Natural User Interfaces

Dilman Salih, 1426091
School of Computer Science, University of Birmingham
Research Topics in HCI
Second Semester

Abstract—Computing devices have been changed and developed continuously since the early age of computing, and also many new kind of technology have been invented during that time. But user Interfaces, on the other hand, have not changed a lot since the graphical user interfaces created in 1980s. Most of the todays devices use GUIs and they rely mainly on pointing and keyboard devices for interaction purpose, which are not more natural for users. The future technologies should use an interface that not require long periods of time for learning and adaptation. Recently, a new form of interaction is provided, which uses human innate features, such as touch, speech, and gestures, and it is called Natural User Interface (NUI). This interfaces is proposed to be the next generation of user interfaces and improve user experiences. In this project, a short history of user interfaces and then natural user interface will be discussed. Finally, a vision based NUIs and the main issues in this approach will be explained.

Keywords—User Interface, User Experience, Command line, Graphical User Interface, HCI, Computer Vision, Voice, Touch, Gesture.

I. INTRODUCTION

COMPUTER and the way users interact with them have always evolved since the first computer was invented. The development of the computer hardware and software has made the ability to use computers in almost every area in our modern life, including education, business and entertainment. These developments have mostly been supported and driven by Moores law, which says that the number of transistors that can be placed inexpensively on an integrated circuit will double approximately every two years. This means that the future technology will be faster and smaller in size. It has been nearly half century since this law claimed by Gordon Moore, one of the founders of Intel. In spite of the fact that this increasing of transistors may not continue forever and the limits to this law may be reached over the next decade or two [1], but probably it has been an accurate and reliable calculation or prediction about how faster, better and cheaper high-technology products will be in the future [2]. While the computing power has been increased more continuously, the interface between computers and human has not evolved continuously [3]. In other words, Moores law is not true for user interface and user interfaces fundamentally have not changed a lot for nearly three decades [4].

As computing machines has became faster and smaller in size, new form of technologies or platforms have been invented, including smart phones, digital cameras, computer games and the Internet [3]. Inventing all these new kind of technologies, made the necessarily of having new ways of interaction between machines and users. Consequently, many new forms of interaction have been designed and some of them have been used widely while some of them have been used just in a specific area or for specific applications.

In this report, the most prominent forms of interaction and their main characteristics will be shortly explained in the next section in order to get the idea why new form of interaction is required for todays technology. The rest of the report will be about natural user interface and how it is differ from other interfaces, then the main issues for this kind of interface will be discussed .Finally, the possibility of using computer vision applications for creating natural interfaces will be described.

II. USER INTERFACE

User Interface (UI) is the environment that is provided for users to use when they want to communicate with an application or a machine. The environment will be anything including display screen, keyboard, mouse, light pen, the appearance of a desktop, illuminated characters, help messages, and how an application program or a Web site invites interaction and responds to it [5]. The communication will either be sending data (input) or be receiving information (output). Because users use UIs to deal with a product (machines or applications), these UIs must be designed in a way that help users to use the product in an efficient way. Moreover, they must make users feel happy when they use the UI. In early stages of UI, interfaces were very simple in design and they were not so user friendly, but as time drifted away they became more complex in design and became more user friendly.

A. The Development of User Interfaces

As mentioned before, the earliest user interfaces were simply designed and they were not easy to use. Also, developing hardware caused to invent new kind of technologies that required designing new form of interaction. Therefore, there has been a growing interest in designing and creating new kind of user interfaces and many movements and developments in user interfaces have taken place. These developments can be divided into many different steps. According to [3], these steps are loosely defined and they can be divided into the phase of commands (the command line), followed by the graphical user interface (GUI). It also mentioned that some other user interfaces were existing during that time but because they designed and used only for some specific applications, they did not get public attention. Menu-only
system can be seen as one example that was designed for ATMs or televisions.

Because command line interfaces (CLIs) allow users to interact with machines or computers by typing some predefined keywords into a screen, it can be completely overwhelming especially for novice users [6]. It might be mainly because in this kind of interfaces users have to learn and memorise a lot of commands and they also need to memorise the order or syntax of the commands. This means any misspelling of the commands or any changes in their order will cause to do not execute them by the system. The only advantage that the CLI might have is that it does not need too much memory (RAM) or processor (CPU) compared to other interfaces, for example GUI. However, todays devices have a powerful hardware and using memory and processor from interfaces will not be a big problem.

In the earliest of 1980s, the graphical user interface (GUI) and its associated desktop metaphor arrived [7]. GUI, which is still in use, provides a direct and predictability way of interaction by using a set of user interface elements, commonly referred as WIMP (Windows, Icons, Menus and Pointers) [3] [4] as a mechanism of input and output. WIMP paradigm has provided a stable and global face to computing, and also provided a direct feedback of what users do by using pictures, graphics, icons, pull down menus, dialog boxes and among others. Another successful concept in GUI, which is used in document processing, is WYSIWYG (What You See Is What You Get). In the early days of document processors, you typed your essay or letter on the screen, but it could look completely different on the printer. A GUI normally tries to ensure that whatever you create on the screen will be very similar to what appears on the printer or World Wide Web [8]. These properties provide the user a clear model of what commands and action are possible and what their affects will be; they allow users to have a sense of accomplishment and responsibility about their interactions with computer applications [4]. Compared to the CLI, the WIMP GUI represented a lower barrier for user because recognizing and choosing are easier than remembering and then typing [3].

In spite of the fact that GUI has been very successful and dominated both the marketplace and HCI research for two decades, but they are not the most natural one, because users must learn how to work with it and, there are many people who feel disoriented by their first encounter with a mouse [22]. Also, this paradigm might not match the form of futures computing devices [9]. There have been always attempts to predict the future of user interfaces and many applications or products with new paradigm of interfaces have been build, for example, Nielsen (1993) in [9] has argued some features of the next generation interfaces, Bolt (1980) announced Put-That-There project, which uses voice command and gesture recognition together to move an object on a wall-sized screen and RAND tablet (1963) , which was funded by RAND Corporation as part of an ARPA-sponsored project, can be seen as the first pen-driven computer interface that allowed engineers to draw flowcharts using a combination of drawing, text recognition and pen gestures [9]. However, they did not become very popular because at that times hardwires that were available for public use, were not so powerful. Therefore, these products just tested in a specific situation or in the lab environment. Figure 1 shows the evolution of user interface paradigm.

<table>
<thead>
<tr>
<th>When</th>
<th>Implementation</th>
<th>Paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960s</td>
<td>Switches, punch cards, lights</td>
<td>None</td>
</tr>
<tr>
<td>1970s</td>
<td>Command-line interface</td>
<td>Typewriter</td>
</tr>
<tr>
<td>1980s</td>
<td>WIMP-based graphical user interface</td>
<td>Desktop</td>
</tr>
<tr>
<td>2000s</td>
<td>Perceptual interfaces</td>
<td>Natural interaction</td>
</tr>
</tbody>
</table>

Fig. 1. The Evolution of User Interface. Taken from [7]

In spite of the fact that GUI has been very successful and dominated both the marketplace and HCI research for two decades, but they are not the most natural one, because users must learn how to work with it and, there are many people who feel disoriented by their first encounter with a mouse [22]. Also, this paradigm might not match the form of futures computing devices [9]. As the computing devices change in term of physical size and become faster, smaller and ubiquitous, users need more general and intuitive ways of interaction. Pointing, clicking, and typing would still be an appropriate way of interaction for many applications in the future, but this will not be the way that most people will use for interacting with computing devices in the foreseeable future. This means new paradigm of interaction will be needed in order to interact with future technologies.

B. Natural User Interface

As mentioned before, many kind of user interface have been designed and some of them, such as GUI, became widely used because they reduced the barrier between users and computers. Also, it is mentioned that the development of user interfaces compared to computing hardware was quite slow. The questions are: Is it time to design and use new paradigm of user interface, what the next generation of user interface will be, will it replace GUI and other kink of interfaces that we are using now.In 2010 Bill Buxton, who is currently a principal researcher at Microsoft Research, in an interview about NUI [10] said that the biggest change of technology is not just the hardware and software, of course they are becoming faster, cheaper and smaller, but the biggest thing that is really changed and it is changing rapidly now is who is doing what, where, with whom, how and when. He tried to say that computing devices have been changed a lot and they are becoming more and more ubiquitous. But UIs, on the other hand, have not changed a lot since 1980s. We need to create an UI that takes a full advantage of users bandwidth and also takes a full advantage of modern input/output technologies, in order to make devices more
natural or make users feel fully integrated to the device that they use.

Now, we are witnessing the beginning of coming new kind of user interfaces, which seems to be the way that people will use to interact with computing devices in the future. It is called Natural User Interface or NUI. In 2010, Steve Ballmer, CEO Microsoft, said [11].

\[ I \text{ believe we will look back on 2010 as the year we expanded beyond the mouse and keyboard and started incorporating more natural forms of interaction such as touch, speech, gestures, handwriting, and vision--what computer scientists call the "NUI" or Natural User Interface. Steve Ballmer, 2010}\]

The effort of creating a friendly user interface under the name natural started a quarter of a century ago, however, it is still difficult to provide a complete set of rules in order to be used for creating a successful NUI, because NUI is not yet exactly defined and it is not yet at a state of evolution or standardization, but like GUIs, NUIs promise to reduce the barrier between users and computing devices and at the same time improve user experience [12] [3].

So, what are NUIs? In [13] NUI is defined as a kind of interfaces that enables users to interact with computing devices in the same way they interact with the physical world, through using voice, hands, and body movements. Unlike GUI, which uses mouse and keyboard, NUI allows users to use their natural behaviour for interaction purpose, for example they can speak with the machines, stroke their surface or gesture at them in the air in order to perform a specific task. One of the most common examples of a natural user interface is a touch screen interface, which allows users to move and manipulate objects by tapping and dragging their fingers on the screen. The digital objects on the screen respond to the movement of their fingers, much like physical objects in the real world. This direct feedback provided by a touch screen interface makes it seem more natural than using a keyboard and mouse to interact with the objects on the screen [14].

Voice, touch and gesture are the main components of NUIs; however, just using these components for creating an interface does not necessarily make the interface natural [10] [3]. In other words, NUIs have a set of strengths based on what they make easier, how they make those things easier, how they shape the users interaction with technology, which niches they fit in, and whether or not these niches expand to dwarf the space occupied by traditional GUIs [3].

Everything is best for something and worst for something else. Bill Buxton, 2007

Voice, for example, may be work very well or may be natural in a certain situation, but it may be not appropriate or natural in another situation. To explain this further, lets use an example, assume you want to write and sent a text message. GUIs probably are more natural when you are at a library or an office, but when you are driving speech or sound recognition system will be most natural. Again, if you are at a noisy crowded place or in some places that privacy is important speech recognition will not be natural [10] [15]. In fact, the word natural in NUI is not about interfaces at all, it is about the way that users interact with the product, what they do and how they behave or feel when they use the interface [10] [15] [3]. This means, for creating a more natural relationship between user and technology is not merely a matter of simply removing mice, keyboards, buttons, and knobs, or adding new input methods such as speech, touch, and in-air gestures [15]. In short, natural refers to users behaviour and feeling during experience [3].

The trick of elegant design is making sure you do the right thing the right way for the right person at the right time. What's right here may be wrong there.

Bill Buxton

III. A REAL WORLD EXAMPLE

The following story is taken from [3] : In 1989, Apple Computer started to create a new kind of plat form (Newton Message Pad), which supposed to recognize regular hand writing. Unfortunately, the result was not satisfactory. The product was fraught with problems, among them that that the hand writing recognition was insufficient robust. Later in 1997, Jeff Hawkins and his team at US Robotics made Palm Pilot, which had also a hand written recognition as entry method. Unlike the Newton, the Palm Pilot became very popular and several versions were produced and millions were sold. The key difference was that Hawkins and his team recognized the limitation of their technology. Instead of trying to build a robust recognizer for regular hand writing, they developed a special input language, known as Graffiti variation of the Unistroke technique invented by David Golberg of Xerox Palo Alto Research Center, Figure 2. The Palm was no less natural than the Newton, but because Graffiti, which was similar enough to standard Roman characters, provided a very good recognizer, it enabled an experience that could be felt to be natural.

Why this example? In [3] two important things, based on this story, are mentioned:

- In spite of the fact that the Pilot shifted the burden of learning from the device to the user, it does not mean that the engineers or designers are free to ignore the user needs or prior knowledge. It means that just mimicry of an existing experience or phenomenon will not make the interface to be a natural user interface.
- Also it tells us that NUIs need learning.

IV. MULTIMODALITY VERSUS UNIMODALITY

As mentioned before, todays computing devices are so powerful and their advanced input/output features enable as
to create more natural UIs. For creating an NUI, many input and output techniques may be used together at the same time (Multimodal Interface), or may be just one technique used (Unimodal Interface). The term multimodal has been used in many contexts and across several disciplines. For our interests, a multimodal HCI system is simply one that responds to inputs in more than one modality or communication channel such as speech, gesture, writing, and others [16]. More information about Multimodal Interfaces can be found in [16] [17] [18].

V. REAL APPLICATIONS

In November 2006, Nintendo released its fifth home video game console, the Nintendo Wii [19]. Nintendo Wii has a handheld remote controller or Wiimote that uses combination of motion sensing technology as well as pointer system to create a good level of interactivity in games. This was an innovative step in the interaction style of playing games. Players can use and move Wiimote in the air to control objects in the games. A few Year later in November 2010, Microsoft released Kinect device, which is a motion sensing Universal Serial Bus (USB) input device that provide a full natural from of interaction by tracking body movements and by using gesture and speech recognition technique [20]. In spite of the fact that the original intend of these devices were home entertainment and video game controller, but they have many advanced features, especially Kinect, that can be used to design and create natural user interfaces. For example, in [20], Kinoogle project is introduced, which is a natural interface that uses Microsoft Kinect device to interact and control Google Earth using a series of hand and full-body gestures. In other words, these motion-based interfaces are considered natural user interfaces because they respond to users natural motions [14].

Microsoft has released, and is continually developing, a number of products that incorporate touch, gestures, speech, and more to make user-computer interaction more natural

more like the way humans interact with each other. For example Microsoft Surface, which is a table that acts as a collaborative massive multi-touch-screen computer, or a voice-enabled Windows phone device, or a Windows 7 laptop that lets users navigate files or the Web using their fingers or a pen tool, or an in-car communications and entertainment system, Microsoft Autos Ford SYNC, that responds to a drivers voice commands, playing favorite songs or answering text messages so drivers can keep their eyes on the road, hands on the wheel, and mind on their driving [15].

VI. NUIs CONSIDERATIONS AND ISSUES

In [19], Bill Buxton explained an example; he said if I want to show you a document in my mobile phone, instead of showing it in my mobile screen, which is very small, why not just tap the mobile screen or make a gesture to send the document from my mobile to that screen, pointed to a flat screen behind the interviewer, and then control it from my mobile. He said, even for this simple example, we have to consider about many things. For example, the screen and mobile should have a compatible hardware infrastructure in order to make a connection between them (network issue). Also, we need a secure network between those two devices because I do not want my document appear on the other screens or I do not want other people see my document (security issue). In fact, if we talk about another example or scenario, we might face other problems and, in general, for designing NUIs, like other UIs, many factors or characteristics need to be considered, including functionality, usability, cost, and so on. But the main issues in NUIs that mentioned in the literature, for example [21] [22], are the artificiality of NUIs and lack of feedback.

As mentioned in the previous example, Palm Pilot, NUIs need learning. For example, in order to use a multi-touch screen device, as an example of NUI, users need to learn a set of predefined gesture commands. Also, different companies may have different guideline for using gestures and two different companies may use the same gesture for different tasks. Furthermore, gestures depend on the cultures or societies, for example may be a gesture gives a happiness meaning in one culture but it gives totally opposite meaning in another culture or society. However, these problems can be solved by using those gestures that are mostly used in real world life and they are familiar for users, or by standardizing gestures, or by participating non-technical user in the design of gestural languages. These reduce learning time, but users still need to learn and use artificial gesture vocabularies [22].

Natural User Interfaces Are Not Natural. Don Norman.

The main aim of natural interfaces should be to break down the technology driven approach to interaction. Alessio Malizia and Andrea Bellucci.

About feedback problem, Norman in [21] proposed:
Because gestures are ephemeral, they do not leave behind any record of their path, which means that if one makes a gesture and either gets no response or the wrong response, there is little information available to help understand why. The requisite feedback is lacking. Moreover, a pure gestural system makes it difficult to discover the set of possibilities and the precise dynamics of execution.

He also proposed a solution and said this problem can be fixed by adding conventional interface elements, such as menus, help systems, traces, tutorials, undo operations, and other forms of feedback and guides.

VII. VISION BASED NUIs

Computer vision can be defined as changing data, which is transmitted from a camera(s), into either a decision or a new representation. The decision may be there is a motion or not (motion detection) or there is a face or not (face detection) or identify the detected face (face recognition) in the video and the representation may be removing the motion in the video [23]. This means, the input is an image or a sequence of images (video) and the output is the decision based on understanding those input images. Vision has the potential of carrying a wealth of information in a non-intrusive manner and at a low cost; therefore it constitutes a very attractive sensing modality for developing perceptive user interfaces [24].

The vision based HCI 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 [25].

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

They all can be used to create an intuitive interface, but they should be used in the right context, for example for disable people with limited body movements, probably Eyes Movement Tracking is the best choice. In this section hand gesture recognition will be taken as an example of vision based NUIs. Also, the main issues for vision based NUIs will be discussed in the next section.

Utilizing hand gestures to interact with computers is an extremely characteristic type of interaction and it can be used like a methodology in Human Computer Interaction in order to provide a Natural User Interface. The first time that gesture-based interaction is used as a new form of Human Computer interaction, was in the middle of 1970s by M.W. Krueger [26], and it has been studied and developed over the past 30 years [27]. In [24], several existing methods for vision based hand gestures is explained and three steps for creating a complete hand interactive system is proposed.

- Detection: The primary step in gesture recognition systems is the detection of hands and the segmentation of the corresponding image regions [2]. This step is responsible for defining and extracting the region of interest, hands.
- Tracking: The purpose of this step is to keep detecting the movement of users hand, or to tell the system what is where.
- Recognition: This step is responsible for analyzing the data that is extracted in the previous steps and assigning the resulting groups with labels associated to particular classes of gestures.

VIII. VISION CHALLENGES

Vision based approach provide an inexpensive user friendly interface, in the most cases just a camera required to create a vision based NUI. However, some big issues arise when this approach is used to create NUIs. In [28] some issues is presented, including:

- Robustness: Most vision technologies have the problem of the robustness. For example, small changes in lighting or camera position may cause fail or may give a wrong result. Some technique can be used to increase the robustness, for example wearing coloured gloves, but this increase the obstacle to the user.
- Speed: Because in vision there is too much video data coming from a camera(s) in real time, sometimes the application require time to respond, which decrease the interactivity of the application.
- Usability: This issue related to NUIs in general. As mentioned before, NUIs need learn. Novice users often have difficulty when they use the system, for example in hand gesture recognition system they may use a gesture that is not defined in the system. Moreover, when they use undefined gesture the system should provide some feedback to help the user understand why the system does not reply.

IX. CONCLUSION

With the development of computer hardware and software and the advent of more powerful computing facilities, the more Natural User Interfaces have gained growing attention. Voice, touch and gestures are the main components for creating NUIs. The important thing is NUIs need learning and the term natural in NUIs is not about interfaces, it is about user experience. Vision based approach can be used to provide a more natural form of interaction with the lowest price but there are some limitations and issues. A primary issue hampering this approach is that it depends on several underlying assumption including assuming high contrast stationary backgrounds and ambient lighting conditions. In this project, the development of user interfaces and the main issues for creating natural user interfaces are shortly explained, then vision based approach for creating NUIs is explained.
• The IEEEtran.cls \LaTeX Class File is used to write this paper, for further information please visit: [http://www.ieee.org/publications_standards/publications/authors/author_templates.html](http://www.ieee.org/publications_standards/publications/authors/author_templates.html)

**References**


