Human-Computer Interaction: Overview on State of the Art

Article *in* International Journal on Smart Sensing and Intelligent Systems · January 2008

DOI: 10.21307/ijssis-2017-283

CITATIONS
218

READS
8,444

4 authors, including:

Fakhri Karray
University of Waterloo
615 PUBLICATIONS **12,142** CITATIONS
SEE PROFILE

Milad Alemzadeh
Linkedin
11 PUBLICATIONS **560** CITATIONS
SEE PROFILE

Mo Nours Arab
ZeusProtocol
5 PUBLICATIONS **531** CITATIONS
SEE PROFILE

Human-Computer Interaction: Overview on State of the Art

Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh and Mo Nours Arab
Pattern Analysis and Machine Intelligence Lab., Department of Electrical and Computer Engineering
University of Waterloo, Waterloo, Canada
karray@uwaterloo.ca, malemzad@uwaterloo.ca, jaabousa@uwaterloo.ca, mnarab@uwaterloo.ca

Abstract

The intention of this paper is to provide an overview on the subject of Human-Computer Interaction. The overview includes the basic definitions and terminology, a survey of existing technologies and recent advances in the field, common architectures used in the design of HCI systems which includes unimodal and multimodal configurations, and finally the applications of HCI. This paper also offers a comprehensive number of references for each concept, method, and application in the HCI.

Keywords: Human-Computer Interaction, Multimodal HCI, Ubiquitous Computing

1 Introduction

Utilizing computers had always begged the question of interfacing. The methods by which human has been interacting with computers has travelled a long way. The journey still continues and new designs of technologies and systems appear more and more every day and the research in this area has been growing very fast in the last few decades.

The growth in Human-Computer Interaction (HCI) field has not only been in quality of interaction, it has also experienced different branching in its history. Instead of designing regular interfaces, the different research branches have had different focus on the concepts of multimodality rather than unimodality, intelligent adaptive interfaces rather than command/action based ones, and finally active rather than passive interfaces.

This paper intends to provide an overview on the state of the art of HCI systems and cover most important branches as mentioned above. In the next section, basic definitions and terminology of HCI are given. Then an overview of existing technologies and also recent advances in the field is provided. This is followed up by a description on the different architectures of HCI designs. The final sections pertain to description on some of the applications of HCI and future directions in the field.

2 Human-Computer Interaction: Definition, Terminology

Sometimes called as Man-Machine Interaction or Interfacing, concept of Human-Computer Interaction/Interfacing (HCI) was automatically represented with the emerging of computer, or more generally machine, itself. The reason, in fact, is clear: most sophisticated machines are worthless unless they can be used properly by men. This basic argument simply presents the main terms that should be considered in the design of HCI: functionality and usability [1].

Why a system is actually designed can ultimately be defined by what the system can do i.e. how the functions of a system can help towards the achievement of the purpose of the system. *Functionality* of a system is defined by the set of actions or services that it provides to its users. However, the value of functionality is visible only when it becomes possible to be efficiently utilised by the user [2]. *Usability* of a system with a certain functionality is the range and degree by which the system can be used efficiently and adequately to accomplish certain goals for certain users. The actual effectiveness of a system is achieved when there is a proper balance between the functionality and usability of a system [3].

Having these concepts in mind and considering that the terms computer, machine and system are often used interchangeably in this context, HCI is a design that should produce a fit between the user, the machine and the required services in order to achieve a certain performance both in quality and optimality of the services [4]. Determining what makes a certain HCI design good is mostly subjective and context dependant. For example, an aircraft part designing tool should provide high precisions in view and design of the parts while a graphics editing software may not need such a precision. The available technology could also affect how different types of HCI are designed for the same purpose. One example is using commands, menus, graphical user interfaces (GUI), or virtual reality to access functionalities of any given computer. In the next section, a more detailed overview of existing methods and devices used to interact with computers and the recent advances in the field is presented.

3 Overview on HCI

The advances made in last decade in HCI have almost made it impossible to realize which concept is fiction and which is and can be real. The thrust in research and the constant twists in marketing cause the new technology to become available to everyone in no time. However, not all existing technologies are accessible and/or affordable by public. In the first part of this section, an overview of the technology that more or less is available to and used by public is

presented. In the second part, an outlook of the direction to which HCI research is heading has been drawn.

3.1 Existing HCI Technologies

HCI design should consider many aspects of human behaviours and needs to be useful. The complexity of the degree of the involvement of a human in interaction with a machine is sometimes invisible compared to the simplicity of the interaction method itself. The existing interfaces differ in the degree of complexity both because of degree of functionality/usability and the financial and economical aspect of the machine in market. For instance, an electrical kettle need not to be sophisticated in interface since its only functionality is to heat the water and it would not be cost-effective to have an interface more than a thermostatic on and off switch. On the other hand, a simple website that may be limited in functionality should be complex enough in usability to attract and keep customers [1].

Therefore, in design of HCI, the degree of activity that involves a user with a machine should be thoroughly thought. The user activity has three different levels: physical [5], cognitive [6], and affective [7]. The physical aspect determines the mechanics of interaction between human and computer while the cognitive aspect deals with ways that users can understand the system and interact with it. The affective aspect is a more recent issue and it tries not only to make the interaction a pleasurable experience for the user but also to affect the user in a way that make user continue to use the machine by changing attitudes and emotions toward the user [1].

The focus of this paper is mostly on the advances in physical aspect of interaction and to show how different methods of interaction can be combined (Multi-Modal Interaction) and how each method can be improved in performance (Intelligent Interaction) to provide a better and easier interface for the user. The existing physical technologies for HCI basically can be categorized by the relative human sense that the device is designed for. These devices are basically relying on three human senses: vision, audition, and touch [1].

Input devices that rely on vision are the most used kind and are commonly either switch-based or pointing devices [8] [9]. The switch-based devices are any kind of interface that uses buttons and switches like a keyboard [10]. The pointing devices examples are mice, joysticks, touch screen panels, graphic tablets, trackballs, and pen-based input [11]. Joysticks are the ones that have both switches and pointing abilities. The output devices can be any kind of visual display or printing device [3].

The devices that rely on audition are more advance devices that usually need some kind of speech recognition [12]. These devices aim to facilitate the interaction as much as possible and therefore, are much more difficult to build [13]. Output auditory devices are however easier to create. Nowadays, all kind of non-speech [14] and speech signals and messages are produced by machines as output signals. Beeps, alarms, and turn-by-turn navigation commands of a GPS device are simple examples.

The most difficult and costly devices to build are haptic devices [15]. “These kinds of interfaces generate sensations to the skin and muscles through touch, weight and relative rigidity [1].” Haptic devices [16] are generally made for virtual reality [17] or disability assistive applications [18].

The recent methods and technologies in HCI are now trying to combine former methods of interaction together and with other advancing technologies such as networking and animation. These new advances can be categorized in three sections: wearable devices [19], wireless devices [20], and virtual devices [21]. The technology is improving so fast that even the borders between these new technologies are fading away and they are getting mixed together. Few examples of these devices are: GPS navigation systems [22], military super-soldier enhancing devices (e.g. thermal vision [23], tracking other soldier movements using GPS, and environmental scanning), radio frequency identification (RFID) products, personal digital assistants (PDA), and virtual tour for real estate business [24]. Some of these new devices upgraded and integrated previous methods of interaction. As an illustration in case, there is the solution to keyboarding that has been offered by Compaq’s iPAQ which is called Canesta keyboard as shown in figure 1. This is a virtual keyboard that is made by projecting a QWERTY like pattern on a solid surface using a red light. Then device tries to track user’s finger movement while typing on the surface with a motion sensor and send the keystrokes back to the device [25].

Figure 1: Canesta virtual keyboard [26]

3.2 Recent Advances in HCI

In following sections, recent directions and advances of research in HCI, namely intelligent and adaptive interfaces and ubiquitous computing, are presented. These interfaces involve different levels of user activity: physical, cognitive, and affection.

3.2.1 Intelligent and Adaptive HCI

Although the devices used by majority of public are still some kind of plain command/action setups using not very sophisticated physical apparatus, the flow of research is directed to design of intelligent and adaptive interfaces. The exact theoretical definition of the concept of intelligence or being smart is not known or at least not publicly agreeable. However, one can define these concepts by the apparent growth and improvement in functionality and usability of new devices in market.

As mentioned before, it is economically and technologically crucial to make HCI designs that provide easier, more pleasurable and satisfying experience for the users. To realize this goal, the interfaces are getting more natural to use every day. Evolution of interfaces in note-taking tools is a good example. First there were typewriters, then keyboards and now touch screen tablet PCs that you can write on using your own handwriting and they recognize it change it to text [27] and if not already made, tools that transcript whatever you say automatically so you do not need to write at all.

One important factor in new generation of interfaces is to differentiate between using intelligence in the making of the interface (Intelligent HCI) [28] or in the way that the interface interacts with users (Adaptive HCI) [29]. Intelligent HCI designs are interfaces that incorporate at least some kind of intelligence in perception from and/or response to users. A few examples are speech enabled interfaces [30] that use natural language to interact with user and devices that visually track user's movements [31] or gaze [32] and respond accordingly.

Adaptive HCI designs, on the other hand, may not use intelligence in the creation of interface but use it in the way they continue to interact with users [33]. An adaptive HCI might be a website using regular GUI for selling various products. This website would be adaptive -to some extent- if it has the ability to recognize the user and keeps a memory of his searches and purchases and intelligently search, find, and suggest products on sale that it thinks user might need. Most of these kinds of adaptation are the ones that deal with cognitive and affective levels of user activity [1].

Another example that uses both intelligent and adaptive interface is a PDA or a tablet PC that has the handwriting recognition ability and it can adapt to the handwriting of the logged in user so to improve its performance by remembering the corrections that the user made to the recognised text.

Finally, another factor to be considered about intelligent interfaces is that most non-intelligent HCI design are passive in nature i.e. they only respond whenever invoked by user while ultimate intelligent and adaptive interfaces tend to be active interfaces. The example is smart billboards or advertisements that present themselves according to users' taste [34] [35]. In the next section, combination of different methods of HCI and how it could help towards making intelligent adaptive natural interfaces is discussed.

3.2.2 Ubiquitous Computing and Ambient Intelligence

The latest research in HCI field is unmistakably *ubiquitous computing* (Ubicomp). The term which often used interchangeably by *ambient intelligence* and *pervasive computing*, refers to the ultimate methods of human-computer interaction that is the deletion of a desktop and embedding of the computer in the environment so that it becomes invisible to humans while surrounding them everywhere hence the term ambient.

The idea of ubiquitous computing was first introduced by Mark Weiser during his tenure as chief technologist at Computer Science Lab in Xerox PARC in 1998. His idea was to embed computers everywhere in the environment and everyday objects so that people could interact with many computers at the same time while they are invisible to them and wirelessly communicating with each other [27].

Ubicomp has also been named the Third Wave of computing. The First Wave was the mainframe era, many people one computer. Then it was the Second Wave, one person one computer which was called PC era and now Ubicomp introduces many computers one person era [27]. Figure 2 shows the major trends in computing.

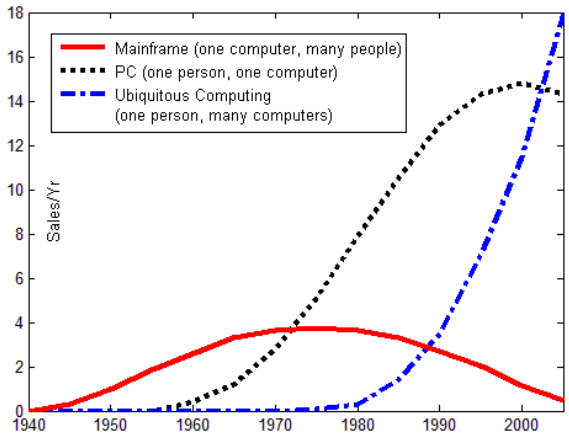

Figure 2: Major trends in computing [27]

4 HCI Systems Architecture

Most important factor of a HCI design is its configuration. In fact, any given interface is generally defined by the number and diversity of inputs and outputs it provides. Architecture of a HCI system shows what these inputs and outputs are and how they work together. Following sections explain different configurations and designs upon which an interface is based.

4.1 Unimodal HCI Systems

As mentioned earlier, an interface mainly relies on number and diversity of its inputs and outputs which are communication channels that enable users to interact with computer via this interface. Each of the different independent single channels is called a modality [36]. A system that is based on only one modality is called *unimodal*. Based on the nature of different modalities, they can be divided into three categories:

1. Visual-Based
2. Audio-Based
3. Sensor-Based

The next sub-sections describe each category and provide examples and references to each modality.

4.1.1 Visual-Based HCI

The visual based human computer interaction is probably the most widespread area in HCI research. Considering the extent of applications and variety of open problems and approaches,

researchers tried to tackle different aspects of human responses which can be recognized as a visual signal. Some of the main research areas in this section are as follow:

- Facial Expression Analysis
- Body Movement Tracking (Large-scale)
- Gesture Recognition
- Gaze Detection (Eyes Movement Tracking)

While the goal of each area differs due to applications, a general conception of each area can be concluded. Facial expression analysis generally deals with recognition of emotions visually [37] [38] [39]. Body movement tracking [31] [40] and gesture recognition [41] [42] [43] are usually the main focus of this area and can have different purposes but they are mostly used for direct interaction of human and computer in a command and action scenario. Gaze detection [32] is mostly an indirect form of interaction between user and machine which is mostly used for better understanding of user's attention, intent or focus in context-sensitive situations [44]. The exception is eye tracking systems for helping disabilities in which eye tracking plays a main role in command and action scenario, e.g. pointer movement, blinking for clicking [45]. It is notable that some researchers tried to assist or even replace other types of interactions (audio-, sensor-based) with visual approaches. For example, lip reading or lip movement tracking is known to be used as an influential aid for speech recognition error correction [46].

4.1.2 Audio-Based HCI

The audio based interaction between a computer and a human is another important area of HCI systems. This area deals with information acquired by different audio signals. While the nature of audio signals may not be as variable as visual signals but the information gathered from audio signals can be more trustable, helpful, and is some cases unique providers of information. Research areas in this section can be divided to the following parts:

- Speech Recognition
- Speaker Recognition
- Auditory Emotion Analysis
- Human-Made Noise/Sign Detections (Gasp, Sigh, Laugh, Cry, etc.)
- Musical Interaction

Historically, speech recognition [12] and speaker recognition [47] have been the main focus of researchers. Recent endeavors to integrate human emotions in intelligent human computer interaction initiated the efforts in analysis of emotions in audio signals [48] [49]. Other than the tone and pitch of speech data, typical human auditory signs such as sigh, gasp, and etc helped emotion analysis for designing more intelligent HCI system [50]. Music generation and interaction is a very new area in HCI with applications in art industry which is studied in both audio- and visual-based HCI systems [51].

4.1.3 Sensor-Based HCI

This section is a combination of variety of areas with a wide range of applications. The commonality of these different areas is that at least one physical sensor is used between user and machine to provide the interaction. These sensors as shown below can be very primitive or very sophisticated.

1. Pen-Based Interaction

2. Mouse & Keyboard

3. Joysticks

4. Motion Tracking Sensors and Digitizers

5. Haptic Sensors

6. Pressure Sensors

7. Taste/Smell Sensors

Some of these sensors have been around for a while and some of them are very new technologies. Pen-Based sensors are specifically of interest in mobile devices and are related to pen gesture [30] and handwriting recognition areas. Keyboards, mice and joysticks are already discussed in section 3.1. For more information consult references: [8] [9] [10] [11]. Motion tracking sensors/digitizers are state-of-the-art technology which revolutionized movie, animation, art, and video-game industry. They come in the form of wearable cloth or joint sensors and made computers much more able to interact with reality and human able to create their world virtually. Figure 3 depicts such a device. Haptic and pressure sensors are of special interest for applications in robotics and virtual reality [15] [16] [18]. New humanoid robots include hundreds of haptic sensors that make the robots sensitive and aware to touch [52] [53]. These types of sensors are also used in medical surgery application [54]. A few

research works are also done on area of taste and smell sensors [55]; however they are not as popular as other areas.

Figure 3: Wearable motion capture cloth for making of video games (Taken from Operation Sports)

4.2 Multimodal HCI Systems

The term multimodal refers to combination of multiple modalities. In MMHCI systems, these modalities mostly refer to the ways that the system responds to the inputs, i.e. communication channels [36]. The definition of these channels is inherited from human types of communication which are basically his senses: Sight, Hearing, Touch, Smell, and Taste. The possibilities for interaction with a machine include but are not limited to these types.

Therefore, a multimodal interface acts as a facilitator of human-computer interaction via two or more modes of input that go beyond the traditional keyboard and mouse. The exact number of supported input modes, their types and the way in which they work together may vary widely from one multimodal system to another. Multimodal interfaces incorporate different combinations of speech, gesture, gaze, facial expressions and other non-conventional modes of input. One of the most commonly supported combinations of input methods is that of gesture and speech [56].

Although an ideal multimodal HCI system should contain a combination of single modalities that interact correlatively, the practical boundaries and open problems in each modality oppose limitations on the fusion of different modalities. In spite of all progress made in MMHCI, in most of existing multimodal systems, the modalities are still treated separately and only at the end, results of different modalities are combined together.

The reason is that the open problems in each area are yet to be perfected meaning that there is still work to be done to acquire a reliable tool for each sub-area. Moreover, roles of different modalities and their share in interplay are not scientifically known. “Yet, people convey multimodal communicative signals in a complementary and redundant manner. Therefore, in order to accomplish a human-like multimodal analysis of multiple input signals acquired by different sensors, the signals cannot be considered mutually independently and cannot be combined in a context-free manner at the end of the intended analysis but, on the contrary, the input data should be processed in a joint feature space and according to a context-dependent model. In practice, however, besides the problems of context sensing and developing context-dependent models for combining multisensory information, one should cope with the size of the required joint feature space. Problems include large dimensionality, differing feature formats, and time-alignment [36].”

An interesting aspect of multimodality is the collaboration of different modalities to assist the recognitions. For example, lip movement tracking (visual-based) can help speech recognition methods (audio-based) and speech recognition methods (audio-based) can assist command acquisition in gesture recognition (visual-based). The next section shows some of application of intelligent multimodal systems.

5 Applications

A classic example of a multimodal system is the “Put That There” demonstration system [57]. This system allowed one to move an object into a new location on a map on the screen by saying “put that there” while pointing to the object itself then pointing to the desired destination. Multimodal interfaces have been used in a number of applications including map-based simulations, such as the aforementioned system; information kiosks, such as AT&T’s MATCHKiosk [58] and biometric authentication systems [56].

Multimodal interfaces can offer a number of advantages over traditional interfaces. For one thing, they can offer a more natural and user-friendly experience. For instance, in a real-estate system called Real Hunter [24], one can point with a finger to a house of interest and speak to make queries about that particular house. Using a pointing gesture to select an object and using speech to make queries about it illustrates the type of natural experience multimodal interfaces offer to their users. Another key strength of multimodal interfaces is their ability to provide redundancy to accommodate different people and different circumstances. For instance, MATCHKiosk [58] allows one to use speech or handwriting to specify the type of

business to search for on a map. Thus, in a noisy setting, one may provide input through handwriting rather than speech. Few other examples of applications of multimodal systems are listed below:

- Smart Video Conferencing [59]
- Intelligent Homes/Offices [60]
- Driver Monitoring [61]
- Intelligent Games [62]
- E-Commerce [63]
- Helping People with Disabilities [64]

In the following sections, some of important applications of multimodal systems have been presented with greater details.

5.1 Multimodal Systems for Disabled people

One good application of multimodal systems is to address and assist disabled people (as persons with hands disabilities), which need other kinds of interfaces than ordinary people. In such systems, disabled users can perform work on the PC by interacting with the machine using voice and head movements [65]. Figure 4 is an actual example of such a system.

Figure 4: Gaze detection pointing system for people with disabilities (taken from www.adamfulton.co.uk)

Two modalities are then used: speech and head movements. Both modalities are active continuously. The head position indicates the coordinates of the cursor in current time moment on the screen. Speech, on the other hand, provides the needed information about the meaning of the action that must be performed with an object selected by the cursor.

Synchronization between the two modalities is performed by calculating the cursor position at the beginning of speech detection. This is mainly due to the fact that during the process of pronouncing the complete sentence, the cursor location can be moved by moving the head, and then the cursor can be pointing to other graphical object; moreover the command which must be fulfilled is appeared in the brain of a human in a short time before beginning of phrase input. Figure 5 shows the diagram of this system.

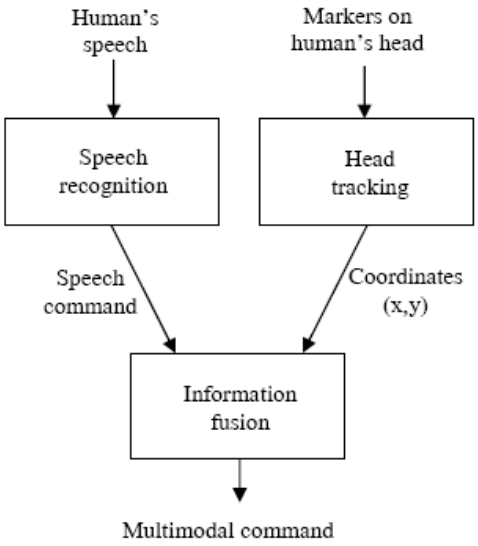

Figure 5: Diagram of a bimodal system [65]

In spite of some decreasing of operation speed, the multimodal assertive system allows working with computer without using standard mouse and keyboard. Hence, such system can be successfully used for hands-free PC control for users with disabilities of their hands.

5.2 Emotion Recognition Multimodal Systems

As we move towards a world in which computers are more and more ubiquitous, it will become more essential that machines perceive and interpret all clues, implicit and explicit, that we may provide them regarding our intentions. A natural human-computer interaction cannot be based solely on explicitly stated commands. Computers will have to detect the various behavioural signals based on which to infer one's emotional state. This is a significant piece of the puzzle that one has to put together to predict accurately one's intentions and future behaviour.

People are able to make prediction about one's emotional state based on their observations about one's face, body, and voice. Studies show that if one had access to only one of these modalities, the face modality would produce the best predictions. However, this accuracy can

be improved by 35% when human judges are given access to both face and body modalities together [66]. This suggests that affect recognition, which has for the most part focused on facial expressions, can greatly benefit from multimodal fusion techniques.

One of the few works that has attempted to integrate more than one modality for affect recognition is [67] in which facial features and body posture features are combined to produce an indicator of one's frustration. Another work that integrated face and body modalities is [68] in which the authors showed that, similar to humans, machine classification of emotion is better when based upon face and body data, rather than either modality alone. In [69], the authors attempted to fuse facial and voice data for affect recognition. Once again, remaining consistent with human judges, machine classification of emotion as neutral, sad, angry, or happy was most accurate when the facial and vocal data is combined.

They recorded the four emotions: "sadness, anger, happiness, and neutral state". The detailed facial motions were captured in conjunctions with simultaneous speech recordings. Deducted experiments showed that the performance of the facial recognition based system overcame the one based on acoustic information only. Results also show that an appropriate fusion of both modalities gave measurable improvements.

Results show that the emotion recognition system based on acoustic information only give an overall performance of 70.9 percent, compared to an overall performance of 85 percent for a recognition system based on facial expressions. This is, in fact, due to the fact that the cheek areas give important information for emotion classification.

On the other hand, for the bimodal system based on fusing the facial recognition and acoustic information, the overall performance of this classifier was 89.1 percent.

5.3 Map-Based Multimodal Applications

Different input modalities are suitable for expressing different messages. For instance, speech provides an easy and natural mechanism for expressing a query about a selected object or requesting that the object initiate a given operation. However, speech may not be ideal for tasks, such as selection of a particular region on the screen or defining out a particular path. These types of tasks are better accommodated by hand or pen gestures. However, making queries about a given region and selecting that region are all typical tasks that should be accommodate by a map-based interface. Thus, the natural conclusion is that map-based interfaces can greatly improve the user experience by supporting multiple modes of input, especially speech and gestures.

Quickset [70] is one of the more widely known and older map-based applications that make use of speech and pen gesture input. Quickset is a military-training application that allows users to use one of the two modalities or both simultaneously to express a full command. For instance, users may simply draw out with a pen a predefined symbol for platoons at a given location on the map to create a new platoon in that location. Alternatively, users could use speech to specify their intent on creating a new platoon and could specify vocally the co-ordinates in which to place the platoon. Lastly, users could express vocally their intent on making a new platoon while making a pointing gesture with a pen to specify the location of the new platoon.

A more recent multimodal map-based application is Real Hunter [24]. It is a real-estate interface that expects users to select objects or regions with touch input while making queries using speech. For instance, the user can ask “How much is this?” while pointing to a house on the map.

Tour guides are another type of map-based applications that have shown great potential to benefit from multimodal interfaces. One such example is MATCHKiosk [58], the interactive city guide. In a similar fashion to Quickset, MATCHKiosk allows one to express certain queries using speech only, such as “Find me Indian restaurants in Washington.”; using pen input only by circling a region and writing out “restaurants”; using bimodal input by saying “Indian restaurants in this area” and drawing out a circle around Alexandria. These examples illustrate MATCHKiosk’s incorporation of handwriting recognition that can frequently substitute for speech input. Although speech may be the more natural option for a user, given the imperfectness of speech, especially in noisy environments, having handwriting as a backup can reduce user frustration.

5.4 Multimodal Human-Robot Interface Applications

Similar to some map-based interfaces, human-robot interfaces usually have to provide mechanisms for pointing to particular locations and for expressing operation-initiating requests. As discussed earlier, the former type of interaction is well accommodated by gestures, whereas the latter is better accommodate by speech. Thus, the human-robot interface built by the Naval Research Laboratory (NRL) should come as no surprise [71]. NRL’s interface allows users to point to a location while saying “Go over there”. Additionally, it allows users to use a PDA screen as a third possible avenue of interaction, which could be resorted to when speech or hand gesture recognition is failing. Another

multimodal human-robot interface is the one built by Interactive System Laboratories (ISL) [72], which allows use of speech to request the robot to do something while gestures could be used to point to objects that are referred to by the speech. One such example is to ask the robot, "switch on the light" while pointing to the light. Additionally, in ISL's interface, the system may ask for clarification from the user when unsure about the input. For instance, in case that no hand gesture is recognized that is pointing to a light, the system may ask the user: "Which light?"

5.5 Multi-Modal HCI in Medicine

By the early 1980s, surgeons were beginning to reach their limits based on traditional methods alone. Human hand was unfeasible for many tasks and greater magnification and smaller tools were needed. Higher precision was required to localize and manipulate within small and sensitive parts of the human body. Digital robotic neuro-surgery has come as a leading solution to these limitations and emerged fast due to the vast improvements in engineering, computer technology and neuro-imaging techniques. Robotics surgery was introduced into the surgical area [73].

State University of Aerospace Instrumentation, University of Karlsruhe (Germany) and Harvard Medical School (USA) has been working on developing man-machine interfaces, adaptive robots and multi-agent technologies intended for neuro-surgery [54].

The neuro-surgical robot consists of the following main components: An arm, feedback vision sensors, controllers, a localization system and a data processing centre. Sensors provide the surgeon with feedbacks from the surgical site with real-time imaging, where the latter one updates the controller with new instructions for the robot by using the computer interface and some joysticks.

Neuro-surgical robotics provide the ability to perform surgeries on a much smaller scale with much higher accuracy and precision, giving access to small corridors which is completely important when a brain surgery is involved [73].

6 Conclusion

Human-Computer Interaction is an important part of systems design. Quality of system depends on how it is represented and used by users. Therefore, enormous amount of attention has been paid to better designs of HCI. The new direction of research is to replace common regular methods of interaction with intelligent, adaptive, multimodal, natural methods.

Ambient intelligence or ubiquitous computing which is called the Third Wave is trying to embed the technology into the environment so to make it more natural and invisible at the same time. Virtual reality is also an advancing field of HCI which can be the common interface of the future. This paper attempted to give an overview on these issues and provide a survey of existing research through a comprehensive reference list.

7 References

[1] D. Te'eni, J. Carey and P. Zhang, *Human Computer Interaction: Developing Effective Organizational Information Systems*, John Wiley & Sons, Hoboken (2007).

[2] B. Shneiderman and C. Plaisant, *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (4th edition), Pearson/Addison-Wesley, Boston (2004).

[3] J. Nielsen, *Usability Engineering*, Morgan Kaufman, San Francisco (1994).

[4] D. Te'eni, "Designs that fit: an overview of fit conceptualization in HCI", in P. Zhang and D. Galletta (eds), *Human-Computer Interaction and Management Information Systems: Foundations*, M.E. Sharpe, Armonk (2006).

[5] A. Chapanis, *Man Machine Engineering*, Wadsworth, Belmont (1965).

[6] D. Norman, "Cognitive Engineering", in D. Norman and S. Draper (eds), *User Centered Design: New Perspective on Human-Computer Interaction*, Lawrence Erlbaum, Hillsdale (1986).

[7] R.W. Picard, *Affective Computing*, MIT Press, Cambridge (1997).

[8] J.S. Greenstein, "Pointing devices", in M.G. Helander, T.K. Landauer and P. Prabhu (eds), *Handbook of Human-Computer Interaction*, Elsevier Science, Amsterdam (1997).

[9] B.A. Myers, "A brief history of human-computer interaction technology", *ACM interactions*, 5(2), pp 44-54 (1998).

[10] B. Shneiderman, *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (3rd edition), Addison Wesley Longman, Reading (1998).

[11] A. Murata, "An experimental evaluation of mouse, joystick, joycard, lightpen, trackball and touchscreen for Pointing - Basic Study on Human Interface Design", *Proceedings of the Fourth International Conference on Human-Computer Interaction 1991*, pp 123-127 (1991).

[12] L.R. Rabiner, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs (1993).

[13] C.M. Karat, J. Vergo and D. Nahamoo, "Conversational interface technologies", in J.A. Jacko and A. Sears (eds), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Application*, Lawrence Erlbaum Associates, Mahwah (2003).

[14] S. Brewster, "Non speech auditory output", in J.A. Jacko and A. Sears (eds), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Application*, Lawrence Erlbaum Associates, Mahwah (2003).

[15] G. Robles-De-La-Torre, "The Importance of the sense of touch in virtual and real environments", *IEEE Multimedia 13(3), Special issue on Haptic User Interfaces for Multimedia Systems*, pp 24-30 (2006).

[16] V. Hayward, O.R. Astley, M. Cruz-Hernandez, D. Grant and G. Robles-De-La-Torre, "Haptic interfaces and devices", *Sensor Review* 24(1), pp 16-29 (2004).

[17] J. Vince, *Introduction to Virtual Reality*, Springer, London (2004).

[18] H. Iwata, "Haptic interfaces", in J.A. Jacko and A. Sears (eds), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Application*, Lawrence Erlbaum Associates, Mahwah (2003).

[19] W. Barfield and T. Caudell, *Fundamentals of Wearable Computers and Augmented Reality*, Lawrence Erlbaum Associates, Mahwah (2001).

[20] M.D. Yacoub, *Wireless Technology: Protocols, Standards, and Techniques*, CRC Press, London (2002).

[21] K. McMenemy and S. Ferguson, *A Hitchhiker's Guide to Virtual Reality*, A K Peters, Wellesley (2007).

[22] Global Positioning System, "Home page", http://www.gps.gov/, visited on 10/10/2007.

[23] S.G. Burnay, T.L. Williams and C.H. Jones, *Applications of Thermal Imaging*, A. Hilger, Bristol (1988).

[24] J. Y. Chai, P. Hong and M. X. Zhou, "A probabilistic approach to reference resolution in multimodal user interfaces", *Proceedings of the 9th International Conference on Intelligent User Interfaces*, Funchal, Madeira, Portugal, pp 70-77 (2004).

[25] E.A. Bretz, "When work is fun and games", *IEEE Spectrum*, 39(12), pp 50-50 (2002).

[26] ExtremeTech, "Canesta says "Virtual Keyboard" is reality", http://www.extremetech.com/article2/0,1558,539778,00.asp, visited on 15/10/2007.

[27] G. Riva, F. Vatalaro, F. Davide and M. Alaniz, *Ambient Intelligence: The Evolution of Technology, Communication and Cognition towards the Future of HCI*, IOS Press, Fairfax (2005).

[28] M.T. Maybury and W. Wahlster, *Readings in Intelligent User Interfaces*, Morgan Kaufmann Press, San Francisco (1998).

[29] A. Kirlik, *Adaptive Perspectives on Human-Technology Interaction*, Oxford University Press, Oxford (2006).

[30] S.L. Oviatt, P. Cohen, L. Wu, J. Vergo, L. Duncan, B. Suhm, J. Bers, T. Holzman, T. Winograd, J. Landay, J. Larson and D. Ferro, "Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions", *Human-Computer Interaction*, 15, pp 263-322 (2000).

[31] D.M. Gavrila, "The visual analysis of human movement: a survey", *Computer Vision and Image Understanding*, 73(1), pp 82-98 (1999).

[32] L.E. Sibert and R.J.K. Jacob, "Evaluation of eye gaze interaction", *Conference of Human-Factors in Computing Systems*, pp 281-288 (2000).

[33] Various Authors, "Adaptive, intelligent and emotional user interfaces", Part II of *HCI Intelligent Multimodal Interaction Environments, 12th International Conference, HCI International 2007 (Proceedings Part III)*, Springer Berlin, Heidelberg (2007).

[34] M.N. Huhns and M.P. Singh (eds), *Readings in Agents*, Morgan Kaufmann, San Francisco (1998).

[35] C.S. Wasson, *System Analysis, Design, and Development: Concepts, Principles, and Practices*, John Wiley & Sons, Hoboken (2006).

[36] A. Jaimes and N. Sebe, "Multimodal human computer interaction: a survey", *Computer Vision and Image Understanding*, 108(1-2), pp 116-134 (2007).

[37] I. Cohen, N. Sebe, A. Garg, L. Chen and T.S. Huang, "Facial expression recognition from video sequences: temporal and static modeling", *Computer Vision and Image Understanding*, 91(1-2), pp 160-187 (2003).

[38] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey", *Pattern Recognition*, 36, pp 259-275 (2003).

[39] M. Pantic and L.J.M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art", *IEEE Transactions on PAMI*, 22(12), pp 1424-1445 (2000).

[40] J.K. Aggarwal and Q. Cai, "Human motion analysis: a review", *Computer Vision and Image Understanding*, 73(3), pp 428-440 (1999).

[41] S. Kettebekov and R. Sharma, "Understanding gestures in multimodal human computer interaction", *International Journal on Artificial Intelligence Tools*, 9(2), pp 205-223 (2000).

[42] Y. Wu and T. Huang., "Vision-based gesture recognition: a review", in A. Braffort, R. Gherbi, S. Gibet, J. Richardson and D. Teil (eds), *Gesture-Based Communication in Human-Computer Interaction, volume 1739 of Lecture Notes in Artificial Intelligence,* Springer-Verlag, Berlin/Heidelberg (1999).

[43] T. Kirishima, K. Sato and K. Chihara, "Real-time gesture recognition by learning and selective control of visual interest points", *IEEE Transactions on PAMI*, 27(3), pp 351-364 (2005).

[44] R. Ruddaraju, A. Haro, K. Nagel, Q. Tran, I. Essa, G. Abowd and E. Mynatt, "Perceptual user interfaces using vision-based eye tracking", *Proceedings of the 5th International Conference on Multimodal Interfaces*, Vancouver, pp 227-233 (2003).

[45] A.T. Duchowski, "A breadth-first survey of eye tracking applications", *Behavior Research Methods, Instruments, and Computers*, 34(4), pp 455-470 (2002).

[46] P. Rubin, E. Vatikiotis-Bateson and C. Benoit (eds.), "Special issue on audio-visual speech processing", *Speech Communication*, 26, pp 1-2 (1998).

[47] J.P. Campbell Jr., "Speaker recognition: a tutorial", *Proceedings of IEEE*, 85(9), pp 1437-1462 (1997).

[48] P.Y. Oudeyer, "The production and recognition of emotions in speech: features and algorithms", *International Journal of Human-Computer Studies*, 59(1-2), pp 157-183 (2003).

[49] L.S. Chen, *Joint Processing of Audio-Visual Information for the Recognition of Emotional Expressions in Human-Computer Interaction*, PhD thesis, UIUC, (2000).

[50] M. Schröder, D. Heylen and I. Poggi, "Perception of non-verbal emotional listener feedback", *Proceedings of Speech Prosody 2006*, Dresden, Germany, pp 43-46 (2006).

[51] M.J. Lyons, M. Haehnel and N. Tetsutani, "Designing, playing, and performing, with a vision-based mouth interface", *Proceedings of the 2003 Conference on New Interfaces for Nusical Expression*, Montreal, pp 116-121 (2003).

[52] D. Göger, K. Weiss, C. Burghart and H. Wörn, "Sensitive skin for a humanoid robot", *Human-Centered Robotic Systems (HCRS'06)*, Munich, (2006).

[53] O. Khatib, O. Brock, K.S. Chang, D. Ruspini, L. Sentis and S. Viji, "Human-centered robotics and interactive haptic simulation", *International Journal of Robotics Research*, 23(2), pp 167-178 (2004).

[54] C. Burghart, O. Schorr, S. Yigit, N. Hata, K. Chinzei, A. Timofeev, R. Kikinis, H. Wörn and U. Rembold, "A multi-agent system architecture for man-machine interaction in computer aided surgery", *Proceedings of the 16th IAR Annual Meeting*, Strasburg, pp 117-123 (2001).

[55] A. Legin, A. Rudnitskaya, B. Seleznev and Yu. Vlasov, "Electronic tongue for quality assessment of ethanol, vodka and eau-de-vie", *Analytica Chimica Acta*, 534, pp 129-135 (2005).

[56] S. Oviatt, "Multimodal interfaces", in J.A. Jacko and A. Sears (eds), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Application*, Lawrence Erlbaum Associates, Mahwah (2003).

[57] R.A. Bolt, "Put-that-there: voice and gesture at the graphics interface", *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, Seattle, Washington, United States, pp 262-270 (1980).

[58] M. Johnston and S. Bangalore, "MATCHKiosk: a multimodal interactive city guide", *Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions*, Barcelona, Spain, Article No. 33, (2004).

[59] I. McCowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard and D. Zhang, "Automatic analysis of multimodal group actions in meetings", *IEEE Transactions on PAMI*, 27(3), pp 305-317 (2005).

[60] S. Meyer and A. Rakotonirainy, "A Survey of research on context-aware homes", *Australasian Information Security Workshop Conference on ACSW Frontiers*, pp 159-168 (2003).

[61] P. Smith, M. Shah and N.D.V. Lobo, "Determining driver visual attention with one camera", *IEEE Transactions on Intelligent Transportation Systems*, 4(4), pp 205-218 (2003).

[62] K. Salen and E. Zimmerman, *Rules of Play: Game Design Fundamentals*, MIT Press, Cambridge (2003).

[63] Y. Arafa and A. Mamdani, "Building multi-modal personal sales agents as interfaces to E-commerce applications", *Proceedings of the 6th International Computer Science Conference on Active Media Technology*, pp 113-133 (2001).

[64] Y. Kuno, N. Shimada and Y. Shirai, "Look where you're going: a robotic wheelchair based on the integration of human and environmental observations", *IEEE Robotics and Automation*, 10(1), pp 26-34 (2003).

[65] A. Ronzhin and A. Karpov, "Assistive multimodal system based on speech recognition and head tracking", *Proceedings of 13th European Signal Processing Conference*, Antalya (2005).

[66] M. Pantic, A. Pentland, A. Nijholt and T. Huang, "Human computing and machine understanding of human behavior: a survey" *Proceedings of the 8th International Conference on Multimodal Interfaces*, Banff, Alberta, Canada, pp 239-248 (2006).

[67] A. Kapoor, W. Burleson and R.W. Picard, "Automatic prediction of frustration", *International Journal of Human-Computer Studies*, 65, pp 724-736 (2007).

[68] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures", *Journal of Network and Computer Applications*, 30, pp 1334-1345 (2007).

[69] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.M. Lee, A. Kazemzadeh, S. Lee, U. Neumann and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information", *Proceedings of the 6th International Conference on Multimodal Interfaces*, State College, PA, USA, pp 205-211 (2004).

[70] M. Johnston, P.R. Cohen, D. McGee, S.L. Oviatt, J.A. Pittman and I. Smith, "Unification-based multimodal integration", *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics*, pp 281-288 (1997).

[71] D. Perzanowski, A. Schultz, W. Adams, E. Marsh and M. Bugajska, "Building a multimodal human-robot interface", *Intelligent Systems, IEEE*, 16, pp 16-21 (2001).

[72] H. Holzapfel, K. Nickel and R. Stiefelhagen, "Implementation and evaluation of a constraint-based multimodal fusion system for speech and 3D pointing gestures", *Proceedings of the 6th International Conference on Multimodal Interfaces*, pp 175-182 (2004).

[73] Brown University, Biology and Medicine, “Robotic Surgery: Neuro-Surgery”, http://biomed.brown.edu/Courses/BI108/BI108_2005_Groups/04/neurology.html, visited on 15/10/2007.