Science and Engineering Handbook, pages 1440–1460. 1997. (Cited on page 14.)

[Norman 1999] Donald A. Norman. The invisible computer: Why good products can fail, the personal computer is so complex, and information appliances are the solution. The MIT Press, August 1999. (Cited on pages 10 and 17.)

[Norman 2010a] Donald A. Norman. *Natural user interfaces are not natural*. interactions, vol. 17, pages 6–10, May 2010. (Cited on page 93.)

[Norman 2010b] Donald A. Norman and Jakob Nielsen. *Gestural interfaces: a step backward in usability*. interactions, vol. 17, pages 46–49, September 2010. (Cited on page 94.)

[Noyes 2001] J Noyes. *Talking and writing-how natural in human-machine interaction?* International Journal of Human-Computer Studies, vol. 55, no. 4, pages 503 – 519, 2001. (Cited on page 94.)

[Oberholzer-Gee 2007] F. Oberholzer-Gee and K. Strumpf. *The effect of file sharing on record sales: An empirical analysis*. Journal of Political Economy, vol. 115, no. 1, pages 1–42, 2007. (Cited on page 9.)

[Parkhill 1966] D.F. Parkhill. The challenge of the computer utility. Addison-Wesley Pub. Co., 1966. (Cited on page 12.)

[Peitz 2004] M. Peitz and P. Waelbroeck. *The effect of internet piracy on music sales: cross-section evidence*. Review of Economic Research on Copyright Issues, vol. 1, no. 2, pages 71–79, 2004. (Cited on page 9.)

[Raman Kumar 2006] K. Raman Kumar, V. Apte and Y.A. Powar. *Improving the accuracy of wireless lan based location determination systems using kalman filter and multiple observers.* In Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE, volume 1, pages 463 –468, april 2006. (Cited on page 47.)

[Raskin 1994] Jef Raskin. *Viewpoint: Intuitive equals familiar.* Commun. ACM, vol. 37, no. 9, pages 17–18, September 1994. (Cited on pages 15 and 70.)

[Rekimoto 1997] Jun Rekimoto. *Pick-and-drop: a direct manipulation technique for multiple computer environments.* In Proceedings of the 10th annual ACM symposium on User interface software and technology, UIST '97, pages 31–39, New York, NY, USA, 1997. ACM. (Cited on page 19.)

[Rekimoto 1999] Jun Rekimoto and Masanori Saitoh. *Augmented surfaces: a spatially continuous work space for hybrid computing environments.* In Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, CHI '99, pages 378–385, New York, NY, USA, 1999. ACM. (Cited on pages 19 and 71.)

[Rekimoto 2003a] J. Rekimoto, Y. Ayatsuka and M. Kohno. *SyncTap: An interaction technique for mobile networking.* Human-Computer Interaction with Mobile Devices and Services, pages 104–115, 2003. (Cited on page 19.)

[Rekimoto 2003b] Jun Rekimoto, Yuji Ayatsuka, Michimune Kohno and Hauro Oba. *Proximal interactions: A direct manipulation technique for wireless networking.* In In Proceedings of INTERACT2003, Sep.-Oct, pages 511–518. Press, 2003. (Cited on page 17.)

[Roth 2010] Daniel Roth. *Time Your Attack: Oracle's Lost Revolution.* Wired, 2010. (Cited on page 12.)

[Rumbolt 1988] R.B. Rumbolt and W.R. McIntyre Jr. *Universal remote control unit*, September 27 1988. US Patent 4,774,511. (Cited on page 19.)

[Said 1990] C. Said. *DYCAM Model 1: The first portable Digital Still Camera*, 1990. (Cited on page 10.)

[Sease 2009] Robin Sease and David W. McDonald. *Musical fingerprints: collaboration around home media collections.* In Proceedings of the ACM 2009 international conference on Supporting group work, GROUP '09, pages 331–340, New York, NY, USA, 2009. ACM. (Cited on pages 74 and 109.)

[Seifried 2009a] Thomas Seifried, Michael Haller, Stacey D. Scott, Florian Perteneder, Christian Rendl, Daisuke Sakamoto and Masahiko Inami.

CRISTAL: a collaborative home media and device controller based on a multi-touch display. In Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, ITS '09, pages 33–40, New York, NY, USA, 2009. ACM. (Cited on page 20.)

[Seifried 2009b] Thomas Seifried, Christian Rendl, Florian Perteneder, Jakob Leitner, Michael Haller, Daisuke Sakamoto, Jun Kato, Masahiko Inami and Stacey D. Scott. *CRISTAL, control of remotely interfaced systems using touch-based actions in living spaces*. In ACM SIGGRAPH 2009 Emerging Technologies, SIGGRAPH '09, pages 6:1–6:1, New York, NY, USA, 2009. ACM. (Cited on page 20.)

[Sharlin 2004] Ehud Sharlin, Benjamin Watson, Yoshifumi Kitamura, Fumio Kishino and Yuichi Itoh. *On tangible user interfaces, humans and spatiality*. Personal Ubiquitous Comput., vol. 8, pages 338–346, September 2004. (Cited on page 70.)

[Sheridan 1992] Thomas B. Sheridan. *Musings on telepresence and virtual presence*. In Papers from SRI's 1991 Conference on Virtual Reality on Virtual worlds : real challenges: real challenges, SRI '91, pages 55–65, Westport, CT, USA, 1992. Meckler Corporation. (Cited on page 8.)

[Slater 1995] Mel Slater, Martin Usoh and Yiorgos Chrysanthou. *The Influence of Dynamic Shadows on Presence in Immersive Virtual Environments*. In Computer Science, editor, Virtual Environments '95, pages 8–21. Springer, 1995. (Cited on page 8.)

[Slater 2003] M. Slater, A. Brogni and A. Steed. *Physiological responses to breaks in presence: A pilot study*. In Presence 2003: The 6th Annual International Workshop on Presence, volume 157. Citeseer, 2003. (Cited on page 15.)

[Stiefelhagen 2002] Rainer Stiefelhagen. *Tracking Focus of Attention in Meetings*. In Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, ICMI '02, pages 273–, Washington, DC, USA, 2002. IEEE Computer Society. (Cited on pages 56 and 57.)

[Stoakley 1995] Richard Stoakley, Matthew J. Conway and Randy Pausch. *Virtual reality on a WIM: interactive worlds in miniature*. In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '95, pages 265–272, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co. (Cited on page 20.)

[Strachan 2010] Steven Strachan, Benjamin Mazoin and Agnes Gimeno. *Pot à musique: tangible interaction with digital media*. In Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, CHI EA '10, pages 3313–3318, New York, NY, USA, 2010. ACM. (Cited on page 76.)

[Talbott 2000] S. Talbott. *The trouble with ubiquitous technology pushers, or: Why we'd be better off without the MIT Media Lab.* NETFUTURE: Technology and Human Responsibility, 2000. (Cited on page 65.)

[Tedre 2008] Matti Tedre. *FEATURE: What should be automated?* interactions, vol. 15, pages 47–49, September 2008. (Cited on pages x, 66 and 67.)

[Van de Walle 2003] G. Van de Walle, P. Turner and E. Davenport. *A Study of Familiarity.* In Human-Computer Interaction–INTERACT, volume 3, pages 463–470. Citeseer, 2003. (Cited on pages 16 and 70.)

[Wallace 1991] G.K. Wallace. *The JPEG still picture compression standard.* Communications of the ACM, vol. 34, no. 4, pages 30–44, 1991. (Cited on page 10.)

[Wang 2006] Avery Wang. *The Shazam music recognition service.* Commun. ACM, vol. 49, pages 44–48, August 2006. (Cited on page 75.)

[Weber 2005] W. Weber, J.M. Rabaey and E.H.L. Aarts. Ambient intelligence. Springer Verlag, 2005. (Cited on page 66.)

[Weiser 1991] M. Weiser. *The computer for the 21st century.* Scientific American, vol. 265, no. 3, pages 94–104, 1991. (Cited on page 16.)

[Whittaker 1996] Steve Whittaker and Candace Sidner. *Email overload: exploring personal information management of email.* In Proceedings of the SIGCHI conference on Human factors in computing systems: common ground, CHI '96, pages 276–283, New York, NY, USA, 1996. ACM. (Cited on page 18.)

[Winograd 1997] T. Winograd. *From computing machinery to interaction design.* Beyond calculation: The next fifty years of computing, pages 149–162, 1997. (Cited on page 14.)

[yi Hong 2009] Jong yi Hong, Eui ho Suh and Sung-Jin Kim. *Context-aware systems: A literature review and classification.* Expert Systems with Applications, vol. 36, no. 4, pages 8509 – 8522, 2009. (Cited on page 44.)

[Zelkha 1998] Eli Zelkha, Brian Epstein, Simon Birrell and Clark Dodsworth. *From Devices to "Ambient Intelligence".* In Digital Living Room Conference, 1998. (Cited on page 43.)

[Zeltzer 1992] David Zeltzer. *Autonomy, interaction, and presence.* Presence: Teleoper. Virtual Environ., vol. 1, pages 127–132, January 1992. (Cited on page 8.)

[Zigelbaum 2008] Jamie Zigelbaum, Adam Kumpf, Alejandro Vazquez and Hiroshi Ishii. *Slurp: tangibility spatiality and an eyedropper.* In CHI '08 extended abstracts on Human factors in computing systems, CHI EA '08, pages 2565–2574, New York, NY, USA, 2008. ACM. (Cited on pages 19 and 71.)

[ZÃ ruba 2007] G. ZÃ ruba, M. Huber, F. Kamangar and I. Chlamtac. *Indoor location tracking using RSSI readings from a single Wi-Fi access point*. Wireless Networks, vol. 13, pages 221–235, 2007. 10.1007/s11276-006-5064-1. (Cited on page 47.)

Mathieu Hopmann
Renens, Switzerland
+41(0)78 922 60 96
mathieu.hopmann@gmail.com

STRENGTHS

- Ph.D in **Human Computer Interaction** at EPFL
- Strong skills in **Java** and **C++**

EDUCATION

2012	Ph.D in Human Computer Interfaces – *VRLab, EPFL*
2007	Master of Computer Science and Multimedia Technology – *ENSEIRB Bordeaux*
2005	Master 1st year of Computer Science, Exchange year – *Université de Montréal*
2004	Bachelor of Computer Science – *Université de Bordeaux 1*
2001	Baccalauréat Scientifique – *Bordeaux*

EXPERIENCES

2009 - 2011 **System Administrator**
Place*: VRLab*, *EPFL* and *Logitech Incubator*
Role: Installation, support and maintenance of computers and servers

2007 **Research Assistant** - 5 months
Place: *VRLab, EPFL*
Role: Integration of the work developed during my Master thesis in the InterMedia European project

2006 **Master semester project** - 4 months internship
Place: *Thales Systèmes Aéroportés – Pessac, France*
Role: Developing a program to translate automatically a binary file into a human readable text, using an XML dictionary

COMPETENCES

Software Development

Programming languages	(Daily used) Java, C/C++
	(Academic) Lisp, Prolog, Smalltalk, SQL
Operating Systems	(Daily used) Windows, Mac OSX, Linux Ubuntu
Prototyping	(Daily used) Processing
	(Occasionally) Android SDK, iOS SDK

Human-Computer Interaction

Development of several prototypes during my thesis: from the concept definition to the **fast prototyping** and **user experiments**.

PROJECTS

Ph.D thesis in Human-Computer Interface (Jan. 2008 to Mar. 2012)

Title: Towards a continuum between digital and physical worlds

Goal: Increasing the presence of our digital collections (music, photo, video) in our physical world

Main developments:

- **Sound Will Follow**: Prototype of a multi-room audio system which locates the user (using a smartphone and Wi-Fi networks) and plays music in the room where he is situated
- **Tangible Drag-and-Drop**: Adaptation of the well-known graphical interface paradigm to the physical world. Development of several prototypes based on the Nintendo's wiimote, such as an interface to easily transfer digital pictures from a television to a digital photo frame
- **Natural Activation for Gesture Interfaces:** Prototype replacing the need of an unnatural gesture to trigger a gesture recognition system by an automatic system based on gaze estimation
- **Vintage radio music browsing:** Visual interface inspired from the old analog radios to easily browse a music collection

Master thesis project (Feb. to Jun. 2007, extended to Dec. 2007 as Research Assistant)

Goal: Developing an indoor localization API based on• Wi-Fi signals. This API has been integrated to an indoor, mobile mixed reality guidance system.

PUBLICATIONS

Journal Paper

- M. Hopmann, M. Gutiérrez Alonso, D. Thalmann and F. Vexo, **Bridging Digital and Physical Worlds Using Tangible Drag-and-Drop Interfaces**, Lecture Notes in Computer Science, Vol. Transactions on Computational Science XII, 2011.

Conference Papers

- M. Hopmann, M. Gutiérrez Alonso, F. Vexo and D. Thalmann, **Natural Activation for Gesture Recognition Systems**, Extended abstracts on Human factors in computing systems (CHI'12), 2012.
- M. Hopmann, P. Salamin, N. Chauvin, F. Vexo and D. Thalmann, **Natural Activation for Gesture Recognition Systems**, Extended abstracts on Human factors in computing systems (CHI'11), 2011.
- M. Hopmann, D. Thalmann and F. Vexo, **Virtual Shelf: Sharing Music Between People and Devices**, 2010 International Conference on Cyberworlds, pp. 53-59, 2010.
- M. Hopmann, D. Thalmann and F. Vexo, **Tangible drag-and-drop: Transferring digital content with a remote control,** Edutainment'09, Volume 5670/2009, pp. 306-315, 2009.
- X. Righetti, A. Peternier, M. Hopmann and D. Thalmann, **Design and Implementation of a wearable, context-aware MR framework for the Chloe@University application**, Proceedings of the 13th IEEE International Conference on Emerging Technologies and Factory Automation, pp. 1362-1369, 2008.
- M. Hopmann, D. Thalmann and F. Vexo, **Thanks to geolocalized remote control: the sound will follow**, Cyberworlds'08, 2008.

Posters

- Peternier, X. Righetti, M. Hopmann, D. Thalmann, G. Papagiannakis, P. Davy, M. Lim, N. Magnenat-Thalmann, P. Barsocchi, A. Kirykou and M. Repetto, **Chloe@University: An indoor, HMD-based mobile mixed reality guide**, Presented at: ACM symposium on Virtual reality software and technology, Newport Beach, California, November 05 - 07, 2007.

LANGUAGES

French	Mother tongue
English	Fluent (level C1)
Spanish	Academic knowledge

INTERESTS

Sports	Tennis, Badminton, Golf, Mountain Hiking
Other	Landscape photography, Motorcycle

SITUATION

29 years old, Single, French nationality