Interactivity: evolution and emerging trends

CONSTANTINE STEPHANIDIS\textsuperscript{1,2}, VASSILIS KOUROUMALIS\textsuperscript{1} AND MARGHERITA ANTONA\textsuperscript{1}

\textsuperscript{1}Institute of Computer Science
Foundation for Research and Technology - Hellas (FORTH)

Heraklion, Crete, GR-70013 GREECE
cs@ics.forth.gr

\textsuperscript{2}Department of Computer Science
University of Crete, Greece

Table of Contents

Table of Contents ................................................................. 1
1. Introduction ........................................................................... 2
2. The concept of interactivity .................................................... 3
   2.1 Interactivity in Human Communication ............................... 3
   2.2 Interactivity in HCI ......................................................... 8
3. Theoretical frameworks ......................................................... 12
   3.1 Human Factors and Ergonomics ....................................... 12
   3.2 Model Human Processor / GOMS ................................... 13
   3.3 User centred design ...................................................... 15
   3.4 UX, POET and Emotional Design ................................... 17
   3.5 Universal Access and Design for All ............................... 19
4. The evolution of interaction .................................................... 22
   4.1 Early stages .................................................................. 23
   4.2 Command-based interaction .......................................... 24
   4.3 Direct manipulation ...................................................... 25
   4.4 Conversational Interfaces .............................................. 28
   4.5 Web-based interaction .................................................. 32
   4.6 Mobile interaction ....................................................... 37
   4.7 Multimodal Interfaces .................................................. 43
   4.8 Virtual and Augmented Reality ..................................... 46
   4.9 Augmented Reality (AR) ................................................. 50
   4.10 Interaction in intelligent environments ............................ 52
   4.11 Summary of interaction paradigms ................................. 55
5. Emerging challenges, future trends and conclusions ................ 58
1. Introduction

In its short history, Human-Computer Interaction (HCI) is characterised by a trend towards elaborating, designing and establishing more human-oriented, natural and intelligent forms of interaction, progressively addressing the needs and requirements of a wider, less experienced and more naive user basis (Stephanidis et al., 1998). This path is intrinsically linked with (i) the progressive emergence of new, more general and systematic frameworks for studying and designing interaction, and (ii) technological evolution supporting the establishment of richer alternative interaction techniques and user interface styles.

The concept of interactivity between humans and computers, contextually defined in this Chapter as the extent to which the characteristics of an interactive system affect the communication behaviour of both the user(s) and the system itself, plays a crucial role in this respect. Interactivity is not a new concept, as it has been investigated in the literature with respect to both human communication and various types of interactive systems.

In order to shed light on interactivity in the context of HCI, understand its evolution and outline emerging challenges, this Chapter first reviews briefly existing accounts of interactivity, focusing on identifying dimensions of the concept which can be meaningfully used for analysing and explaining its various instantiations in HCI.

Second, the Chapter looks at how the various research disciplines that contribute to HCI have a different focus that leads to different theoretical models of design and evaluation. To understand the different perspectives in more depth, this Chapter examines the two parties of interaction, the human and the computer, and how insights from the contributing disciplines have shaped understanding of the interaction between them and of the interactivity of user interfaces.

It also looks into the evolution of interaction, starting from the 1950s until today, by concentrating on technological and research advancements that have led to the current interaction models and styles. A brief outline of the most important interaction paradigms is presented, highlighting the interactivity dimensions addressed in each of them.

Finally, it attempts to present the latest developments and emerging trends in HCI and address several issues that concern them.
2. The concept of interactivity

2.1 Interactivity in Human Communication

The concept of interactivity originates in the context of human communication, and has been addressed in various related disciplines, such as philosophy of language, linguistics, semiotics, and communication psychology. While a review of such accounts is beyond the scope of this Chapter, some basic aspects of interactivity in human communication naturally lend themselves to provide terms of comparison in analysing interactivity in HCI.

In the famous work on communication theory by Shannon and Weaver (1949), communication has been defined as a process whereby information is enclosed in a package and is channeled and imparted by a sender to a receiver via some medium. The receiver then decodes the message and provides feedback to the sender. All forms of communication require a sender, a message, and an intended recipient. However, the receiver does not need to be present or aware of the sender's intent to communicate at the time of communication in order for the act of communication to occur. Communication also requires that all parties share a common code or language for message exchange.

Interactive communication is commonly defined as a process involving at least two participants, where the content of a particular message is determined in part by the content of the prior messages from all participants, i.e., by the communication context (Chapanis, 1988). Interactive communication can take place through a symbolic system, notably natural language in spoken or written form, and complementary through gestures, facial expressions, and actions. Natural language is a unique symbolic system. Some of the most important distinguishing characteristics of human language are (Hockett, 1960):

- Vocal-auditory channel. Standard human language occurs as vocal communication (i.e., producing sounds with the mouth), which is perceived by hearing it. Exceptions are writing and sign language, which are examples of communication in the visual and manual channel respectively.
• Rapid fading (transitoriness). The human language signal does not persist over time. Speech waveforms fade rapidly and cannot be heard after they have faded. Writing and audio-recordings can be used to record human language, so that it can be reproduced at a later time.

• Interchangeability. The speaker can both receive and broadcast the same signal. This is distinctive from some form of animal communications.

• Total feedback. The speakers can hear themselves speak and can monitor their language performance. This differs from some other simple communication systems, such as traffic signals, which are not normally capable of monitoring their own functions.

• Semanticity. This is a fundamental aspect of all communication systems, implying that specific signals can be matched with specific meanings. Speakers of a language recognize the meaning to which a signal is associated.

• Arbitrariness. There is no necessary connection between the form of the signal and its meaning.

• Discreteness. The basic units of speech (such as sounds) can be categorized as belonging to distinct categories. There is no gradual, continuous shading from one sound to another in the linguistics system, although there may be a continuum in the real physical world.

• Displacement. The speaker can talk about things which are not present, either spatially or temporally. For example, human language allows speakers to talk about the past and the future, as well as the present. Speakers can also talk about things that are physically distant (such as other countries, the moon, etc.), or even refer to things and events that do not actually exist.

• Productivity. Human languages allow speakers to create novel, never-before-heard utterances that others can understand. Human beings are unrestricted in what they can talk about; no area of experience is accepted as necessarily incomunicable. This includes language and communication themselves. Thus, human language allows metalinguistic discourse.

• Learning: Human language is not something inborn. Although humans are probably born with an ability to do language, they must learn, or acquire, their native language from other speakers. This is different from many animal
communication systems where the animal is born knowing their entire system, e.g., bees.

Another inherent characteristic of natural language, which distinguishes it from formal languages, such as programming languages and command languages, is underspecification of meaning, which may take two forms, namely ambiguity and vagueness. Ambiguity refers to the fact that natural language words and utterances may be interpreted in different ways depending on context. Vagueness refers to the fact that natural language may refer to events and entities at an abstract level, omitting details that are not relevant in a specific context. Undespecification is often mentioned as a “defect” of natural language, which constitutes an obstacle to precise communication. On the other hand, it can be seen as an economy mechanism which allows human communication to be specific enough for a particular purpose with the minimum necessary effort (Wasaw et al., 2005).

Apart from the characteristics of language, human spoken dialogue can be analysed along a number of dimensions which appear to be relevant also in the wider context of HCI. These include (Petukhova & Bunt, 2009):

- Dialogue purpose and domain of discourse. Dialogues are usually motivated by goals, tasks, or activities which are non-communicative in their nature, e.g., to obtain certain information, to solve a problem, to act in a game, etc.
- Contact, presence, and attention. A basic requirement of communication is that the parties are in contact and stay so. For some types of dialogue this aspect is of a particular importance, namely when there is no or limited visual contact between the participants. For example, telephone conversations are dependend on the quality of the communication channel. But also, when dialogue participants have direct visual contact, they tend to permanently check the attention of their interlocutors and their readiness to continue the conversation. To this purpose, they utilise both their bodies and facial expressions (e.g., gaze) and a variety of vocal phenomena to show attention, as well as the type of reaction they expect from others.
• Grounding and feedback. Successful dialogue is based on shared knowledge and beliefs (Clark, 1996). Towards establishing such a common communication basis, speakers and addressees during dialogue attempt to confirm that each of them has understood what is uttered. This process is called grounding. Grounding includes feedback (Traum, 1999), i.e., during dialogue the speaker provides information on his/her own processing of the partner’s previous utterance(s).

• Taking turns. Turn management is another essential aspect of interactive conversation, and is defined as the distribution of the right to occupy the sender role in dialogue. Turn taking is usually understood as obeying normative rules, depending on the speaker’s needs or motivations and beliefs, and on the rights and obligations in a conversational situation.

• Social obligations and politeness. Participating in a dialogue is a social activity, where one is supposed to do certain things and not to do others, and to act in accordance with the norms and conventions regulating social behaviour. Each participant in dialogue not only has functional but also ethical tasks and obligations, and performs social obligation acts to fulfill them. Such obligations include politeness rules, such as do not imposing anything to the communication partner, offering alternative options, and encouraging positive feelings (Lakoff, 1973).

• Dialogue structure. Dialogue participants may at several dialogue stages indicate their view of the state of the dialogue and make the hearer acquainted with his/her plans for the continuation of the conversation. The speaker can give indications that he is going to close the discussion of certain topic(s); or that he/she wants to concentrate the hearer’s attention on a new topic. Dialogue structuring is based on the speaker’s view of the present linguistic context, on his/her plan for continuing the dialogue, and on the assumed need to structure the discourse for his/her partner.

• Handling errors. Speakers continuously monitor the utterance that is currently being produced or prepared to produce (Clark & Krych, 2004), and when problems or mistakes are discovered, they stop the flow of the speech and signal to the addressee that there is trouble and that a repair follows (error signalling). Human conversations contain large numbers of phenomena such
as disfluencies, interruptions, confirmations, anaphora, and ellipsis (Glass, 1999).

- Timing. Another aspect of communication which is concerned with disfluent speech production is time management, where the speaker suspends the dialogue for one of several reasons and resumes it after minor (stalling) or prolonged (pause) delay. Delays take place at all major levels of planning - from retrieving a word to deciding what to talk about next (Clarck & Fox Tree, 2002).

- Adaptation. One of the most robust findings of studies of human-human dialogue is that people adapt their interactions to match their conversational partners’ needs and behaviors (Pennebaker & King, 1999). People adapt the content, the syntactic structures of their utterances, as well as their lexical choices to match their partners’. They also adapt their speaking rate, amplitude, and clarity of pronunciation (Walker at al., 2007). Adaptation is also a crucial aspect of intercultural communication, i.e., people adjust their communication styles toward or away from each other during cross-cultural interactions (Cai & Rodriguez, 1996).

Besides speech dialogue, other aspects of human communication are also important. For example, recent phenomenological views on language and communication emphasise action associated with speech (Tripathi, 2005). These actions into which language is woven are inseparable from communicative meaning. Thus, language has an extra dimension associated with social conventions and actions, such as gestures, pointing and body language. Human beings have the ability to utilise their entire bodies for the purpose of communication (rather than simply voice or writing), thus implying multimodality. The tone of the voice, body language, and gaze, all constitute communicative meaning, either consciously or subconsciously (Bunt & Beun, 1998).

Human communication is also supported through semiotic systems other than natural language, namely iconic languages. Icons are semiotic signs which directly resemble the objects they refer to. In contrast to natural language, iconic languages are not arbitrary. Because of their communicative power, which transcends different languages and cultures, icons are used in a variety of real-life situations to inform
people about particular conditions or give instructions. Typical examples appear in 
public information spaces, trains, airplanes, cars, and printed books (Barker, 2000). 
The human ability to communicate through action and iconic languages is at the 
centre of the notion of direct manipulation (see section 4.3).

Finally, emotion plays a central role in human communication, especially when 
disagreement between participants emerges. Emotional reactions represent an 
important type of feedback on the effects of utterances on dialogue participants. In 
dialogue, emotional reactions can be signalled by response speed, reiteration of 
claims, lexical choice, response avoidance, sentence length, etc. Likely emotional 
reactions are defensiveness, indignation, frustration, anger, regret, guilt, and 
enthusiasm (Anderson & Guerrer, 1998).

2.2 Interactivity in HCI

Many definitions of interactivity have been provided in the HCI literature, especially 
with reference to web services, computer-supported communication, computer-
supported work, electronic advertising, eLearning, interactive TV, electronic games 
and virtual reality.

User-machine interaction was the focus of early definitions of interactivity, in which 
the emphasis was on human interaction with computers. To be interactive, a computer 
system must be responsive to users' actions. In this context, interactivity has generally 
been measured in terms of input or output devices, for instance the number of “point 
and click”opportunities on a computer screen (Shneiderman, 1998). Norman (1990) 
suggested that the interactive process is a repeated loop of decision sequences of a 
user’s action and the environment’s reaction.

However, though user-machine interaction is an important aspect of interactivity, it is 
not adequate to fully capture the concept, especially since the emergence of more 
advanced technology such as the Internet. As a result, researchers have started 
investigating interaction in a broader technological context, also considering other 
types of interaction, such as user-user interaction and user-message interaction.
User-user interaction is usually discussed from an interpersonal communication perspective. In this respect, the more communication in a computer-mediated environment resembles human interpersonal communication, the more interactive such environment is considered (Ha & James, 1998). However, a medium such as the Internet offers many possibilities to break the boundaries of traditional interpersonal communication. Through the Internet, people no longer need to be at the same place, and to be communicating at the same time. Research has shown that computer-mediated communication and face-to-face communication are not functional alternatives (Flaherty et al., 1998), as each has distinctive characteristics and addresses different needs.

From a user-message interaction perspective, interactivity is defined as the ability of the user to control and modify messages (Steuer, 1992). Whereas people have little control over messages in traditional media, the Internet gives users much more freedom in controlling the messages they receive, and allows to customize messages according to the users’ own needs. Based on the above, Liu & Shrum (2002) define interactivity as the degree to which two or more communication parties can act on each other, on the communication medium, and on the messages and the degree to which such influences are synchronized.

Other definitions have focused on two distinct aspects of interactivity: reciprocal communication and control (Liu, 2003). Reciprocity implies that interaction should allow two-way flow of information, and the information being exchanged in a sequence should closely relate to each other (Rafaeli & Sudweeks, 1997). Additionally, the exchange of information should happen in real-time. The control dimension implies that participants should be able to exert control on both sent and received information (Jensen, 1999), as well as over the communication medium. Both control and reciprocal communication are important aspects of interactivity. Control helps ensure a reciprocal exchange that satisfies the needs of all communicating parties, while reciprocal communication provides an effective channel for exerting control. Melding the two aspects, Liu & Shrum (2002) proposed three dimensions of interactivity: active control, which describes a user’s ability to voluntarily participate in and instrumentally influence a communication; two-way communication, which captures the bi-directional flow of information; and
synchronicity, which corresponds to the speed of the interaction. Based on the above, Liu (2003) defines a framework for measuring interactivity in web sites.

Interactivity has been discussed also in relation to new media and educational technologies. Rice (1984) defined "new media" as consisting of communication technologies that allow or facilitate interactivity among users or between users and information. Heeter (1989) describes six dimensions of interactivity in new media: (i) complexity of available choice, meaning the amount and variety of user choices; (ii) the effort that any user of a media system must exert to access information; (iii) responsiveness: interactivity is a continuous variable measuring how "actively responsive a medium is to users"; (iv) information use monitoring, that is, how well information selection can be monitored across a population of users; (v) ease of adding information, meaning the degree to which users can add information for access by the audience; and (vi) interpersonal communication facilitation, which comes in at least two forms: asynchronous (allowing users to respond to messages at their convenience) and synchronous (allowing for concurrent participation).

Downes & McMillan (2000), following an interview-based study, propose a conceptual definition of interactivity identifying six main dimensions: direction of communication, time flexibility, sense of place, level of control, responsiveness, and perceived purpose of communication.

In the domain of eLearning, Chou (2003) investigates a technical framework for interactivity following an empirical methodology. The framework includes the following dimensions: (i) choice, i.e., the amount and multimedia type of information users (learners and instructors) have access to, as well as other types of user options; (ii) non sequential access, the users can access information in a nonlinear way; (iii) responsiveness to learner, i.e., the system responds to a user's request without delay; (iv) monitoring information use, i.e., the system can collect data on the users themselves, their selections, their use of information, etc., and the users can monitor the personal information which is collected; (v) personal-choice helper, i.e., information is available to help learners who choose instructional content; (vi) adaptability, i.e., the interaction process and the exchange of information are adapted to individuals; (vii) playfulness, i.e., information stimulates users’ curiosity and fun;
(viii) facilitation of interpersonal communication, i.e., users can communicate asynchronously and/or synchronously; (ix) ease of adding, i.e., users can add information and content to the system.

Based on the instructional quality of the interaction, Schwier and Misanchuk (1993) identified three levels of interaction, namely reactive, proactive, and mutual interactions. A reactive interaction is a response to a given question. Proactive interaction involves learner’s construction and generation activities during the learning process. In a mutual interactive environment, the learner and system are mutually adaptive in reactions with each other. The relationships among the three levels of interaction are hierarchical in terms of quality of interaction. Therefore, the quality of a mutual-level interaction is higher than that of a proactive-level interaction, and the quality of a proactive-level interaction is higher than that of a reactive-level interaction. Consequently, the higher levels of interaction provide a greater opportunity for mental engagements and learner involvements than the lower ones.

In the context of Virtual Reality, Steuer (1992) defines interactivity in telepresence as a concept which “refers to the degree to which users of a medium can influence the form or content of the mediated environment” (p. 80), and can be further understood in terms of degrees of speed, range and mapping. Speed relates to how responsive the system is to the user’s actions. Range refers to how many possibilities for manipulation there are in the mediated environment, including intensity (loudness, brightness, etc.), spatial organization (where objects appear, etc.), and temporal ordering. Mapping in this context refers to how closely actions taken on the mediated environment are mapped to corresponding “natural” actions in the human physical environment, thus contributing to a sense of telepresence.

New media are usually intended to include electronic games. In such a context, interactivity is considered a very important design dimension. Friedl (2002) distinguishes three levels of interactivity in games: player to player, player to game, and player to computer. By understanding these different types of interactivity, game designers can identify the factors that affect each, and use appropriate techniques and methods in their own game designs. These factors also provide a basis for continuous
evaluation during the development process, and can be used to classify and analyse online game designs.

The above variations on the theme of interactivity lead to two observations. First, the concept of interactivity in HCI, while somehow commonly understood, is still subject to research in order to achieve consensus on its constituting dimensions across perspectives and application domains. Second, many of the aspects of interactivity which have been taken into account in existing work in the HCI field are closely interrelated with aspects of human communication as briefly outlined in section 2.1.

The next section will present an analysis of popular design approaches, as they have evolved in the HCI field over the years, attempting to highlight how emphasis on different design concerns impacts the interactivity of the design outcomes.

3. **Theoretical frameworks**

This section looks into the theoretical frameworks that influenced interaction design, and consequently the development process, the degree of human-centredness of the proposed interaction paradigms, and eventually the evolution of the way users use and perceive interactive technology. Therefore, this section looks at the most influential works in the field, dealing with how consciously the research community has observed the interaction happening between the computer and the human.

The underlying assumption is that the more human-centred the design process, the more interactivity assumes a fundamental role, and its various dimensions are catered for in the resulting designs.

### 3.1 Human Factors and Ergonomics

In 1911 Frederick Taylor published his work “The Principles of Scientific Management” (Taylor, 1911). This was arguably the first publication aimed at improving the work practice, using the new technologies of the time in the industry. The prime motivation was improving efficiency, and although this was more of a management oriented work than actual human factors, it served as the starting point of
looking scientifically at the work process and the people who executed the necessary tasks.

Efficiency and cutting down time and costs was the primary motivation, but during World War II there was a shift of priorities. It was important to improve the safety and efficiency of aircraft cockpits and weapons systems in order to cut down on human loss. In the work done to improve aircraft equipment and controls, the focus was on the human side, and instead of trying to suit people to the work task and equipment, the opposite was the case; the US army employed psychologists who studied the behaviour of pilots and made measurements to determine human capabilities and limitations. An obvious example that was derived from this research domain was the work of Fitts, made famous for his Law, which was part of a study aimed at finding optimal designs and solutions for cockpit controls that would be easier to use by pilots. From the point of view of interactivity, early work in Human Factors and ergonomics can be seen as a first attempt to match the system’s characteristics to the human user and design systems which are better suited to human abilities.

After World War II, the aviation psychologists would form the Human Factors Society (HFS, later to be named Human Factors and Ergonomics Society) in the US in 1957. This new discipline thrived in the following years (Grudin, 2008).

3.2 Model Human Processor / GOMS

The most influential work that emerged from the marriage between Cognitive Psychology and Human Factors, as well as Computer Science, was the Model Human Processor (MHP), first presented in detail in the seminal book “The Psychology of Human-Computer Interaction”, published in 1983 by Card, Moran and Newell (1983). It is interesting to note the background of the authors. Card was working in Human Factors in Xerox, with a background in Psychology, and Newell was already a well respected researcher in Artificial Intelligence, who took an interest in studying human behaviour (McCarthy 1988).
The model is based on looking at the interaction between the human and the computer fundamentally as an information processing task. That is, treating the two parties of the interaction, the human and the computer, as two information processing systems, each with its own properties, performance capabilities and limitations. The human user is the party that has specific goals and attempts to accomplish them by feeding commands to the computer system. Output from the computer is processed and reviewed, and the cycle continues until the user’s goal is accomplished. In the model human processor, the human information processing system is treated in terms analogous to those of the computer system; the human cognitive architecture consists of 3 processors (Perceptual, Cognitive and Motor), the memory (working and long-term), 19 parameters and 10 principles of operation.

The presentation of visual information on the computer display is perceived by the perceptual processor (basically the eyes and ears). The cognitive processor processes information chunks from memory that was put there by the perceptual processor. The motor processor acts after those chunks have been processed and evaluated and a decision towards accomplishing a goal has been taken. An action may be performed by the motor processor and so on. All these events can be isolated and ultimately analysed to find the optimal and most efficient solutions.

The authors did not imply that human users do actually operate fundamentally as computers do, in fact they explicitly stressed that this is not the case. However, the model provided tangible and quantifiable measures of performance for some of the tasks and consequently allowed measuring (in some cases at least) how different presentations and parameters of the interaction affect the efficiency and usability of the communication between the two systems and how the user may benefit in accomplishing his goals.

In the same book, the GOMS (Goals, Operators, Methods, and Selection rules) analysis framework was also presented. A GOMS analysis of a task describes the hierarchical procedural knowledge a person must have to successfully complete that task. Based on that, and the sequence of operators that must be executed, it is possible to make quantitative predictions about the execution time for a particular task. Other analyses, such as predictions of error, functionality coverage, and learning time, are
also sometimes possible. Since the original formulation presented in Card et al. (1983), a number of different forms of GOMS analysis have been developed, each with slightly different strengths and weaknesses (John & Kieras, 1996; Byrne, 2008). The above approaches have led to further analysis of humans as information processing systems. The fundamental assumption and identification of the human user as a system that can be divided into further subsystems which can be analyzed has produced a significant volume of work (mainly from the Human Factors side of the community), which in large part has been applied successfully in HCI. For example, under the human-information processing system approach, extensive work has been conducted on issues of human attention, distraction, memory performance, problem solving, response time, etc. While the scientific validity behind this work is irrefutable, there are however differing opinions on the success of the application of such models among HCI specialists, specifically regarding the overall interaction experience from the perspective of the user and the general context of use. A complementary approach to the human-information processing model that partially addresses this issue is the ecological approach (Wickens & Carswell, 2006), which takes into account the environment and views the information flow not in distinct stages, but as an integrated flow.

Three years after the publication of “The Psychology of Human-Computer Interaction”, another approach came into focus from the cognitive science discipline, which tried to address the lack of accounting for the personal experience of the user in the above models, named user-centred design, which is addressed in the next section.

### 3.3 User centred design

User Centred Design is among the most influential design philosophies in the HCI community, and had a major impact in the design and development of information systems. Traditionally, the software development process was seen as a series of distinct stages of activity that reminisce a waterfall, in which each activity naturally leads into the next. This is generally known as the waterfall model (Royce, 1970).

Inherited from the traditional engineering industry, the waterfall model divided the process into neat and manageable sections, ideal for setting and monitoring deadlines,
as well as producing a rich amount of tangible deliverables. However, its monolithic nature is not really suited for the development of software, especially when usability issues are taken seriously. User Centred Design became the alternative that emerged from the HCI community, shifting the focus from a technology-driven approach to the user being the center of each development phase. In addition, it responded to the unrealistic distinction between each stage of development by calling for the blending of each development phase and the need for iteration in the various stages of the process life-cycle, with each iteration loop ending with an evaluation of the outcome based on user feedback.

As a result of this approach, every stage in the life cycle is characterized by the strong involvement of users, in the beginning, as the main source of requirements specification, later as the main providers of feedback, to the final evaluation of the product with user testing. Each stage of the development cycle has been a subject of research and the past years have seen the introduction of many different techniques to enhance the quality of the outcome. For example, in the requirements specification stage, analysts employ the use of field observation, focus groups, personas, diary keeping and more. In the design stage, rapid prototyping and evaluation techniques with user testing or expert evaluation techniques are adopted.

In 1986 Draper and Norman were the editors of a collection of papers under the title “User Centred System Design” (UCSD), which they described as “the design of computers, but from the user’s point of view”. In his paper in the book, entitled “Cognitive Engineering”, Norman presented his model of human-computer interaction, based on cognitive science. This model decomposes human action into seven distinct stages, starting from establishing a goal to evaluating the outcome from the computer in relation to this goal. The precise number of stages in the model can vary. The author nevertheless claimed that any theory of action involved a continuum of stages in the action/execution side and similarly in the perception/evaluation side of the full interaction cycle.

In the same paper, Norman also mentions mental models, a concept that he also discussed, among others, in the book “Mental Models” (Gentner & Stevens, 1983): “people form internal, mental models of themselves and of the things and people with
which they are interacting. These models provide predictive and explanatory power for understanding the interaction” (p. 7). Mental models were another cognitive psychology construct that was primarily used to explain how people perceived the world around them and how these affect cognition and reasoning.

3.4 UX, POET and Emotional Design

User Experience (UX) is a term made popular by Norman, Miller and Henderson, when they were working for Apple in the 1990s. In 1995, they published a paper which dealt with the cross-organizational process that Apple used in interface research and development. The overall process was called User Experience. The defining feature of the UX process was to view the user’s interaction not just in terms of hands-on experience with the company’s product, but more broadly, encompassing all interaction with the company itself, including marketing, retail, support and services. In practice, this meant the bridging of various departments within the company, keeping them inside the loop of development and emphasizing intercommunication and collaboration.

In the beginning, the term did not have a well-defined meaning, and was subject to diverse interpretations. For example, as it coincided with the explosion of the web and the dot.com bubble (see section 4.5), companies used the term as “user-centered design for the web” (Morville, 2010). In time and with adequate clarifications given by its influential originators, User Experience has been established as a well-understood concept. ISO 9241-210 gives this definition: "...a person's perceptions and responses that result from the use or anticipated use of a product, system or service". The definition’s notes explain that user experience includes the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours and accomplishments that occur before, during and after use. Three factors are listed that influence UX: system, user and the context of use.

The Nielsen-Norman Group’s definition is given from a company’s perspective:
“"User experience" encompasses all aspects of the end-user's interaction with the company, its services, and its products. The first requirement for an exemplary user experience is to meet the exact needs of the customer, without fuss or bother. Next
comes simplicity and elegance that produce products that are a joy to own, a joy to use. True user experience goes far beyond giving customers what they say they want, or providing checklist features. In order to achieve high-quality user experience in a company's offerings there must be a seamless merging of the services of multiple disciplines, including engineering, marketing, graphical and industrial design, and interface design.”

UX views HCI in its broader context. When an organization creates technology, it embodies in a product its idea for a solution to an end-user problem, with the goal that this will ultimately help the organization itself. HCI is how the end-user interacts with the product, but this symbiotic relationship between the end-users and the organization lies at the core of how that interaction is structured. Understanding the user experience, therefore, is the process of understanding the end-user needs and the organization needs with the goal of maximizing the benefit to both.

In his “Psychology for Everyday Things” (POET), a book highly influential among the design community, expanding on the ideas presented in “UCSD”, Don Norman emphasized usability and making products easier to use. Design aesthetics were not really considered, in fact there were remarks about designers winning awards for products that lacked usability. However, aesthetics and beauty are always major factors while considering a product. The book is highly influential as noted, but it did receive criticism regarding this. Norman included the gist of this criticism in his 2002 essay called “Emotion and Design: Attractive things work better”:

“If we were to follow Norman’s prescription, our designs would all be usable, but they would also be ugly”.

Seeing that this critique was in fact valid, Norman looked into how emotion and affect influence the user experience, acceptance, and ultimately preferences. The 2002 essay became the preface for his 2004 book “Emotional Design”, which again proved to be influential in the design community. In the book, the author presented three aspects of design that deal with human response towards it, based on psychology’s ABC model of attitudes. The ABC model stands for Affect, Behavior and Cognition. Translated to

1 http://www.nngroup.com/about/userexperience.html
design aspects these became Visceral, Behavioral and Reflective. The Behavioral aspect can be thought of as “traditional” HCI territory; effectiveness of use, how well the design fulfills its purpose. The other two aspects were the so-to-speak missing pieces for a more holistic approach to design, and namely emotion and rationalization. The Visceral aspect deals with the design’s appearance and beauty and how these affect users. It is part of human nature, a system to make rapid judgements of what is good or bad, safe or dangerous and of course if something is beautiful and desirable. The Reflective aspect is the rationalization and intellectualization of a product. A product can be totally unusable but still be desirable, to the point that users will forgive its shortcomings in usability. A telling example was the teapot for masochist, famous from the cover of POET. Not particularly beautiful (visceral), certainly not useful but scores highly on the reflective aspect; it can become an object of discussion, it tells a story, it is unique and therefore desirable. What was more interesting, however, was that all these three aspects are influencing each other and a designer may actually take advantage of this fact.

The point of these aspects regarding design is that it was not enough to focus on usability alone. A user will begin evaluating a product (including interfaces) and form an opinion about it from the moment he/she looks at it. The response of the user on the visceral and reflective level will influence the experience of usage. The design must therefore be appealing aesthetically, which will make the user invest more time in interaction with it to learn how to use it. The same applies for the reflective aspect.

Don Norman’s work was not the first to underline these aspects of design, although it certainly helped bring them into the spotlight. Research into emotions and affect and their influence on cognition was being done years before. The essay, for example, cites the experiments conducted by Kurosu and Kashimura (1995), and then duplicated by Tractinsky (1997), which proved that indeed aesthetics and affect play a role in the usability of interfaces.

### 3.5 Universal Access and Design for All

The emergence of the web and of the so called “Information Society” in the nineties brought about radical changes in the way people work and interact with each other
and with information. In this context, the “typical” computer user can no longer be identified. In the past, the “typical” computer user was often considered as a professional, capable and willing to use technology in the work environment in order to increase productivity and performance. In the new environment, interactive artifacts are used by diverse user groups, including people with different cultural, educational, training and employment background, novice and experienced computer users, the very young and the elderly, and people with different types of disabilities. Accordingly, while in the past computer-mediated human activities were mainly oriented towards the business application domain, in the context of the Information Society existing applications undergo fundamental changes, and new ones appear. The latter include access to on-line information, e-communication, digital libraries, e-business, on-line health services, e-learning, on-line communities, on-line public and administrative services, e-democracy, tele-work and tele-presence, on-line entertainment, etc.

Finally, technological proliferation increases the range of systems or devices facilitating access to information resources. These devices include personal computers, but also standard telephones, cellular telephones with built-in displays, television sets, information kiosks, and various types of information appliances. Depending on the context of use, users may employ any of the above to review or browse, manipulate and configure information artifacts, at any time.

The above radical changes brought about the need to revise HCI frameworks and approaches to cater for a much larger and diversified user base and context of use, leading to the concepts of Universal Access and Design for All (see also Chapter ++ of this Handbook).

In Stephanidis et al. (1998) Universal Access is defined as follows: “Universal access in the Information Society signifies the right of all citizens to obtain equitable access to, and maintain effective interaction with, a community-wide pool of information resources and artifacts” (p. 6). Accessibility has been a term traditionally associated with elderly individuals, individuals with disabilities and more in general individuals with functional limitations (Stephanidis et al., 1999). However, because of the current influx of new technologies into the market, the population of users who
may have particular interaction requirements is growing. As a result, accessibility has taken on a more comprehensive connotation. This connotation implies that all individuals with varying levels of abilities, skills, requirements, and preferences be able to access information technologies (Stephanidis et al., 1999). Universal access also implies more than just adding features to existing technologies. Rather, the concept of universal access emphasizes that accessibility be incorporated directly into the design (Stephanidis et al., 1998).

The term Design for All denotes an effort to unfold and reveal the challenges of accessibility and usability, as well as to provide insights and instrument appropriate solutions in the Human-Computer Interaction (HCI) field (Stephanidis et al., 1998). The fundamental vision is to offer an approach for developing computational environments that cater for the broadest possible range of human abilities, skills, requirements and preferences.

Design for All in the Information Society is the conscious and systematic effort to proactively apply principles and methods, and employ appropriate tools, in order to develop interactive products and services which are accessible and usable by all citizens thus avoiding the need for posteriori adaptations, or specialized design. Design for All in HCI recognizes, respects, values and attempts to accommodate the broadest possible range of human abilities, requirements and preferences, eliminates the need for ‘special features’ and fosters individualization and end-user acceptability.

Design for All fosters a proactive strategy, postulating that accessibility and quality of interaction need to be embedded into a product at design time. This entails a purposeful effort to build access features into a product, as early as possible (e.g., from its conception, to design and release). In the context of HCI, a proactive paradigm is advocated for the development of systems accommodating the broadest possible end-user population. In other words, design approaches are required that seek to minimize the need for a posteriori adaptations and deliver products that can be adapted for use by the widest possible end-user population (adaptable user interfaces).

This implies the provision of alternative interface manifestations depending on the abilities, requirements and preferences of the target user groups. The main objective
in such a context is to ensure that each end-user is provided with the most appropriate interactive experience at run-time. Producing and enumerating distinct interface designs through the conduct of multiple design processes would be an impractical solution, since the overall cost for managing in parallel such a large number of independent design processes, and for separately implementing each interface version, would be unacceptable (Stephanidis, 2001).

The scope of design for diversity is broad and complex, since it involves issues pertaining to context-oriented design, diverse user requirements, as well as adaptable and adaptive interactive behaviors. This complexity arises from the numerous dimensions that are involved, and the multiplicity of aspects in each dimension. In this context, designers should be prepared to cope with large design spaces to accommodate design constraints posed by diversity in the target user population and the emerging contexts of use in the Information Society. Moreover, user adaptation must be carefully planned, designed and accommodated into the life-cycle of an interactive system, from the early exploratory phases of design, through to evaluation, implementation and deployment. Additionally, design for diversity is anticipated to be an incremental process, in which designers need to invest effort in anticipating new as well as changing requirements, and accommodating them explicitly in design through continuous updates.

In terms of interactivity, Universal Access and Design for All introduce two important dimensions previously overlooked in HCI: the individual diversity of users, as well as the need to adapt interaction behaviour to such diversity. This implies that there is no best interaction style, but different interaction styles may be appropriate in different circumstances, depending on the involved users and the context. For example, universal access fosters the view that both visual and non-visual (e.g., speech-based) rendering of an interface dialogue can be provided (either alternatively or multimodally) to cater for the interaction requirements of sighted and blind users.

4. The evolution of interaction

The history of HCI is punctuated by the debate among approaches based on the conversation metaphor, which tries to emulate human dialogue, and the model-world
metaphor, which emphasizes the user’s direct action mediated through a visual language (Hutchins at al., 1986). The conversational metaphor favours textual language, whereas the model-world paradigm relies more on iconic languages (although text may also be present).

The conversation paradigm lies at the basis of interaction paradigms such as command-based interfaces and speech interfaces, whereas the model-world metaphor informs direct manipulation interaction and all its subsequent evolutions. Whereas the conversation paradigm addresses interactivity by progressively developing methods and tools to better understand the communication context, the model-world paradigm adopts a radically different approach, whereby the context is reconstructed visually to ground communication.

In more recent years, these two paradigms appear to merge in multimodal interfaces, virtual environments, and Ambient Intelligence environments.

4.1 Early stages

HCI started in the forties with the construction of the first computers. At that time, interaction was very cumbersome and limited to trained scientists. ENIAC, arguably the first general purpose electronic computer, was a massive machine that occupied a large room and needed weeks to program via punch cards. Such were the machines for about a decade. While the main focus was to keep the machines working correctly, i.e., on functionality, there was also an effort to make the interaction easier for the operators by formatting printouts and reducing the programming tasks by creating machines that could store programs, usually on tape (Grudin, 2008). In the strictest sense, the interaction was limited to basic operation of the machine. Programming and using the output was a separate work that was done away from the computer itself. In 1955, transistors replaced tubes in computer hardware, but the basic mode of interaction was basically the same. Until the mid-sixties this was the norm; the only people who had hands-on access were operators and data-entry personnel who dealt with switches, knobs and dials. It was not until the invention of the microchip and later the microprocessor that the way people interact with computers would change
radically, making HCI as a discipline relevant and essential to the design of computer applications.

This has been called the first wave of computing (Weiser and Brown 1996), characterized by many people making use of a single computer. Computers were large mainframe machines that people booked time on to run their programs and get back their results, without actually dealing with the machine itself.

The second wave of computing is the current one, the era of the personal computer (PC), when advancement in technology made it possible to have small personal computers for individual discreet use. During this period, there was an explosion in the evolution of interaction, starting from the command based interfaces, to GUIs and direct manipulation, to web-based and mobile interaction.

The third wave has been given many names, such as ubiquitous computing, the invisible computer, ambient intelligence, calm technology, each name pointing to the presence of computers in the environment, in the background, where one person now has access and use of many computers distributed in the surrounding environment.

4.2 Command-based interaction

Command-based interaction was typical of command-line interfaces. These were the defacto interfaces that people used in the seventies and a significant part of the eighties. Command-line interfaces are based on the conversation model, and provide a mean of passing instructions to the computer, by typing them in with the keyboard in the form of commands, which could be whole words or abbreviations. An example is the Windows Command Prompt, where the user can type in DOS type of commands. Their advantages are that they provide direct access to system functionality and can be very efficiently used by expert users. They can also be very flexible with the use of parameters that allow the user to perform complex tasks with one command.

Their first incarnation was in the form of teletypes, where the operator would type in commands and receive one-line output from the computer, such as feedback or status messages, on scrolling paper. These were then replaced by glass teletypes or video
display units, available to very few at the beginning until the CRT (Cathode Ray Tube) monitors that became more widespread.

On the other hand, command-line interface users complained about the slower speed and lack of flexibility in entering multiple commands. But the most serious problem with command-based interaction is the slow learning curve, as the user must learn the various, often arbitrary, commands. Indeed, one task of HCI at the time was the design of command names, in order to make them more easily memorable. So, even if expert users tend to prefer working with command-line interfaces, most people viewed this type of interaction baffling and the sight of a blinking cursor on a screen did not give any clues to the proper use of the system. Command-based interfaces, although they attempt to establish a simple form of dialogue, present very limited interactivity, as there is no conversational context, and exchanges of command-feedback are independent from each other.

Until the eighties, computers were reserved for use in work by trained staff, performing often tedious data entry tasks, and for the computer enthusiasts. It would not be until the commercial success of the Mac in 1985, and more prominently of the Windows 3.0 environment in 1990, that command line interfaces would gradually disappear or start being integrated in Graphical User Interfaces (GUIs). In fact, many applications currently still offer a dual mode of interaction by incorporating a command line interface complementary to the GUI.

4.3 Direct manipulation

Direct Manipulation (Schneiderman, 1983) was the next major step in the way people interacted with computers. Related research started in the sixties and brought out novelties such as the mouse or the GUI (Engelbart, 1963, 1968; Sutherland, 1963), was further developed into prototypes in the seventies, until finally Xerox’s Palo Alto Research Centre completed the Alto and then the Star system. It would take a few more improvements and additions by Apple in the Lisa computer and more importantly in the Macintosh, which became a commercial success to bring the Desktop Metaphor and Direct manipulation to the mainstream. This shift was solidified with the global success of Microsoft’s Windows, starting in 1990 (Windows
3.0) and exploding by the time Windows 95 were released. This new means of interaction offered easiness of use, more intuitive interfaces and a richer experience. It marked the beginning of the second wave of computing, where Personal Computers were available to the wider audience for discreet use and with the addition of the more friendly interaction that GUIs and direct manipulation interaction offered, suddenly computer use was becoming the norm instead of a sophisticated work tool that needed extensive training to operate.

In Direct Manipulation Interaction users perform actions directly on visible objects with the use of a pointing device, most usually a mouse. Objects include window controls, menus, icons, buttons and other elements in the so-called WIMP interfaces (such as those mentioned above), but there are other types of direct manipulation interfaces, such as 3D interfaces (in VR environments) or haptic interfaces.

The advantages were obvious compared to command-line interfaces. The new interfaces were easier to learn and remember; the user received immediate visual feedback and more accurate representations of what he/she was working on, i.e., WYSIWYG (What You See Is What You Get), and it was easier to reverse actions (undo), which meant that the new interfaces were also less error-prone. Finally, these visual-rich interfaces exploited more fully the human use of visual-spatial cues.

On the other hand, the new paradigm was a much more challenging job for the designer; instead of formatting screens of text, the designer should consider color choices, layout choices, GUI control choices and a lot more. Even many years after the establishment of the WIMP style of interfaces and extensive HCI research on the subject, it is not an easy job.

Direct Manipulation interfaces can be distinguished in two main categories, namely WIMP (Windows, Icons, Mouse, Pointer) and Post-WIMP.

WIMP are the classic interfaces that have been the standard GUI for most computers in the past 25 years. The name comes from the elements that characterize the interface, i.e., the windows and icons on the screen manipulated with the mouse pointer. As described above, this style of interaction originated in the late seventies,
when the Graphical User Interface was combined with the mouse in the Alto machine and then the Star. The Macintosh would make the match a commercial success, introducing the now familiar Desktop Metaphor to the public, where the screen is metaphorically viewed as the top of a desk and various objects that lie on it may be manipulated by the mouse pointer.

While WIMP style interfaces still dominate as the interaction style, there are several other examples of direct manipulation interfaces that either expand on the classic WIMP interfaces or are entirely different altogether. An example of the former would be the recent multi-touch interfaces developed by Apple\(^2\), where the use of a touchpad (on the macbook) or the screen itself (like in the iPad) allows the user to manipulate objects via different gestures and movements.

An example of the latter is the Google Earth’s\(^3\) Zooming User Interface (ZUI), where the interaction is focused in the zooming in and out of a 3D photorealistic representation of the Earth and clicking points of interest.

These two examples also serve to underline another important distinction between direct manipulation interfaces, which concerns the hardware available. It is more common to visualize a typical personal computer with a keyboard and a mouse for input and a monitor for the output, but the example of Apple’s latest hardware points to new input techniques because of the novel hardware. There are more similar examples, for example Surface Computing hardware, which are typically large surfaces that act as touch screens, where the user directly manipulates objects by touching them. Another example would be Virtual Reality environments, where the hardware could include a VR helmet and a dataglove or data wand to manipulate virtual objects (see section 4.8).

Although they radically differ from the conversational paradigm of interaction, direct manipulation interfaces introduce several elements of interactivity. First, they establish a visual context, which plays the role of grounding the human-computer communication process. Second, they provide more structured dialogues and more

\(^2\) http://www.apple.com/magictrackpad/
\(^3\) http://earth.google.com
articulate feedback with respect to command-based interfaces. An important aspect of direct manipulation is its reliance on metaphors. In this respect, visual languages used in direct manipulation interfaces fundamentally differ from natural language. However, it is exactly this metaphoric value which allows the direct manipulation paradigm to ground communication and enrich dialogue, by offering a real-world context in which users’ actions can be rooted.

4.4 Conversational Interfaces

During the late seventies and eighties, when significant advances in the fields of Artificial Intelligence, Natural Language Processing (NLP) appeared to be maturing, a trend towards the explicit provision of human-like communication in user interfaces emerged. In the context of HCI, NLP applications range from various speech recognition systems, to natural language interfaces to database, expert and operating systems (Manaris, 1998).

As the use of computers expanded throughout society, affecting various aspects of human life, the realisation that the number and heterogeneity of computer users was dramatically increasing, and that many of these users were not computer experts, the provision of interaction means exploiting natural language was conceived as a potential path towards increasing the user friendliness of interactive computer systems, and their eventual acceptability to users.

NLP offers mechanisms for incorporating natural language knowledge and modalities into user interfaces. As NLP tools started becoming more powerful in terms of functionality and communicative capabilities, their contribution to HCI also become more significant.

The history of NLP can be very briefly summarised into three main phases. The first phase started in the mid 1940s and lasted until the early 1960s. It is characterized by an emphasis on algorithms relying mostly on dictionary-lookup techniques, used in conjunction with empirical and stochastic methods. During this phase, some NLP application areas began to emerge. An example is speech recognition, employing speaker-dependent, template-based architectures.
The second phase in NLP spanned approximately from early 1960s until late 1980s. It was characterized by (a) a strong emphasis on language theory, including the lexicon, syntax, semantics and pragmatics; (b) the construction of “toy” systems that demonstrated particular principles; and (c) the development of an NLP industry which commercialized many of the achieved results. In terms of applications, this phase is mainly characterized by question-answering systems and database interfaces (1960s), as well as interfaces to other interactive systems. During this phase, it became apparent that symbolic approaches to NLP problems were not adequate when attempting to widen linguistic coverage, or apply the developed systems to a different domain. This realization motivated the development of non-symbolic approaches, mainly based on statistics, connectionism, or on the analysis of language corpora.

In the late 1980s, NLP entered an empirical and more “user-centered” phase. Major advances and tangible results from the last fifty years of NLP research were reinvestigated and applied to a wider spectrum of “real-life” tasks, including, for example, spelling checkers, grammar checkers, and limited-domain, speaker-independent, continuous-speech recognizers for various computer and telephony applications.

During this phase, human-computer interaction entered the mainstream of computer science. This was a result of the major advances in graphical user interfaces during the 1980s and early 1990s (see section 4.3), the proliferation of computers, and the emergence of the World-Wide-Web (see section 4.5). Accordingly, the evolution of NLP reflected the continued growth of research and development efforts directed towards performance support, user-centered design, and usability testing. Emphasis was mainly placed on systems integrating speech recognition and traditional natural language processing models, as well as hybrid systems – systems combining results from symbolic, stochastic, and connectionist NLP approaches.

A conversational agent is a human-computer dialogue system that interacts with the user turn by turn using natural language. Efforts were initiated by Alan Turing (1950) in his famous paper “Computing Machinery and Intelligence”. He suggested that within 50 years a computer would pass a comparison test if it is a human or a
machine. Historically, the first conversational agent was the ELIZA system (Weizenbaum, 1966). ELIZA featured the dialogue between a human user and a computer program representing a psychotherapist. ELIZA is based on simple stimulus-response architecture (i.e., patterns and related responses).

LUNAR (Woods et al., 1972), developed in late the sixties, was the first Natural Language Interface to a DataBase (NLIDB). The LUNAR DataBase contained chemical analyses of moon rocks and had a significant influence on subsequent computational approaches to natural language. Subsequently, several NLIDB appeared exploiting different approaches towards handling Natural Language. By the mid-1980s, Natural Language Interfaces to DataBases become a very popular research area, and numerous research prototypes were implemented.

These initial attempts identified a series of practical difficulties in the development of Natural Language Interface to DataBases, including extensibility of lexica and grammars and portability of systems across different application domains. Other criticisms came from the HCI community, since natural language was considered too ambiguous to provide effective communication in user interfaces. On the other hand, when restricted to some sub-language to limit ambiguity, natural language looses its distinctive feature of allowing free expression and becomes more similar to a command or formal language which needs to be learnt to be used (Hill, 1983).

Additionally, natural language interfaces were criticised for leading users to anthropomorphize the computer, or at least to attribute more intelligence than is warranted to it. This leads to unrealistic users’ expectations regarding the capabilities of the system, and in turn such expectations lead to disappointment when the system fails to perform accordingly (Shneiderman, 1998). Various experiments were conducted to find out how users can adapt to system’s restrictions in vocabulary and syntax, and the results appeared to confirm that humans are keener to learn a command language than to restrict their use of natural language to conform to the system’s limited abilities (Slator et al., 1986; Ogden & Bernick, 1997).

In subsequent years, natural language interfaces have been applied to operating systems (Manaris, 1994) and information retrieval (Jacob & Rau, 1988). The focus in
more recent approaches has been more on using speech, as well as on achieving more natural forms of dialogue communication. Systems that use speech interfaces range from call routing to navigation systems, to VoiceXML-type applications which enable speech interfaces on the web (Jokinen, 2009). The common technology is based on recognizing keywords in the user utterance and then linking these to appropriate user goals and further to system actions. Speech-based conversational interfaces, besides recognizing speech-input, also provide speech-output (Zue and Glass, 2000).

Besides the capability to understand and generate linguistic expressions, some systems include cooperation and planning of complex actions on the basis of observations of the communicative context, i.e., communicative competence (Jokinen, 2009). A well-known example of plan-based system is TRAINS (Allen et al. 1995), a train route planner, where a human manager and the system must cooperate to develop and execute plans.

The notion of natural interaction in this context refers to the spoken dialogue system's ability to support functionality that the user finds intuitive and easy. The challenge that speech and language technology faces in this context is not only in producing tools and systems that enable interaction in the first place, but to design and build systems that allow interaction in a natural way, i.e., to provide models and concepts that enable experimentation with complex natural language interactive systems, and to test hypothesis for flexible human-computer interaction.

In addition to improved interaction strategies, natural language interfaces are also required to be extended in their knowledge management and reasoning capabilities, so as to support inferences concerning the user's intentions and beliefs behind the observed utterances. The goal of building natural interactive systems thus comes close to studying intelligent interaction in general. Towards this end, research efforts have attempted to achieve context understanding (Jokinen, 2009), as well as to exploit to the notion of grounding (see section 2.1) which is inherent in human communication (Traum, 1999).

Further evolution of research towards natural dialogue-based communication has led to developing on the one hand the concept of multimodality (see section 4.7), and on
the other hand to embody conversational systems in anthropomorphic representations (Cassell, 2001). Such anthropomorphism, implemented in the form of animated avatars, is targeted to make explicit the system’s intelligent behaviour, and allows representing the system’s knowledge to humans in multiple ways on multiple modalities (e.g., speech and hand gestures).

### 4.5 Web-based interaction

The Internet materialized in late 1969, when ARPANET was deployed, connecting four academic institutions in the US. Soon, with the implementation of several network protocols, islands of networked computers appeared, leading to the introduction of services that had been developed in previous years, such as hypertext, e-mail and eventually the World Wide Web (Hafner, 1998). The latter suffered at the beginning from the same problems of command-line interfaces compared to GUIs. It was not until 1994, when Mosaic was released, the first graphical web browser, very similar to the browsers we use today, that Internet use exploded beyond academia and government. Three years earlier the Internet had been made available to unrestricted commercial use, but it would take the graphical web browser to provide the missing piece to exploit the full implications that the Internet presented.

From an interaction perspective, the Web in its first years consisted of web pages that contained primarily text, images and links to other web pages. The web pages themselves were not highly interactive; apart from clicking on links and navigating through web pages, the primary interaction style was the familiar form-filling paradigm (Grudin, 2008), including elements such as entering text, selection from menus, checkboxes and option buttons and submitting the forms via command buttons. Content was primarily static, but a major difference in the context of use was the discretionary nature of interaction. This had already begun with the introduction of home computers in the early 80s, where users could choose to use a computer instead of the alternative “traditional” method (for example, consider the typewriter - word processor options) but soon, as computers became the standard tool in the workplace, their use was not a matter of discretion anymore. The web brought discretionary use back into focus and the vast number of competitive content choices
and the notorious impatience with slow speeds meant that user discretion and preference became a major research theme in HCI.

Perhaps the most striking change brought by the introduction of the Web into the wide audience was precisely its penetration into everyday households. The computer was technology used by people with specific tasks and discreet needs, whether it was a work tool or an entertainment device. With the Internet and the Web, the computer became an interactive communication window to the world, as revolutionary a change as the television, only with much more potential and with much richer interaction.

This journey into the household began with the introduction of affordable microcomputers. The Macintosh and Windows provided the appropriate interface to make the technology more accessible to the novice user, and finally the Internet and the Web to make it indispensable to the citizen of the information age. The new avenue of e-commerce made possible by the web also meant that virtually all business environments had to employ computers, if anything for communication and co-ordination purposes. From a sociological perspective, the web also gave rise to social computing, the creation of virtual communities in the form of fora and newsgroups, and more recently social networking sites such as Facebook or LinkedIn.

If interactivity between computers and humans was a secondary issue before, its research dependent on technology developments, the Web definitely brought it to the forefront of hot topics, a fact evident by the sudden flourish of activity in the HCI discipline.

Naturally, a major proportion of this activity revolved around web design practices and evaluation techniques. The web brought many new interaction design issues to the table. This was a natural consequence for two reasons. Firstly, because the web was a new technology, it suffered from a lack of concrete design guidelines and practices. Secondly, creating a web page was considerably easier compared to programming an application, so content was created by many people outside the computing community. A lot of people started publishing web pages, many of which suffered from serious, amateur design issues, like excessive use of colors and images, unreadable content and inconsistency.
Matters were complicated by the inconsistencies across web browsers, the platform-independent nature of the web, the lack of control from the part of the designer in the way pages render on the user’s end (e.g., browser window size, font settings, laptops vs desktop PCs, etc.) (Nielsen, 1997), the lack of common standards and the introduction of new web technologies that were developed to provide richer interaction. The latter, as all new technologies, suffered from misuse and abuse of featurism over usability.

Some of these problems were addressed by the formation of the W3C, the World Wide Web Consortium, founded in 1994. The task of the consortium was to develop international standards for the World Wide Web, including its primary language, HTML or the Cascading Style Sheets (CSS) language\(^4\), which seeks to separate content from presentation. Another important activity of the W3C was the development of accessibility guidelines, aiming to reduce as much as possible the exclusion of users with disabilities from accessing the web\(^5\).

Less formally but perhaps more influentially, Jakob Nielsen’s work on Web Design through his articles on correct practices and criticizing bad practices (such as his famous top ten web design mistakes articles), along with a number of books on usability for the web, played (and still play) a major role in reducing design mistakes which were so frequent in the early web, such as poor layout, bad choice of colors, abuse of distracting and annoying animations, poor content management, innappropriate writing style and typography, excessive size of web pages and consequent slow loading times and more.

HTML was problematic for the creation of good web pages. It was simply inadequate from an interaction perspective, offering very little beyond simple form controls. In response to this, CGI-scripts, mostly written in Perl, “the duct-tape of the internet”, according to Hassan Schroeder Sun’s first webmaster\(^6\), were employed at the beginning to add more programming power to developers. By the end of the century,

\(^4\) [http://www.w3.org/standards/webdesign/htmlcss](http://www.w3.org/standards/webdesign/htmlcss)

\(^5\) [http://www.w3.org/standards/webdesign/accessibility](http://www.w3.org/standards/webdesign/accessibility)

languages had been introduced specifically for the web, such as, for example, PHP, ASP or Flash. This finally allowed developers to create applications for the web that could match, at least to a certain extent, the interactive richness of standard software applications.

HTML was also inappropriate as a layout tool, having been implemented as a hypertext language, aiming to link content, not present it. In response to this problem, web developers relied excessively on elements such as tables or graphics hacks to realize their designs, leading to accessibility issues. The development of CSS by the W3C answered that problem to an extent.

Regardless of specific interaction issues of the web, the main challenge users faced on the Web was locating the specific content they were looking for. The Web is a vast space of information and it was soon apparent that one of the most important issues was searching. The names of the first popular browsers were a clear indication of this problem. Navigator and Explorer, both hinted at an ocean of information that the user could surf through. Search Engines therefore became a critical application and usually the starting point of the web interaction experience.

4.5.1 The .com bubble and Web 2.0

One of the most important events in the history of the web was the so-called .com bubble and burst (Cassidy, 2003). It is not only important as a socio-economic phenomenon, but also as a case study to gain lessons about the nature of the web and what actually worked or not. Not to be underestimated also is the significant exposure the new technology received through mainstream media, which also played a part in the growth of Internet and computer usage.

The period covered between the beginnings of the bubble until the burst is generally considered to be between 1995 and 2001. In that time, Internet use exploded in numbers and a lot of companies tried to exploit this new exciting medium. Everyone was certain that the web was changing society and that a new huge potential market was made available, but there was definite uncertainty on how exactly to exploit this market. Many companies were founded without a specific business model and
consequently went bust without ever making any actual profit, erroneously thinking
that traffic would somehow generate revenue through advertising or that the
elimination of the traditional brick and mortar model would translate to profit. The
hype in the stock market and the soaring of stock values of these so-called .coms,
sustained this illusionary impression until roughly the end of the millennium, but at
the end the bubble burst and only those who had understood the nature of the web
survived, indeed thrived.

One explanation for the success of these companies, such as Google, eBay or
Amazon, was that they matched and took advantage of the specifications of the so-
called Web 2.0. This concept, first articulated in 2001 after the .com burst, sought to
explain the common factor between the aforementioned companies and propose a new
approach of understanding the Web. The underlying principle was that the Web
should be seen as a platform, as opposed to a medium for which standard desktop
applications should be developed for. The example cited by the originators of the
concept is Netscape vs. Google (O’Reilly, 2005). Netscape began by trying to replace
the desktop with the “webtop” (their browser) and planned to populate that webtop
with information updates and applets pushed to the webtop by information providers
who would purchase Netscape servers. In reality, value was transferred up the stack to
services delivered over the web platform. Google, on the other hand, was such a
service, and customers were paying directly or indirectly for its use. There were no
issues of new software releases or OS/hardware-specific editions. Netscape relied on
the classic software paradigm, Google on the concept of the service running between
the two computers (the Google server and the user’s computer), in the web space. The
real challenge was managing the data and turning it into useful information for the
end-user. Similarly, Amazon did not offer a particularly different catalogue of
products than its competitors, but it invested heavily in the management of all sort of
data, so it could make it into useful information for its customers that would lead them
to a purchase.

This focus on the data and its processing into useful information also pointed to
another significant factor, namely users add value. In services such as Google,
Amazon or Wikipedia, it is the users that provide the actual content and this
ultimately means that if an application can provide a useful service with a critical
mass of users, then it can be successful. The underlying principle is in essence a paraphrase of the well-known Open Source mantra “given enough eyeballs, all bugs are shallow” (Raymond, 2000); given enough users, the content is valuable. Typical examples of this principle, apart from those mentioned above are blogs, wikis, media-sharing sites (such as YouTube), social networking sites (such as Facebook or Myspace), etc.

Collective user generated information also proved to be the best solution regarding the web’s primary challenge, that of searching. Google became the defacto search engine by exploiting user sponsorship of web sites by considering links as sponsorships of approval, with great success. The same principle also applies to Amazon, as it exploited its users’ selections as an indication of what is the most probable content they were looking for, in the form of suggestions for related content. A key feature of a successful web service is to provide the easiest route to desirable content, and managing collective user data has proved to be a very efficient way to achieve this.

In terms of interactivity, it can be argued that not much has changed in the way users physically interact with the computer, nor in what type of interaction controls are used. The input and output devices remain the same and only the link is a web-specific interaction element in the interface. However, there is undoubtedly a major transformation in the context(s) of use, as well as the environment space that the user moves through, resulting in an overall different user experience that has a profound effect in the way humans perceive computers. The Web 2.0 paradigm of harnessing collective intelligence and user-generated content is one example of how computer services have changed significantly in the course of a decade. The concurrent developments in communication technologies, specifically wireless and mobile communications, as well as the widespread availability of high-end mobile devices, have provided a set of emerging tools which are starting being used in augmented reality and ambient intelligence applications.

4.6 Mobile interaction

Mobile interaction is a relatively new field of research in the HCI community, but it has become one of the most interesting, since the current generation of mobile devices
has reached a technological maturity that allows much more sophisticated interaction than when they first appeared. Furthermore, the inherent property of mobility and their personal nature, coupled with the advanced processing power and multimedia capabilities, makes them a good candidate for playing a major part in the way people generally interact with computers in the future. Before examining the characteristics of mobile interaction and the design challenges that stem from the device properties, this section looks into the historical context of mobile devices, focusing on mobile telephony and Personal Digital Assistants (PDAs), the combination of which (the smartphone), will be the primary focus of the chapter.

In 1979 Sony released the Walkman. It was an instant success because it managed to take one activity, listening to music, which was confined to the home, and take it anywhere. It’s obvious advantage and appeal was that it was small and easy to carry around and offered the same service, more or less, with another much larger cassette player. Interestingly, before the product was launched, critics thought it would be a commercial failure because it did not offer a recording function. It turned out of course that most people did not need that specific function, but were very happy with being able to listen to music anywhere. The Walkman also marked a trend towards miniaturization and portability.7

Personal Digital Assistants are basically handheld mini computers. The term was first used by Apple in 1992 to describe the Apple Newton,8 the company’s first attempt at creating a mobile device, which also featured a touchscreen. Before the Newton, the line was blurred between small handheld electronic organizers (such as the very minimal Psion Electronic Organizer and the quite sophisticated Sharp’s Wizard series) and portable PCs, which were closer to the size of what we now refer to as notebooks. The latter trace their roots back in 1972, when Alan Kay proposed the design of Dynabook (Kay, 1972), which however was never built into a working prototype. The Dynabook is considered the ancestor of the laptop or the tablet PC, and was a huge influence on the Palo Alto by Xerox, which Kay had joined in 1970.

---

7 http://en.wikipedia.org/wiki/Sony_Walkman
8 http://en.wikipedia.org/wiki/Apple_Newton
Personal Digital Assistants were quite popular in the 90s, but the market was fragmented and the devices never really caught on as more than electronic organizers in the market, offering calendar support, note taking, etc., despite efforts from major players such as Microsoft to support the medium by releasing a PDA specific OS, CE Windows. However, PDAs were used extensively in business and healthcare.

Mobile telephony had a similar evolution. Telephony started as a fixed-location service, then moved into the car (although mostly in the US), until the mobile phone at the mid-90s. Mobile telephony took off in the mid-90s and the first generation cellphones were large and rather cumbersome devices, with a minimal screen. The primary use of mobile phones at the time was limited to calling and answering, as well as text messaging. The standard interface was the keypad plus some buttons, without any standard configuration across the various devices. Technological advancements led to devices with much better processing power and screens, to the point that these devices were matching the capabilities of personal digital assistants. Roughly by the early 00s, mobile phones offered colour screens, better GUIs and smoother interaction than the first generation phones. Usability matured in getting the right design for the handsets, as well as establishing design guidelines for the specific challenges posed by the nature of the devices, in regards to interface styles, text-entry, etc.

Appropriately designing the hardware controls was not a trivial task either. Nokia took four years to reduce the number of buttons (besides the standard numeric keypad) from eight to four, from 1994 to 1998. It took extensive research and user feedback to realize that the mobile phone user at the time primarily did two fundamental tasks: dialling from the phone book and answering the phone. Therefore, the elegant and successful solution they came with, having simplicity in mind, was to use one big prominent button (the Navi-Key) directly below the screen. The button was used to answer and hang the phone, as well as confirming a selection (Jenson, 2002).

New generation mobile phones are also referred to as smartphones, since the next logical step was to combine the functionality of the mobile phone with that of the PDA. The first official smartphone is arguably the Simon, a device designed by IBM
in 1992 and released in 1993 by BellSouth. It featured a stylus operated touchscreen and combined a mobile phone with many PDA capabilities including e-mail. Its initial price was too high to penetrate the market. In 1996 Nokia released the Communicator, a mobile phone literally combined with a PDA device, as the first prototypes were a Nokia phone hinged together with a HP PDA. The 9000 Communicator as it was called, effectively marked the beginning of the smartphones. It was a very cumbersome device, and it did not affect the PDA market much, but eventually smartphones rendered stand-alone PDAs almost obsolete. The term smartphone itself was probably used for the first time in 1997 when Ericsson released the GS88 phone.

Mobile interaction is radically different from PC interaction. The reasons for this can be grouped into two categories, device characteristics and context of use. The most obvious device characteristic is of course its size. This includes the size of the screen, as well as the size of the controls, usually the keypad and navigation buttons of the mobile phone. The fact that there is no standard screen size or hardware controls further complicates matters. Furthermore, the input/output interface consists of the screen and the keypad, i.e., no mouse or keyboard is available. These two characteristics are the main factors influencing interface design in mobile devices.

Because of the lack of a pointing device, although modern mobile devices are capable of rendering quality graphical user interfaces, albeit in a small scale, the majority of the devices do not offer direct manipulation. Instead, the major interaction paradigm is scroll and select, where the interface is presented most commonly in the form of a list-based layout. This layout also takes advantage of the fact that most mobile screens are portrait oriented (height bigger than width). Several variations or features of list-based layouts can be found according to the task at hand. For example, a common implementation is fish-eye lists, where the item selected expands to reveal more information. Such a solution works well for long lists, such as contacts or e-mail messages. Another helpful solution to long lists is circular scrolling, where the list loops after reaching its end. This is helpful because users of mobile phones cannot use the scrollbar as they would on a normal PC with a mouse.

The second challenge in mobile interaction design is data entry, and specifically text entry. Text entry with a keypad is a notoriously tedious process and users avoid it as
much as they can. Solutions to this problem were the use of autocomplete functions, predictive text, either dictionary based or pure predictive algorithms (which seem to perform better overall) (McKenzie et al, 2001) and other novel solutions, such as gesture recognition, shapewriting (both solutions for touchscreen equipped devices), and voice recognition.

It should be noted however that although these paradigms and styles regarding mobile interaction concern approximately 79% of devices (Entner, 2010), the new generation of smartphones featuring touchscreens overcome some of those problems by allowing direct manipulation interaction. They are still affected by the obvious size differences, and many of the guidelines certainly apply, but the difference in the overall user experience is very notable. This is important because the market share of these devices is growing very rapidly, and it is not a wild speculation to assume that this trend will continue in the following years (Entner, 2010), especially since all major players, including Google, Microsoft, Sony Ericsson and Nokia among others, as well as Apple of course, the first that made a huge market impact with the iPhone, are focusing on the development of such devices.

In essence, smartphones appear to be creating a paradigm shift in mobile interaction that can be compared to the paradigm shift of command-line interfaces to the direct manipulation of GUIs. The old paradigm is scroll and select using hardware buttons, and the new one is direct manipulation interfaces through touchscreens.

The context of using a mobile device is inherently different from the use of a static personal computer. Mobile phones are primarily communication devices but also, since the merging of PDA capabilities, cameras and media players, as well as the introduction of mobile web technologies such as 3G, they have become part of a larger ecosystem of networked devices and an important personal appliance for the user. Today the owner of a modern smartphone does a lot more than simply calling or texting another person. The smartphone is also used as a music player, a web browser (which involves many of the issues we discussed in the Web Interaction section), a GPS, a gaming device and more, usually on the go.
Evaluating mobile device usage is particularly challenging because it is not possible to conduct studies in the laboratory as the context of use is entirely different. Mobile users are by definition usually on the move and, more importantly, interaction occurs infrequently and in bursts. Observing the interaction in real circumstances however is equally difficult (if not more), since there are too numerous environments and circumstances to take into account and there is no way to predict when the user will actually use a device. Furthermore, interaction is mostly private, so the act of observation would contaminate the results as there is no way to determine if the user has altered his/her behaviour (Jones and Marsden, 2006).

Regarding the context of use of mobile web interaction, it merits a closer examination, because modern smartphones have rapidly changed its use.

4.6.1 Evolution of mobile web interaction

Access to the Internet through a mobile phone has been available since 1996, when Nokia released the 9000 Communicator model. However, the technical limitations of mobile devices made browsing the Internet almost impossible, as web pages did not render gracefully on the small screen or at all, making them unreadable. As an attempt to address this problem the Wireless Application Protocol (WAP) was developed in 1998 as an open international standard, based on which WAP browsers could provide the basic services of a desktop computer browser but in a simplified form to overcome device limitations. Slow speeds, pricing and the notably poorer experience compared to the familiar desktop browsing were the reasons why the WAP based mobile web did not really catch on, with the notable exception of Japan. In the latter, a rival system to WAP was developed and released in 1999 by NTT DoCoMo, the i-mode, which became a huge commercial success. Soon afterwards, its two major rivals in the mobile market in Japan offered WAP based services, with considerable success as well. Mobile web use in Japan has been far more widespread than the rest of the world until very recently. There are several reasons for this; primarily favourable flat-rate plans, extremely high 3G handset penetration and excellent network quality and signal coverage, and the carriers approach of “open garden” as opposed to Western carriers who try to keep consumers on their portals (Billich, 2010).
3G mobile web speeds are considerably better, but still the experience is not even remotely similar to the desktop counterpart. The main reason is that instead of trying to fit everything in the tiny space of a mobile device screen, designers began to customize and reduce the functionalities offered through the mobile sites to make the experience more appropriate with the context of use of mobile users. This also followed the specifications by the W3C Mobile Web Initiative\(^9\), which addressed the issue of balancing the homogenous nature of the web and the specific circumstances of the mobile context of use.

Regarding the latter, Google identified three primary contexts of use for mobile consumers of its services:

- The Casual Surfer or “Bored now” user. Users who find themselves with spare time (such as waiting in lines, while travelling by train, sitting in cafés, etc). These users resemble the casual web surfers. Since mobile phones cannot match the robust user input of a desktop PC, applications for these users should be tailored.
- The Repeat Visitor or “Repetitive now” user. Users who seek the same information on a regular basis, such as stock market prices, weather reports, sports scores, etc. Catering to their needs would be ensuring that repetitive steps or search queries can be eliminated by “remembering” each user’s preferences, in the same manner that cookies work in desktop browsers.
- The “Urgent, Now!” visitor. Users who seek specific information, fast, such as directions to the airport or the nearest ATM. The key issue in this case being location, mobile services catering to such situations should emphasize location-awareness.

### 4.7 Multimodal Interfaces

Multimodal systems process combined natural input modes, such as speech, pen, touch, hand gestures, eye gaze, and head and body movements, in a coordinated manner with multimedia system output (Oviatt, 1999). Multimodal systems represent a paradigm shift from conventional WIMP interfaces toward providing users with greater expressive power and naturalness. The goals are twofold: to achieve an

\(^9\) http://www.w3.org/Mobile/
interaction closer to natural human-human communication, and to increase the robustness of the interaction by using redundant or complementary information in different modalities (Reeves et al., 2004).

One of the first multimodal systems was Bolt’s Put that there -system (Bolt, 1980), where the users interacted with the world through its projection on the wall and using speech and pointing gestures. Subsequent attempts in the domain of multimodal interaction rely on advances in speech and natural language processing, computer vision, and gesture analysis. Major progress has occurred in both the hardware and software for component technologies like speech, pen, and vision, as well as on progress in the development architectural components and design frameworks (Oviatt & Cohen, 2000). Additionally, applications have been built that range from map-based and virtual reality systems for simulation and training, to field medic systems for mobile use in noisy environments, to Web-based transactions and standard text-editing applications. An overview of architectures and applications for multimodal interfaces is provided in (Oviatt et al., 2000).

Multimodal systems integrate complementary modalities to yield a highly synergistic blend in which the strengths of each mode are capitalized upon and used to overcome weaknesses in the other. Whereas traditional interfaces support sequential and unambiguous input from devices such as keyboard and conventional pointing devices (e.g., mouse, trackpad), multimodal interfaces relax these constraints. For example, they can support asynchronous, ambiguous, and inexact input by applying more sophisticated analysis of input. They can also detect and correct errors utilizing models of the media, user, discourse, and task (Maybury, 1998).

Systems that process multimodal input also aim to give users better tools for controlling embedded visualization and multimedia output capabilities, as opposed to the limited possibilities offered by keyboard and mouse input, in particular when dealing with complex environments (Oviatt, 1999).

In the context of multimodal interfaces, as the center of human-computer interaction shifts toward natural multimodal behavior, human communication patterns are used to control computers in a more transparent interface experience than ever before. Such
interface designs become more conversational in style, rather than limited to command and control, because many of the modes being processed are language-oriented (speech, manual gestures, pen input) or involve communication broadly defined (gaze patterns, body movement) (Oviatt & Cohen, 2000).

Achieving natural patterns of multimodal input is however not as straightforward as it would appear. A dominant issue in this respect concerns the integration and synchronization requirements for combining different modalities into a system. Oviatt (1999), in the seminal paper “Ten Myths of Multimodal Interaction”, on the basis of empirical evidence as well as experience, analyses common pitfalls which may impact negatively multimodal interaction design. Examples of such assumptions are that users will always interact multimodally if they have the possibility to do so, that speech and point is the dominant modality integration pattern, that speech is the primary input mode in multimodal systems, and that multimodal languages do not differ from unimodal languages. Regarding this last point, which presents particular interest from the point of view of interactivity, Oviatt claims that multimodal languages are briefer, syntactically simpler, and less disfluent than natural unimodal speech, as multimodality allows to eliminate linguistic complexity resulting from the need to express verbally elements to which the user can refer deictically (through gestures) using a different modality. A related issue concerns redundancy between modalities.

Whereas it is commonly claimed that the content conveyed through different modalities in multimodal communication contains a high degree of redundancy, it appears that users tend to use different modalities in a complementary rather than redundant way.

To optimize human performance in multimodal systems, principles regarding how to integrate multiple modalities or how to support multiple user inputs (for example, voice and gesture) have been elaborated based on cognitive science literature on intersensory perception and intermodal coordination (Reeves et al., 2004).

Perceptual User Interfaces (PUI), introduced by Turk & Robertson (2000), are multimodal user interfaces which combine active input, such as speech, pen-based
gestures, or other manual input, with passive input mode that requires no explicit user command, such as vision-based tracking that unobtrusively monitors user behaviour and senses a user’s presence, gaze, and/or body position. Through perceptual interfaces, multimodality permeates into intelligent interaction environments where interaction is to a large extent implicit and continuous (see section 4.10).

4.8 **Virtual and Augmented Reality**

Virtual and Augmented Reality (VR and AR respectively) are two areas that share many common aspects. Their difference lies in the degree to which they replace (in the case of VR systems) or enhance/augment (AR) the real world. The similarity is that this is performed with the use of computer hardware and software. Practically, virtual and augmented reality systems have common research backgrounds, but the applications of each field are distinctly different. Both are in the stage where mass use of these technologies is not widespread and research is ongoing. Technological developments and costs are major factors for this, and the fluid nature of computing developments in general will determine the role these technologies will play in everyday life. Below, a brief historical account of both technologies is presented, describing the major features of each, in terms of hardware, software, user experience and applications.

Virtual Reality is a term that encompasses a broad range of research directions and applications. In the popular mind, virtual reality is the classic futuristic technology, being showcased in science fiction works such as Star Trek, where crew members of a spaceship can experience realistic worlds simulated by a sophisticated computer. A technical definition is offered in (Heim, 1998):

“Virtual Reality is an immersive, interactive system based on computable information” (p. 6).

The term itself was first used by Jaron Lanier in 1986 (Behr, 2002), replacing various descriptions such as virtual worlds. The latter was used by Ivan Sutherland, who built what is considered the first head mounted display (HMD), an early VR prototype called the Sword of Damocles attempting to realize what he had called earlier “the ultimate display” (Sutherland, 1965). Other terms included synthetic environments,
tele-existence, artificial reality and immersive computing, the latter emphasizing a key feature of Virtual Reality systems, namely immersion. Immersion refers to the feeling of being present in another reality apart from the real world. According to Heim, this goes beyond physical input and output because it involves psychological components, but it also surpasses purely mental imagination because of the sensory input involved. A successful VR system should at least partially offer this user experience of being someplace else, in a reality created artificially. Total immersion is what most researchers refer to as strong VR.

Virtual Reality differs significantly from the “regular” computer interaction in terms of input and output devices. As the whole paradigm revolves around the immersive experience of being present in an artificial world, the way users interact with this world depends on these devices. The most important device is the medium that provides (primarily) the visual information that attempts to replace the physical world perceived through the eyes. There are two basic ways that are employed for this: Head Mounted Displays and CAVE-type environments.

Head Mounted Displays, as their name suggests, are displays that fit on the user’s head, often as a helmet, which block the view of the physical world and present the computer generated images to replace it. The displays themselves are miniaturized and the technology used is usually CRT or LCD types of screens. Issues that affect the quality of the user experience are the weight of the device, the resolution of the displays and the field-of-view (FOV). The more sophisticated devices include a head-tracking system, which sends data to the computer system about the positioning of the user’s head (and consequently gaze), so the image displayed is refreshed accordingly. This is essential to create an immersive experience, as the user freely moves his head about and views the scenery changing appropriately around him. This feature makes the HMD a significant input device as well. Some HMDs also include headphones to provide the audio. HMDs are also known as goggles and, together with datagloves, form the goggles and glove VR paradigm. The problems with HMDs, despite significant progress since the Sword of Damocles system, are still their cumbersome nature, suboptimal resolution and limited FOV (Stanney et al, 2008).
The data glove is an input device that is worn like a glove that tracks finger and hand movements and gestures, and allows its wearer to manipulate virtual objects, thus bringing direct manipulation interactivity in a virtual environment. Its history can be traced back to 1977 with the Sayre Glove, which was an inexpensive, lightweight glove that was developed to track hand movements. Over the next years, this technology was enriched to include more sensors for tracking finger flexure and introducing tactile feedback to the fingertips, making the dataglove an output device as well. Notable incarnations were the Digital Data Entry glove (1983), which was developed to recognize the Single Hand Manual Alphabet for the American Deaf and the 1987 Nintendo Power Glove, which was a crude dataglove in terms of precision but it was the first that was widely available to the public (Sturman et al, 1994). Datasuits are the whole-body equivalent of the dataglove. The user wears a suit that contains sensors that track the entire body movement. The data is then sent to the computer which appropriately updates the virtual equivalent of the user, sometimes called a cyberbody.

CAVE-type environments are systems that are based on a configuration of large displays. They are rooms where usually projectors are employed to combine images to create a large scene. CAVE stands for Cave Automatic Virtual Environment, which was the name of the first such system, developed by the University of Illinois in 1992, the name also pointing to Plato’s allegorical cave from the Republic. CAVE-type systems offer some significant advantages over HMDs. They allow multi-user immersion and collaboration, without cumbersome wearable devices, user mobility (which is only achievable through expensive omni-directional treadmills in the case of HMD-based VR), high resolution and wider FOV (Cruz-Neira et al, 1992).

While vision and hearing are the two senses that dominate interaction with computer systems, there are however efforts to exploit another human sense, touch. This is an area that interests particularly VR researches, since touch is a channel that provides rich information in real life, particularly in situations where VR is being used, such as healthcare. Doctors often feel with their hands bodily areas to make a diagnosis. In order to realize this sort of tasks in a virtual space, tactile interfaces or haptics have been developed. These devices use pins or small electrical currents to stimulate the
nerves that cause the sensation of touching. Another method used is by using motor mechanisms to provide resistance to the user’s probings (Iwata, 2008).

Perhaps Virtual Reality has failed to fulfill the expectations of enthusiastic fans, as well as more moderate (as they seemed at the time) expectations of widespread VR use for the average user. VR is still very costly, so its main users remain limited to large business organizations or the military. But there are many applications it has found a place in (Heim, 1998).

Training is the most established application area for VR (Stanney et al, 2008). The first VR applications were for simulation and training, especially in aviation and the military, and still today flight or combat simulations are typical examples of VR technology use.

Entertainment has always been a major force behind VR research and development. Arcade Rooms often have a VR system that provides a much more exciting gaming experience than regular gaming systems. Disney is another example. In (Mine, 2003), their work in creating Virtual Reality attractions is described, where guests to Disney World can enjoy 4-5 minutes rides in CAVE-type systems, allowing a whole family to be simultaneously immersed in the game.

Another major consumer of VR technology is the building industry. It is quite common these days to construct 3D computer representations of building projects that are not only used to determine the design, but also for showcasing and presenting the project to prospective customers, who can have a virtual walkthrough or fly-by in their future home. They can then offer feedback and express their preferences before any costly construction begins. One construction company specializing in luxury homes called their VR presentations “our greatest marketing tool”¹⁰. The automotive, aerospace, naval and other industries also have been exploiting VR in product design and manufacturing. Virtual Wind Tunnels are favoured over their physical expensive counterparts. NASA was the first to combine the HMD and Dataglove technology in

¹⁰ http://www.thefreelibrary.com/Builder+Spectrum+Skanska+using+virtual+reality+tours'++to+sell+homes...--a097876792
their VIEW (Virtual Interactive Environment Workstation) system, used among other things to develop a virtual wind tunnel and according to Mark Pesce, it was then that “Virtual Reality was born” (Pesce, 2000), since it was one of the first instances that the users commented they actually felt being inside the virtual environment (tele-presence).

A recent VR application that has been utilized is in the treating of psychological conditions, namely phobias (Stanney et al, 2008) and posttraumatic stress disorder (Rothbaum et al, 1999). While these treatments are fairly new and expensive, it is expected that they will become more common. Another area of healthcare that benefits from VR is the treatment of patients who undergo painful procedures, such as skin scraping in burn victims. The immersive quality of VR actually distracts patients from the pain when they become absorbed in the VR program (Mueller, 2002).

Finally, VR has been used successfully in surgery education, where the surgeon examines and “operates” on 3D holographic images of a patient, obtained through normal MRI scans and computer tomographic images, before performing the procedure on the real patient (Versweyveld, 2001). This technique also takes advantage of force feedback, where the controls relate the pressure needed to be applied by the surgeon. VR has also been used in education and learning in various ways. One common example is in the reconstruction of places from history (Acevedo et al, 2001), where users can walk through sites that are either far away or have been destroyed. The exciting and entertaining aspect of VR definitely contributes to the learning experience.

### 4.9 Augmented Reality (AR)

Augmented Reality began at roughly the same time as Virtual Reality research, sharing a common origin and diverging in the application domain. As mentioned, augmented reality does not seek to replace reality like VR does, instead it enhances it (or augments it) by superimposing computer generated information over the primary visual field. In some cases, the physical view is processed to remove irrelevant data with the goal to enhance not reality itself but the user’s perception of it. The term itself is relevantly recent, dating back to 1990. It was coined by Tom Caudell while he
was working at Boeing. He used the term to describe a system utilizing a HMD to assist workers assembling cables into aircrafts (Mizell, 2001).

HMDs have been described above as the classic devices used in VR applications. In fact, HMDs are used extensively in AR applications, with the difference being that the displays employ semi-transparent mirrors that do not block the physical view but allow computer generated imagery to be projected on them. Alternatively, non-see-through displays can be used but they use real-time video of the outside world with superimposed graphics, effectively achieving the same result. They were first developed for use in fighter jet pilots helmets (and the ‘H’ in the acronym actually stands for ‘Helmet’ instead of ‘Head’), to replace heads up displays (HUDs) (Defense Industry Daily, 2010), which in turn were one of the first applications of augmented reality, long before the term was even thought of. Heads Up Displays were developed to eliminate the need for pilots to look down on their instruments (hence to keep their “heads up”). When the technology became available, the same idea was moved into the pilots’ helmet and integrated with the weapons systems. Like in VR HMDs, the helmet contains sensors to track head position, with the effect of the pilots’ gaze being used to lock on enemy targets.

The field of augmented reality really boomed in the 90s and the past decade. Recent advances in technology, particularly in the capabilities and features of mobile devices and the fusion of different technologies in robotics and VR have yielded a host of different applications with very promising prospects. Those are briefly summarized further down. Before looking into them however, it must be noted that Augmented Reality is now more clearly defined, not as an offshoot of Virtual Reality but as a particular field that blends computing and reality in real time with the ultimate goal of enhancing user perception and performance through this blending.

Augmented Reality has been employed successfully in healthcare in various tasks, including surgery preparation, brain surgery and laparoscopic surgery. In this context, images obtained through CT or MRI scans are projected directly on the patient’s body, providing to surgeons the three-dimensional vision and feeling (Kania, 2001).
Another popular example of augmented reality is Google Earth\textsuperscript{11}. The user views satellite pictures of the earth and by setting preferences can superimpose additional information, such as borders between countries or states, roads, names of places, extra pictures submitted by users, calculate distances between two geographical points etc. Similarly, taking advantage of the recent advances in mobile devices, particularly the use of cameras and GPS, there are applications for mobile devices that can display useful information over the video view of the world provided by the phone’s camera. Examples include navigational directions for GPS applications or the display of points of interest in a particular area over the view the user points the camera at. Such applications are expected to become more common in the future, as there is growing demand from the smartphone user base to exploit the capabilities of their devices (Chen, 2009).

\textbf{4.10 Interaction in intelligent environments}

Ambient Intelligence (AmI) is an emerging technological paradigm which envisages an environment populated by several interoperating computing-embedded devices of different size and capabilities, which are interweaved into “the fabric of everyday life” and are indistinguishable from it (see Chapter ++ of this Handbook).

From a technological point of view, AmI targets to distribute, embed, coordinate and interactively deliver computing intelligence within the surrounding environment. AmI technologies integrate sensing capabilities, processing power, reasoning mechanisms, networking facilities, applications and services, digital content, and actuating capabilities distributed in the surrounding environment. AmI will have profound consequences on the type, content and functionality of the emerging products and services, as well as on the way people will interact with them, bringing about multiple new requirements for the development of interactive technologies (e.g., Butz, 2010).

While a wide variety of different technologies is involved, the goal of AmI is to either hide the presence of technology from users, or to smoothly integrate it within the surrounding context as enhanced environment artifacts. This way, the computing-oriented connotation of technology essentially fades-out or disappears in the

\textsuperscript{11}http://www.google.com/earth/index.html
environment, providing seamless and unobtrusive interaction paradigms. Therefore, people and their social situation, ranging from individuals to groups, and their corresponding environments (office buildings, homes, public spaces, etc), are at the centre of the design considerations.

The pervasiveness of interaction in AmI environments requires the elaboration of new interaction concepts that extend beyond the current user interface concepts like the desktop metaphor and menu driven interfaces (Aarts & De Ruyter, 2009).

AmI environments will integrate a wide variety of interactive devices, in many cases equipped with built-in facilities for multimodal interaction and alternative input/output (e.g., voice recognition and synthesis, pen-based pointing devices, vibration alerting, touch screens, etc), or with accessories that facilitate alternative ways of use (e.g., hands-free kits), thus addressing a wider range of user and context requirements than the traditional desktop computer. Devices will also vary in the type and specialization of the functionality they offer, ranging from “personal gadgets” (e.g., wrist-watches, bracelets, personal mobile displays and notification systems, health monitors embedded in clothing), to “general-purpose appliances” (e.g., wall-mounted displays). Regarding personal devices, an important role will be played also by smart mobile phones, which already offer sensing and location-awareness facilities, as well as augmented reality applications which goes in the direction of distributed and natural interactivity.

AmI will bring about new interaction techniques, as well as novel uses and multimodal combinations of existing advanced techniques, such as, for example, gaze-based interaction (Gepner et al., 2007), gestures (Ferscha et al., 2007), and natural language (Zhou et al., 2007). Progress in computer vision approaches largely contribute to the provision of natural interaction in AmI environments, making available, amongst other things, techniques for facial expression, gaze and gesture recognition, face and body tracking, activity recognition, etc.

Additionally, interaction will be embedded in everyday objects and smart artifacts. This concept refers to interfaces that use physical artifacts as objects for representation and interaction, seamlessly integrating the physical and digital worlds.
Such objects serve as specialized input devices that support physical manipulation, and their shape, color, orientation, and size may play a role in the interaction.

The interaction resulting from tangible user interfaces is not mediated and it supports direct engagement of the user with the environment. Consequently it is considered more intuitive and natural than the current keyboard and mouse-based interaction paradigm (Aarts & De Ruyter, 2009).

Interaction in AmI environments inherently relies on multimodal input, implying that it combines various user input modes, such as speech, pen, touch, manual gestures, gaze and head and body movements, as well as more than one output modes, primarily in the form of visual and auditory feedback. In this context, adaptive multimodality is prominent to support natural input in a dynamically changing context-of-use, adaptively offering to users the most appropriate and effective input forms at the current interaction context. Multimodal input is acknowledged for increasing interaction accuracy by reducing uncertainty of information through redundancy (Lopez-Cozar & Callejas, 2010).

However, AmI is also anticipated to introduce increased complexity for its users. As technology ‘disappears’ to humans both physically and mentally, devices will be no longer perceived as computers, but rather as augmented elements of the physical environment (Streitz, 2007). The nature of interaction in AmI environments will change radically, evolving from human-computer interaction to human-environment interaction (Streitz, 2007) and human-computer confluence (Ferscha et al., 2007). These concepts emphasize the fusion of the technology and the environment, as well as the inextricable role of interaction in all aspects of everyday life.

AmI environments will be very interaction intensive, and humans will be constantly “surrounded” by a very large number of devices of different shapes and sizes. Therefore, interaction shifts from an explicit paradigm, in which the users’ attention is on computing, towards an implicit paradigm, in which interfaces themselves drive human attention when required (Schmidt, 2005). Interaction in the emerging environment will no longer be based on a series of discrete steps, but on a continuous input/output exchange of information (Faconti & Massink, 2001). Continuous
interaction differs from discrete interaction since it takes place over a relatively longer period of time, in which the exchange of information between the user and the system occurs at a relatively high rate in real-time. A first implication is that the system must be capable of dealing in real time with the distribution of input and output in the environment. This implies an understanding of the factors which influence the distribution and allocation of input and output resources in different situations for different individuals.

Due to the intrinsic characteristics of the new technological environment, it is likely that interaction will pose different perceptual and cognitive demands on humans compared to currently available technology (Gaggioli, 2005). It is therefore important to investigate how human perceptual and cognitive functions will be engaged in the emerging forms of interaction, and how this will affect an individual’s perceptual and cognitive space (e.g., emotion, vigilance, information processing, and memory). The main challenge in this respect is to identify and avoid forms of interaction which may lead to negative consequences such as confusion, cognitive overload, frustration, etc. This is particularly important given the pervasive impact of the new environment on all types of everyday activities and on the way of living.

### 4.11 Summary of interaction paradigms

Command based interaction (section 4.2), most prominently (but not exclusively) demonstrated by command line interfaces, was the first major interaction style in human-computer interaction. It is still widely used, often preferred by expert users for its high speed and flexible nature and complements very well direct manipulation interfaces. The keyboard remains the most efficient text-entry medium, so it is unlikely that it will go away and as a result, the same is true for command line interaction.

Direct Manipulation interaction (section 4.3) encompasses a wide array of styles, starting from WIMP interfaces to tangible and haptic interfaces and Virtual Reality interaction styles. The characteristic of direct manipulation interfaces is that it allows users to perform actions on objects directly, either through a pointing device (like in a WIMP interface) or through any other way, such as a gesture or directly by hand (for
example in touchscreens). It is the most widely used interaction style, specifically in the form of the GUI paradigm of the desktop metaphor that is found almost universally across desktop computers. Direct manipulation interfaces are also featured in new technologies that are growing in the market, such as tablet PCs and smartphones featuring a touch screen.

Natural language and speech dialogue (section 4.4) remains an unrealized promise in HCI, at least regarding the vision of natural interaction for all users, who seem to prefer an all natural language interaction or nothing. However, it has provided useful task-specific applications that have found a place in interaction with computer systems, for example situations where the hands are busy or for people with disabilities. It is also an important input mode in multimodal systems, which complement speech input with gestures or some other input mode to overcome strict NLP limitations. It should be expected that the speech channel as input and output will be an important part of multimodal interaction in intelligent environments.

Web interaction (section 4.5) is unique as context of use, which is a vital parameter in HCI, while the physical interaction itself in web use does not present a specifically different paradigm from conventional application use. The web information space and the choices offered, as well as the technology that opens vast communication capabilities, such as wireless connectivity, make web interaction (along with mobile interaction perhaps) the most important evolution in the way people interact with computers.

Mobile interaction (section 4.6) brought miniaturization and mobility into the fold. Starting from non-interactive tasks such as listening to music (with the Walkman) and branching into much more interactive activities such as making phone calls and surfing the web, mobile interaction is a major step in human-computer interaction that has opened many new possibilities, including the role of the device in ubiquitous computing. Mobile interaction is characterized (and constrained) by the physical size of the device and the context of use, which varies significantly from desk-top interaction. At this point in time, mobile interaction is undergoing a noticeable paradigm shift from scroll and select interfaces, dependent on the mobile phone
keypad interface, to the more sophisticated direct manipulation interface offered by the latest smartphones.

Multimodal interaction (section 4.7) integrates natural input modes, such as speech, pen, touch, hand gestures, eye gaze, and head and body movements. It represents a paradigm shift from conventional WIMP interfaces toward providing users with greater expressive power and naturalness and achieving forms of interaction closer to natural human-human communication. Recent advances in multimodal interaction also cater for with passive input that requires no explicit user command, such as vision-based tracking. This way, multimodality permeates into intelligent interaction environments where interaction is to a large extent implicit and continuous.

Virtual Reality (section 4.8) is a special case of human computer interaction, distinguished by the goal of providing an immersive experience to the user by replacing the real word with a computer-generated environment. This is achieved primarily with the use of head mounted displays, which isolate the user’s visual perception of the real world and provide an artificial alternative and with CAVE-type environments, which use large displays that dominate the user’s vision and allow for the immersive experience. How the user interacts in the various VR environments differs significantly, dependent of the nature of the VR application. For example, a training VR application mimics the real world controls of interaction (e.g., the plane’s cockpit) but a virtual environment that has no real world equivalent (such as a fantasy world) could use a direct manipulation interface with a “magic wand”, a tangible interface and so on. Due to the high cost of Virtual Reality, its applications are still limited to the military or large business organizations and widespread adoption of this interactive paradigm seems difficult and unlikely to become a reality in the near future.

Augmented Reality (section 4.9) blends real world information with computer generated visualization to facilitate user perception and user tasks. The main interactive element that distinguishes AR interaction, is in the form of visual feedback provided by the computer. Unlike VR, simple augmented reality applications (such as GPS based navigation) are starting to penetrate the market, mainly through the exploitation of the capabilities of modern smartphones. It is expected that with the
rapid growth of the market share presented by these devices, more applications of this sort will be quite common. Otherwise, augmented reality is successfully being employed in healthcare.

Ambient Intelligence (section 4.10) is arguably the next major step in the overall computing paradigm and consequently, the next major evolutionary step in human-computer interaction. What is fundamentally changing is that users no longer interact with a single machine but with the environment surrounding them, the computer becoming a more abstract entity, fusing intellegent artifacts through an invisible network of sensors. Interaction does not occur through a single interface and can be achieved via a multimodal paradigm that can include any of the interaction styles described in this chapter, such as gestures, speech, direct manipulation and combinations of those. At the moment, AmI environments are largely in the research stage, since they present a host of challenges that need to be overcome.

5. **Emerging challenges, future trends and conclusions**

There are several trends and challenges that determine the direction of future HCI research and development, and as a consequence the future evolution of interactivity. These include trends and developments in mobile interaction, concerns about universal access, multi-touch based devices, the evolution of web interaction, and the progressive emergence of Ambient Intelligence Environments.

Mobile web use has evolved into a totally different paradigm from the desktop counterpart. Similarities can be drawn with some add-ons or plugins of web browsers, which offer specific web services rather than a free net roaming experience. Examples would include a toolbar which displays the weather or stock market prices or such. Mobile users download web applications that provide specific services as well. For example, Apples App Store has thousands of applications that use the web, each with its own interface, without any particular uniform design paradigm (still in its infancy) and users consume bits of information from the web without really noticing it. Web usage is therefore spread over different small applications, from different manufacturers and providers. The competition is not any longer on the website level,
but on the mini-application level. There is no homogenous paradigm of web usage, the mobile web is “balkanized”, divided into thousands of service points with thousands of interfaces. Search is not the main issue, as it is on desktop web surfing. According to Google user experience designer Leland Rechis "The Pangaea of the Web is gone." (Wellman, 2007). Because of carrier portals and off portal applications, there is no mobile standard to develop for. In the mobile world developers have to be prepared to optimize for different devices, browsers, languages, carriers, countries and cultures. While the international and cultural factors affect desktop web surfing as well (Marcus and Rau, 2009) and have been examined quite well, the portal and device variety remain the most challenging issue in mobile development. An open question regarding the future of mobile devices is if there will be a convergence of devices, i.e., devices that combine the capabilities of more, or an increased diversity of appliances. Most mobile phones combine the capabilities of a telephone, a camera (including video recording), an electronic organizer and more often than not, a music device. Taking into account the recent trend and market acceptance of high-end touch-based phones that offer these features, as well as Internet connectivity, in a considerably more user-friendly manner, it seems that the convergence of devices case has more merit as the safest prediction, at least on the mobile phone size-scale.

This potential of mobile devices in general needs further investigation. There are many ongoing research projects that explore the mobile device as an ubiquitous computing medium, an augmented reality device, new interaction technique opportunities as presented by high-end phones with touchscreens, accelerometers and gyroscopes, cameras, etc.

Another imortal concern that influences HCI is the rapid aging of the population in the developed world. This has not only socio-economic but also accessibility implications in (especially mobile) computer interaction. Mobile devices are particularly susceptible to accessibility issues because of their inherent small size. Furthermore, mobile interaction is challenged by situational impairments and disabilities (SIIDs), due to the plethora of the contexts of use. A major challenge therefore in future HCI research, is to provide a solution to the accessibility problem regarding mobile use. Universal Access is a constant challenge in HCI but desktop
computing (for which most related work has been done) presents more alternative solutions and is, arguably, easier to respond to compared to mobile interaction.

The design approach of the new smartphones has been recently transferred into the tablet PC platform. While tablet PCs have been around for sometime and did feature a touchscreen, these devices are not merely (or, arguably, fully) PCs with a touchscreen. The Apple iPad, for example does not offer the same capabilities as a normal netbook or laptop, such as multitasking and running any type of application. Instead it is designed to offer specific services, such as multimedia, electronic reading and connection to the web. A major difference is the multi-touch interface, which bears significant advantages over older touch interfaces, as it offers a much richer and enjoyable user experience. This type of appliance, despite its obvious shortcomings compared with a standard (and much cheaper, for now at least) netbook, seems likely to become in the future the computing interaction of choice for a significant number of users. Its multitouch interface, the enjoyable user experience and its inherent mobility are a strong combination that covers many market needs. Moreover, there is a conscious effort to make smart phones and tablet PCs, and information sharing between them easy. This can be argued as a first step towards a more ubiquitous computing that is not constricted into one device or location. While there are still major difficulties to overcome, this trend is anticipated to move closer to a situation where abstract cloud computing is available and compatible through a host of different device types and platforms.

Augmented Reality in general is expected to be more available and accessible to the everyday user, through smartphone cameras, the aforementioned tablet PCs (also equipped with HD cameras) and other promising projects, such as SixthSense\textsuperscript{12}, which is a wearable device that blends the digital world with the physical.

Interacting with the web has changed significantly since the first days of the ARPANET. The text-based presentation of the first era was mostly confined in academia and business. The wide public became familiar with the web through GUI-based browsers, which soon were capable of providing a richer interactive experience.

\textsuperscript{12}http://www.ted.com/talks/pranav_mistry_the_thrilling_potential_of_sixthsense_technology.html
Today, the web is accessible by far more devices and in a plethora of contexts of use that is significantly altering the overall user experience. Service-specific information consumption in chunks, which characterizes mobile interaction, is also affecting desktop and laptop web interaction. An example is RSS feeds or browser plugin (addons, extensions) services. What this means in essence is that web interaction is now expanded to include more than the usual browser-based surfing style that dominated web interaction for more than 15 years.

Gesture-based interaction, mostly familiar to the wide audience through gaming (the Wii console is a particular example\(^\text{13}\)) is also being researched in the context of ambient intelligent environments and 3D interfaces (in VR applications or information visualization and navigation of large data sets—for example, see Underkoffler’s g-Speak SOE presented at TED in 2010\(^\text{14}\)) and is gaining more attention.

Finally, considering the technology available, it seems that the so-called third wave of computing mentioned earlier in the chapter, where the computer will gradually disappear in the background and be embedded in the surrounding environment, is approaching. Ambient Intelligence Environments may take many forms. For example, in Chapter ++ of this Handbook, various applications in different domains are discussed, where each installation differs in the actual size of the environment, the hardware and software involved, in the interaction styles employed, etc. Similarly, various ongoing research projects present different approaches and manifestations of the same idea of the disappearing computer, with different emphasis on the intelligent behaviour, the natural interaction involved, accessibility, aesthetics, design methodology and so on, depending on the specific domain.

Looking at the evolutionary history of interaction, the first observation is that none of the interaction styles that appeared from the first days of computing until today have actually disappeared. The oldest and most basic interaction style, command-line interfaces, is still widely used and the same holds true for every interaction style presented in this chapter. Following the ongoing research and development of new

\(^{13}\) http://www.nintendo.com/wii/console/controllers

\(^{14}\) http://www.ted.com/talks/lang/eng/john_underkoffler_drive_3d_data_with_a_gesture.html
interaction styles, such as gesture interfaces or haptic interfaces, there is no evidence that these will render any of the current popular styles obsolete. The conclusion that can be drawn is that there is no one optimal interaction method that can be devised. On the contrary, the paradigm shift towards an ambient intelligence environment that seeks to replace current methods of interaction with computing favours a multimodal approach, where every interaction style is utilized according to its appropriateness to the context of use and the task at hand. Therefore, in the AmI context, an important challenge is to approach every interaction style and determine its optimal use and appropriate place in the services offered by the intelligent environments.

More pragmatically, a huge challenge involved in Ambient Intelligence Environments, escaping the laboratory and entering the public, is the complexity inherent in such a large scale. There are issues regarding integration, interoperability and synergy between various devices and artefacts. This also includes security and privacy concerns, increased with the presence of more than one platform and devices that must all conform to the same high standards of safety. Not to be underestimated is the cost of such an environment, including the cost of use (parallels may be drawn by the Western/Japanese contrast of mobile web use adoption).

The technology to realize ubiquitous computing exists. Theoretically and to a certain degree practically, the wide availability of the Internet everywhere (through Wi-Fi or mobile access) and the corresponding network capable devices can form an environment in which they can communicate and exchange information to offer ubiquitous services. The question is whether (or more to the point, when) the interaction we experience with these components individually will expand to include all in a loosely perceived abstract supersystem that intelligently observes, adapts and responds to human needs and desires.
References
Butz, A. (2010). User Interfaces and HCI for Ambient Intelligence and Smart Environments. In H. Nakashima, H. Aghajan and J. C. Augusto (Eds.). Handbook of Ambient Intelligence and Smart Environments (pp 535-558). Springer US.


Lopez-Cozar, R., Callejas, Z. (2010). Multimodal Dialogue for Ambient Intelligence and Smart Environments. In H. Nakashima, H. Aghajan and J. C. Augusto (Eds.), Handbook of Ambient Intelligence and Smart Environments (pp 559-579). Springer US.


