SEACW DELIVERABLE D.3.5

"State of the Art in interfaces human-machine"

Project Acronym	**SEACW**
Grant Agreement No.	325146
Project Title	Social Ecosystem for Anti-aging, Capacitation and Well-Being
Deliverable Reference Number	SEACW_D3.5
Deliverable Title	State of the Art in interfaces human-machine
Revision Number	1.1
Deliverable Editors *(main redactors)*	NIB

Project co-funded by the European Commission within the ICT Policy Support Programme	
Dissemination Level	
P	PUBLIC

Revision	Date	Description
0	2/09/2013	Definition of objectives and expected results
0.1	4/09/2013	Assignment of paragraphs to the contributors
0.2	21/10/2013	Collection of contributions and first elaboration of document
0.3	25/10/2013	First review of the document
0.4	28/10/2013	Elaboration of the document
1.1	1/11/2013	Second review of the document

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Abstract

The Deliverable 3.5 is a first attempt to sketch the different interaction techniques and input/output devices in the communications human-machine. It tries to specify the current trends and to give some directions of the most appropriate user interfaces. They may provide some guidelines about the proper conceptual models and interaction techniques to be used in the work on the ecosystem.

Table of contents

Index of figures

1. Introduction

Human Machine Interface or Interacting, more commonly known as Human Computer Interacting (HCI) is concerned with the joint performance of tasks by humans and machines; the structure of communication between human and machine; human capabilities to use machines (including the learnability of interfaces); algorithms and programming of the interface itself; engineering concerns that arise in designing and building interfaces; the process of specification, design, and implementation of interfaces; and design trade-offs (Kumar, 2005). The ACM HCI provides the following definition "Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them". Because it relates both the human and the machine side in their communication, HCI is an interdisciplinary field of study that combines knowledge from a variety of scientific fields like computer science, cognitive science, psychology, human factors, linguistics and anthropology, neuroscience, sociology, industrial design, engineering.

Human-computer interaction (HCI) is the study of how people design, implement, and use interactive computer systems and how computers affect individuals, organizations, and society. This encompasses not only ease of use but also new interaction techniques for supporting user tasks, providing better access to information, and creating more powerful forms of communication. It involves input and output devices and the interaction techniques that use them; how information is presented and requested; how the computer's actions are controlled and monitored; all forms of help, documentation, and training; the tools used to design, build, test, and evaluate user interfaces; and the processes that developers follow when creating interfaces (Tripathi, 2011). Research in HCI helps to reveal new aspects of human tasks and activities and effective ways to provide better human performance, to understand better the technology and how it may affect the human's life. Its goals are to produce usable and safe system with maximal functionality.

Tufte (1998) characterized the human-computer interaction as communication between two powerful information processors (human and computer) attempting to communicate with each other via a narrow-bandwidth, highly constrained interface. Research in this area attempts to increase the useful bandwidth across that interface. Faster, more natural—and particularly less sequential, more parallel—modes of user-computer communication will help remove this bottleneck. On the user's side of the communication channel, interaction is constrained by the nature of human attention, cognition, and perceptual-motor skills and abilities; on the computer side, it is constrained only by the technologies and methods that we can be invented. Therefore, the efforts are on studying new modes of communication and inventing new devices and technologies, as well as a better understanding of the human cognitive and perceptive abilities and how they are used in performing a certain task in order to find better fit between technological achievements and human abilities.

This document is inevitably limited as it has to cover a very broad and dynamic field of research that is strongly affected by the fast development of technology. Our attempts were to touch all facets of human-machine interaction research, and the different components of user interfaces.

The user interface is the communications media between the human being and the machine. We begin with the presentation of the most conventional input devices as they have shaped the interaction style for some period of time. The communication was restricted to command lines, menus or forms filling-in and a restrict set of manipulations. However, the technology rapidly changes and this can be clearly seen in the novel opportunities for us to discover a

virtual world offered by the new displays i.e. from the typical output of the system; by the possibilities never existing before due to the sensors in the mobiles; by the growing potential for independent life of the disabled people. The technological development changes the ways of communication with the machines. Thus, in our presentation of the human-computer interaction we tried to follow this sometimes invisible thread from the most conventional to the most visionary, from the devices to the ideas like ubiquitous computing, affective, pervasive or transparent computing.

We will not be able to discuss the role of human cognitive and perceptual capabilities on molding the human-computer interaction or the software and hardware elements that are involved in different interaction techniques and as a result our presentation of user interfaces will be limited.

2. General terminology (Hinckley & Widgor, 2002)

- **Input device:** A transducer that senses physical properties of people, places, or things.
- **Conceptual Model:** Coherent model that users form about the function of a system: what it is, how it works, and how it will respond to their input.
- **Interaction Technique:** The fusion of input and output, consisting of all hardware and software elements, that provides a way for the user to accomplish a task, given a particular conceptual model. Interaction techniques typically vary across input devices, based on the strengths of the device sensing capabilities, its ability to really incorporate state transitions such as button-presses into the design of the device, and the user's physical abilities and hand comfort when using the device.
- **User Interface:** The representation of the system – the summation of all its input devices, conceptual models, and interaction techniques – with which the user interacts. It is responsibility of the user´s interface to represent and reinforce the user's conceptual model of the system, in concert with the input device and interaction techniques, as well as to present affordances and constraints that make it clear to users how to achieve key tasks. User interfaces despite being stereotyped as graphical user interfaces, also include auditory, tactile, and kinesthetic qualities, even if such secondary feedback results only from the passive mechanical feedback from the physical input devices.

3. Conceptual models (based on Preece et al, 2002)

We present the main conceptual models in HCI as they will provide a more general framework for understanding the interaction techniques and the interaction devices in human-computer interaction.

3.1 Conceptual models based on activity

The most common types of activities that users are likely to be involved when interacting with systems are:

1. Instructing

This kind of conceptual model describes how people carry out their tasks through instructing the system what to do. Major benefit: supports quick and efficient interaction. Example: print a file.

2. Conversing

This conceptual model is based on the idea of a person conversing with a system, where the system acts as a dialog partner. Main benefit: It allows people, especially novices, to interact with a system in a way they are already familiar with. Examples: search engines, help facilities.

3. Manipulating and navigating

This conceptual model describes the activity of manipulating objects and navigating through virtual spaces by exploiting user's knowledge of how they do this in the physical world. Example: moving, selecting, closing of virtual objects.

A special case of this model is Direct manipulation (Shneiderman, 1983). He described this model as having the following physical properties:

- Continuous representation of object and actions of interest.
- Rapid reversible incremental actions with immediate feedback about the object of interest.
- Physical actions and button pressing instead of issuing commands with complex syntax.

Main benefits:

- Help beginners learn basic functionality rapidly;
- Experienced user can work rapidly on a wide range of tasks;
- Infrequent users can remember how to carry out operations over time;
- No need for error messages, except very rarely as the users could see immediately whether the results of their actions corresponds to their goals and if not do something else; users experience less anxiety;
- Users gain confidence and mastery and feel in control

4. **Exploring and browsing**

This conceptual model is based on the idea of allowing people to explore and browse information, exploiting their knowledge of how they do this with existing media (e.g. books, magazines, TV, etc.).

3.2 Conceptual models based on objects

This category of conceptual models is based on an object or artefact, such as a tool, book, or a vehicle. This type of models tend to be more specific than the models based on activity, focusing more on the way a particular object is used in a particular context. They are often based on analogy with something in the physical world. Example: spread sheet.

3.3 Metaphors and interaction paradigms

Another way of describing conceptual models is in terms of interface metaphors. By this is meant a conceptual model that has been developed to be similar in some way to aspects of a physical entity (or entities) but that also has its own behaviours and properties. Such models can be based on an activity or an object or both. As well as being categorized as conceptual models based on objects, the desktop and the spread sheet are also examples of interface metaphors. Another example of an interface metaphor is a "search engine." Benefits of interface metaphors: they provide users with a familiar orienting device and helping them understand and learn how to use a system. People find it easier to learn and talk about what they are doing at the computer interface in terms familiar to them-whether they are computer-phobic or highly experienced programmers.

By interaction paradigm is meant a particular philosophy or way of thinking about interaction design. It is intended to orient designers to the kinds of questions they need to ask. For many years the prevailing paradigm in interaction design was to develop applications for the desktop-intended to be used by single users sitting in front of a CPU, monitor, keyboard and mouse. A number of alternative interaction paradigms include:

- Ubiquitous computing (technology embedded in the environment)

Another development that has evolved from this paradigm:

  - Tangible bits or interfaces (Ishii & Ullmer, 1997): the integration of the digital and the physical worlds by embedding computation in physical artifacts and environments (see Section Tangible interfaces for more information);
  - Augmented reality: superposition of virtual representations on physical devices and objects (see Section Virtual and Augmented reality for more information)
  - Physical/virtual integration

- Pervasive computing (seamless integration of technologies)
- Wearable computing (wearables)
- Attentive environments and transparent computing: the computer attends to user's needs through anticipating what the user wants to do (see Section: Background sensing for more information)

The terminology presented in this section will be used throughout the rest of the document or presented in more detail.

4. Input / Output devices

To perform a communication, the two sides should share information. The input to the computer connects the inner world of bits to the real world perceptible to human senses; it consists of sensed information about the physical environment; the output can comprise any modification of the physical environment, such as a display (including the cathode ray tube (CRT), flat-panel displays, or even light emitting diodes), speakers, or tactile and force feedback devices (sometimes referred to as haptic devices) (Hinckley, Jacob, Ware, 2004). A common input is a mouse click, a common output – a text on a display. An interaction technique provides the user a way to perform certain task by combining the input with a certain feedback. An interaction technique is a way of using a physical input/output device to perform a generic task in a human-computer dialogue (Foley, 1990).

4.1 Properties of the input devices

Here, we will only briefly enumerate the most important properties of input devices. For more information, see Hinckley, Jacob & Wade, 2004.

- **Physical property sensed:** Most of traditional pointing devices sense position, motion or force. Examples: tablet (position), mouse (motion) and isometric joystick (force). For a rotary device, the corresponding properties are angle, change in angle, and torque.
- **Transfer function:** A device, in combination with the host operating system, typically modifies its signals using a mathematical transformation that scales the data to provide smooth, efficient, and intuitive operation. An appropriate mapping is a transfer function that matches the physical properties sensed by the input device. Appropriate mappings include force-to-velocity, position-to-position, and velocity-to-velocity functions. For example, the force applied to a joystick is transformed into velocity of cursor movement.
- **Number of dimensions:** This property represents how many linear or angular dimensions the device could measure. For example, the mouse measures two linear positions while a six-degree magnetic tracker measures three linear and three angular dimensions (e.g. Ware & Jessome, 1988).

Pointing speed and accuracy

- **Input devices state:** tracking, (cause a cursor movement), dragging (selection of objects e.g. by clicking) and out-of-range (the device is out of the physical limits where it can be sensed (Buxton, 1990).
- **Direct vs. Indirect control:** A mouse is an indirect input device, while a touch-screen is a direct input device. The direct input devices may be less accurate due to parallax errors, reduced transmissivity of the screen introduced by a sensing layer, or by occlusion of the displays by the user's hand.
- **Device acquisition time:** the time needed to pick or put down a device.
- **Latency:** the delay between the user's physical movement and the feedback provided by the system to the user.

4.2 Direct input devices

A direct input device has a unified input and display surface. Direct devices such as touch-screens, or display tablets operated with a pen, are not necessarily easier to use than indirect devices, like the mouse. Occlusion is also a major design challenge: the finger or pen covers the area where the user is pointing or other important information on the screen like visual feedback, dialogs, status indicators, or other controls.

The direct input devices usually have a system's feedback for user input localized to the physical point(s) of contact. Some direct input devices can only sense a bare finger. Others such as resistive touchscreens can sense either a plastic stylus or a bare finger, but cannot distinguish one type of contact from the other. Transparent electromagnetic digitizers, such as those found on Tablet PC's, require the use of a special pen, and cannot sense touch unless the system integrates a second touch-sensitive digitizer. Some commercially available digitizers based on capacitive coupling can sense a special pen, and simultaneously differentiate it from multi-touch inputs (Engelhard 2008). The pen by itself introduces little occlusion, than a finger or a palm used as input on the touchscreens, but still the screen is occluded by the hand. It provides higher precision, than using a finger and is appropriate for writing and sketching tasks. Pen-based input can be done only with the preferred hand while the touch input devices may be manipulated by both hands. However, the use of pen slows the input in comparison with the use of hands as it needs time to pick it. Though different, both types of input devices have problems with false inputs.

Therefore, while pen and touch share common ground as direct input modalities, they also exhibit differences in many important ways.

- **Single Touch vs. Multiple Touch.** Single-touch devices, such as traditional resistive touch screens, are adequate for emulating a mouse, and for detecting most of the gestures employed in commercial 'multi-touch' software today (Potter, Weldon, and Shneiderman 1988). Multi-touch devices can be further classified by the number of finger contacts they are able to detect and track. Multiple contacts are required for a user to perform true multi-touch gestures, such as pinch-to-zoom or multi-fingered grabbing (Krueger, Gionfriddo, and Hinrichsen 1985; Moscovich and Hughes 2006). For table-top interfaces, still more contacts must be tracked to enable multiple users to perform multi-touch gestures at once.
- **Pressure and Contact Area Sensing.** Pressure is the measure of force that the user exerts on an input device. Pressure sensing is often confused with contact area sensing. True pressure sensing is supported by many pen-operated devices, but typically only contact area can be sensed by touch-screen interfaces. With contact area, rather than relying on an absolute degree of contact, one should instead emphasize changes in contact area as a more controllable parameter that users can modulate, such as by rolling one's finger in contact with the display (Benko, Wilson, and Baudisch 2006). Some laptop touchpads include sensors for measuring the force exerted on the pad, while multi-touch screens give measure the total force applied at all contact points and do not give an independent measure of pressure.
- **In-Air Hand Postures and "Direct" Sensing Beyond the Display.** Recent approaches have also demonstrated the utility of differentiating in-air hand postures, as opposed to those that occur while in contact with a device (Grossman, Wigdor, and Balakrishnan 2004; Hilliges et al. 2009; Wilson and Benko 2010). For example, the SecondLight system can see through the display, and can both project light and sense interactions in the volume above the display itself. This enables to separate purposed actions from incidental

movements of hands in the air. Moreover, it shows that not only the direct contact with the display should be considered a direct input device. The positions of users relative to the display (Ballendat, Marquardt, and Greenberg 2010), the motions of their hands above the display (Tang et al. 2010), and the posture of the hand as it comes into contact with the displays (Holz and Baudisch 2010) are all also forms of direct input that can extend and enrich user interfaces.

- **Finger Differentiation, User Differentiation, and User Identification.** Some touch devices, like Microsoft Surface (Lepinski, Grossman, and Fitzmaurice 2010) are able to determine which of the user's fingers are in contact with the display. Using, for example, capacitive coupling techniques (Dietz and Leigh 2001), other systems can distinguish the touch from one user from the touch of another user. Identification of users, or distinguishing which fingers of the user's hand are touching the display, based on visual features, was demonstrated by computer vision technologies.

Major problems of the direct input devices are the mismatch between the sensed input position and the viewing angle (i.e. the parallax error) and the longtime elapsed between the moment of the physical action performed by the user and the time the feedback produced by the system is perceived. This latency has strong negative effects if it exceeds 100 ms.

4.3 Indirect input devices

An indirect input device is one which does not provide input at the same physical space as the output. Indirect devices do not introduce occlusion of the screen, but usually produce higher cognitive load and require more explicit feedback.

The properties of some indirect devices are shortly given below.

- **Mouse.** Douglas Englebart and colleagues (English, Englebart, & Berman, 1967) invented the mouse in 1967 at the Stanford Research Institute (Figure 1).

Figure 1.The first mouse, invented by Douglas Engelbart

Pros:

- Reduced hand fatigue. Users exert force on mouse buttons in a direction orthogonal to the mouse's plane of motion, thus minimizing inadvertent motion.
- Accurate: All of the muscle groups of the hand, wrist, arm, and shoulder contribute to pointing, allowing high performance for both rapid, coarse movements as well as slow, precise movements (Guiard, 1987; Zhai, Milgram, & Buxton, 1996).
- Familiar
- Widely available

- Low cost
- Ease to use

Cons:

- Higher cognitive load
- Requires desk space
- Long motions aren't easy or obvious
- Requires some acquisition time

Trackball. A trackball senses the relative motion of a partially exposed ball in two degrees of freedom. It is often regarded as an inverted mouse. Trackballs have a small working space, and may be used on an angled surface. Trackballs may require frequent clutching movements because a user must lift and reposition their hands after rolling the ball a short distance. The buttons are located to the side of the ball, which can make them awkward to hold while rolling the ball (MacKenzie, Sellen, & Buxton, 1991).

Isometric joysticks. An isometric joystick (e.g., the IBM Trackpoint) is a force-sensing joystick that returns to center when released. Most isometric joysticks are stiff, offering little feedback of the joystick's displacement. The rate of cursor movement is proportional to the force exerted on the stick; as a result, users must practice in order to achieve good cursor control.

Isotonic joysticks. Isotonic joysticks sense angle of deflection.

Indirect tablets. Indirect tablets report the absolute position of a pointer on a sensing surface. Touch tablets sense a bare finger, whereas graphics tablets or digitizing tablets typically sense a stylus or other physical intermediary. Tablets can operate in absolute or in relative mode, in which the tablet responds only to motion of the stylus. Absolute mode is generally preferable for tasks such as drawing, handwriting, tracing, or digitizing, but relative mode may be preferable for typical desktop interaction tasks such as selecting graphical icons or navigating through menus.

Touchpads are small, touch-sensitive tablets often found on laptop computers. Usually they use relative mode for cursor control because they are too small to map to an entire screen, but most touchpads also have an absolute mode to allow features such as sliding along the edge of the pad to scroll. Touchpads support clicking by recognizing tapping or double-tapping gestures, but accidental contact (or loss of contact) can erroneously trigger such gestures (MacKenzie & Oniszczak, 1998). Like trackballs, the small size of touchpads necessitates frequent clutching, and touchpads can be awkward to use while holding down a button, unless the user employs his or her other hand.

Multi-Touch Pads. Most modern touchpads are multi-touch devices. The indirect touchpad must support relative cursor control, and typically single touch is mapped to moving the cursor. Thus two fingers are required to scroll or pan documents. With a direct-touch input device, cursor tracking is not necessary, and hence single touch can pan and scroll.

Indirect multi-touch pads can also support additional novel techniques, such as targeting many degrees of freedom to a single cursor position, or representing each point of contact with a separate cursor (Moscovich and Hughes 2006). Such models have also been explored in the context of mice augmented with multi-touch input surfaces (Villar et al. 2009; Benko et al. 2010

The iPad introduced recently by Apple is one of the many implementations of full multi-touch displays whereas it is a completely new way how people can interact with their computer. It

allows that the user is able to use up to four fingers at the same time to navigate through the interface. For example two fingers can be used to zoom and four fingers to browse through windows. With using the screen as a big touchpad the techniques of the normal touchpad have been enhanced.

The 10/GUI system which Miller (2009) invented, is an enhanced touchpad for desktop computer purpose which can recognize 10 fingers. With this opportunity human beings can interact with the computer with both hands and use it as a tracking and maybe also as a keyboard device. This touchpad except recognizing more fingers is remarkable with pressure detection of each finger, which is directly stated on the screen. It allows also to use every finger as a pointing device.

Some other multi-touch devices will be considered later in the section Displays..

4.4 Keyboards, Text Entry, and Command Input

Keyboards and typewriters have been in use for more than 140 year and still they are the preferred choice for text entry. The principal virtue of the standard QWERTY key layout is that common pairs of letters tend to occur on opposite hands. This allows very efficient hand movement patterns as during a key press by on hand, the other hands move to the next key (MacKenzie and Soukoreff 2002).Efforts to replace this layout were not successful even though other layouts, like Dvorak keyboard have provided quicker typing by about 5% (Lewis, Potosnak, and Magyar 1997). Split-angle ergonomic QWERTY keyboards are close enough to the standard layout as they preserve much of a user's existing skill for typing. They have also been shown to help maintain neutral posture of the wrist and thus reduce fatigue and the negative consequences of prolonged work. At least for heavy text entry, mechanical keyboards are unlikely to be supplanted by new key layouts, speech recognition technologies, or other techniques any time soon.

Two-Thumb Mechanical Keyboards: Many designs for cell phones and other handheld devices, such as the RIM Blackberry, offer two-thumb keyboards with QWERTY key layouts. Using two-thumb keyboards reduces the text entry rates (Clarkson et al. 2005).

Touch-Screen Keyboards. Modern multi-touch-screens enable text entry that is adequate for mobile interaction. One of the main problems with such graphical keyboards is that they require dividing the user attention between the workspace, where the text appears and the keyboard itself. This is a serious problem for large form-factors, like slates as the distance between the graphical keyboard and the point of text insertion could be very big. The graphical keyboards also provide less feedback as compared to physical keyboards. The graphical keyboards reduce the visible portion of the screen where occupied by the document. Furthermore, because the user typically cannot rest their fingers in contact with the display (as one can with mechanical keys), and also because one must carefully keep other fingers pulled back so as to not accidentally touch keys other than the intended one, extended use of touchscreen keyboards can be fatiguing.

Other text entry mechanisms. One-handed keyboards can be implemented using simultaneous depression of multiple keys; such chord keyboards can sometimes allow one to achieve high peak performance (e.g. court stenographers), but take much longer to learn how to use (Noyes, 1983; Mathias, MacKenzie & Buxton, 1996; Buxton, 1990a). They are often used in conjunction with wearable computers (Smailagic & Siewiorek, 1996) to keep the hands free as much as possible (but see also Voice and Speech below). With complex written languages, such as Chinese and Japanese, key chording and multiple stages of selection and disambiguation are currently necessary for keyboard-based text entry (Wang et al., 2001).

Handwriting and Character Recognition. Handwriting recognition technology on the Tablet PC has improved markedly over the past decade. Nonetheless recognizing natural handwriting remains difficult and error prone for computers, and demands error correction input from the user. Handwriting recognition works well for short phrases such as search terms (Hinckley et al. 2007), or for background tasks such as indexing handwritten documents for search, but converting lengthy handwritten passages to error-free text remains a tedious process. Hence, while handwriting recognition is an important enabling technology, in our view pen-operated devices can best avoid the graveyard by emphasizing those user experiences that make minimal demands of recognition, and instead emphasize the virtues of ink as a uniquely expressive data type.

To make performance more predictable for the user, some devices rely on character recognition, often implemented as single-stroke ("unistroke") gestures (Goldberg and Richardson 1993). Unistroke alphabets attempt to strike a design balance such that each letter is easy for a computer to distinguish, yet also straightforward for users to learn (MacKenzie and Zhang 1997). With the widespread adoption of touch-screen keyboards, coupled with the large strides made in handwriting recognition, such approaches have fallen out of favor in most contexts.

5. Modalities of interaction

In the search for designs that enhance interfaces and enable new usage scenarios, researchers have explored many input modalities and interaction strategies that transcend any specific type of device.

5.1 Bimanual input

Except for typing, more computer input devices and modes of operation use only one hand. In real-world situations, however, people usually use both hands with the non-preferred hand performing a supportive role. For example, in handwriting the preferred hand performs fine, small movements, while the non-preferred hand performs large-scale infrequent movements, like orienting the paper. This asymmetric role of the hands is also observed in compound navigation/selection tasks such as scrolling a web page and then clicking on a link (Buxton & Myers, 1986), command selection using the non-preferred hand (Bier, Stone, Pier, Buxton & DeRose, 1993; Kabbash, Buxton & Sellen, 1994), as well as navigation, virtual camera control, and object manipulation in three-dimensional user interfaces (Kurtenbach, Fitzmaurice, Baudel & Buxton, 1997; Balakrishnan & Kurtenbach, 1999; Hinckley et al., 1998b).

In hand-held devices users usually hold the device with non-preferred hand while entering text, or command selection, tapping etc. The role of the non-preferred hand is increased in devices where spatial manipulation is achieved by moving or tilting the device (Fitzmaurice and Buxton 1994; Hinckley and Song 2010). This approach leaves the preferred hand free to point or sketch at the content thus revealed.

Integrating additional buttons and controls with keyboards to encourage bimanual interaction can also improve the efficiency of some common tasks (MacKenzie and Guiard 2001; McLoone, Hinckley, and Cutrell 2003)). Non-preferred-hand could be used for mode switching, such as by holding down a mode button, as has also been demonstrated to be a particularly effective means of changing mode in pen-based interfaces (Li et al. 2005).

5.2 Gesture recognition vs. physics-based manipulation

Gestures are expressive, meaningful body motions. They represent physical movements of the fingers, hands, arms, head, face, or body that have the intent to convey information or interact with the environment. Messages can be expressed through gesture in many ways. For example, an emotion such as sadness can be communicated through facial expression, a lowered head position, relaxed muscles, and lethargic movement. Similarly, a gesture to indicate Stop! can be simply a raised hand with the palm facing forward, or an exaggerated waving of both hands above the head. People use gestures a lot when talking, often subconsciously. The idea behind gestures is the fact that they are used to communicate or to support communication. Therefore, using gesture as human-computer interaction seems more natural than using other input devices. Gestures can be static, where the user assumes a certain pose or configuration, or dynamic, defined by movement. Some gestures have both static and dynamic elements, where the pose is important in one or more of the gesture phases; this is particularly relevant in sign languages. When gestures are produced continuously, each gesture is affected by the gesture that preceded it, and possibly by the gesture that follows it.

An often-cited taxonomy is that of Rime and Schiaratura [1991]:

- Symbolic gestures have a (single) verbal and often cultural dependent meaning, for example the OK sign, or sign language for deaf people.
- Deictic gestures are made by pointing or motioning to direct attention to some object or event.
- Iconic gestures are gestures that display information about the size, shape or orientation of objects, spatial relations, and actions, for example using hands to indicate the size of fish that one caught).
- Pantomimic gestures consist of manipulating an invisible imaginary object or tool, for example making a fist and moving to indicate a hammer.

Symbolic gestures can be identified most easily by a gesture recognition system. Deictic, iconic, and pantomimic gestures usually require additional information (context) and thus are harder to recognize. Deictic gesture in particular has received much attention, with several efforts using pointing (typically captured using instrumented gloves or camera-based recognition) to interact with “intelligent” environments (Baudel & Beaudouin- Lafon, 1993; Maes, Darrell, Blumberg & Pentland, 1996; Freeman & Weissman, 1995; Jojic, Brumitt, Meyers & Harris, 2000) as well as deictic gesture in combination with speech recognition (Bolt, 1980; Hauptmann, 1989; Lucente, Zwart & George, 1998;).

However, in general, there exists a many-to-one mapping from gesture to concept. In particular, with most forms of gestural input, errors of user’s intent and errors of computer interpretation seem inevitable. The sole exception is physical manipulation: moving an object from one place to another or otherwise changing its position or orientation. Physics-based systems extend this by mapping inputs to a virtual world governed by Newtonian physics, leading to user experiences described as 'natural' (Agarawala and Balakrishnan 2006; Wilson et al. 2008; Wilson 2009). This approach has been characterized as "reality-based interaction", which advocates for using naïve physics combined with awareness of the body, surrounding environment, and social interaction as the bases for successful user interfaces (Jacob et al. 2008).This is in contrast to an approach where gestures are specifically recognized to differentiate system responses, such as assigning functions to particular hand postures (Baudel and Beaudouin-Lafon 1993). Physics-based systems have the advantage of 'feeling' like the real

world, but have not yet been demonstrated to scale to enable the range of functions expected of an interactive system. Recent exploration of tangible interaction techniques (Ishii & Ullmer, 1997) and efforts to sense movements and handling of sensor-enhanced mobile devices represent examples of sensing manipulation (that is, ergotic gestures) (Hinckley, Pierce, Horvitz & Sinclair, 2003; Hinckley, Pierce, Sinclair & Horvitz, 2000; Harrison, Fishkin, Gujar, Mochon & Want, 1998)

There are several ways of making a gesture but the most important methods for HCI are hand gestures (e.g. pointing), pen- or mouse-created gestures (e.g. handwriting), and human-body motion gestures (e.g. nodding yes). Popular gesture input methods are touch-screens, mouses, computer vision (image differencing), electromagnetic fields (field distortions, and data-gloves). Several taxonomies have been devised that categorize the different types of gestures.

5.2.1 Pen and Pen-Based Gesture Input

Pen-based gestures can indicate commands, such as crossing out a word to delete it, or circling a paragraph and drawing an arrow to move it. Such gestures support cognitive chunking by integrating command selection with specification of the command's scope (Buxton, Fiume, Hill, Lee & Woo, 1983; Kurtenbach & Buxton, 1991). Marking menus use directional pen motion to provide extremely rapid menu selection (Kurtenbach & Buxton, 1993; Kurtenbach, Sellen & Buxton, 1993). With pen-based interfaces, designers often face a difficult design trade-off between treating the user's marks as ink that is not subject to interpretation, versus providing pen-based input that treats the ink as a potential command (Kramer, 1994; Moran, Chiu & van Melle, 1997; Mynatt, Igarashi, Edwards & LaMarca, 1999). Pen input, via sketching, can be used to define 3D objects (Zeleznik, Herndon & Hughes, 1996; Igarashi, Matsuoka & Tanaka, 1999). Researchers have also explored multimodal pen and voice input; this is a powerful combination as pen and voice have complementary strengths and weaknesses, and can disambiguate one another (Cohen et al., 1997; Cohen & Sullivan, 1989; Oviatt, 1997).

Pen-Based sensors are specifically of interest in mobile devices and are related to pen gesture [44] and handwriting recognition areas. Although recognition is far from perfect, there are now hand-held devices on the market that are capable of handwriting recognition. It is reported that these devices can reach a recognition rate of 95%, although in every day practice it is very hard to achieve this rate. Also, these devices require the use of slightly altered and simplified characters in order to successfully identify the hard-to-recognize characters (Ehlert, 2003).

Recently, there is a renewed interest in paper-based interfaces. While previous paper-based interfaces such as DigitalDesk [Wellner, 1993], Xax [Johnson, et al., 1993] and PaperPDA [Heiner, et al., 1999] required either a complex setting to capture strokes made on paper in real time, or relied on users to scan paper documents for post-processing, the Anoto digital pen [Anoto, 2002] adopts a highly portable fountain-pen-like digital pen as its interface, and captures not only the shape of the strokes but also the pressure and timing information in real time. The system also provides an ID of the page on which the strokes have been made, making it easy to merge captured data onto the digital version of a printout [Guimbretiere, 2003]. These new features led to the design of several interactive paper systems. For instance, PapierCraft (Figure 2) [Liao, et al., 2008] and PaperPoint [Signer and Norrie, 2007] import the captured pen marks from printouts into the corresponding digital documents for active reading and slide annotation, respectively. Extending such 2D pen interaction into 3D space, ModelCraft [Song, et al., 2006] captures annotations on 3D models. The digital pen input can also be integrated with other devices like a mobile projector such as PenLight [Song, et al., 2009] and MouseLight [Song, et al., 2010] to create a highly portable descendent of the

original Digital Desk system.

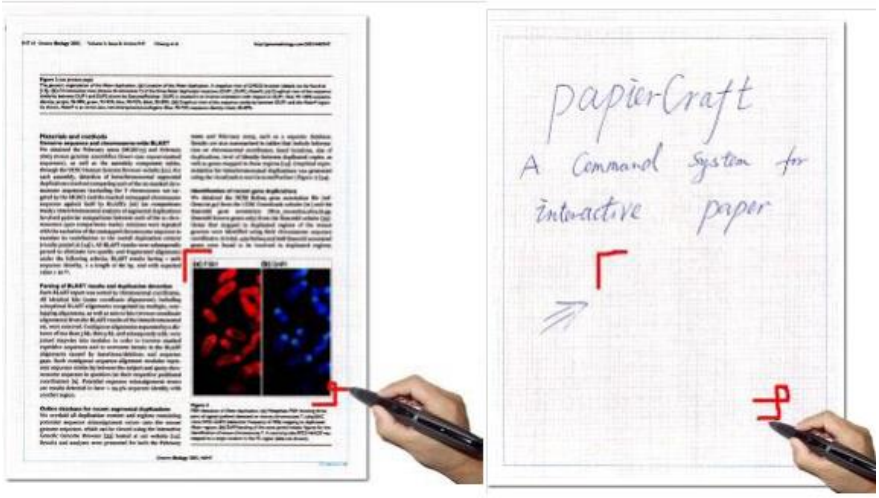

Figure 2. An example of the Copy command and of Paste command in PapierCraft (taken from Liao, et al., 2008)

For more flexibility in issuing commands, PapierCraft [Liao, et al., 2008] introduces to paper a generic pen-gesture-based command system with which users can draw pen, gestures to select arbitrary paper document content and choose a digital command to be applied. There are also other paper-based gesture commands customized for specific fields. PaperProof [Weibel, et al., 2008] supports a set of gestures for proof-editing Word documents on printouts. Upon synchronization, gestures are interpreted as Word editing commands on the corresponding digital document. CoScribe [Steimle, et al., 2009] permits a "stitching" gesture to attach personal paper book marks to a printout of shared documents and supports collaborative reading. For other applications of paper-based interaction see (Liao & Guimbrerièere, 2012)

An interesting application is the Biometric Smart Pen (BiSP), which is a multi-sensor pen that is used for the acquisition of neuro-motor features by measuring the kinematics and dynamics of hand movements during handwriting, i.e. the pressure applied to the pen and the tilt angles. This system is used for the authentication of individuals where query samples are compared against stored references using some matching algorithms (Saleh, 2008).

5.2.2 Whole Body Input

Whole-body input is also possible, typically utilizing computer vision techniques (Krueger, Gionfriddo, and Hinrichsen 1985). When a single camera is used, the vocabulary of whole-body input to those things that are clearly observable in the imagery is significantly limited and restricted mostly to 2-dimensional manipulations within the viewing plane (Krueger, Gionfriddo, and Hinrichsen 1985). More recent technologies have augmented this sensor stream with the distance of objects from the camera, enabling more subtle, 3D interactions. This has been demonstrated by processing two camera images simultaneously and observing binocular disparity. (Ko and Yang 1997; Matsushita and Rekimoto 1997). It has also been demonstrated with techniques such as time-of-flight cameras (Wilson 2007), as well as structured light techniques. Commercial products such as Microsoft Kinect provide depth information for each pixel, which has the potential to enable richer interaction, such as correcting the projected image to allow for consistently-sized projection of objects onto moving surfaces (Wilson 2007). Person tracking is a specific area of the more general object tracking problem. Yilmaz et al. (2006) recently reported a survey on object tracking. They

defined object tracking as “the problem of estimating the trajectory of an object in the image plane as it moves around a scene”. Naturally, in the case of person tracking the object is a person.

Human activity recognition refers to the recognition and analysis of human activities during a certain time period. Activity can be walking, sitting, throwing a ball, using hand gestures, and so on. Reflective markers are also commonly used in human motion analysis. In this case human motion is modelled by tracking the markers that are placed around the body. The very first marker based-system was the Moving Light Display (MLD) system by Johansson (1975). Nowadays there are many commercial systems (Ascension, 2007; Qualisys, 2007; Vicon, 2007) available that track humans using markers.

Yilmaz et al. (2006) mentioned that robust real-time trackers that work in simple scenarios have been developed in the last few years. Yilmaz et al. (2006) stated that the possible problems for object tracking are the following:

1. “Loss of information caused by projection of the 3D world on a 2D image
2. Noise in the images
3. Complex object motion
4. Non-rigid and articulated nature of the objects
5. Partial and full object occlusions
6. Complex object shapes
7. Scene illumination changes
8. Real-time processing requirements.”

5.3 Face and emotion recognition

Face and emotion recognition

The face provides vast amount of information about the other person and his/her feelings. It is natural to look at another person’s face and eyes during conversation. Facial expressions mirror our emotions, mental activities, and physiological activities (Fasel and Luetin, 2003), and help other people to understand us. On the other hand, communication between humans and computers still mostly happens using keyboard, mouse, and a display. Face recognition is a promising tool to change the way of interaction with computers and to turn them in more effective, versatile, and user friendly. This is especially true for adaptive interfaces i.e. when the operation of the computer changes depending on the emotional state, stress level or task load of the user. It could facilitate e-learning by adapting the level of difficulty of the material depending on its acceptance by the user. Emotion recognition could also be applied in usability studies.

Emotion recognition is tightly coupled with face recognition as the first stage of emotion recognition requires the detection of face. For this reason the two topics will be intermixed in this presentation. Face could provide additional information about the person different from the emotional state like personal identity, age, gender, etc. A typical application in the security and surveillance is face detection and recognition from a video image. Another possibility is to use face recognition as a biometric authentication method instead or with finger print based or iris based methods. Zhao et al. (2003) also listed other face recognition applications than those related to security and surveillance like building access, computer access, or even authentication for bank transactions and on-line shopping. For example, Lenovo has included

biometric face recognition authentication in some new laptops that it sells in India.

Security and surveillance have gained importance recently. One typical application in this field is, for example, suspects can be found at airports and other public places by analysing captured video automatically and reporting possible matches to the security personnel. Some of the specific applications were games, virtual reality, training programs, and human-robot interaction. The Face Recognition Homepage (2013) lists 21 vendors of face recognition software. Omron released the first commercial face recognition software for mobile phones in 2005 (Omron, 2005). These examples show the commercial potential of face recognition.

With the increase use of digital cameras and mobiles the amount of photographs and digital media greatly increases. Information search and data mining is an area where face analysis is already in use. Google included an option to search for images containing faces to their image search in May 2007. Microsoft followed in July and included face and portrait filtering in their "Live Search Images" engine. Exalead engine also allows searching images with faces. Face recognition will assist to organize the large collection of photos that one has or to find a specific person in them.

Face analysis has a lot of potential in computer and console games. Probably the simplest use is to include a user's face (or the whole body of the user) in the game character. For example, the Xbox 360 game "Rainbow Six Vegas" has this option. The more complex examples include controlling a game with face or head movements. Gorodnichy and Roth (2004) presented two such games: BubbleFrenzy and NousePong. The head movements were registered by tracking the player's nose. In the BubbleFrenzy game the purpose was to drop bubbles of the same colors by shooting them with bubbles. The user turned a bubble turret for a desired direction by rotating the head to the left or to the right. The controlling was implemented so that even small head rotations were sufficient to turn the turret over the whole 180° range. The users preferred to control the turret by head rotation instead of the mouse. Pressing the spacebar launched a bubble from the turret but it would be possible to use, for example, eye blinking as Gorodnichy and Roth did in a painting application also presented in the article. In the NousePong game two players played against each other. There was a ball bouncing over the table and the players tried to bounce the ball back to the other player by moving a club horizontally with their head movements.

Ayoob et al. (2003) implemented drowsiness detection by observing the driver's eyes and the driver was alerted if the eyes were closed often and for many seconds at the time. Facial expression analysis can be used in user tests for observing user emotions. Noldus provides the FaceReader™ tool (FaceReader, 2007) that classifies facial expressions and other facial attributes and they mention usability testing and market research among other application areas for the FaceReader™.

Face recognition is tightly coupled also with body analysis and modelling. One approach to face recognition is based on the extraction of features and to compare their relative position, size, shape and used them for search in other images for potential match. The features could be the eyes, the mouse, the nose, the check-bones, the overall shape of the face, the jaw. This approach requires powerful methods like Support Vector Machine, Hidden Markov Model, Multilinear Subspace learning, Principal Component analysis, Linear Discrimination analysis, etc., for classification and pattern recognition. One initial stage for feature extraction could be the location of the eyes as the darkest elements in the face as used by Krestinin & Seredin, (2009) are applied for face analysis based on feature extraction. Faces in images are commonly represented by ellipses, rectangles, by coordinates of eye centers or by the center of the face region and its radius. Other approach is based on the 3D modelling of the face. This approach requires 3D sensors to capture information of the shape of the head. Face recognition benefits

also by texture analysis of the skin. A challenge in face analysis is the variable lighting conditions, shadows, occlusions.

Automatic face expression recognition systems are divided into three modules: 1) Face Tracking and Detection, 2) Feature Extraction and 3) Expression Classification. Suwa et al. (1978) made the first attempt to automatically analyse expressions from a sequence of images (movie frames) using twenty tracking points. However, up to 1990 there was no interest in this topic. Emotion recognition mostly tries to recover the basic emotion types, specified by Ekman and Friesen (1971) and their corresponding facial expressions that seem to be universal among all cultures in the world. These emotions are: happiness, sadness, fear, disgust, surprise, and anger. However, in addition to the six basic expressions there are a lot of expressions that do not belong to the basic ones (Tian et al., 2001) and there are differences among cultures in showing and interpreting facial expressions (Fasel and Luetin, 2003; Matsumoto, 1993). Facial expressions can also be caused on purpose, in which case they can be considered gestures.

The various facial behaviours and motions can be parameterized based on muscle actions. Up to date, analysis of emotion expressions is based on two important and successful parameter sets:

1. The Facial Action Coding System (FACS) (Ekman and Friesen, 1977) and
2. The Facial Animation parameters (FAPs) which are a part of the MPEG-4 Synthetic/Natural Hybrid Coding (SNHC) standard, 1998.

Facial Action Coding is a muscle-based approach. It involves identifying the various facial muscles that individually or in groups cause changes in facial behaviours. These changes in the face and the underlying (one or more) muscles that caused these changes are called Action Units (AU). For example, if a person's is angry he will squint, moving the action units placed at the top and bottom eyelids, close together. (Figure 3). AUs are said to be additive if the appearance of each AU is independent and the AUs are said to be non-additive if they modify each other's appearance (Cohn, Ambadar &. Ekman, 2005). This system consists of a taxonomy of 44 AUs with which facial expressions can be described. Ekman (1982) evaluate the possible combinations of AUs to be about 7000. In addition, there are still differences in facial expressions and their intensities between different individuals. Since facial expressions are dynamic they have temporal characteristics. Each expression has onset (attack), apex (sustain), and offset (relaxation). These can be analysed from an image sequence but not from a single image.

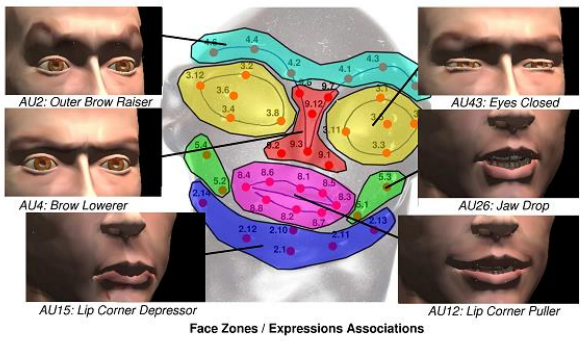

Figure 3. Example of action units. Taken from Jacquemin, 2007

Tian et al. (2001) have developed the Automatic Face Analysis (AFA) system which can automatically recognize six upper face AUs and ten lower face AUs. However real time applications may demand recognition of AUs from profile views too. Pantic (2001) used a

training set of AUs that not only classified facial expression but also tried to rate the intensity of the six basic emotional categories e.g. 20% nervous, 70% happiness. FACS have also been used to achieve the opposite, generating facial expressions in a computer-animated face (Wojdel and Rothkrantz 2001). Pantic and Rothkrantz (2004, 2006) have worked on the automatic AU coding of profile images.

In an effort to choose a better parameterization of the facial movement, the Moving Pictures Experts Group (MPEG) introduced the Facial Animation (FA) specifications in the MPEG-4 standard. Version 1 of the MPEG-4 standard (along with the FA specification) became the international standard in 1999. Cowie et al. (2008) indicate the relationship between the MPEG-4 FAPs and FACS AUs: "MPEG-4 mainly focusing on facial expression synthesis and animation, defines the Facial Animation parameters (FAPs) that are strongly related to the Action Units (AUs), the core of the FACS" .The MPEG-4 defines a face model in its neutral state to have a specific set of properties like a) all face muscles are relaxed; b) eyelids are tangent to the iris; c) pupil is 1/3rd the diameter of the iris and so on. Key features like eye separation, iris diameter, etc are defined on this neutral face model. The standard also defines 84 key feature points (FPs) on the neutral face (Figure 4). The movement of the FPs is used to understand and recognize facial movements (expressions) and in turn also used to animate the faces. The MPEG-4 standard defines six primary facial expressions: joy, anger, sadness, fear, disgust and surprise

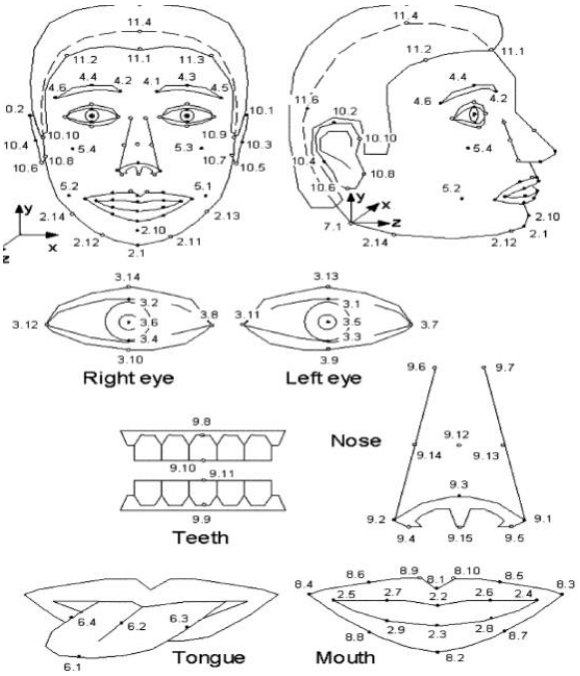

Figure 4. The 84 Feature Points (FPs) defined on a neutral face. Figure taken from MPEG Video and SNHC, 1998

Emotions greatly affect human behaviour and it has been shown that emotions affect our memory and other mental processes (Lewis and Critchley, 2003). For example, humans cannot make decisions if their emotional functions have been damaged (Damasio, 1994). Emotions and facial expressions also have an extremely important role in communication between humans. Computers must also gain this ability as suggested by one direction of development in HCI - Affective Computing. Expression recognition plays a significant role in recognizing one's affect and in turn helps in building meaningful and responsive HCI interfaces. Expression recognition systems have applications in robotics and affect sensitive HCI,

telecommunications, behavioural science, video games, animations, psychiatry, automobile safety, affect sensitive music juke boxes and televisions, educational Software, etc. Practical real-time applications have also been demonstrated. Bartlett et al. (2003) have successfully used their face expression recognition system to develop an animated character that mirrors the expressions of the user (called the CU Animate). They have also been successful in deployed the recognition system on Sony's Aibo Robot and ATR's RoboVie [13]. Another interesting application has been demonstrated by Anderson and McOwen (2006), called the 'EmotiChat'. It consists of a chat-room application where users can log in and start chatting. The face expression recognition system is connected to this chat application and it automatically inserts emoticons based on the user's facial expressions.

5.4 Gaze and Eye tracking

Eye movement-based input, properly used, can provide an unusually fast and natural means of communication, because we move our eyes rapidly and almost unconsciously. The gaze also reflects our attention, intention and desire. Thus, detection of the gaze direction makes possible to extract such information that is valuable in Human-Computer Interaction. Computers integrated with gaze tracking function must potentially provide an intuitive and effective interactive system. Eye movement input is distinctly faster than other current input media (Ware, 1987, Sibert & Jacob, 2000) as when using any pointing device the users usually first look at the intended direction of motion. The eye reaches the goal before using the pointing device. Pointing at an object with eyes is quite natural and no training is needed. A benefit of eye tracking could be reduced stress for hand and arm muscles by transferring the computer input from the hand to the eyes. In environments with high hygienic demands, like an operation room for surgery, an eye-gaze interface would be useful because it allows interacting without anything to touch. Also for public interfaces, especially in times of pandemic threats, hygienic interaction is desirable.

Gaze tracking systems can be intrusive (Topal et al, 2008; Kumar et al, 2009; Li & Wee, 2009; Tunhua et al, 2010;) or non-intrusive (Ebisawa, 2009; Yang et al, 2010; Zhank et al, 2010;). Intrusive systems are head-mounted devices or the head is fixated. One such method is based on the detection of the boundary between the white sclera and the darker iris. This boundary is easily detectable. One disadvantage of this method in addition to the relative determination of the eye position with respect to the head is that the boundary is not fully visible because of the eyelid. This makes the method more appropriate for measuring only the horizontal gaze position. The main non-intrusive methods are based on glint detection, electrooculography (EOG) and 3D modelling. The glint occurs when infrared is reflected off the back of the pupil and magnified by the lens. This method allows the estimation of both the horizontal and vertical eye movements and provides higher resolution. The EOG method is based on the electric field of the eye which is an electric dipole. Sensors are attached at the skin around the eyes. The main advantage of the method is its ability to detect of eye movements even when the eye is closed, e.g. while sleeping. However, the measurements are sensitive to electro-magnetic interferences. The last approach is using 3D model of face, based on mouth, eyes and nostrils position to evaluate face position and gaze direction. Most of the gaze detection methods require calibration.

Figure 5 Sensors attached at the skin around the eyes (taken from [metrovision@])

Gaze reflects our attention, intention and desire. This characteristic may be used to prevent the activation of a mobile phone by accidental button press while in a pocket; a system can demand that a warning text is read before allowing the user to continue with other functions. When a speech recognition system is used the gaze direction may permit to distinguish whether the user interacts with the computer or with other persons in the room. It might be the only possible channel of communication for a paralysed person. Besides fixations, the eye tracker can measure the fixation duration, the scan pattern (and its randomness), pupil diameter and blink rate. These measures provide useful information about the user's workload in a certain task. Therefore, gaze detection may provide an indirect form of interaction between user and machine which is mostly used for better understanding of user's attention, intent or focus in context-sensitive situations. Gaze tracking methods can be applied in many fields. In civil field, these methods can be used, for example in medical diagnostics, to help disabled people to interact with world, or to improve comfort of controlling computer, also they can be used for preventing drivers from falling asleep.

The use of gaze detection as pointing devices has some shortcomings. First, gaze tracking accuracy is low, about 1 degree of visual angle, which corresponds to the size of the fovea. Eye movements are non-intentional; a sudden appearance of an object in the periphery of the visual field may provoke reactive saccade. This could cause the so-called "Midas Touch" problem (Jacob, 1990). For the eye-gaze interface it will be difficult to decide whether our gaze is on an object just for inspection or for invoking an action. Misinterpretation by the gaze interface can trigger unwanted actions wherever we look. Also, eye movements are always "on" and there is no way to indicate when to engage the input device, as there is with grasping or releasing the mouse. Eyes cannot press buttons. One solution of this problem is to stare at a given position for a certain time (dwell time). This solution is time consuming, typical dwell times are 500 to 1000 milliseconds, and eats up the speed benefit that could result from the quick movement of the eyes. Ware and Mikaelian (1987) tested three different input methods which they called "dwell time button", "screen button", and "hardware button". The dwell-time button method is the standard method used today for eye typing. The gaze has to stay for a certain time, the dwell time, on the button to trigger the action associated with the button. The screen button method is a two-target task. The gaze moves to the chosen button and afterwards to the screen key to trigger the action. The hardware button method uses a key to press with the finger in the moment when the gaze is on the chosen button. The first two methods are gaze-only while the hardware button uses an extra modality. Similar proposal for combination of eye detection and other input device was proposed also by Bolt (1981). He proposed two solutions: the first, equivalent to the dwell-time method, while the second – a combination of gaze and joystick or speech input.

Gaze pointing is also a direct method and does not need feedback. The reason to desire

feedback comes from possible calibration error. Jacob (1995) pointed another critical aspect of gaze interaction: providing a feedback cursor. Of course the user knows where she or he is looking but not with a one-pixel-accuracy and additionally there may be calibration errors which cause the gaze position to differ from the position reported by the eye tracker. Providing a gaze feedback cursor can result in chasing the cursor across the screen or as Jacobs expressed it: "If there is any systematic calibration error, the cursor will be slightly offset from where the user is actually looking, causing the user's eye to be drawn to the cursor, which will further displace the cursor, creating a positive feedback loop." Distraction by moving or blinking objects might also cause conflicts.

To reduce the negative effect of gaze detection inaccuracy, several solutions were proposed. In some of them the gaze is used for crude positioning, while another device – mouse (Zhai, Morimoto, Ihde 1999) or speech command (Miniotas, Špakov, Tugoy, MackKenzie 2005) to refine and disambiguate the positioning. Salvucci and Anderson (2000) used a gaze key analogous to a mouse key to trigger the action and the system evaluates the most probable position where the user is looking. To find these items the system uses a probabilistic algorithm which determines the items by the location of the gaze, i.e. the items close to the reported gaze.

Eyetracking is a promising direction of human-computer interaction. Its value is further increased when combined with other input devices.

5.5 Speech recognition, Sound and Voice Interaction

Speech as an interface to computers has several aspects, namely, speech recognition, control by speech and natural language processing. Natural language processing combines speech recognition with language understanding. The speech interface has the long-term potential to free users from the necessity of using touch and gesture interfaces. Speech-to-text combines speech recognition and control by speech to create text on the computer. Speech-to-text is important for hands-free applications and for uses where a keyboard cannot be used to input text. Examples include the transcription of doctors' comments during hospital rounds into patients' medical records and the ability to send text messages or email while driving. A computing device, such as a tablet, can capture what was said and send it as a message. Speech is emerging as a feature on search engines as providers compete to attract more users. Search engines will include options for spoken queries and use natural language processing to return more relevant answers.

Any good natural language system would require at least the following components [Wyard et al 1996]:

1. **Speech recognition;** conversion of an input speech utterance into a string of words.
2. **Language understanding;** analysis of the string of words (as much as possible) to extract a meaning representation for the recognized utterance.
3. **Dialogue management;** controlling the interaction or dialogue between the system and the user, which includes coordination of other components of the system.
4. **Database query;** retrieving the information requested by the user.
5. **Response generation;** specification of the text that is to be the output message of the system.
6. **Speech output;** actual generation of the output message using text-to-speech synthesis or pre-recorded sentences.

Speech recognition is the process of transforming a continuous speech signal into a form that can be understood by a computer (usually text). A typical approach is to analyse the acoustic signal obtained by an acoustic and a language model. The existing models differ mostly by the building blocks (e.g. phonemes) used to analyse the acoustic signal data. Algorithms for stochastic modelling are applied to decode a sequence of symbols. For example, Florez-Choque, et. al (2007) used genetic algorithms self-organizing map to recognize the presence of phonemes in spoken Spanish. The system analyses the resulting string of building blocks using a language model that contains a base vocabulary. The probability of possible words is calculated and those with highest probability are the result of the recognition process.

While progress is being made, it is slower than optimists originally predicted, and daunting unsolved problems remain. For limited vocabulary applications with native English speakers, speech recognition can excel at recognizing words that occur in the vocabulary. Error rates can increase substantially when users employ words that are out-of-vocabulary (i.e. words the computer is not “listening” for), when the complexity of the grammar of possible phrases increases, or when the microphone is not a high-quality close-talk headset. The more words are stored, the higher the chance that the recognition process will make mistakes and the slower the system becomes. A problem with speech recognition based on an acoustic signal is that it functions very badly when there is a lot of noise. The reason for this is that it is very hard to distinguish the speaker’s voice from other sounds. One way to reduce the effect of noise and of multitude sound sources on speed recognition is to use an array microphone that combines multiple microphones. Through the process of beamforming the output of the multiple microphones in an array is combined to form a single audio signal in which all but the dominant speaker’s signal has been removed. Beamforming can also reveal information about the position of the speaker (Tashev & Malvar, 2005).

Even if the computer could recognize all of the user’s words, the problem of understanding natural language is a significant and unsolved one. It can be avoided by using an artificial language of special commands or even a fairly restricted subset of natural language. But, given the current state of the art, the closer the user moves toward full unrestricted natural language, the more difficulties will be encountered. For computers to embed themselves naturally within the flow of human activities, they must be able to sense and reason about people and their intentions: in any given dialog, multiple may people come and go, they may interact with system or with each other, and they may interleave their interactions with other activities such that the computational system is not always in the foreground (Bohus, Horvitz, and 2010).

Yet even without understanding the content of the speech, computers can digitize, store, edit, and replay segments of speech to augment human-human communication (Arons, 1993; Stifelman, 1996; Buxton, 1995b). Conventional voice mail and the availability of MP3 music files on the web are simple examples of this. Computers can also infer information about the user’s activity from ambient audio, such as determining if the user is present, or perhaps engaging in a conversation with a colleague, allowing more timely delivery of information, or suppression of notifications that may interrupt the user (Schmandt, Marmasse, Marti, Sawhney & Wheeler, 2000; Sawhney & Schmandt, 2000; Horvitz, Jacobs & Hovel, 1999). Recording simultaneous speech and handwritten annotations is also a compelling combination for human-human communication and collaboration (Levine and Ehrlich 1991).

Currently, speech recognition is used in fields such as: government and private industries, aviation maintenance, medical fields where fast access to information is important, and hands free banking. The application of speech recognition in these fields is limited to simple, predefined commands that the user has to memorize. Speech recognition currently does not work for people who have heavy accents, who want to use it in noisy environments, and who

do not want to spend time training software. Another reason that people may not want use speech recognition applications is because of privacy concerns.

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware. A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech. Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units. Systems that store single units of speech sounds (phones) or pairs of phones provide the largest output range, but may lack clarity. Any reasonable phoneme text-to-speech output system can cover an entire language or sometimes even multiple languages. Synthetic speech may sound unnatural and, for example, long series of numbers may be problematic for the synthesizer. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output.

Non-speech audio is also a useful way of conveying information. Humans perceive sounds often without even noticing it. We can tell that there is someone moving in the corridor when we hear his footsteps and sometimes we may know who the person is without seeing him or her. On the other hand, sound can also alert us. For example, we focus our attention immediately to the direction of a crashing sound. Non-speech audio has been used as an output modality with computers. Sounds can take many forms beeps bongs (deep ringing sounds) , clonks (loud thudding sounds), whistles, whirrs (like the sound of rapidly vibrating wings) to indicate error (e.g. incorrect command/input) or to alert for a risky situations (e.g. when deleting files) confirmation of actions e.g. keyclick notification of events/status e.g. a new email has arrived, download has been completed. Humans can collect considerably more meaning from simple sounds, both natural sound effects and artificially-conceived tones (Bly, 1982, Buxton, 1989, Gaver, 1986). When the event sounds do not have a natural counterpart, the user has to learn their meaning. Such sounds are called earcons (Brewster, 1993). If the sound has a natural counterpart it is called an auditory icon. Audio feedback may be crucial to support tasks or functionality on mobile devices which must take place when the user is not looking at the display (for some examples, see (Hinckley et al., 2000)).

There are also ways to convey information using sounds that have been used less with computers: music, soundscapes, and sonifications. Music is commonly listened to while using computer but it is usually listened to only to entertain. However, information specific to the computer usage situation can be presented by varying musical properties such as pitch, timbre, and rhythm. Soundscape creates an auditory picture of the environment using background and foreground sounds. It can be used to “encompass environments and create atmospheres and identities to locations, and tie different interaction elements together” (Kainulainen et al., 2007). Sonifications represent data and data relations as sounds similarly to the visualizations.

6. Displays

Output from a personal computer in most cases means output of visual data, and therefore, the most common type of output device comprises of different displays, including the cathode ray tube (CRT), liquid crystal display (LCD), or specialized devices like a pilot's head-up display. The speeded developments in technology provide different new type of displays.

Autostereoscopic displays: These types of displays are based on the ability of the human brain

to recover 3D depth based on two 2D images presented separately to the two eyes as is typical for binocular vision. This technique is known as Stereoscopy or 3D imaging./Latest trend in visualization aims to furnish the “illusion of depth” in an image by presenting two offset images separately to the left and right eye of the viewer. The brain is then able to combine these two-dimensional images and a resulting perception of 3-D depth is realized. Three are the main techniques developed to present two offset images one for eye: the user wear eyeglasses to combine the two separate images from two offset sources; the user wear eyeglasses to filter for each eye the two offset images from a single source; the user’s eyes receive a directionally split image from the same source. The latter technique is known as AutoStereoscopy (AS) and does not require any eyeglasses. Current AutoStereoscopic systems are based on different technologies which include lenticular lens (array of magnifying lenses), parallax barrier (alternating points of view), volumetric (via the emission, scattering, or relaying of illumination from well-defined regions in space), electro-holographic (a holographic optical images are projected for the two eyes and reflected by a convex mirror on a screen), and light field displays (consisting of two layered parallax barriers). An improvement of the AS refers of AutoMultiscopic (AM) displays which can provide more than just two views of the same image. So, the AS realized by AM displays is undoubtedly one of the really new frontier that must be consider for the near future to realize the “illusion of depth”, since leaves aside the uncomfortable eyeglasses and realizes a multi-point view of the same image. In such a way the user has not only the “illusion of depth” but the “illusion to turn around” the visualized object just moving his/her head position with respect to the source.

Figure 6. Autostereoscopic Display Image taken from SIGGRAPH 2001

OLED Displays Organic Light Emitting Diodes (OLEDs) are a flat display technology, made by placing a series of organic thin films between two conductors. OLEDs are called organic because they are made from carbon and hydrogen. When electrical current is applied, a bright light is emitted. Because OLEDs produce (emit) light they do not require a backlight (OLED Info, 2011). Some key advantages are that they can be ultra-thin, flexible and transparent, have low power consumption, a greater brightness, a fuller viewing angle and can operate in a broader temperature range. This provides the potential for curved OLED displays, placed on non-flat surfaces; wearable OLEDs; and transparent OLEDs as windows.

Figure 7. A Kinect-driven prototype desktop environment by the Microsoft Applied Sciences Group allows users to manipulate 3D objects by hand behind a transparent OLED display (www.microsoft.com/appliedsciences)

Multiple Displays Researchers have recently recognized that some very interesting design issues arise when multiple displays are considered, rather than the traditional single display of desktop computers. Having multiple monitors for a single computer is not like having one large display (Grudin, 2001). Users employ the boundary between displays to partition their tasks, with one monitor being reserved for a primary task, and other monitors being used for secondary tasks. Secondary tasks may support the primary task (e.g. reference material, help files, or floating tool palettes), may provide peripheral awareness of ongoing events (such as an e-mail client), or may provide other background information (to-do lists, calendars, etc.). Switching between applications has a small time penalty (incurred once to switch, and again to return), and perhaps more importantly, it may distract the user or force the user to remember information while switching between applications. Having additional screen space "with a dedicated purpose, always accessible with a glance" (Grudin, 2001) reduces these burdens (Czerwinski et al., 2003), and studies suggest that providing multiple, distinct foci for interaction may aid users' memory and recall (Tan, Stefanucci, Proffitt & Pausch, 2001; Tan, Stefanucci, Proffitt & Pausch, 2002). Finally, small displays can be used in conjunction with larger displays (Myers, Stiel & Gargiulo, 1998; Myers et al, 2000; Rekimoto, 1998), with controls and private information on the small device, and shared public information on the larger display. This shows how displays of different dimensions support completely different user activities and social conventions. It is also possible to dynamically join multiple displays for collaboration or to create a larger but temporary tiled display (Tandler, Prante, Müller-Tomfelde, Streitz & Steinmetz, 2001; Hinckley, 2003b; Hinckley, 2003a).

Large-Format Displays Trends in display technology suggest that large-format displays will become increasingly affordable and common. There are several types of large screen display,

some using gas plasma technology to create large flat bitmap displays. They are appropriate for lectures and large-scale meetings. Large displays often implicitly suggest multiple simultaneous users, with many applications revolving around collaboration (Swaminathan & Sato, 1997; Funkhouser & Li, 2000) and giving a large-scale physical presence to virtual activities (Buxton, Fitzmaurice, Balakrishnan & Kurtenbach, 2000). To support input directly on whiteboard-size displays, researchers have explored gestural interaction techniques for pens or touchscreens (Guimbretiere, Stone & Winograd, 2001; Moran et al., 1997). Unless life-size viewing of large objects is necessary (Buxton et al., 2000), in general it is not yet clear what performance benefits a single large display may offer as compared to multiple monitors with the same screen area partitioned by bezels (Czerwinski et al., 2003). One recent study suggests that he increased field-of-view afforded by large-format displays can lead to improved 3D navigation performance, especially for women (Czerwinski, Tan & Robertson, 2002). Gouin et al. (2009) provide a number of human factor guidelines for the use of large displays, including human perception/legibility, information organization and display control.

The Interactive DataWall (Figure 6) developed at the Air Force Research Laboratory (AFRL) is a good example of how multi-modal interaction can apply to LGDs. It is built using three horizontally tiled video projectors each displaying 1280 x 1024 pixels for a combined resolution of 3840 x 1024 pixels across a 12' x 3' screen area. The system also features speaker-independent voice activation and a wireless pointing device using camera tracked laser pointers (AFRL, 2001)

Desktops can be expanded with multiple monitors. Research has shown that users have a strong tendency to partition tasks between discrete physical displays; for example, users often dedicate monitors to particular tasks or types of applications (Grudin 2001). The presence of discontinuities between the monitors (monitor bezels, Hutchings and Stasko 2004; Tan and Czerwinski 2003) have both positive and negative effect depending on the user's task influence the way that users arrange application windows and lead to discontinuities in information displays. Augmented desktops with additional peripheral displays for direct pen input on horizontal and vertical surfaces have also been explored (Morris, Brush, and Meyers 2008)

DisplaxTMMultitouch Technology (Future Labs, DisplaxTMCompany) is a technology that allows to turn "...any surface into an interactive multitouch surface." (InteractiveSystems, 2010). They are using a very thin transparent paper that is attached to the DisplaxTMMultitouch controller that could turn any surface into a up to 50 inch big touchscreen. With this possibility you can work directly on a big screen by just using your hands. Additionally this interface allows a usage of 16 fingers at the same time so that more than just one user can work on the screen simultaneously. With the weight of just 300g it is also a very transportable tool beside the fact that it is well durable as the film is placed on the back of the surface to protect it from scratches and other damage.

Figure 8. Displax's thin transparent multitouch surface. Image taken from D. InteractiveSystems, http://www.displax.com/en/future-labs/multitouch-technology.html#/en/future-labs/multitouch-technology.html.

Nomadic video A different way of visualization comes from the Nomadic Video (NV) approach (Huber et al., 2011) based on a pico-projector and Kinect capabilities (motion tracking and depth sensing). This technology allows every surface to become a display. The input could be manipulated by everyday objects (i.e. they turn to be tangible devices. More information about tangible interaction will be given later). The level of detail displayed by the projector can also be altered dynamically, with respect to the amount of display surface available.

7. Dimensional Graphics and Virtual Reality(VR)

Virtual reality (VR) is a technology that refers to computer-generated, interactive and three-dimensional environments into which users are immersed, or which add graphical information to the perceived natural environment that is updated according to the movements and position of the user (Augmented Reality). Virtual reality systems rely combinations of the 3-D devices, typically a magnetic tracker to sense head position and orientation to determine the position of the virtual camera for scene rendering plus a glove or other 3-D hand input device to allow the user to reach into the displayed environment and interact with it. This technology gives the opportunity to the users to use as input actions similar to the habitual ones that the user employs in the real world like pointing, grabbing, moving objects in space. Virtual reality interfaces, too, exploit the pre-existing human abilities and expectations. Instead of inputting strings of characters, users interact with a virtual reality in more natural and expressive ways—moving their heads, hands, or feet. The research in three-dimensional information visualization and virtual reality is motivated by the observation that humans naturally operate in physical space, and can intuitively move about and remember where things are (an ability known as spatial memory). VR extends the traditional 3D graphics world in order to include stereoscopic, acoustic, haptic and even other feedbacks, like smell and taste to create a sense of immersion (Sundgren et al.1992; Kalawsky 1993; Burdea and Coiffet 1994; Fuchs et al.2006a). The scientific community is able to exploit VR for visualizing scientific data, modelling, animating complex engineering systems and for traditional applications such as medicine, education, arts, entertainment, defense and robotics (Stanney 2002; Fuchs et al.2006b; Burdea and Coiffet 2003). The attempts to make human operations more natural in

artificially generated graphical environments are limited by the display and interaction technologies.

To improve the immersion in the virtual reality, head- mounted Display (HMD), also known as Helmet Mounted Display, are used. It provides stereo vision by projecting different images to the two eyes. The HMD is considered to be the center-piece for early visions of VR. In fact, the first VR system also highlighted the first HMD. A recent product (Personal 3D Viewer HMZ-T1) was developed by Sony, that takes the wearer into a 3D cinema of videos, music and games. The Sony's personal 3D viewer is being targeted at people who prefer solitary entertainment rather than sitting in front of a television with family or friends. But probably the more interesting HMD was developed by Sensics Inc. (www.sensics.com) with the SmartGoggles™ technology, based on which was realized the "Natalia", a highly immersive 3D SmartGoggle available as a development platform to content and device partners, with the expectation that it will be available to consumers later in 2012. Generally speaking, the HMDs have the advantages to be lightweight, compact, easy to program, 360° tracking, generally cheap, and let's experience a cinema-like viewing. However, they have low resolution, low field of vision (Arthur, 2000), apparent aliasing problems, high latency between the time a user repositions his/her head and the time it takes to render an update to the scene (Mine et al, 1993), level-of-detail degradation in the periphery (Watson et al., 1997) are serious drawbacks. Head-mounted displays with wide-angle optics can also provide some of the same benefits. We do anticipate rapid advances in small head-mounted displays over the next few years. Major challenges will be in the physics and optics of getting correct field of view and stereoscopy parameters, as well as just getting a large number of colour pixels into a small package. An emerging form of heads-up display is a retinal display that 'paints' a picture directly on the sensitive part of the user's retina. Although the image appears to be on a screen at the user's ideal viewing distance, there is no actual screen in front of the user, just special optics (for example, modified eyeglasses) that reflect the image back into the eye. Some HMDs incorporate inertial sensors to determine direction and movement (for example, to provide context-sensitive geographic information) or as the interface to an immersive virtual reality application" (Gartner, 2010).

Figure 9. Virtual retinal display. Image taken from SIGGRAPH

Augmented reality (mixed reality) superimposes information on the surrounding environment rather than blocking it out. Thus, while in VR everything surrounding the viewer is illusory, the semitransparent display worn by the user allows providing additional information, like labels and diagrams onto objects in the real world. It has been suggested that this may be useful for training people to use complex systems, or for fault diagnosis. For example, when repairing an

aircraft engine the names and functions of parts could be made to appear superimposed on the parts seen through the display together with a maintenance record if desired (Caudell & Mizell, 1992; Feiner et al., 1993). The computer must obtain a detailed model of the environment; otherwise it is not possible to match the synthetic objects with the real ones. Even with this information, correct registration of computer graphics with the physical environment is an extremely difficult technical problem due to measurement error and system latency. This technology has been applied to heads-up displays for fighter aircraft, with semi-transparent information about flight paths and various threats in the environment projected on the screen in front of the pilot (Stokes et al., 1990), as well as digitally augmented desk surfaces (Wellner, 1993). With the potential for displaying a larger image in view, tablets with two video cameras, a powerful processor and access to the Internet will make AR applications exciting for a range of uses including, tourism, architecture, engineering, medicine, and education. Today, a foreign tourist can take a picture of a restaurant sign and gain access to the menu in his or her own language. It would also be possible to provide the specials of the day and the local critic's reviews all translated in real time. Similarly, AR could provide a tourist with a guided tour through an historic neighbourhood and learn about the people and events that happened in the past. Filters could be added to confine the information to recent history or perhaps, to provide the architectural history of significant buildings in the area (Benko, Ishak & Feiner 2003, 2004; Bimber 2005; Gutiérrez, Vexo & Thalmann 2008a, 2008b).

Figure 10. Google glasses. Augmented reality

Wide varieties of applications of VR and mixed reality have emerged and span many areas of human needs such as product design, interactive computer applications, medical trainers, and rehabilitation. Other opportunities include the training of sensory-motor skills in general (Crison et al. 2005).

Immersive Video (IV) technology stands for 360° video applications, such as the Full-Views Full-Circle 360° camera. IV can be projected as multiple images on scalable large screens, such as an immersive dome, and can be streamed so that viewers can look around as if they were at a real scenario. This technology allows the user to navigate in any direction while looking at a video. The scenario in generally available for a 360° view, but it is visible in a reduced portion at time, changeable according to the user's preference. A pentagon-shaped room (StarCave), with three screens on both sides and two screens at the top and bottom was created by the UC San Diego division of the California Institute for Telecommunications and Information Technology (Calit2, www.calit2.net). The system has a resolution of 68 million pixels - 34 million per eye - distributed over 15 rear-projected walls and two floor screens and allows that scientific models and animations are projected in stereo on 360- degree screens surrounding the viewer, and onto the floor as well.

Multitouch Tables and Screens Not only the screen, but the screen orientation could affect the

usability of different devices. Research has demonstrated that large touch screens mounted horizontally afford uses distinct from screens mounted vertically. “Surface computers are large-screen displays that support direct interaction via touch or gesture (as opposed to external devices, such as mice or keyboards). They are typically horizontal, often built into the furniture, such as a table top, but may be delivered as vertical wall-mounted or free-standing displays. The displays incorporate much of the style of interaction (such as rotate, pinch, zoom and flick movements) found in multitouch devices but can typically recognize more than one set of touches at a time, enabling multiple users to interact or work collaboratively. Some also have the capability to recognize physical objects marked with a special identification tag, allowing context-sensitive information to be provided when items are placed on the display. Their size is constrained by the ability to physically reach across the surface. Larger displays may require a noncontact approach involving a gestural interface, where the user does not need to physically touch the surface” (Gartner, 2010).

Large horizontal screens seem better suited for multi-user usage scenarios, because multiple users could sit or face one another (Shen, Everitt, and Ryall 2003). However, the different viewing angle of the users creates problems (Shen et al. 2004). One solution to this problem is to alter the view seen from each side of the table (Matsushita et al. 2004), by using head-worn displays as the sole display mechanism (Agrawala et al. 1997), or in combination with other information displays, such as projection directly onto an input device (Benko, Ishak, and Feiner 2004).

Sharing a single tabletop, however, leads to unconscious separation of the tabletop nto multiple territories: personal, shared, and storage (Scott et al. 2004), varying by the size of the table (Ryall et al. 2004). It is not clear how much of this separation is due simply to issues of comfort of reach, as described by anthropometricists as the kinetosphere (Toney and Bruce 2006), versus those dictated by the mores of social distance as studied in the field of proxemics (Ballendat, Marquardt, and Greenberg 2010). Viewing information horizontally and from different sides has been shown to lead to perceptual differences (Wigdor, Shen, et al. 2007) and to improve visual search efficiency (Forlines et al. 2006).

An example of large tabletop displays is Microsoft© Surface. The multi-touch surface allows interaction using fingers of any other objects when placed on the screen. With this opportunity the de- vice supports recognition of human's natural hand gestures as well as interaction with real objects and shape recognition. The large 30 inch display allows more than one person to interact with the screen at the same time. The recognition of the objects placed on the tabletop pc then provides more information and interaction. So it is perhaps possible to browse through different information menus about the placed item and obtain more digital information.

8. Haptic interfaces

Our touch (haptic) sense is such an integral part of our everyday experience that few of us really notice it. Touch is unlike any other human sense in that sensory receptors related to touch are not associated to form a single organ. Humans perceive haptic sensory information using skin, muscles, tendons, joints, and mucosae (Klatzky and Lederman, 2002). Within and beneath our skin lie layers of ingenious and diverse tactile receptors comprising our tactile sensing subsystem. These receptors enable us to parse textures, assess temperature and material, guide dexterous manipulations, find a page’s edge to turn it, and deduce a friend’s mood from a touch of his hand. Intermingled with our muscle fibres and within our joints are load cells and position transducers making up our proprioceptive sense, which tell our nervous

systems of a limb's position and motion and the resistance it encounters. In addition to tactile and kinesthetic sensing, the human haptic system includes a motor subsystem. Exploratory tasks are dominated by the sensorial part of the sensory motor loop, whereas manipulation tasks are dominated by the motor part (Jandura & Srinivasan, 1994).

Haptic modality is present in several of the input devices, already considered in this document like keyboard, mouse, touchpad, touchscreens, and joysticks. Internal sensations of body posture, motion, and muscle tension (Burdea 1996; Gibson 1962) may allow users to feel how they are moving an input device without looking at the device or receiving visual feedback on a display, therefore providing proprioceptive (or force) or kinesthetic feedback. This is important when the user's attention is divided between multiple tasks and devices ((Balakrishnan and Hinckley 1999; Fitzmaurice and Buxton 1997; Mine, Brooks, and Sequin 1997). Muscular tension can help to phrase together multiple related inputs (Buxton 1995) and may make mode transitions more salient to the user (Hinckley et al. 2006; Raskin 2000; Sellen, Kurtenbach, and Buxton 1992; Hinckley et al. 2010). The tactile and force feedback devices are sometimes referred as haptic displays. Specifically, tactile cues, such as vibrations or varying pressures applied to the hand or body, are effective as simple alerts, while kinesthetic feedback is key for the more dexterous tasks that humans carry out (Biggs & Srinivasan, 2002; Hale & Stanney, 2004). Active haptic devices are interfaces to computers or networks that exchange power (e.g., forces, vibrations, heat) through contact with some part of the user's body, following a programmed interactive algorithm. Cell phone vibrators and force feedback game joysticks are also active haptic interfaces; whereas the vibrator is only a display, the joystick is both an input and an output device, and its control is considerably more complex.

Most haptic devices share the same principles: on the one hand, it is an input device, providing the application the position and (possibly) the orientation of a certain point in space. On the other hand, the device also is capable of generating forces that are felt by the user. This dual function of the haptic devices – as input and as output devices, providing feedback is the main reason to label this section "haptic interfaces". The user interface represents the fusion of the input and output, the hardware and software elements with a coherent model of the function of the system. Haptic interface hardware consists of the physical mechanism that is used to couple the human operator to the virtual or remote environment. This hardware may be a common computer gaming joystick, a multiple-degree-of-freedom (DOF) stylus, a wearable exoskeleton device, or an array of tractors that directly stimulate the skin surface.

Tactile based technology is the only technology that physically requires you to touch something, all the others in some sense can operate hands free. The most classically touch technology is the button, as used on the keyboard, which is the most popular HCI device ever created, we use it in an incredible amount of applications ranging from the computer keyboard, mobile phones, to personal entertainment devices. Contrary to vast number of force feedback devices on the market, there are not many commercially available tactile interfaces. Nowadays, almost all mobile phones have a vibrating mode. Nintendo Wii (2012) and Logitech Driving Force™ GT (2012) are two examples of tactile interfaces used in computer games for better realism and immersion. Currently, tactile technology in touch screens and mobile phones is going beyond the primitive haptics and presenting the boundaries or surface properties of an object on screen as you move your finger over it. TouchSense® tactile technology from Immersion Corp. (2012) is claimed to provide "HD haptics" using piezo actuators. This technology is already integrated in Immersion's touch screens and some mobile phones such as Synaptics Fuse (2012). It is also used in cars to facilitate drivers to select an icon on the control menu. They are so popular tactile devices that we have even woven small versions (using conductive materials) into our clothing, now called smart clothing or smart textiles (Marculescu, Marculescu, & Jayaraman, 2003), so that we can control our

entertainment devices no matter where we are and what we are doing. Some HCI devices are quite unique and imaginative, Digital clay (Ngoo, 2009) for instance, is clay that can be moulded, and the shape digitally transferred to a computer, it has great future potential if a truly 3D application can be developed and then combined with a 3D printer. Interestingly enough motion sensing gloves, which you might have thought fit in the Motion section, actually fits under the tactile category, although their main function is to provide motion feedback they are still mostly operated through tactile interaction.

8.1 Force Haptic Displays

Wearable haptic interfaces: They are worn by the user and could be classified as arm exoskeletons or hand masters. Arm exoskeletons are typically attached to a back plate and to the forearm. Hand masters, on the other hand, are attached to user's wrist or palm. As compared to point contact devices, exoskeletal devices are capable of measuring location of various human joints and can provide feedback at multiple locations. Thus, with an exoskeleton- type interface the user is no longer restricted to interact with a single point in the workspace, but can use the whole arm as with an arm exoskeleton, or grasp and manipulate multidimensional objects using a hand master. In addition, wearable devices have a workspace that is comparable to the natural human workspace. In the field of robotics research, exoskeletons have often been used as master manipulators for tele-operations. However, most master manipulators entail a large amount of hardware and therefore have a high cost, which restricts their application areas. The first example of a compact exoskeleton suitable for desktop use was published in 1990 (Iwata, 1990). The device applies force to the fingertips as well as the palm. Lightweight and portable exoskeletons have also been developed (Burdea, Zhuang, Roskos, Silver, & Langlana, 1992).

8.2 Tool-Handling Type of Force Display

The tool-handling type of force display is the easiest way to realize force feedback. The configuration of this type is similar to that of a joystick. Unlike the exoskeleton, the tool-handling type force display is free from the need to be fitted to the user's hand. It cannot generate a force between the fingers, but has practical advantages. A typical example of this category is the pen-based force display (Iwata, 1993). Another example of this type is the Haptic Master (2002). It is the only is the only admittance controlled haptic interface on the market i.e. based on the measured force applied by the user the device is controlled to move proportionally to this force. The Phantom device is an example of a 3D force feedback device. Recently, a new handle design for the 6-DOF family of haptic devices permits attaching interchangeable new end effectors providing pinch functionality.

Figure 11. PHANTOM Omni® from Sensable Technologies, Inc.®. Image taken from Samur, 2012

8.3 Object-Oriented Type of Force Display

The object-oriented type of force display is a radical idea for the design of a haptic interface. The device moves or deforms to simulate the shapes of virtual objects. A user can make physical contact the surface of the virtual object. (e.g. Tachi et al., 1994).

Passive Prop

A passive input device equipped with force sensors is a different approach to the haptic interface. Murakami and Nakajima used a flexible prop to manipulate a three dimensional (3D) virtual object (Murakami & Nakajima, 1994). The force applied by the user is measured and the deformation of the virtual object is determined based on the applied force. These passive devices allow users to interact using their bare fingers but they could not represent the shape of virtual objects.

8.4 Proprioception and Full-Body Haptics

One of the new frontiers of haptic interface is full-body haptics that includes foot haptics. Force applied to a whole body plays a very important role in locomotion. The most intuitive way to move about the real world is walking on foot. Locomotion interface is a device that provides a sense of walking while the walker's body is localized in the real world. Examples of this type of interface are the sliding device (Iwata & Fujii, 1996), treadmill (e.g. Noma, Sugihara, & Miyasato, 2000), foot pad BiPort (http:// www.sarcos.com) and others.

Haptic devices are used for stroke rehabilitation, as a Braille haptic display and for Braille navigation aid, in gaming. They have serious application in medicine like surgery simulators and in surgical robotics.

9. Background Sensing Techniques

Background sensing techniques are a result from the technology development that permits to passively detect different user's characteristics and to use them in the interaction with the user. The intensive use of sensors, predominately in the mobile device is closely related to the idea of adaptive interfaces as the existing and occurring sensors provide better context awareness about the user. Researchers are currently exploring ways that will allow the technology to interpret the context of a situation and to respond more appropriately using the information obtained through location sensing, ambient sensing of light, temperature, and other environmental qualities, movement and handling of devices, detecting the identity of the user and physical objects in the environment, and possibly even physiological measures such as heart-rate variability, skin conductance or others (Schilit, Adams & Want, 1994; Schmidt, 1999; Dey, Abowd & Salber, 2001; Hinckley et al., 2003). Background interaction can ensure better fit between individual human activities and the technology making intelligent use of passive behavioural measurements, such as observation of typing speed, manner of moving the cursor, sequence and timing of commands activated in a graphical interface (Horvitz et al., 1998), and other patterns of use. For example, a carefully designed user interface could adapt itself to provide an appropriate interaction with the user based on inferences about the user's alertness or expertise. Such possibilities do not necessary need additional sensors as the information is already in the input stream. These are sometimes known as intelligent or adaptive user interfaces, but mundane examples also exist. For example, cursor control using the mouse or scrolling using a wheel can be optimized by modifying the device response depending on the velocity of movement (Jellinek & Card, 1990; Hinckley et al., 2001).

Input can also become a by-product of our activities in the world at large. For example, our location can be sensed through GPS and our movements can be captured using CCTV cameras, providing inputs to a range of interactive technologies. Low-cost Radio Frequency Identification (RFID) tags can also be tracked and provide new forms of information that can be fed into supply chains.

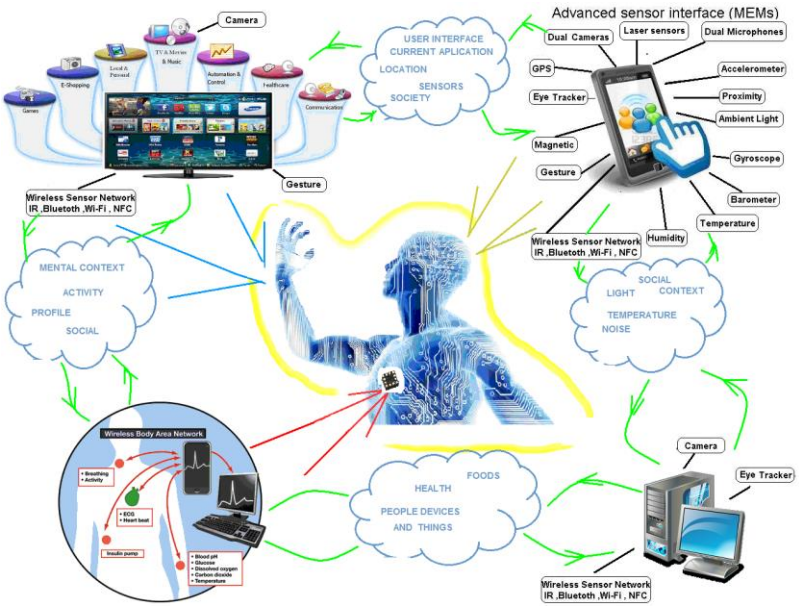

Figure 12. Illustration of Mobile phone sensing abilities. Image taken from Lane et al, 2010

The background sensing technique allows that the computer attend to user's needs through anticipating what the user wants to do. Instead of users being in control, deciding what they want to do and where to go, the burden should be shifted onto the computer. IBM's BlueEyes project (2000) is an example of using sensor-rich environment to track and identify users' actions. The information obtained by non-obtrusive sensing technology is then analysed with respect to where users are looking, what they are doing, their gestures, and their facial expressions. In turn, this is coded in terms of the users' physical, emotional or informational state and is then used to determine what information they would like. For example, a BlueEyes-enabled computer could become active when a user first walks into a room, firing up any new email messages that have arrived. If the user shakes his or her head, it would be interpreted by the computer as "I don't want to read them," and instead show a listing of their appointments for that day.

10. Biosensors - Direct Muscle-Based Input and Brain-Computer Interfaces

Traditional input devices can be thought of as secondary sensors, in that they sense a physical action that is the consequence of cognition and muscle movements. An alternative approach is to attempt directly sense the brain activity and muscle movements directly. Brain-computer interfaces could be invasive, with microelectrodes directly implanted in the gray matter of the brain during neurosurgery in an effort to capture brain activity more accurately and non-invasive, using external systems, such as electroencephalography (EEG) or functional near-infrared spectroscopy (fNIRS), to measure brain activity. EEG measures brain activity with electrodes placed on the surface of the scalp that registers the electrical activity caused by neuronal firing. The electrical signal is difficult to monitor as it is only few microVolts and it needs to be amplify up to 100 000 times. This makes the EEG sensitive to eye, face and body movements and to the presence of near-by electronic devices. However, as it measures electricity—a direct result of neuronal activity—it has very high temporal is thus very fast, even though it is also spatially indeterminate. A recent example of a BCI that uses EEG is a wheelchair that can be controlled through brain activity, created by Rebsamen et al. (2007). The researchers created a list of paths to locations in a small apartment and then presented those target locations to users. To select a target, the users were instructed to focus on that target when it was presented to them. After several minutes of training with a participant, the system could detect the desired location with almost perfect accuracy. A positive peak occurs in the central and parietal cortex about 300 ms after the presentation of infrequent stimulus. This response is termed “P300”. A P300 BCI was successfully used in a speller (Farwell & Donchin, 1988). They evaluated the performance of this technique and achieved high transfer rates up to 97.57 bits/min (47.26 bits/min). It is the highest bit rate for EEG-based BCIs that we found in literature.

The state of the art is that correct decoding of EEG signal is possible to a very large extent. It is still not good enough for applications since the erroneous responses in a remaining 10% can lead to completely wrong actions. The main challenge in BCIs EEG-based is to identify the particular EEG signal components (features) that can be successfully used as control commands. The main, but not unique, problem in these approaches is that single trial EEG data is very noisy, with data stemming from many sources. The characteristic responses to specific events are usually obtained by averaging signals from many trials, like in evoked signals. To successfully match single trial data, the relevant source of the signal needs to be separated out before it can be matched to average templates. There are few new concepts in the design of EEG measurement systems like miniaturized, battery-powered front-end close to patient, with

fiber optic data transfer to the signal processing PC (see Farshchi et al, 2004; Weinmann, & Schroeder, 2003), or use of active electrodes, which have the property that the first amplifier stage is integrated within the electrode. Future progress will depend on (Wolpaw et al, 2002):

1. Identification of those signals, whether evoked potentials, spontaneous rhythms, or neuronal firing rates, that users are best able to control;
2. Development of training methods for helping users to gain and maintain that control;
3. Delineation of the best algorithms for translating these signals into device commands;
4. Attention to elimination of artifacts as electro-myographic and electro-oculographic activity;
5. Adoption of precise and objective procedures for evaluating BCI performance.

The fNIRS measures changes in the blood flow. It uses optical wires to emit near-infrared light and the sensors detect the reflected light from different tissues of the head, including the brain. Sensors in the system detect changes in the oxygenated and deoxygenated blood in that region (Chance et al, 1998). The basic technology is common to all systems, but the measured signal differs depending on the location of the probe and the amount of light received. There are many possible placements of fNIRS probes, allowing the study of multiple brain regions. The most common placements are on the motor cortex (Sitaram, et al.), and the prefrontal cortex (PFC) (Ehlis, Bähne, Jacob, Herrmann, & Fallgatter, 2008; Mappus, Venkatesh, Shastry, Israeli, & Jackson, 2009), although other regions have also been explored (Herrmann, et al., 2008). This technology provides better spatial resolution, but less sensitivity to the temporal changes in the brain. Benefits of using fNIRS include ease of use, short setup time, and portability, making it a promising tool for HCI researchers. Additionally, the part of the fNIRS system placed on the scalp or forehead is typically small and therefore less bothersome to users than other brain measurement technologies. Using fNIRS it was possible to detect workload and user engagement (Hirshfield, Chauncey, et al. 2009) in order to conduct usability studies, as well as to explore the possible dynamic adaptation of user interfaces based on such metrics (Hirshfield, Solovey, et al. 2009).

There are many successful direct control paradigms using EEG signal generally used to move mouse cursors or type on a keyboard. Direct control involves a structured mental activity that results in an explicit command to the computer. To perform the action, you have to imagine it. These direct-control interfaces rely on the fact that the brain activity occurring when you move your hand to the right is very similar to the activity that occurs when you imagine moving your hand to the right. This consistency can be used to pair mental “movements” with commands: when participants imagine waving their arms up and down, for example, the volume on their phone might mute, or the zoom level on their screen might change. To perform direct control, however, requires a lot of training and this reduces the used of these technologies in ordinary computer environment. Related applications have been used as an alternative means of communication for motor-impaired people in order to provide them with the basic tools that help them to communicate with the others and to implement some programs. Interestingly, some studies on BCI have been recently extended into areas of entertainment such as interactive human-computer games for healthy users. BCI applications have appeared in entertainment and the game industry as a (Blankertz et al, 2006; Pfurtscheller et al, 2006; Nijholt et al 2008; Nijholt, 2009).

Passive BCI were used with healthy individual to detect brain activity that occurs naturally during task performance. Thus, it allows focusing on the brain as a complementary source of information, as additional, not as a single input to the computer. The principal advantage of passive BCIs is that they do not add to the user’s task. Such approach was successfully

implemented to detect task difficulty (Peck et al, 2010). This allows adapting and controlling applications using brain signals.

Magnetoencephalography (MEG) and even functional magnetic resonance imaging (fMRI) have both been used successfully as rudimentary BCIs, in the latter case allowing two users being scanned in real-time to play Pong against one another by altering their haemodynamic response through various biofeedback techniques.

The muscle activity sensing is accomplished through electromyography (EMG). It uses electrodes to measure the electrical activity caused by muscle contraction. However, the EMG provides much strong signal than EEG (in the range of miliVolts), so it requires less amplification. Saponas et al. (2009) demonstrated its use to enable sensing of muscle activation as fine-grained as detecting and identifying individual fingers, and used in combination with touch-screen input to provide a richer data stream (Benko et al. 2009).

EMG based interfaces generally involve signal acquisition from a number of several electrodes, signal processing (feature extraction) and real-time pattern classification (Zecca et al, 2003; Crawford et al, 2005). Classification methods based on both statistical and neural network approaches have been made with satisfactory results. However, given the complexity of the task and the variability of the EMG signals these systems usually require calibration for each user or training of the pattern recognition algorithms. In a different fashion EMG signals have been used in conjunction with other physiological signals (skin conductivity, blood pressure and respiration) to detect the affective state of the user. Experimental results from a preliminary study show that even with simple processing techniques it is possible to detect brief muscle contractions in data acquired from moving subjects. The results encourage further development of this kind of interface (Costanza et al, 2004). EMG has a great potential for control of prosthetic limb, predominantly for control of active prosthetic hands or for enhancing body strength using exoskeleton (Kawamoto, & Sankai, 2002). Another interesting area is EMG motionless gestures (Costanza & Perdomo, 2004), which gives the user a more private experience as it can provide “invisible” input to a mobile device.

11. Multimodal interfaces

When people communicate they use multiple ways to convey information. Research has shown that speech is not the most important method of communication between people, but research has shown that body language (gestures and facial features) are just as important. Mehrabian (1968) indicated that the verbal part (i.e., spoken words) of a message contributes only for 7 percent to the effect of the message as a whole, the vocal part (e.g., voice intonation) contributes for 38 percent, while facial expression of the speaker contributes for 55 percent to the effect of the spoken message. With multimodal interfaces, the idea is to use multiple input channels in human-computer interaction. Multimodal interfaces describes interactive systems that seek to imitate the natural human-human interaction and rely on the natural human capabilities to communicate via speech, gesture, touch, facial expression, and other modalities. Modality is a sense that is used in human-computer interaction (or in human-human interaction) while communication channel conveys information using a modality in a specific way. The definition of these channels is inherited from human types of communication which are basically his senses: vision, hearing, touch, smell, and taste. A channel has direction from human to computer or from computer to human. The former direction is denoted as input and the latter as output. Examples of communication channels are face analysis, gaze tracking, speech recognition, display, auditory information channel, and keyboard.

The goal of research in multimodal interaction is to develop technologies, interaction methods,

and interfaces that remove existing constraints on what is possible in human–computer interaction, towards the full use of human communication and interaction capabilities in our interactions. The literature on formal assessment of multimodal systems suggests that multimodal interfaces provide better flexibility and reliability, can offer interaction alternatives to better meet the needs of diverse users with a range of usage patterns and preferences and are in general preferred over the unimodal alternatives Xiao et al., 2002; Xiao et al., 2003; Oviatt et al., 2005; Bohus and Horvitz, 2010). Humans may process information faster and better when it is presented in multiple modalities (van Wassenhove et al., 2005). Other potential advantages of multimodal interfaces include the following (Oviatt et al., 2000):

- They permit the flexible use of input modes, including alternation and integrated use.
- They support improved efficiency, especially when manipulating graphical information.
- They can support shorter and simpler speech utterances than a speech-only interface, which results in fewer disfluencies and more robust speech recognition.
- They can support greater precision of spatial information than a speech-only interface, since pen input can be quite precise.
- They give users alternatives in their interaction techniques.
- They lead to enhanced error avoidance and ease of error resolution.
- They accommodate a wider range of users, tasks, and environmental situations.
- They are adaptable during continuously changing environmental conditions.
- They accommodate individual differences, such as permanent or temporary handicaps.
- They can help prevent overuse of any individual mode during extended computer usage

Multimodal interfaces also allow the usage of the devices by a wider audience that may not have been able to use them before. For example, people with disabilities or temporary illness may not have full use of their motor control and can benefit from alternative input methods. Young or old people may also face challenges in using a device that was not intended for them, as well as people that speak other languages (Oviatt, 2002).

Multimodal systems and architectures vary along several key dimensions or characteristics, including the number and type of input modalities; the number and type of communication channels; the ability to use modes in parallel, serially, or both; the size and type of recognition vocabularies; the methods of sensor and channel integration; and the kinds of applications supported. Key issue in multimodal integration (or fusion) is how and when modalities should be integrated (see Johnston et al., 1997; Johnston, 1998; Wu et al., 1999; Nakamura, 2002; Chai et al., 2004; Johnston and Bangalore, 2005; Wasinger, 2006; Portillo et al., 2006; Mendonca et al., 2009; Song et al., 2012). Modalities have different characteristics and may not have obvious points of similarity and straightforward ways to connect e.g., speech and eye gaze or facial expression and haptics input. Different modalities may have different temporal constraints and different signal and semantic endurance. To describe the possible way of integrating the different modalities Nigay and Coutaz (1993) offered a classification that depends on the fusion method (combined or independent) and the use of modalities (sequential or parallel). In an exclusive multimodal system, the modalities are used sequentially and are available separately but not integrated by the system. In an alternative multimodal system, modalities are used sequentially but they are integrated to some degree (across time). In a concurrent multimodal system, modal information is available in parallel, but separately (not integrated). Finally, in a synergistic multimodal system, the modes are available in parallel and fully integrated. While synergistic multimodal systems are the

assumed goal here, there are still possible benefits of the other styles of multimodal interfaces over unimodal systems. The combination of the modalities could be early, during the (pre-) processing of the input signals or after the processing the signals from each modality. The first method is known as feature-based fusion and it is appropriate for combining two related modalities like speech processing and lip movement. The advantage of this method is that the information from the two channels could complement each other and disambiguate the incoming information. However, any change in one of the input modalities requires a "retraining" of the system. The late integration of the signals from different modalities is also known as semantic-level fusion. It is much easier to create and extend, but it does not have the benefit of direct complementary information. Semantic-level integration is usually done either through unification of existing data or by looking for missing data. All multimodal (semantic-level) input systems need some kind of time stamping in order to combine and interpret a combination of input measures. Time stamping should occur at least at the beginning and end of each input signal.

It is still a question of debate whether the fusion of different modalities should be early or late. Another open question is whether the task of multimodal integration to produce a multimodal event or it is to produce a more complex representation of perceptual activity that may better match the human interaction which the system is intended to support.

A classic example of a multimodal system is the "Put That There" demonstration system (Bolt, 1980). This system allowed one to move an object into a new location on a map on the screen by saying "put that there" while pointing to the object itself then pointing to the desired destination.

Another common multimodal interface combines pen and speech. This combination is useful in noisy environments where speech input is primary; if the computer cannot distinguish between similar sounding words, it can display a simple selection box for confirming the correct word. This combination suppresses errors by 19 to 41 percent when compared to unimodal speech inputs.

Combining speech with lip movement is valuable in noisy environments, but does not offer many additional benefits in quiet conditions (Oviatt, 2003).

Successful implementation of multimodal interface is the emotion recognition using auditory and visual modality. In one of the first bimodal emotion recognition studies, De Silva et al. (1997) found that some emotions were better recognized by humans through the auditory modality than the visual modality, and vice versa: anger, happiness, surprise and dislike were more visual dominant, and sadness and fear were more audio dominant.

Although face analysis can be applied alone in many applications, even more applications become available when it is used with other perceptual technologies such as speech recognition and haptic feedback. Person identification based on multiple input channels is a topic that has been under intensive research. As an example, Brunelli and Falagvina (1995), Chibelushi et al. (1997), and Faraj and Bigun (2007) used speech along with facial cues for person identification. Ali et al. (2006) integrated face and fingerprint biometrics. In all the studies the identification accuracy was improved by combining the input channels together. Bevacqua et al. (2006) proposed an interactive agent that would recognize a user's facial expressions, head movements, and hand gestures and would act according to the interpretation and behaviour model of the agent.

An example of combining 3 different interaction types: Multi-touch, Video gesture and pointing, and speech recognition, with the computer could look like was introduced by Microsoft in their vision of a future home. It was originally developed for the usage in a future

kitchen but describes perfect how the future Human Computer Interaction may look like. In this approach they are multi-modal interfaces with combining video and speech recognition. Therefore they are using video to detect goods in the kitchen and video projection to display the user interface directly on the kitchen surface. The detection, for instance, can be imagined like that the systems recognized which ingredient is placed on the surface. For the navigation through the interface they are then combining this video detecting method with speech detection.

It does not mean that just because a device is designed for multimodal inputs people will interact with it multi-modally. People tend to interact multi-modally when their context allows for it, which may not be all of the time. They know which input methods are best for their context, and use those inputs accordingly.

12. Tangible interfaces

Tangible interaction is an extension of Direct Manipulation (Schneiderman, 1983). Tangible user interface (TUI) is a physical representation of digital information that one is able to touch in order to manipulate the digital data. It is user interface that augments the real physical world by coupling digital information to everyday physical objects and environments (Ishii & Ullmer, 1997).

Broadly viewed, tangible interfaces give physical form to digital information. The approach has two basic components. First, physical objects are used as representations of digital information and computational operations. Secondly, physical manipulations of these objects are used to interactively engage with computational systems. The most popular application of tangible interfaces has been using physical objects to model various kinds of physical systems, like in the layout of assembly lines (Schäfer et al, 1997; Fjeld et al, 1998), optical systems, buildings (Underkoffler et al. 1999), and furniture (Fjeld et al. 1998). Another approach is based on constructing assemblies by separate building block similar to the LEGO™ concept. It was used for modelling buildings [Aish 1984; Frazer 1994; Anderson et al. 2000], fluid flow [Anagnostou et al. 1989], and other geometrical forms [Anderson et al. 2000]. These systems extend the existing physical representations and work practices with the benefits of computational augmentation. Another type of TUIs are the so-called “tokens and constraints” that could be used to represent abstract information (i.e. with no inherent physical or geometrical representation) using mechanical constraints (Ullmer, Ishii, & Jacob, 2005). Tokens are discrete, spatially reconfigurable physical objects that represent digital information or operations. Constraints are confining regions within which tokens can be placed. Constraints are mapped to digital operations or properties that are applied to tokens placed within their confines. Constraints are often embodied as physical structures that mechanically channel how tokens can be manipulated, often limiting their movements to a single physical dimension (Ishii, 2008).

Tangible user interfaces allow users to interact with digital information through grasping and manipulating physical objects, and through gestures. By allowing users to draw on their natural skills for interacting with digital information, tangible user interface could reduce the cognitive load required for performing a computational task, and offer an intuitive and collaborative interface to support activities such as learning, problem solving, design, and entertainment. However, the application of TUs requires proper choice of metaphors that give physical form to digital information, and to determine which information is best represented digitally and which is best represented physically [Ullmer 2002]. Moreover, as the behaviour of the physical object in a TUI depends not only on its nature, but also by that objects context of use (the

behaviour of a physical interaction object may change when a new physical object is added to the TUI or when it is physically associated with another physical object), it is very important to define each possible context.

12.1 Comparison with GUI

The Graphical user interface (GUI) is based on the “desktop metaphor” represents information (bits) with pixels on a bitmapped screen by simulating a desktop. These graphical representations could be manipulated with generic remote controllers like keyboard and mouse. Thus, the representation (pixels) is decoupled from control (input devices) which provides GUIs with flexibility to emulate a variety of media graphically. GUI made a significant improvement from the command-line user interfaces by utilizing graphical representation and “see, point, and click” interaction. It released the user from the need to” remember and type” characters. Another important design principle is “what you see is what you get” (WYSWYG). It serves as a general-purpose interface by emulating various tools using pixels on a screen.

As GUI uses windows, icons, menus made of pixels on bitmapped displays to visualize information, the representation is intangible. To make the pixels interactive, one has to use general “remote control” like the keyboard, the mouse, tablets, etc. Trying to achieve generality, GUI introduced a deep separation between the representation and the controls, provided by the keyboard and the mouse.

TUI serves as a special-purpose interface for a specific application using explicit physical forms. It uses tangible representations of information that also serve as the direct control mechanism of the digital information. While in GUI all the information representation is intangible, in TUI it has two parts: tangible and intangible that allows the users to more directly control the underlying digital information using their hand. In this way TUI can take advantage of the dexterity or skills for manipulating different physical objects, something unexplored in GUI.

While the keyboards, mice and other input devices in GUI could also be regarded as physical objects, their role is quite different than the role of the physical artifacts in TUI where they could change function depending on the context and make the digital information directly manipulable by our hands and perceptible through our peripheral senses, by physically embodying it (Ishii, 2008).

Other characteristics of TUI (Fitzmaurice, Ishii & Buxton, 1995) are:

- It encourages two handed interactions;
- Shifts to more specialized, context sensitive input devices;
- Allows for more parallel input specification by the user, thereby improving the expressiveness or the communication capacity with the computer;
- Leverages off of our well developed, everyday skills of prehensile behaviours for physical object manipulations;
- Facilitates interactions by making interference elements more “direct; and more “manipulable” by using physical artefacts; takes advantage of our keen spatial reasoning skills;
- Offers a space multiplex design with a one to one mapping between control and controller;
- Affords multi-person, collaborative use.

12.2 Comparisons with augmented reality

Both interaction paradigms mixed virtual and physical objects. In augmented reality the focus is on how the virtual augments/adds up to the real objects, while in tangible interaction the focus is on how the real objects allows us to better manipulate information (which tends to be virtual).

12.3 Examples

Marble Answering Machine (Bishop, 1992). It uses physical marbles as containers and controls for manipulating voice messages. The marbles are moved between different depressions or “wells” to replay marble contents, redial a marble message’s caller, or store the message for future reference. For example, to listen to a message the user picks up a marble and adds it to a special play in¬dentation on the machine. This TUI is an example of “tokens and constraints” type GUI.

The ToonTown system is a ‘‘virtual auditorium’’ with small figures representing users of a chat system [Singer et al. 1999]. It uses physical tokens covered by cartoon characters to represent users within the audio space. By manipulating these tokens upon an array of racks, new users may be added, or some of the old removed. The manipulation allows also audio localization of users; assignment of users to tokens; and the display of information relating to participants. The ToonTown system includes a number of interesting and provocative components. One of these is the physical representation of people, which we believe has powerful potential in future communication systems.

I/O Brush (Ryokai, Marti, Ishii, 2004) - I/O Brush is a drawing tool to explore colours, textures, and movements found in everyday materials by "picking up" and drawing with them. I

Doll’s Head (Hinckley, Pausch, Goble, Kassel, 1994) – provides a head prop – a sphere or a doll’s head for manipulating individual patient’s data by a neurosurgeon. By rotating the prop with hands causes a polygonal model of the patient’s brain to rotate correspondingly on the screen.

Navigational Blocks (Camarata et al, 2002) - Orientation, movement, and relative positions of physical Blocks support visitor querying, retrieving, understanding, navigation and exploration in a virtual gallery.

Beads (Resnick et al, 1998) are designed to engage children in creating dynamic patterns. Each Programmable Bead communicates with its neighbouring beads. String beads together in different ways gives different dynamic patterns of light.

Wacom (Fukuzaki, 1993) used a tabletop interface with devices that have a unique shape and a fixed, predefined function associated with it [5]. The idea is that the form or shape of the device reveals or describes the function it offers. Three character devices were defined: (1) eraser, which functioned to erase electronic ink, (2) ink pot which served to select from a colour palette and (3) a file cabinet which brought up a file browser to retrieve and save files.

Urp (Underkoffler & Ishii, 1999) - Urp uses scaled physical models of architectural buildings to configure and control underlying urban simulation of shadow, light reflection, wind flow, etc. In addition to a set of building models, Urp also provides a variety of interactive tools for querying and controlling the parameters of the urban simulation like a clock tool to change a position of sun, a material wand to change the building surface between bricks and glass (with light reflection), a wind tool to change the wind direction, and an anemometer to measure wind speed. Moving the building allows urban designers to be aware of the relationship

between the building reflection and other infrastructure.

Figure 13. a.ToonTown; b. Wacom; c. Marble Answering Machine; d. Urp; e. Doll's Head; f. Navigational Blocks; g. Beads; h. I/O Brush

12.4 Taxonomy of TUI

Fishkin (2004) provides taxonomy to classify the TU along two axes: metaphor and embodiment. The embodiment is related to the link between the input and output device. Based on the proximity and similarity of the input and output, Fishkin defines 4 levels: full (when the output device is the same as the input); nearby (the output is in close proximity to the input, it is tightly couple with the focus of input); environmental (when the input and output devices are related, but appear somewhat apart. This type is defined like "non-graspable" in Ullmer and Ishii (2001), and distant (when the output device is far from the input one, on another screen or even in another room. In this case the input device is like a remote control). With the increase of embodiment, the ''cognitive distance'' between the input mechanism and the result of that mechanism decreases. Examples: full – Illuminating Clay (Piper, Patti, & Ishii, 2002); nearby - I/O Brush (Ryokai, Marti, Ishii, 2004); environmental - Marble Answering machine (Bishop, 1992) & ToonTown (Singer et al, 1999); distant – Doll's Head (Hinckley et al, 1994).

To quantify the amount of metaphor, Fishkin grouped the metaphors in two groups: those that appeal to the shape of an object (''metaphor of noun''), and those that appeal to the motion of an object (''metaphor of verb''). The more that either type of metaphor is used, the higher the interface on this scale. He specifies four levels, with graduation in each: none (when the physical manipulations of an object are not based on real-world analogy); noun (input object is closely tied to the look of some real-world object, but the analogy ends there) or verb (when the analogy is with the act being performed, largely independent of the object it is being

performed), noun and verb (when there is an analogy between the actions affordable with a real object and the actions with the virtual object (like in “drag-and-drop” interface); full (in the user’s mind the physical and the virtual objects are the same). Examples: none – Beads (Resnick et al, 1998); noun –building-objects in Urp (Underkoffler & Ishii, 1999); verb –moving a building in Urp (Underkoffler & Ishii, 1999); noun + verb - ToonTown (Singer el al 1999).

Tangible user interfaces provide a new way to materialize Mark Weiser’s (1991) vision of Ubiquitous Computing where digital technology weaves into the fabric of a physical environment and make it invisible. However, instead of making pixels melt into an assortment of different interfaces, TUI uses tangible physical forms that can fit seamlessly into a users’ physical environment.

13. Future Trends

The field of human-machine interface continues to go through rapid changes with the introduction of new multi-sensory interfaces (speech, sound, haptics) and metaphors (gestures, avatar in augmented or virtual reality world, shared cognitive spaces). Large interactive displays, smart devices and embedded systems become more and more pervasive. Over the last few years, smart phone technology, has gone through significant evolution. Multitouch, inertial sensors, accelerometers, location awareness, video analysis and even direction are all becoming standard functionalities for high-end smartphones, enhancing familiarity with these non-traditional sensory interfaces and encouraging the move toward useful augmented reality applications. One can think that a number of applications will be available, such as: conferencing, culturally-assisted translation, live status tracking, biometry-based (e.g. facial) recognition, virtual assistant, and with other increasing intelligent applications. Figure 14 shows the status of HCI technology on the hype cycle (Gartner, 2013), a graphic representation of the maturity and adoption of technologies and applications, showing emerging technologies as well as indicating which technologies have gone through convincing focused experimentation (slope of enlightenment) and have found adoption (plateau of productivity). It clearly shows the intensity and speed of the present technology development.

One of the most influential ideas that shapes these developments is with certainty the vision of the ubiquitous computing. It means that technology will be designed so that it will be integrated seamlessly into the physical world in ways that extend human capabilities.

These are the early days of natural user interfaces for products and services, and we're going to see a lot of experimentation and more systematic user research invested in it. Technology aside, the factor that will clearly determine success and effectiveness of the new UI approaches will be the actual user experience and the feeling derived from it. The key fundamental design principles that will allow a successful user experience are: focus on the user task quickly and easily, flexibility to allow users to have a seamless visual experience as they switch between different devices, effortlessness by keeping the look simple, clean and consistent, and, finally, emotional engagement.

People are able to interact with the technology that surrounds them in more accessible, intuitive and less restrictive ways. In a growing ubiquitous computing world, computers will communicate through high speed local networks, over wide-area networks, and via infrared, ultrasonic, cellular, and other technologies. Data and computational services will be portably accessible from many, if not most, locations to which a user travels. Computation will pass beyond desktop computers into every object for which uses can be found. The environment will be alive and the addition of networked communications will allow many of these embedded computations to coordinate with each other and with the user.

With tons of ideas, experimentation, and technologies going around, it seems that, in the near future, more things around us will slowly be replaced by touch-free interaction. We keep on innovating to create new concepts that help us accomplish tasks in a better, faster, more efficient way, and that's the reason technology is important. Moreover the entire system will assist as cooperative partner to the users in accomplishing their intended tasks.

This future will become even more exciting if UI and relevant products get to the core of human nature, encompassing social and business interaction, through all media of communication, from written or spoken language to gestures and facial or conversational emotions, and respond accordingly.

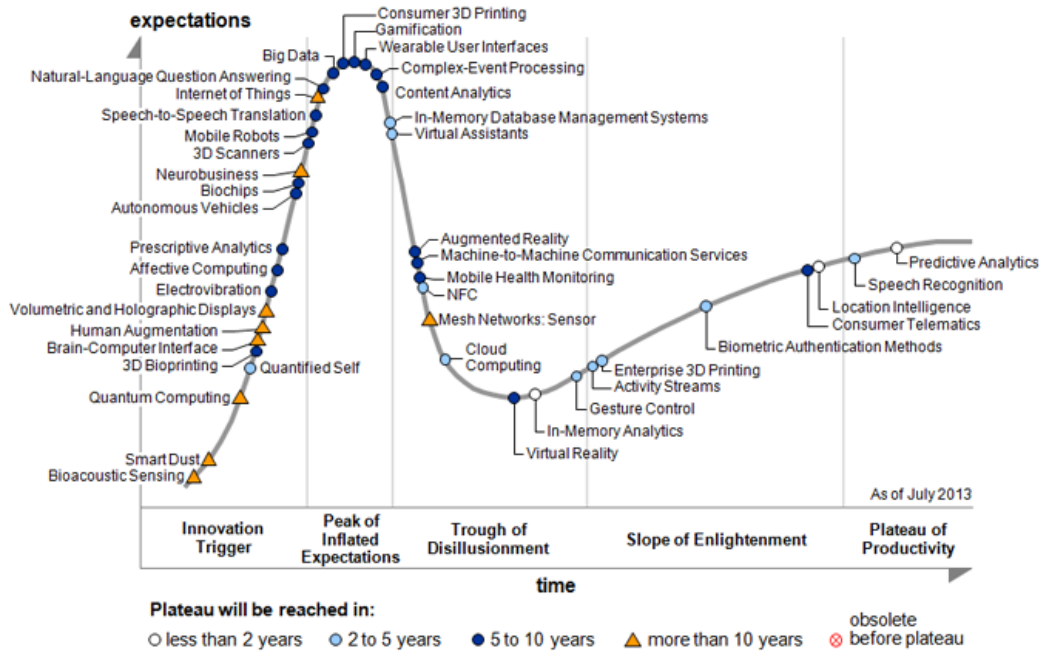

Figure 14. Hype cycle of Human-Computer Interaction, 2013, (Garner ,2013)

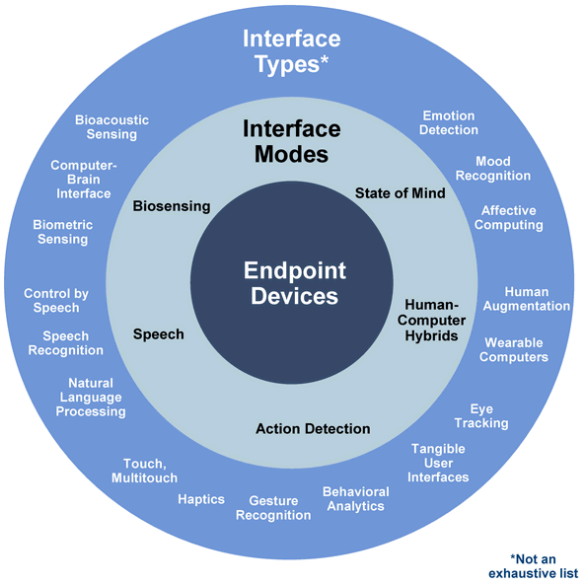

Figure 15. Interface types (Garner, 2011)