Analysing Natural User Interfaces: Literature review

Kapilnath Ranganathan
Department of Electrical and
Computer Engineering
The University of Auckland
New Zealand
kran034@aucklanduni.ac.nz

ABSTRACT

In this paper, we discuss how Natural User Interfaces (NUI) can help bridge the gap between real and digital world. NUIs are based on human touch, vision, voice, motion and doesn't involve in using input devices. The main problem with existing user interfaces is the lack of power to express user intentions. This paper mainly discusses various NUIs developed by Human Computer Interaction (HCI) researchers. Most of us have heard/used Microsoft Kinect in which users are the controller. Kinect is a good example of NUI as the system in intuitive and there are no additional input devices involved. Similarly there are several NUI systems developed through researches.

We discuss how NUIs can be used in Geographic Information System (such as Google Earth, Bing map), exhibition/museum exhibit learning, interacting with mobile phones, instructing robots and helping elderly/disabled people. These discussions are based on the researches done in the NUI area. According to the research experimental results, it is concluded that the NUIs are intuitive (quick user transition from novice to expert) and therefore has significant advantage over current user interfaces (such as GUIs used in computer systems).

Author Keywords

Natural User Interface; Natural Interaction; Robot Interface; Kinect; Human Computer Interaction; Future User Interface;

ACM Classification Keywords

HCI, NUI; GUI; UI;

INTRODUCTION

This literature review discusses about Natural User Interfaces (NUI) focusing on the research done in the area, problems with existing user interfaces, and comparing NUI with other user interfaces.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

An important aspect of developing a human-machine system is the definition and construction of user interface[1]. A common use of user interfaces is to visualize information (InfoVis). In today's world, user interface and interaction technologies are advancing rapidly. A good example would be the touch screen phones which are widely adopted for its ease of use; this shows people's affection on mouse/keyboard less interfaces. Commonly, user interaction can be described as a starting and ending points which can be further divided into: user's intent to do something, user's action and the device's response.

The first type of user interface that we saw at earlier stages was the "Batch Interface" in which punch cards were used as the input device and a line printer as the output device. An important part of user interface is the language used by the user and the machine for interaction. The real beginning of human-computer interaction started with "Command Line Interface" (CLI) that used a common language, but only very few humans can understand. An improved user interface called "Graphical User Interface" (GUI) emerged after the development of ultra large scale integrated circuits, high-resolution displays and the mouse. When comparing GUI to CLI, GUI has visual data representation that command line didn't offer. Although GUIs are easier to use, users still need to learn how to use them. This is where natural user interface outshines as its main focus is on natural interaction that commonly doesn't require any training and can change a beginner to expert in a shorter time period.

Recent technologies such as the Microsoft Kinect, capacitive screen supporting pen and touch interaction, speak recognition are good examples for natural interaction. Advanced input technologies such tangible objects [2], pen based tablets, and large multi-touch surfaces are also being developed. The main aim of these recent technologies is to help reduce/remove the gap between real and digital world[3] and provides high freedom of user expression.

Natural User Interface

The term "Natural User Interface" (NUI) is one of the many different ways to interact with computers. NUI uses human abilities such as touch, vision, voice, motion and higher cognitive functions such as expression, perception and recall[4]. For instance, NUI can allow people to interact with digital objects in physical world with similar gestures that they would use in real world. NUI's intuitive and natural approach makes it the most effective and easiest way to interact with computers. The characteristics of NUI can vary based on the type of interaction, but are usually categorized into seven interfaces as follows: user-centered, multi-channel, inexact, high bandwidth, voice based interaction, image based interaction and behaviour-based interaction[4].

PROBLEM

The major problem with current mainstream user interfaces (GUI) is the cognitive processing (learning process, memory… etc.) involved. Although GUIs are a significant improvement over CLI, it lacks the power to express user intentions (i.e. users need to learn to use). An optimal interaction is only possible if the user interface closely matches the user's characteristics and takes users' limitations into account[1].

No matter how easy-to-learn and easy-to-use; an interface always exists between user's intention and execution of their intent [5]. These additional interface layers introduce cognitive processing for people. These problems are commonly identified as the gap between human and machine.

APPROACHES

NUI is one of the effective ways to solve the cognitive problem in current user interfaces. Natural Interfaces have started to slowly take over current user interfaces. A good example would be the "Microsoft Kinect" [6, 7] which indicates the beginning of the end for the mouse[7]. Figure 1 shows the Kinect device which was mainly developed for Xbox360 (a gaming console) in order to provide users with new ways to interact and play their games. Kinect is a motion sensing input device (connected via Universal Serial Bus interface) that enables users to control and naturally interact with games and applications without having to use any traditional input devices. Microsoft Kinect uses an RGB (Red Green Blue) colour camera, an accelerometer [6], a depth sensor which uses infrared projector (Figure 2) and a camera, and a noise cancelling, multi-array microphone to detect a person's location in three dimensional (3D) space.

Another similar approach to mouse and keyboard is the Nintendo's Wii [6, 8], a console with controller with motion sensing (using inertial sensors) technology. Although these products are related to gaming, several researches are also being done using them to come up with NUI. A group of research students from California Institute of Technology did a study on motion technologies such as depth camera technology of Kinect, inertial sensors of Wii and electromagnetic sensing technology developed by "Sixense Entertainment" and "Razer"[7]. "Peter Ngo" (one of the research student) [7] believes the idea of NUI as a fully tool less interface is exaggerated. "Amir Rubin" (CEO of "Sixense") says "People don't buy motion control, they buy the experience being delivered to them" [7]. There are several researches currently being done in this area to come up with different types of NUI technologies. Some of the NUI researches and their interesting user evaluations are discussed in this paper (described in the following sections).

Figure 1 Microsoft Kinect

Figure 2: Kinect's infrared structured light[7]

NUI in navigation systems

"Kinoogle"[6] is a NUI project developed by students at Texas A&M University. Kinoogle uses Kinect to provide a NUI for Google Earth navigation. Kinoogle and the user control were provided through a series of hand and full-body gestures (detected via Kinect). Kinoogle's main functionality is interpreting point data received from Kinect to detect specific gestures/poses and use it for Google earth navigation. Based on the poses (detected using hand gestures), map commands (as shown in Figure 3) such as panning, zooming, rotation and tilt will be activated. Kinoogle uses skeleton data to interact with Google street view by the means of intuitive "walking" and "turning" gestures. Mouse and keyboard simulating third party software were also used to map hand gesture data in order to control the earth model. Kinoogle also provides a GUI for feature selection and user interaction data representation. Before using the system, users have to go through a calibration process in order to ensure their body parts are properly detected.

Results show that NUI for Geographic Information System (GIS) are highly welcomed since the users don't need to touch any input devices. Kinoogle can be improved further by if it can use speech recognition functionality offered by Kinect as an alternative to gestures. Results also suggest that NUIs(hands-free gesture and speech recognition) are not a full replacement for more conventional interaction methods[6]. Nevertheless, Kinoogle creators believe NUI affordances can improve user experience significantly.

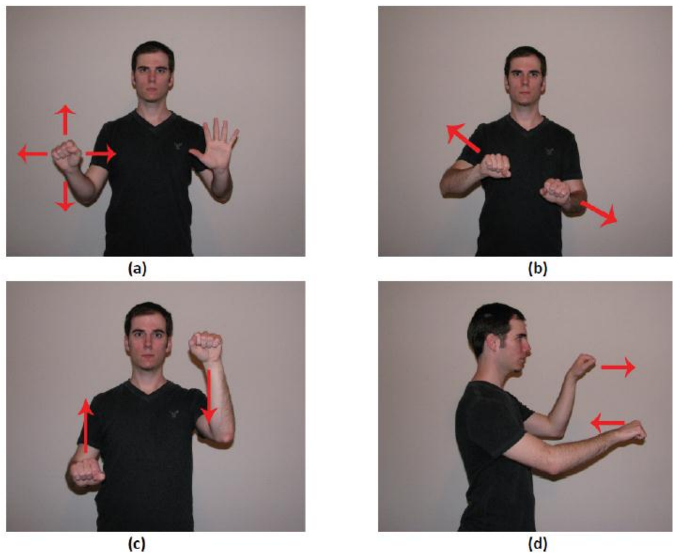

Figure 3 : Gesture Movement used: (a) pan (b) zoom (c) rotate (d) tilt[6]

NUI to enhance visitors' experiences

We know museums and exhibitions is a collection of objects, standing deaf in front of visitors. The Media Integration and Communication centre at the University of Florence have worked on two NUI that can make museums and exhibitions interesting. The first NUI was developed for the museum of "Palazzo Medici Riccardi" [9]. (one of the most famous museum in Florence) and the second NUI was developed for the New City Exhibition of "Palazzo Vecchio"[9]. The project was developed aiming for natural interaction between people and physical object (as in everyday life), seamless computer system, cognitive aspects and aesthetics. The main use of these natural interfaces is to teach visitors about the famous frescos' contents and exhibits.

Figure 4: Learning about a Fresco[9]

The system's NUI is achieved by using two cameras which grabs the points a person is pointing to and an algorithm which calculates and selects the part pointed by the person. The system automatically highlights the important parts of a fresco to make it easier for the visitor/user to identify important parts and get explanations. Figure 4 shows a user pointing at one of the character in a fresco; the system automatically magnifies it and displays information about it. The images used in the system are digitalized photos and are presented in a continuous horizontal image stripes form. The display that projects the frescos is a large, rear-projected display approximately 2.5m[9] wide and only one person can interact with it at a time.

According to the system usage records, visitors were eager to use the system and commented that the whole system is intuitive and easier to use. The system is currently being improved to allow multiple user interact with the system simultaneously[9].

NUI in mobile phones

Mobile technologies are rapidly improving over the past years. Modern smart phones have multi core CPUs and GPUs; enabling developers/researchers to come up with complex applications that can make full use of the new hardware features.

Four Korean University researchers[10] have developed a natural interface model that can allow people to express requests without needing any prior mobile knowledge. Commonly people use mobile phones to call, read/send text messages. The NUI is specifically designed to carry out such common tasks that a user might do with their mobile phone. Figure 5 shows how the system may react to the user's requests. For instance, Figure 5's (a) section shows that the user wants to call Ashley's office number and the system reacts to the request by making the call for the user as shown in Figure 5's (a') section.

The system's input part (commands) is done via voice (speech recognition) or text. Most device control commands are derived from user request that is constructed by a functional and parameter part. For example, calling task's (Figure 5-a section) functional part is calling and the parameter part is the callee's name. The system recognizes user commands by function scoring (model generates possible functions using mathematical derivation process and ranks them by scores) [10, 11] and parameter tagging (using character based recognition). Researchers expressed that NUI command creation for the system is a difficult process since a command can be represented in different sentences.

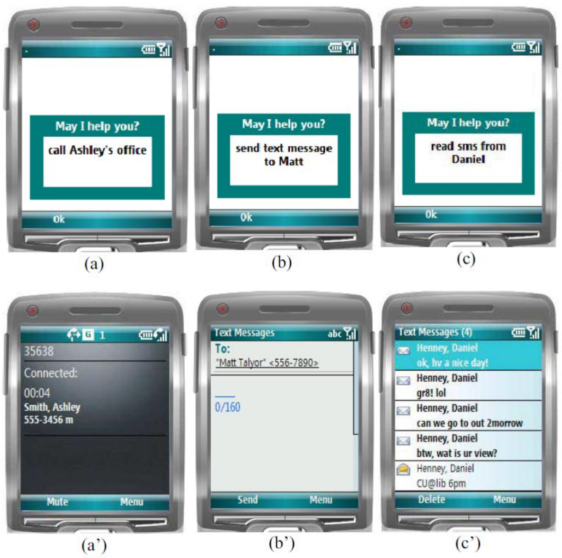

Figure 5: User Requests and response[10]

The proposed model was tested using 3,000 request sentences (a request is composed of 3.28 words and 20.27 characters)[10]. Experimental results show that the system was able to achieve an accuracy of 97.22% [10] for the proposed model. Most of the errors in the model were due to the parameter recognition phase. Although this was a model based research, it clearly shows the difficulties (recognizing the user parameters and the commands that can be given in different sentences) involved in creating a NUI.

NUI and Age groups

A user interface can be used by different age groups. This is an important aspect when developing a user interface, but often most user interface developers fail to consider particular age groups (usually the elderly) due to mythical assumptions (a famous one: most older age groups do not use computer systems). The real reason behind this assumption is the cognitive interactive tasks involved. However user interfaces that are natural (i.e. NUI) are usable by almost all age groups due to its real life like interactions. Another interesting fact about user age groups is that young children can learn to communicate using gestures before they learn to talk[8].

A research group at University of Dhaka (Bangladesh) have developed gesture controlled user interface that can improve the quality of life of older and disabled users along with general users (Figure 6) [8]. The research's main aim was to provide a Gesture Controlled User Interface (GCUI) that is suitable for elderly and disabled people. This is done by open gestures that use low cost hardware such as a webcam and an open source software technology. The open gesture application uses augment reality (AR) [3, 8] application which enables users to perform everyday tasks via television screen. The user interface will be displayed on a pre-configured television channel. The system will display user's image on the screen that is filmed using the web camera. Figure 6 shows the screen with a list of objects (text and icon), users point at objects using hand gestures to perform a selection. Researchers believe the system can be improved by having a multi-mode input (voice and gestures) and feedback (sound feedbacks).

Figure 6: Different Age group people using the Open Gesture application[8]

The developed system was tested using 70 participants (mixed age group 18-80+), and the participants had to fill in a survey at the end of the evaluation. According to the evaluation results, most participants rated the system as simple to use and easier to learn. When asked about elderly people using the system, majority expressed that the system will increase the independence of elderly users (specially disabled). Since the developed system's NUI was used by elderly (who have lower cognitive processing) and disabled people that are not familiar with user interfaces, this research is a good example showing how intuitive NUIs are.

NUI in Robots

In today's world, "Robot" has different meanings. The "Advanced Telecommunication Research Institute International" (ATR)[12] has separated robots into three categories: virtual robot (type of robot that is found in cell phones), unconscious type (works to collect information and recognize situations), and visible type (the ones that talk with people). In order for the robots to fit in human community, a natural robot user interface is essential. Tokyo University Researchers have developed a natural robot interface that is based on the gestures that we humans normally use in conversations.

The research mainly focussed on pointing gesture based robot commands. The robot system was developed in a way such that it is capable of predicting the user's intention (For example, Users may want the robot to pass/move some objects pointed by them or park the robot on the pointed location) by analysing the pointed object. A software called "iSpace" [12] was developed by the research team and it was used to track human motion on the robots. The robots were equipped with two cameras and the software can use the Red-Green-Blue (RGB) information (derived from cameras) to track objects. The software is also used to calculate three-dimensional (3D) coordinates of the user's head/hand positions and robot's position. The system recognizes Pointing gestures using "Fuzzy Associate Memory" (FAM) [12]. A virtual room interface was also developed to simulate the robot's NUI system in a virtual room.

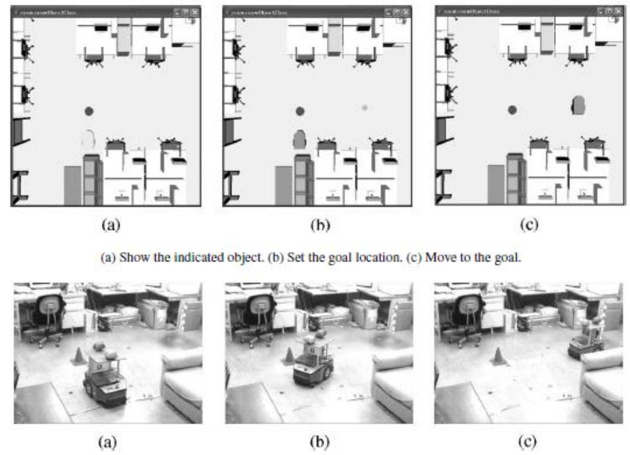

Figure 7: System reactions to virtual room interface[12]

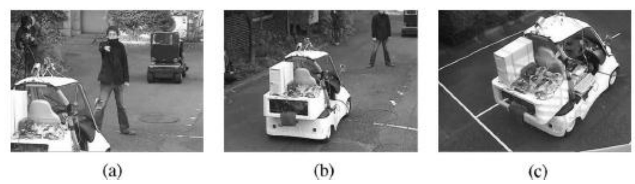

Figure 8: System parking itself on the location pointed by the user[12]

The system was evaluated by several participants. The system evaluation involved in finding position indicated by user (route instructions- moves to the positions indicated by the user), pointing an object and moving it to a location(using virtual room interface) and parking instructions. Figure 7 shows the virtual interface driven system detecting the object position pointed by the user; once detected, the system will wait for the user to confirm by displaying a dot on the detected position (user confirms by nodding/shaking their head); the system will then choose a behaviour based on the type of object detected. Figure 8 shows the system's parking movement directed by the pointing gestures. Evaluation results suggest that the system was able to perform the tasks with some position recognition errors (average maximum error of 22.9 cm and a maximum error of a 40.5cm) [12]. This research clearly takes a step forward into natural human-robot communication/interaction using NUI based object/position recognition (which can vary according to the context and user's intention).

SUMMARY

Based on the researches [6, 8-10, 12] and experimental results, it is clear that NUIs can improve Human Computer Interaction by reducing/removing the cognitive processing (or in other words the gap between real and digital world). Most experimental results suggest that NUIs are found to be intuitive by the participants (including elderly/disabled people).

The NUI growth doesn't imply NUIs can replace existing user interfaces. Some of the NUI researches discussed in this paper have recognizing (gestures, poses, voices … etc.) and implementation (complicated model as commands may vary for each individual) problems – meaning more work to be done. Nevertheless, these researches clearly take a step forward in bringing Natural User Interfaces to reality. It will take few more years for the NUI researchers to come up with perfect (close to 100% recognition accuracy) NUIs.

When comparing NUIs to existing user interfaces (such as GUI), NUIs are not only intuitive but also eliminates additional input devices and are capable of doing everything that existing user interfaces can do. There is no doubt in saying that NUIs will be the future of user interfaces (that can replace most user interfaces provided that new technologies will emerge).

FUTURE WORK

Natural interface's revolution has already begun to replace some of our current user interfaces (Ex: Kinect in gaming industry replacing game controllers). Even though there are successful NUIs, researches are still required to develop new NUI technologies and identify the areas of information technology that are suitable to use NUI. Since current NUI technologies cannot replace everything that our current user interfaces offer, it is not surprising to see that most computers are still using the GUIs. Similar to how key pad mobile phones got replaced by touch screen mobile phones, user interface transformation (to use NUIs) will happen once appropriate NUI technologies (meaning more researches in this area is required) are developed.

In today's world, people have started using natural interactions such as gestures and voice recognition on their mobile phones. In addition, current researches have brought NUIs closer to reality and the work done will help develop new devices/technologies (similar to Kinect) that can understand human gestures and voices. Other than that, NUIs will be used in robots to naturally (similar to communication between humans) interact. NUI is certainly going to be the future of user interfaces (doesn't necessarily mean NUIs can/will replace all existing user interfaces).

REFERENCES

1. Van Der Veer, G.C., *Individual differences and the user interface.* Ergonomics, 1989. **32**(11): p. 1431-1449.

URL: http://www.tandfonline.com.ezproxy.auckland.ac.nz/doi/abs/10.1080/00140138908966916

2. Lee, J., et al., *Tangible user interface of digital products in multi-displays.* The International Journal of Advanced Manufacturing Technology, 2012. **59**(9-12): p. 1245-1259.

URL: http://link.springer.com.ezproxy.auckland.ac.nz/article/10.1007/s00170-011-3575-0

3. Petersen, N. and D. Stricker. *Continuous natural user interface: Reducing the gap between real and digital world*. in *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*. 2009.

URL: http://ieeexplore.ieee.org.ezproxy.auckland.ac.nz/xpl/articleDetails.jsp?reload=true&arnumber=5336502

4. Weiyuan, L. *Natural user interface- next mainstream product user interface*. in *Computer-Aided Industrial Design & Conceptual Design (CAIDCD), 2010 IEEE 11th International Conference on*. 2010.

URL: http://ieeexplore.ieee.org.ezproxy.auckland.ac.nz/xpl/articleDetails.jsp?arnumber=5681374

5. Bongshin, L., et al., *Beyond Mouse and Keyboard: Expanding Design Considerations for Information Visualization Interactions.* Visualization and Computer Graphics, IEEE Transactions on, 2012. **18**(12): p. 2689-2698

URL: http://ieeexplore.ieee.org.ezproxy.auckland.ac.nz/xpl/articleDetails.jsp?arnumber=6327275

6. Blanchard, B.J., et al., *Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation.* International Journal of Health Geographics, 2011. **10**: p. 45.

URL: http://go.galegroup.com.ezproxy.auckland.ac.nz/ps/i.do?action=interpret&id=GALE%7CA265608766&v=2.1&u=learn&it=r&p=AONE&sw=w&authCount=1

7. Goth, G., *Brave NUI world.* Commun. ACM, 2011. **54**(12): p. 14-16.

URL: http://dl.acm.org.ezproxy.auckland.ac.nz/citation.cfm?doid=2043174.2043181

8. Bhuiyan, M. and R. Picking, *A gesture controlled user interface for inclusive design and evaluative study of its usability.* Journal of Software Engineering and Applications (JSEA), 2011. **4**(9): p. 513+.

URL: http://web.ebscohost.com.ezproxy.auckland.ac.nz/ehost/detail?sid=b732e907-db3c-491c-ad5d-0940cf4405a5%40sessionmgr13&vid=1&hid=19&bdata=JnNpdGU9ZWhvc3QtbGl2ZSZzY29wZT1zaXRl#db=iih&AN=67704548

9. Alisi, T.M., A. Del Bimbo, and A. Valli, *Natural interfaces to enhance visitors' experiences.* MultiMedia, IEEE, 2005. **12**(3): p. 80-85.

URL: http://ieeexplore.ieee.org.ezproxy.auckland.ac.nz/xpl/articleDetails.jsp?arnumber=1490500

10. So-Young, P., et al., *Natural language-based user interface for mobile devices with limited resources.* Consumer Electronics, IEEE Transactions on, 2010. **56**(4): p. 2086-2092.

URL: http://ieeexplore.ieee.org.ezproxy.auckland.ac.nz/xpl/articleDetails.jsp?arnumber=5681076

11. Hearst, M.A., *'Natural' search user interfaces.* Commun. ACM, 2011. **54**(11): p. 60-67.

URL: http://dl.acm.org.ezproxy.auckland.ac.nz/citation.cfm?doid=2018396.2018414

12. Sato, E., T. Yamaguchi, and F. Harashima, *Natural Interface Using Pointing Behavior for Human–Robot Gestural Interaction.* Industrial Electronics, IEEE Transactions on, 2007. **54**(2): p. 1105-1112.

URL: http://ieeexplore.ieee.org.ezproxy.auckland.ac.nz/xpl/articleDetails.jsp?arnumber=4140637