Natural Interaction White Paper

Alessandro Valli, PhD
Natural Interaction, iO

Revised April 17th, 2007

PREAMBLE

Researchers fit in two categories: those who are good at doing things and those who are good at writing things. The author belongs to the first one, so don't expect too much from this reading experience.

PREFACE

About twenty years ago Mark Weiser elaborated a vision that was built around two core concepts, and that is still guiding HCI research today. The first concept is *ubiquitous computing*, it's mainly about hardware, and it is almost fully realized today. Google returns more than 3 millions links about the subject, including conferences, research projects, industry initiatives, etcetera.

But what about the second one, *calm technology*? It is about interaction design, poetry, perception, quality, and it is almost completely unrealized. It is about much more subtle properties than quantities, miniaturization and wireless networking. Google returns less than 50 thousands results on the subject. The world forgot the most precious heritage of Weiser. This text is about calm technology [0] in the public space.

SensitiveTable, direct manipulation of digital content.

INTRODUCTION

The history of interface and interaction design is a path from complexity to simplicity, from machines designed for scientific purposes that could be used only from a few technologists, to pervasive devices, supposed to be simple to use for everyone, in everyday life. Such process deeply influences current design practices, still bound to metaphors derived from technological reasons (e.g. the windows, icons, menus, pointing paradigm). Nowadays, people must often use machines and interfaces based on the GUI paradigms or based on cerebral, abstract approaches (e.g. the hypertext), even in contexts where this is not appropriate, such as when having to do with the real world. In this paper an alternative design practice is proposed, grounded in cognitive and perceptual assumptions, that represents a discontinuity in this trend. The author defines this practice *natural interaction.*

This white paper addresses the problem of the relationship between humans and technology-enhanced spaces and physical objects (later defined as *artifacts*), the purpose of which is the communication of a meaning.

Interaction design is the art of instigating and guiding behaviors (or interaction dynamics) by means of proper static or dynamic stimuli (e.g. the shape of a hammer or the audiovisual feedback of an interface). Natural interaction is defined in terms of experience: people naturally communicate through gestures, expressions, movements, and discover the world by looking around and manipulating physical stuff. The key assumption here is that people are meant to interact with technology as they are used to interact with the real world in everyday life, as evolution and education taught them to do.

The creation of new interaction paradigms and alternative media conventions, that exploit the new machines' sensing capabilities and take care of human spontaneous ways to discover the world, is a great challenge for today's designers; at the same time, interactive technology, in terms of sensors, actuators and narrative intelligence, is still a matter of research for engineers and scientists. The author's work is focused on both of these two aspects, conceived as inseparable activities.

In this paper several artifacts are briefly described, as exemplifications of the theory. This practice work covers different needs, depending on the content that has to be communicated. In some cases a completely *visceral* experience is needed, in others a very *analytical* one. The artifacts can be enumerated from the more visceral to the more analytical:

1. *SensitiveAdv and SensitiveFloor*
2. *SensitiveWindow*
3. *SensitiveWall*
4. *SensitiveTable*

From top to bottom the language goes from extremely simple to (moderately) complex, both from the artifact to the person and viceversa. The sixth artifact discussed, the *SensitiveSpaceSystem*, enables a visceral to analytical cross experience.

The approach here proposed is based on spontaneous, straightforward interaction, in order to let the interaction scheme disappear to users' attention, which thus remains focused on content; it is also aimed at creating an aesthetics of interaction, not focusing merely on usability issues.

Historically, the language between people and machines has been determined mainly by technological constraints, and humans had to adapt to such language; it is now possible to make machines able to adapt to humans' languages, in terms of sensing, presentation and narration [1]. This requires a new language paradigm. A simple example is useful to understand the concept of natural interaction.

VISION

A five-months-old child lying in his cradle, looking upwards with curiosity and interest at a group of toy bees hung over the cradle, flowing around in the air; the neonate stretches his arm in order to grasp the bees, but doesn't reach them. Those colored moving stimuli are very attractive for the newborn, and he reacts expressing his interest.

The design strategy here presented is built around this notion. Its purpose is to radically follow the dynamics described: to present information and content in an attractive and clean way, and to let people express their interest and will spontaneously. The key aspect of the framework is to preserve immediacy, to refuse additional elements that could increase the complexity of the interaction dynamics; this is especially important when computing becomes pervasive and meets the physical space, since aesthetics here is an issue, both in terms of appearance and in terms of interaction patterns, that should be coherent with everyday experience.

Natural interaction is achieved through a combination of many factors; such factors must not be considered by themselves, but have to be analyzed as a whole, since the whole defines the overall experience: it is senseless to discuss sensing of human actions without analyzing the feedback sensed by the subject; it is senseless to design a physical space without a deep understanding of the perceived technology it will host.

SensitiveFloor, water.

SensitiveAdv and SensitiveFloor

The SensitiveFloor is a video projected floor section (not necessarily rectangular) observed by an optical sensor that detects people motions. A variety of audiovisual experiences has been developed, including digital water, mosaics, and games. The system runs at 60 frames per second with a resolution of 4 centimeters.

The SensitiveAdv, although technically similar, doesn't rely on a white matte to project onto; high contrast images are pervasively projected onto the environment (e.g. being it wood or concrete). A small (30 centimeters wide) video or animation on black background (thus *invisible* using good quality DLP projectors) is projected on a wall or on the floor. As someone moves across it the animation expands from that point, involving a much larger area, and expressing content; then the animation goes back seamlessly to the small initial stimulus.

LESS IS MORE

Italian artist and designer Bruno Munari used to say that progress is when things are made simpler. *Less is more.* Simplicity leads to an easier and more sustainable relationship with media and technology.

In current interfaces, contents are often immersed in a bunch of audiovisual objects (e.g. widgets, notification sounds) associated with functions and information; this draws people attention away from the content itself, and makes aesthetics and functional integration with the overall environment difficult, if not impossible. Moreover, this is similar to contemporary culture, where things are always immersed in many opinions

and comments: it is necessary to unleash the power of things, the power of contents, by putting these back in the foreground, following a Thomistic approach. The higher the level of abstraction of the interface, the higher the cognitive effort required for mere interaction. The first direction in which simplification takes place is the removal of any kind of mediation between the person and the machine, to achieve the greatest immediacy. This happens at different levels: interaction schemes, representation of content, information organization, disappearing of devices into interaction-related objects (devices not perceived as technology-related devices).

As technology becomes invisible at all such levels, from a perceptual and cognitive point of view, interaction becomes *completely* natural and spontaneous. It is a kind of *magic*. One of the characteristics of a successful natural interface is thus the reduction of cognitive load on people interacting with it. Simplicity is not necessarily obtained through a reduction of information: it can be provided through order and aesthetics, as in the beautiful stained glasses of a Gothic church.

Chartres, stained glasses.

ANALOGY

Many attempts are being made at redesigning interfaces for multi modal purposes; most of these rely on standard GUI paradigms, extended to deal with multi-point input devices, speech etcetera. The natural interaction framework instead suggests acting more radically. The interface is seen only as a *simulacrum of reality*, seamlessly integrated into surrounding reality. Natural interfaces are modeless, i.e. their behavior does not manifest different functional modes; interruptions are not allowed; as an example, state changes are not marked by confirm requests, but these happen as continuous transformations that can be reversed at any time by stopping the human expression that started them, or by starting an other (exclusive) expression; the transformation will become irreversible only as it is finished. Such interfaces are not grounded on metaphors or paradigms, since these are structures that would introduce unacceptable cognitive leaps; on the contrary,

these are based on a faithful simulation of reality. Users are not required to wear or deal with technological devices; such devices are always concealed into everyday objects and everyday interaction modalities [2].

The objective of the framework is to let people spontaneously interact with digital objects as they do with real ones; to achieve this, *digital objects must appear and behave like real ones*. Physical objects obey to the laws of physics; digital objects don't: digital content may manifest and change in ways that are impossible for physical realities (e.g. images may appear and disappear abruptly) so a series of (simulated) physical constraints is applied to the digital content. It helps to think that actually the *digital doesn't exist*, it's just a form of representation. Paul Dourish wrote: "[Embodiment] strikes to make computation (rather than computers) directly manifest in the world so that we can engage it using the same sets of skills with which we, as embodied individuals, encounter an embodied world. So, it exploits our physical skills, the ways in which we occupy and move around in space, and the ways in which we configure space to suit our needs" [3].

Common real objects are persistent, can't teleport, and their appearance transformations are relatively slow. By applying similar behaviors to digital objects, interaction turns into an intuitive experience; moreover, human perception is well suited to track seamless changes, without involving additional cognitive effort. Ungar and Chang wrote: "User interfaces are often based on static presentations - a series of displays, each showing a new state of the system. Typically, there is much design that goes into the details of these tableaux, but less thought is given to the transitions between them. Visual changes in the user interface are sudden and often unexpected, surprising users and forcing them to mentally step away from their task in order to grapple with understanding what is happening in the interface itself" [4]. Seamless evolution of the content manifestation is aimed at making people precisely aware of the progress of the transformation.

The interfaces follow the principle of continuity; *natura non facit saltus*, nature does not progress by leaps. The interface doesn't introduce novelties that would draw persons' attention: the category of novelty is only associated with content.

Digital objects in natural interfaces are made solid, with a mass. Accelerations and decelerations affect the movement of such objects, as well as the changes of other properties, such as video fade in and out or audio volume shifts. These objects are so made cognitively persistent, and the interface gets a seamless behavior. Digital content gives the illusion of being real as the furniture, the physical setup that hosts interaction. The more the interactive space is coherent, the more users will be able to collaborate and solve interaction conflicts.

SENSING

Alex Pentland wrote in the Scientific American: “The problem, in my opinion, is that our current computers are both deaf and blind: they experience the world only by way of a keyboard and a mouse. [...] I believe computers must be able to see and hear what we do before they can prove truly helpful” [5].

In order to let the artifacts sense the described subset of human expressions, the author developed a set of software modules, designed to process the rough input streams from sensors and extract features useful for the interpretation of users’ interest. Such modules couple with three different kinds of sensors:

- Cameras
- Microphones
- RFID

Interactive frescoes, Palazzo Medici Riccardi, Florence.

Cameras

By means of robust computer vision techniques, video feeds from a collection of digital cameras that observe the interaction space are processed in real-time, providing information about presence and location of people, actions performed and behaviors. In order to make vision sensing reliable and robust to changes in the environment, the cameras used are sensitive only to the near infrared (NIR) spectrum and the scene is illuminated with NIR light; this makes functioning possible also in dark settings, suitable for video projection.

Microphones

Statistical audio analysis methods, similar to those used for speech recognition, have been implemented in order to classify and recognize sounds from the environment. Instead of recognizing specific commands, audio is processed in order to detect hints about behaviors, such as silence, chat, scream, applause, tap. This information can be used to estimate the level of attention of the public near the artifact. Microphones are hidden in the environment (e.g. below the table surface or in the ceiling) and are invisible to users. Several speaker independent speech recognition modules have been developed in order to detect words from very small vocabularies in quite silent environments; the modules are activated only when needed and don't provide a continuously ready speech interface.

RFID

Radio frequency identification technology is used to let the artifact recognize physical objects that are part of the interactive setting. Tiny RFID tags are hidden inside such objects, and antennas are integrated in the environment at specific locations: as the users move the objects towards the hidden antennas, the system detects and identifies these and tweaks presentation accordingly. This method allows non-appearance-based recognition. A range of RFID tags and antennas is available from the industry; in particular, there are devices that are designed to be used and worn by people, and devices that allow the management of collisions (multiple tags present simultaneously inside the reader range).

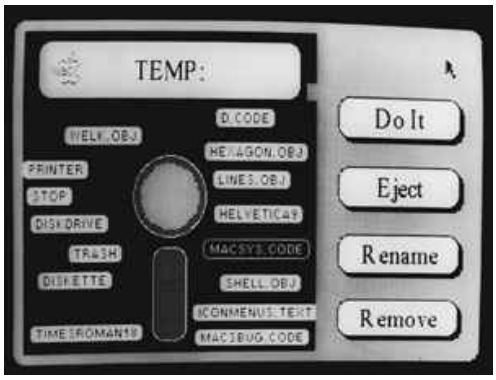
Other sensors have been experimented and used without having to develop custom software modules; as an example, touch sensitive LCD panels or films and OLED buttons allow direct selection and manipulation on small and medium visualizations. All the information provided by the sensing modules has to be integrated and injected into the narrative engine [6].

ACTUATORS

In order to make content manifest in the real world, the artifact exploits bi-dimensional visualization devices, such as LCD panels, plasma screens and video projections; directional speakers, to let people hear audio content in specific locations or from specific directions; odors emitters and controllable light sources; the author also experimented auto stereoscopic displays, holographic (pseudo) three-dimensional visualizations, and micro computers concealed inside everyday objects and tools.

AFFORDANCES

The concept of affordance is a key element in the proposed approach. Mark Weiser wrote: “An affordance is a relationship between an object in the world and the intentions, perceptions, and capabilities of a person. The side of a door that only pushes out affords this action by offering a flat push plate. The idea of affordance, powerful as it is, tends to describe the surface of a design” [0].



Apple Macintosh, an abandoned early concept for the first Finder interface. Files on disk are shown inside a graphical representation of a disk on screen.

Affordances are common to traditional interaction design; nevertheless, the author suggests using these in a very strict, radical sense. Designing things that people can learn to use easily is good, but it’s even better to design things that people find themselves using without knowing how it happened. The interface content and the physical setup are carefully designed so that users can spontaneously and intuitively interact with the space in a successful way; the environment suggests and guides interaction. Ishii and Ullmer wrote: “Our vision is [...] about awakening richly-afforded physical objects, instruments, surfaces, and spaces to computational mediation, borrowing perhaps more from the physical forms of the pre-computer age than the present” [7].

Physical objects that can be grasped and moved play a fundamental role in natural interfaces: since the set of human expressions that is considered spontaneous and general does allow only selection, more complex functions are mapped onto physical tools and objects that can be put in touch or near contents.

Aesthetics of affordances plays a relevant role; science fiction writer Frank Herbert wrote: “A leaping fish had been shaped on the wood with thick brown waves beneath it. [...] If he pushed the fish's one visible eye that would turn on the room's suspensor lamps. One of the waves, when twisted, controlled ventilation. Another changed the temperature” [8]. Why is this portrait so unusual?

DIRECT MANIPULATION

Common computer interfaces made people used to move a mouse on a horizontal surface to control a visual pointer on an almost vertical display surface. On the contrary, natural interaction requires direct manipulation of the objects involved in interaction; the media space and the manipulation space (the space that the person can reach with his limbs) must be coincident or related by deictic projection, in order to give people the illusion of being in a coherent real situation, thus allowing a much easier and satisfactory experience. Gestures are much richer than traditional input methodologies: users are allowed to interact creatively, and express by leaving a sign on the interface surface; moreover, differently from traditional GUI interaction modalities, natural gestures are easily understood by other spectators, and this enables a shared, social interaction experience between actor(s) and spectator(s).

PUBLIC SPACE INTERFACES

The way occasional users approach interactive artifacts in public spaces is very different from the relation between traditional users and personal computers. In this latter case, people are motivated to start interaction, they have a purpose that is clear (e.g. editing a text or checking the e-mail); they know the semantics of the interface, or learn it reading some instructions. Moreover, users are used to deal with a general purpose interface (the operating system's GUI) and voluntarily start the applications they need; after the desired task has been completed, the application is closed. Natural interfaces have a different nature, which can be detailed enumerating some key differences and peculiarities:

- Persons experiencing naturally interactive artifacts are not necessarily active or willing users, they can just be passing by and enjoy passively the encounter, the interface includes a basic reactive behavior for this situation.
- Users are not motivated; the interface must be attractive (like signage) in order to catch the attention and then hold it.
- The interface must suggest that the artifact is interactive, since most people will not think it is; the greatest problem is to convey the initial stimulus, the hint that causes the first voluntary action of the person towards the system; once interaction is engaged, it will be easier for the person to learn the additional interaction capabilities of the system.
- Users don't know how to interact, since there is no common ground of semantics as in GUIs; the interface has to be intuitive and self explaining.
- Duration of interaction is little, a few seconds or a few minutes: one more reason to offer immediate and intuitive access to synthetic content.

- The interface has a *24h* behavior, without splash screens, *begin* and *end* phases of interaction; natural interfaces are not started by the user, these are virtually always on, a user finds the interface as the previous user left it.
- Interactive artifacts are social environments. In public spaces there is often a continuous flow of people: while one or more are actively interacting with the space, there will probably be others looking at them and waiting for their turn. Such spectators *implicitly* train by observing current users while enjoying content presentation. *Imitation* is thus a key dynamics; interaction modalities also need to be learnable imitatively. In order to test prototypes, the author takes as a rule of thumb that a natural interface must be learnable in five seconds of imitation or 20 seconds of intuition maximum.
- Since the artifacts involve even large spaces and large displays, interaction happens at different levels in an extended space: a direct manipulation zone allows active interaction, while a surrounding implicit zone allows more basic interaction or even only public display behaviors.
- Common people are ashamed of trying to interact with artifacts they don't master in front of other people looking at them. Design must address discrete gestures or actions and physical layouts that preserve some sort of intimacy.

SensitiveWindow, outside a bank.

SensitiveWindow

The SensitiveWindow is designed to detect people presence and expressive actions in front of a shop window display (in uncontrolled lighting conditions), and to present contents in a very immediate way, suitable for communication with outdoor passers-by. People positions and walking directions are estimated in order to trigger the appropriate engaging contents (e.g. full screen movies created to *invite* or *surprise* someone walking left to right, etcetera).

The *touchless* display detects people hands up to 20 centimeters from the window glass, and allows detection of content selections events. Hands are tracked at 60 frames

per second; the resolution is 2 centimeters. The interface is based on a series of fullscreen movies, a series of seamless transitions to move from movie to movie, and a series of smaller movies for content selection; this results in a very dynamical shopwindow.

EXPRESSIONS

Interaction, the communication between people and machines, can be described as the play of *human expressions* and *artifact expressions*. These two groups of messages are mutually influenced (i.e. interaction depends on feedback – proper feedback loops can enable spontaneous processes of disambiguation). The next two sections highlight the key features of both categories.

HUMAN EXPRESSIONS

Human expressions can be very rich, subtle, and thus difficult to sense and interpret by computers. In order to facilitate the recognition of behaviors, implicit constraints are introduced through architecture and interface design: people are not instructed how to express themselves, but they are naturally induced to act in ways that can be easily interpreted. This approach makes it possible to avoid introducing explicit instructions or constraints that would catch people attention and distract from the content itself. These implicit constraints range from ergonomic ones to the design of visual elements and continuous interface feedback in time.

Human expressions that are meaningful for the technology-enhanced artifact can be divided into two main categories: *implicit* expressions and *explicit* expressions.

In this framework, the concept of human expression is mainly related to the concept of interest: as the attention of the person gets focused on a particular, this is manifested through expression that the machine can detect and interpret; complex, codified language are thus excluded; the vocabulary is kept as basic as possible.

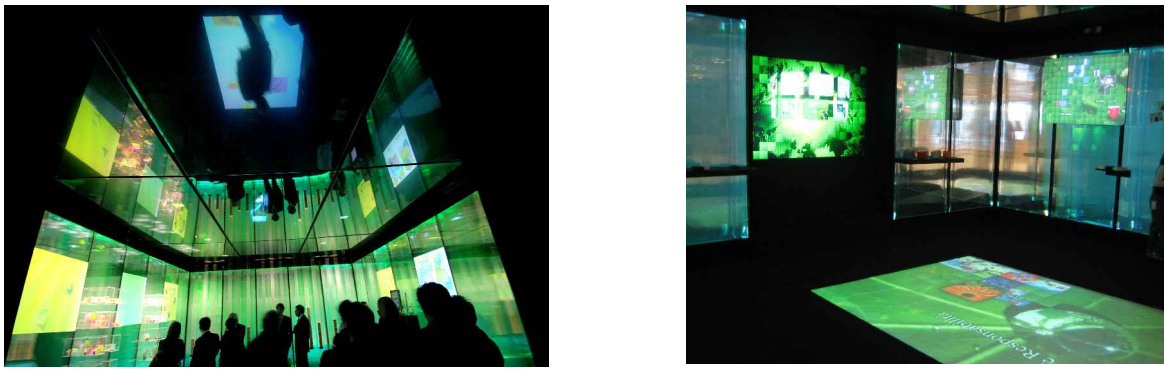
Explicit or voluntary expressions include: touch (for targets inside the manipulation space), deictics (for target outside the manipulation space), manipulation of physical objects (6 degrees of freedom), manipulation of virtual objects (in most cases movements on a 2D surface), and mutual or reciprocal actions on more than one object. Implicit or unconscious expressions include gaze (as a sign of interest, not as a source for visual control), stopping in front of something, getting near something, and affective states, such as calm and anxiety, talking to a fellow or listening to the artifact.

SensitiveSpaceSystem

The SensitiveSpaceSystem is a set of furniture designed to create partitions in public spaces. It is made up of the artifacts described in this document, and computer controlled color wall elements, odour emitters and directional speakers. All the artifacts communicate the sensing and presentation events to a main server, that orchestrates the overall place experience by controlling the color of the wall elements and the odours in the environment, and sending appropriate commands back to the single artifacts, when needed, changing contents or shifting behaviors. These commands cannot take control over contents that are currently used by people, thus leaving a sense of full control to visitors. The stylistic coherence among the artifacts given by this natural interaction framework is fundamental when assembling extensive experience like in the case of these spaces.

WHOLE ARTIFACT

People interaction with technology-enhanced objects or spaces is not simply defined by the nature of the interface in a strict sense; persons are influenced by the physical and social situation they are in (i.e. presence of other people, outdoor or indoor environment, et cetera). For this reason, design must consider such aspects, integrating technology in the overall context: it is the whole context that communicates. To stress the importance of this issue, the author suggests to use the term *whole artifact* to define enhanced spaces and devices integrated in order to be perceived by people as a *coherent* interactive environment. Someway related to this, Bill Buxton's personal mantra states: “Ultimately, we are deluding ourselves if we think that the products that we design are the "things" that we sell, rather than the individual, social and cultural experience that they engender, and the value and impact that they have. Design that ignores this is not worthy of the name” [9].



SensitiveSpaceSystem, two different prototypes.

Note that sensors, computers, and the whole technological infrastructure are not visible to users, concealed in the overall architecture and furniture, so that their attention can be focused on content. The integration of computation and media into physical objects and spaces results in an augmentation: contextualized, intelligent digital content manifest in the environment, thus enabling real objects and places to communicate with people, creating experiences that retain the best of both domains. Glorianna Davenport et al. wrote: "Over the centuries, stories have moved from the physical environment (around campfires and on the stage), to the printed page, then to movie, television, and computer screens. Today, using [...] sensing technologies, story creators are able to bring digital stories back into our physical environment" [10].

Physicality helps people think and learn, and affords interaction modalities. The author implemented interactive floors, tables, walls [11], windows, appliances and rooms [12]; all such settings provide a volume or area to manifest (e.g. visualize) content, and an ergonomic constraint, to *help interaction by limiting the possible actions.*

A visualization of digital information on a table that fits the entire table surface is not perceived as a table with an image on it, but as an enhanced table, as an entity. This shifts the approach of people with it, and involves spontaneous interaction modes that would not be available if it would not fit; it is a kind of experience humans are not used to. Similarly, this happens when content is presented with its actual size in the real world (i.e. a person displayed on a wall with a height of 1.75 meters or a dish displayed on a table with a diameter of 25 centimeters).

Experiments with interactive filmed characters.

William J. Mitchell, from MIT Media Lab, wrote: "Architecture is no longer simply the play of masses in light. It now embraces the play of digital information in space" [13]. The challenge for researchers and designers is to understand that it is something new, which can't be approached with traditional schemes. The author's goal

when designing an experience is to shape the space so that it becomes a place, that has emotional and intellectual impact on people.

By placing wine bottles on the table, their virtues are depicted on the surface.

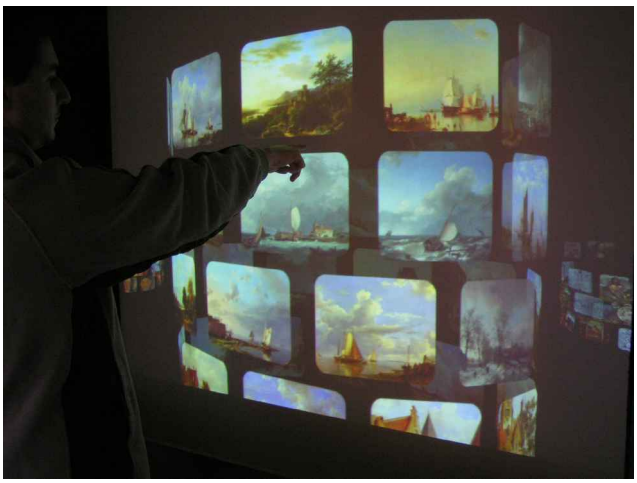
Since the artifact becomes part of a public space, it has to be *aesthetically pleasing*. This requirement also impacts interaction; Donald Norman highlighted how an attractive device improves interaction in terms of usability: since the person is charmed by the object, he will be much more creative in finding out how to interact with it, and will better accept the problems that could arise [14]. Oscar Wilde wrote: "Beauty is a form of genius -- is higher, indeed, than genius, as it needs no explanation" [15]. Aesthetics as an enabler for immediacy: the different features of natural interaction often find unpredictable relations.

PRESENTATION

The way the artifacts express will be described in the following sections. The focus of this text is on visualization on a 2D surface, since it is the most used channel. The purpose of the proposed method is to be in line with the principles enumerated in the previous sections: interface has to be clean, minimal, and support people attention and curiosity.

The problem of natural interface appearance design is closer to the creative problem of film directors and artists than it is to usability engineering. It is the problem of creating an illusory experience, so to enable users' everyday interaction capabilities; for this reason also visual control of pointers is refused. The visual languages to which to refer are photography, cinema, and modern computer games, instead of GUIs; in all such fields content is put in front of users, as immersive as possible; in order to leave people attention on content, functional elements are made less invasive as possible.

Natural interfaces are not static. The audiovisual *liveliness* of the interface, specially when not interacted, conveys a sense of interactivity, and makes the artifact more appealing.



SensitiveWall, Bubbles and Towers interfaces.

SensitiveWall

The SensitiveWall consists in a large vertical *touchless* display able to detect hands' presence in real time, and a set of software templates that present digital content on screen. Hands are tracked in *3D* in front of the display area, up to 30 centimeters from the surface, at 60 frames per second; the resolution is less than 4 millimeters. Three templates (*Bubbles*, *Landscape*, *Towers*) are used to arrange and modulate contents' behavior.

Liveliness is expressed through a fluid continuous motion of contents: attraction and repulsion of bubbles, spin of towers' discs at different speeds, accelerated and decelerated viewpoint translation on a large landscape. Contents can be dynamically added, erased or substituted at runtime through a remote control server. People can move contents just by waving their hands in front of the screen; in order to make contents manifest (e.g. play and zoom for a video) it is sufficient to move a hand close to the content. Special behaviors are enabled when there's no one around the artifact for a long time (it becomes an ambient display) and when someone is passing by (in an area up to 3x2 meters in front of the display), in order to engage people.

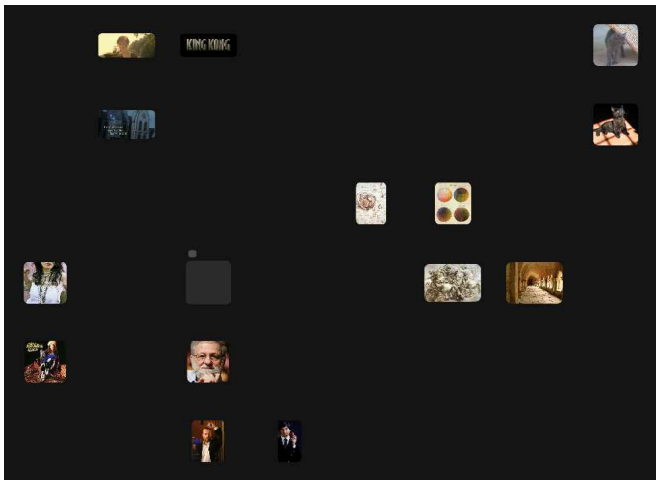
SPATIAL ORGANIZATION

Visuospatial perception is humans' ability to process and interpret visual information about where objects are located in space. It represents the relation between physical space around the person and what the person senses. Human mind is well suited

to deal with information that is spatially located, so digital information can be made more accessible and understandable by a mapping to physical space; for this reason natural interfaces exploit a strict spatial organization.

Contents are arranged spatially, instead of being organized in a series of displays; allowing a simple and coherent navigation which simulates reality. On the contrary, the hypertext navigation paradigm is based on an abstract series of jumps from one piece of information to another, with no spatial reference. All the relations between objects must be actively stored in users' memory, increasing the cognitive effort; in natural interfaces such relations are visualized in front of users. In such a framework, similar contents (according to some context sensitive criteria) are expected to be near, and hierarchical relations are self evident, without requiring to be expressed by additional visual cues.

In order to fulfill such requirements, the author chose to exploit tri-dimensional perspective visualization (among which a 2D orthogonal view is just a particular case); this is well supported by today's hardware, and it also satisfies the fact that the rules that govern the whole representation are unique and minimal. Just a single, full screen (i.e. no windows, no menus, no bars) view is used, where either the whole content world or a single portion that moves coherently (i.e. continuously) is visualized at a time, depending on the particular design (either the objects or the point of view moves, not both). This concept is similar to the zooming user interface [16], but stricter.



Screenshot from a grid based zooming presentation tool.

CONTENTS

The virtual space described in the previous section is populated by pieces of content. A piece of content *c* is defined as a single 3D audiovisual object (possibly associated to smell information), whose appearance changes in time in the general case,

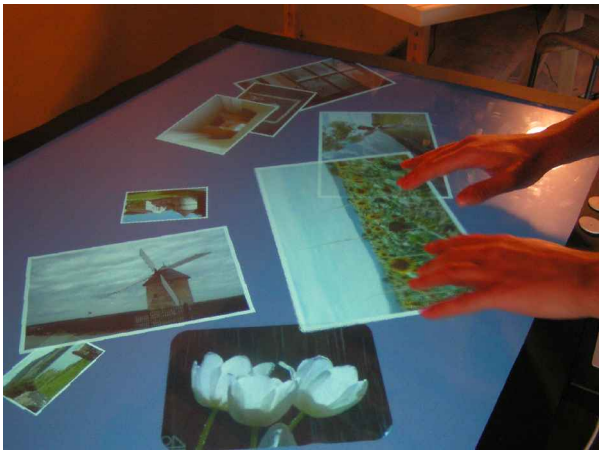
and that can rigidly move in the 3D space. The appearance of the interface I is defined as the sum of all the pieces of content, with no additional elements. Note that there is no mention of widgets or other functional elements, even if some pieces of content may play the same role.

$$I = \sum c$$

Every piece of content has a function to manifest (e.g. play the movie); in addition to this functionality, these can move (3 DOFs), rotate (3 DOFs) and scale (uniform scaling, 1 DOF only, although this is similar to motion along the axis perpendicular to the view plane); their transparency and audio volume can change as well. As already stated, all such transformations are seamless, continuous; every property has an important role towards people attention: size is a natural hint for importance; agitation is a hint for urgency (motion is the sign that something is changing, that a novelty is coming).

Note that all this *removes the concept of icon*, and even the concept of thumbnail, since these will be replaced by the piece of content itself, displayed at different sizes. Moreover, a single piece of content presents only a single face at a time, it cannot be duplicated or shown from different views at a time; the perceived unity of the object is preserved.

As a general criterion, visualization is kept as simple as possible, through a reduction of the graphical elements, fonts and colors; information is split between different channels (e.g. video, sound notification, digitalized speech, written text), since visual, audio and linguistic information is processed in parallel by the human brain, thus reducing the cognitive effort. Whenever possible, high level information is represented by means of elementary sensory stimuli, which can be processed by perceptual intelligence.



Media manipulation on the SensitiveTable.

SensitiveTable

The SensitiveTable is a large multi-touch display that can detects and track the surface of people hands in contact with it at 60 frames per second with a resolution of about 1.5 millimeters. A software application framework allows the creation of custom *natural* experiences. The table is equipped with array microphones and RFID antennas on its edges. The table runs a speaker independent speech recognition engine, based on a very small vocabulary, that is invoked only in specific circumstances. RFID tagged objects are used to populate the interface with contents, activate functions and authenticate users.

Due to its analytical nature (high resolution and multi-point gestures), the table in the public space is used mostly as a form of digital mediation between two or more persons (e.g. consultant and customer): the expert can lead the novice through the more complex and less intuitive dynamics of interaction.

CONCLUSIONS

In addition to the traditional features of interactive digital media, like updatability, freedom of users' to follow their curiosity and interests, logging of people behaviors and real time statistics, the proposed natural interfaces feature a specific architectural aesthetics about how to move computation to the real world, creating immersive experiences that involve people senses in the physical space.

The problem of communication of content is approached from a creative, artistic point of view, in the case of movie pictures, signage and marketing, computer games. It is addressed from an aseptic scientific research perspective in the case of computer interfaces. It has been shown here instead (through theory and practices that work in the real world) how an integrated approach benefices both these domains.

Alan Kay said in 1971: "Don't worry about what anybody else is going to do... The best way to predict the future is to invent it". This is why the approach proposed here is much more involved in progressing in this revolution through working prototypes than through writing scientific papers.

ACKNOWLEDGEMENTS

The author wishes to thank Stefano Baraldi for his work on the SensitiveTable prototype, Marcello Persiani of 3M for providing the exact quote from *Dune*, Daniele

Modesto for his great ideas about the SensitiveAdv, Francesco Buso for suggesting the wonderful quote by Oscar Wilde and all the friends that through their talent and effort make iO a so special company.

Alessandro Valli

ABOUT THE AUTHOR

Alessandro is an Italian interactive systems engineer and experience designer. He received his MS and PhD degrees in Computer Engineering from the University of Florence, in 2000 and in 2004. In 2001 he started focusing on the topic of natural interaction between humans and machines.

After doing academic research and professional consulting for six years, in 2005 Alessandro quit his position at the University to join the iO company, and founded Natural Interaction, a research initiative that investigates intuitive forms of relationship between communicative artifacts and common people in public spaces. Alessandro lives and works in Chianti, Tuscany, Italy.

REFERENCES

[0] Mark Weiser and John Seely Brown, *Designing calm technology*, Xerox PARC, 1995.
[1] Donald Norman, *The design of everyday things*, Doubleday, 1990.
[2] Flavia Sparacino, *Narrative spaces: bridging architecture and entertainment via interactive technology*, 6th International Conference on Generative Art, Milan, Italy, 2002.
[3] Paul Dourish, *A foundational framework for situated computing*, CHI 2000 Workshop on Situated Computing, 2000.
[4] Bay-Wei Chang and David Ungar, *Animation: from cartoons to the user interface*, Technical Report TR-95-33, Sun Microsystems, 1995.
[5] Alex Pentland, *Smart rooms*, Scientific American, Vol. 274, No. 4, April 1996.

[6] Flavia Sparacino, *Sto(ry)chastics: a Bayesian network architecture for user modeling and computational storytelling for interactive spaces*, Proceedings of Ubicomp, The Fifth International Conference on Ubiquitous Computing, Seattle, USA, 2003.
[7] Hiroshi Ishii and Brygg Ullmer, *Tangible bits: towards seamless interfaces between people, bits and atoms*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**,** Atlanta, USA, 1997.
[8] Frank Herbert, *Dune*, Chilton Books, 1965.
[9] http://billbuxton.com
[10] Ali Mazalek, Glorianna Davenport and Hiroshi Ishii, *Tangible viewpoints: a physical approach to multimedia stories***,** Proceedings of the tenth ACM International Conference on Multimedia, Juan-les-Pins, France, 2002.
[11] Carlo Colombo, Alberto Del Bimbo and Alessandro Valli, *Visual capture and understanding of hand pointing actions in a 3D environment*, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 33(4), 2003.
[12] http://naturalinteraction.org
[13] William Mitchell, *e-topia: Urban life, Jim—but not as we know it*, MIT Press, 1999.
[14] Donald Norman, *Emotional design*, Basic Books, 2003.
[15] Oscar Wilde, *The Picture of Dorian Gray*, Lippincott's Monthly Magazine, 1890.
[16] Benjamin Bederson and James Hollan, *Pad++: A zooming graphical interface for exploring alternate interface physics*, Proceedings of User Interface and Software Technology, 1994.