Facilitating Natural User Interfaces through Freehand Gesture Recognition

Sam Kavanagh
University of Auckland
83 Symonds Street, Auckland,
New Zealand
skav012@aucklanduni.ac.nz

ABSTRACT

The Natural User Interface (NUI) can be defined as an interaction paradigm emphasizing human-computer interaction that both comes naturally to the user and reflects natural (real world) elements. Traditional peripherals are innately unnatural to users; their limited affordances requiring the additional utilization of metaphors through the popular Graphical User Interface (GUI). The closeness of freehand gestures to their 'real world' counterparts (and their respective affordances) both lessens the need for an additional metaphorical layer and decreases the learning curve involved in this interaction paradigm. This literature review investigates how freehand gesture recognition facilitates NUIs by analyzing existing techniques, the potential areas of their application and their underlying technologies (including their respective limitations). Furthermore, freehand gesture input is compared and contrasted in terms of accuracy, enjoyment, practicality and overall naturalness to existing peripheral devices. If recent architectural advances in technology such as the Microsoft Kinect and Leap Motion (which allow gesture recognition at the sub-millimeter level) are any indication of the field's direction; designers of the future are only limited by their imagination.

Author Keywords
Natural User Interface; NUI; Freehand Gesture Recognition.

ACM Classification Keywords
Design; Theory; Human Factors.

INTRODUCTION

The disappearance of the mechanical interface is inevitable. Physical buttons have long remained the primary gateway for interacting with the digital realm. Recent technological advancements however have seen new forms of input arise in an attempt to replace traditional interfaces. The traditional user interface has experienced three major evolutions thus far; Batch interfaces, Command Line Interfaces, and more recently the Graphical User Interface [5]. The Natural User Interface is promoted as the next logical evolution in interaction paradigms, attempting to rectify the existing problems of unnaturalness, limited affordances and high learning curves which are prevalent in current GUI-based systems. It does so by attempting to replace current interaction techniques with ones that both come naturally to the user and relate to elements of the real (physical) world.

An alternative form of input suggested to facilitate NUIs to gestural input is voice recognition. However, despite increasing levels of voice-recogntion accuracy, an interaction paradigm emphasizing speaking commands to inanimate objects is innately unnatural to users. Long before humans learn to speak they learn to interact with their environment physically, thusly gestural input is the logical choice for designing an interaction paradigm emphasizing that which comes most naturally to users.

Gesture based input has been an area of research for decades, however the need for further research has been greatly spurred in recent years by a two-pronged increased in the technology. Technological innovations such as Microsoft Kinect have seen the accuracy of freehand gestural recognition technology increasing exponentially in recent years, with the price of such technologies reacting conversely. Naturally this is resulting in a rapidly increasing prevalence of gesture recognition hardware in households [8]. End-user utilization and availability of such technologies is not limited to consumers however; the recent public release of Software Development Kits (SDK) for major gestural recognition platforms such as Kinect in 2011 are allowing hobbyist programmers access to technology that would be otherwise be unavailable [3].

Although gesture recognition technology is now reaching a point where it is both affordable to consumers and accurate at a sub-millimetre level, it has (thus far) failed to replace the mouse-and-keyboard as the dominant desktop interaction paradigm. This is likely attributed to the fact that actions performed in certain digital environments have no equivalent real world counterparts [1].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI'12, May 5–10, 2012, Austin, Texas, USA.
Copyright 2012 ACM 978-1-4503-1015-4/12/05...$10.00.

Despite its failure to thus far replace the GUI as the dominant desktop interaction paradigm, gestural input is nonetheless potentially more appropriate to other areas of digital interaction than existing peripherals. In the following sections the application of gestural input across a diverse range of industries is discussed, comparing and contrasting their commonalities, motivations, underlying technology and respective issues and weaknesses. An emphasis is placed on the overall naturalness of the gestural input and its viability as an interaction paradigm to replace the traditional mouse-and-keyboard peripherals in each respective application/industry.

GESTURAL INTERACTION TECHNIQUES

The following section outlines a diverse range of gestural interaction techniques, describing the basic categories of gestural interaction before investigating unique gestural interaction techniques and evaluating whether certain gestural styles are more suited to a particular task. Several full-body based gestural techniques are described, however emphasis is placed on freehand gestural interaction. Furthermore the limitations and issues of the respective gesture types are discussed.

The ideal (and most natural) gesture recognition system is one where the user need not learn additional gestures or don special apparatus to utilize the system to its full potential; i.e. the *walk in and use* system [1]. Although this is possible for some simple applications in practice, oftentimes virtual interactions simply have no real-world physical counterparts. This unfortunately requires the design and learning of additional unnatural gestures, which in turn lessen the potential gesture interaction has to facilitate Natural User Interfaces. It is therefore critical to their success that gesture designers take into account context of use and attempt to create gestures that are both comfortable to perform and simple to remember [1].

Hand-based gesture interaction is classifiable into two overarching categories: unimanual and bimanual (which utilize a single and both hands for gesture input respectively). Unimanual interaction is commonly applied where only simple interaction (such as pointing) is required, however if gesture sets are well designed unimanual systems can remain effective in complex environments [6]. Nonetheless, bimanual interaction is generally more suitable where complex interaction is required; for example Boussemart et al. [1] designed an interaction style dedicating one hand to selection/pointing, while utilizing the additional (free) hand to perform actions on selected objects.

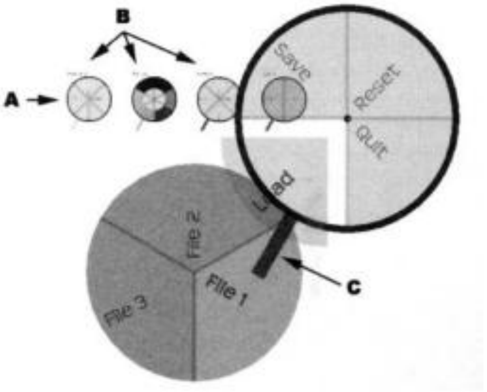

Figure 1. The Generic Pieglass Widget designed by Boussemart et al. [1]

Nancel et al. [6] classify freehand gestures into two further subcategories; linear and circular. The problem linear gestures suffer from (which circular gestures avoid) is that of *clutching*, i.e. being forced to return the hand to a comfortable position in order to perform an additional gesture. However, despite this lack of clutching Nancel et al. [6] found that circular gestures not only perform worse than linear gestures (22% slower on average), but are also generally less comfortable to users.

In order to decrease the need to learn entire new gesture sets for every application, Boussemart et al. [1] have proposed a generic 'pieglass' interaction metaphor for freehand gesture interaction. A rack of pieglasses (each with different functionality) are available to be applied to whatever objects the user has selected. The pieglass functions effectively as a generic menu tool for freehand gesture systems, and can effectively be utilized across a wide range of systems. However the pieglass is undoubtedly unnatural to users, and does nothing towards facilitating an effective NUI through gestural interaction.

Gesture recognition systems which detect areas other than the hands have been popularized recently by Microsoft's Kinect, which has sold over ten million copies [8]. Full-body recognition was proposed by Nancel et al. [6] to allow users to execute zooming simply by moving towards displays. However, the utilization of such systems thus far has been almost exclusively limited to gaming. Given the fact that we interact with the physical world with our bodies in their entirety – it is an area that requires further investigation for its potential to facilitate NUIs.

Zigelbaum et al. [10] propose a gestural interaction paradigm which utilizes a *complimentary* passive 2D surface (containing additional metadata) to make up for the shortcomings of gestural interaction.

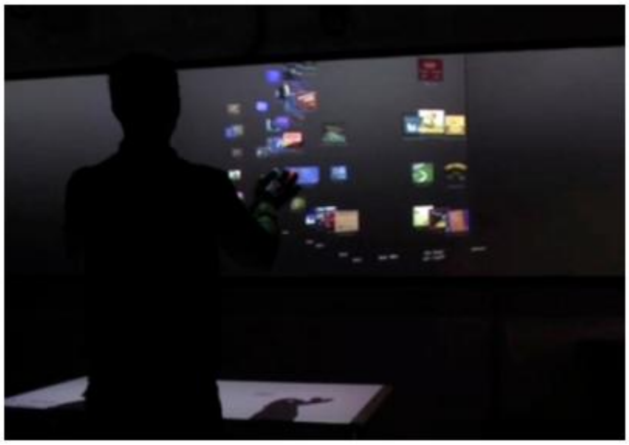

Figure 2. G-stalt [10]. The table in front of the user is can be interacted with in order to apply additional filters to the data.

Additionally Zigelbaum et al. [10] developed an interaction paradigm for cube-shaped data which utilizes the thumb, index and middle figure in order to rotate and translate data on the y, z, and x axis respectively. These gestures although logical, are obviously not self-revealing and their effectiveness and complexity was not sufficiently evaluated by the authors.

UNDERLYING TECHNOLOGIES

The following section similarly describes the various categories of technology that are commonly utilized in gesture based systems. The limitations and common issues encountered when using such technologies are also discussed.

Gesture recognition systems fall into two main categories: those that require motion sensing devices be attached to the user (marker based), and those that utilize spatial tracking (markerless) [8].

Markerless systems are preferable in facilitating NUIs as they do not require the wearing/utilization of apparatuses which users would not use in the physical world. They do however suffer from a myriad of problems. The traditional implementation of markerless gesture systems is via video based image processing and recognition. Boussemart et al. [1] have developed an immersive 3D environment which combines video data from multiple cameras. The system performs background removal and detects skin-like colours before combining data in order to give an accurate 3D position of the user. However, like many video tracking systems it suffers from problems of occlusion and false positives (mainly from skin-like colours). Microsoft Kinect is by far the most popular example of markerless gesture recognition. It improves on existing video recognition based systems by utilizing depth sensing technologies (via infrared rays) which are then detected by an additional RGB camera. This greatly improves on markerless systems that solely utilize image processing. However the infrared depth sensing technology is rendered more or less useless at close distances.

Marker based systems are favourable where accuracy is of a higher priority than naturalness. Unlike markerless systems, which emphasize *walk in and use* interaction [1], marker based systems require users to utilize additional apparatuses which can be detected by the system. A popular marker based system that is utilized across a wide range of applications is Vicon, which tracks retroflective markers with sub-millimetre accuracy [1][10]. These markers are often applied to the fingertips of specially designed gloves, providing suitable interaction to users where high levels of dexterity is required. Marker based gesture systems are generally more accurate, and less prone to errors than their markerless counterparts – however they are less effective at facilitating NUIs [8].

An additional form of technology utilized in gesture based systems is peripherals utilizing accelerometers. Popularized more recently by Nintendo's Wii remote, accelerometer based gestural technology has been utilized effectively across a wide range of industries [4]. However, accelerometer based systems are less effective still at facilitating a natural user interaction paradigm than marker based gestural systems [8].

AREAS OF APPLICATION

The following section outlines a diverse range of proposed applications of gestural input. Initially the motivation for deviating from existing interaction paradigms is questioned, before evaluating its effectiveness and viability as a replacement to existing forms of interaction for that particular application. This is done by combining the limitations and weaknesses identified in the previous two sections. Finally, the overall naturalness of the systems are evaluated in order to determine whether they successfully operate as a NUI (regardless of whether this was developer's intention). This is done in an attempt to evaluate the overall effectiveness of gestural interaction as a facilitator of natural user interaction.

High-resolution wall sized displays have been proposed as an effective means of accommodating very large heterogeneous datasets across various domains [6]. Due to the size of such displays (which can be over 5.5m wide, and 1.8m high [6]) users need to be able to move about freely, as attempting to discern large quantities of (physically spread) data from a single vantage point is impractical. Nancel et al. [6] state that this precludes the utilization of the keyboard and mouse, and resultantly designed a series of interaction techniques (with varying degrees of freedom) to determine the appropriate form of input for mid-air panning-and-zooming on wall sized displays. Surprisingly however Nancel et al. [6] found that when comparing different interaction techniques users generally preferred devices with only a single degree of freedom (such as the mouse-wheel) for zooming actions. Similarly G-stalt, a system developed by Zigelbaum et al. [10] to interact with large displays from a distance developed their own gestural interaction paradigm. G-stalt works with a complex set of gestures controlling cube shaped data. The approach taken by Zigelbaum et al. [10] puts a large emphasis on efficiency, whereas the research performed by Nancel et al. [6] focuses on the quality of user experience. As such, G-stalt is more suitable to professionals and advanced users (due to its higher learning curve). Whereas a system combining the preferred interaction techniques (identified by Nancel et al. [6] in their pilot test) has a higher potential to be utilized by casual end users due to its lower complexity and overall naturalness comparatively. Such a system is an improvement on, and viable replacement to the traditional mouse-and-keyboard interaction paradigm for working with large displays. Regardless, it is apparent that despite its universal appeal, utilizing freehand gestural input *exclusively* may not be best suited for tasks requiring a high degree of accuracy [6].

Immersive 3D virtual reality environments are appropriate applications for gestural input. These environments attempt to create an interaction paradigm that keeps the user feeling completely untethered, something which special apparatuses and traditional peripherals do not allow [1]**.** In order to create a truly immersive 3D environment the user should be able to interact with the virtual world as closely as possible to how they interact with the physical. In order to achieve this vision Boussemart et al. [1] developed a walk-in-and-use wall projection based 3D environment allowing direct manipulation via bimanual image tracking based gestural recognition. Unfortunately as mentioned, not all virtual actions have real-world counterparts. Rather than developing a complex set of gestures to counter this Boussemart et al. [1] opted to develop the aforementioned generic Pieglass widget mapping layer which can be applied to elements of the environment to bring up a virtual menu. This contradicts the papers original intentions however, further distancing users from a truly immersive, realistic and untethered environment. Nonetheless, the goal of creating an immersive and realistic 3D environment that closely resembles the physical is by definition a Natural User Interface.

The industry where gesture recognition is undoubtedly most prevalent currently is gaming. The uptake of gesture-based gaming systems has occurred only recently as a result of the release of technology giants Nintendo and Microsoft's Wii and Kinect respectively. As such, game developers are scrambling to create games that will meet this recent trend [8]. An important consideration for game designers when replacing existing input techniques is the importance of an enjoyable user experience. Gesture based NUIs that closely map the physical world are appealing to users in that they reduce the barriers involved in learning new games, allowing them to focus on the games content [8]. Siratuddin & Wong [8] concisely and effectively outline the existing popular gesture gaming paradigm technologies, describing both their benefits and limitations. Freehand gesture based games, though effective at lowering entry barriers, unfortunately are not viable as a replacement for games requiring high degrees of accuracy and rapid speed, such as traditional micro-based strategy games [8]. Furthermore, users of gesture based games frequently report what is known as *Gorilla Arm Syndrome*; an ache in the shoulders and arms that is often associated with hands being held in front of the body for an extended period of time [8]. It is therefore apparent that despite its natural appeal (even within the gaming industry), free-hand gesture recognition is limited in its potential areas of application. Similarly to the aforementioned 3D immersive environments however, if designed effectively gesture-based NUIs are undoubtedly ideal for gaming environments emphasizing realism [8].

Another area of application (and possible solution to the aforementioned Gorilla Arm Syndrome) proposed by Freeman et al. [2] is the use of freehand *pose*-based gestural interaction in typical household environments. Concluding that the amount of mental and physical effort required to perform traditional gesture sets was an excessive ask of home users, Freeman et al. [2] set out to develop a set of gestures which could comfortably (and accurately) be performed in a typical household setting – such as a couch. However two main problems were quickly identified; naturally by performing gestures in a non-standing position a user's range of motion was highly constricted. This in turn led to the aforementioned problem of clutching, as well as increasing levels of false positives. Furthermore, when placed in an informal environment users rapidly began modifying/relaxing gestures in order to make them more comfortable, eventually reaching a point where they could not effectively be recognized (even in a Wizard of Oz study [2]). Although gesture recognition systems are likely to become commonplace in households of the future, the complexity and range of gestures that can be performed in relaxed environments will likely be limited.

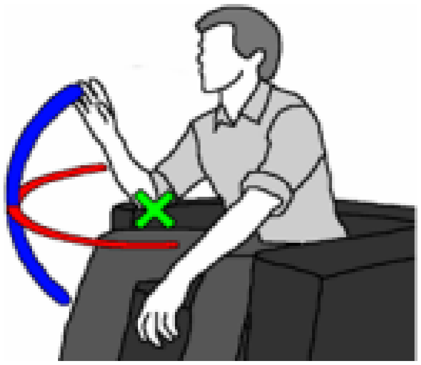

Figure 3. Performing gestures in relaxed states greatly limited user's range of motion [2].

An additional (and long studied) area of gestural recognition is that of sign language. Although it has the potential to replace the keyboard as the main source of input, the primary motivation for sign language recognition is as an educational tool for the deaf [9].

The creation of 3D objects via 3D gestures is motivated not only by the similarity of the interaction to the output, but because traditional 3D data input is a tedious and time-consuming task [7]. Nishono et al. [7] developed a bimanual gestural interaction system which creates complex geometric shapes through the combination and deformation of geometric primitives. The system utilized the aforementioned glove-based technology in order to maximise the dexterity/accuracy of the user. Despite the issues in terms of facilitating NUIs that come with donning special apparatuses, Nishono et al. [7] have developed an efficient and intuitive system allowing users with limited training to develop complex geometric shapes.

A final area where gestural input is being utilized is healthcare. Consumer electronics now contain sensors that can detect acceleration, orientation, location etc. To take advantage of this Khan et al. [4] have developed Gesthaar, an accelerometer-based gesture recognition tool for pervasive health care. The motivation behind Gesthaar was to develop an activity diary that can be easily updated to help patients keep track of their lifestyle [4]. Gestures were programmed to work with an iPod touch to represent the onset of various activities with a 99% success rate. Despite this high level of accuracy, the gestural interaction paradigm developed by Khan et al. [4] is neither intuitive nor do the gestures closely reflect the real-world counterparts they are describing; as such it cannot be classified as a NUI.

SUMMARY AND FUTURE WORK

This literature review investigates the potential freehand gesture recognition holds in facilitating Natural User Interfaces. This is done by subdividing the body of this review into three major subsections. Gestural interaction techniques are first categorized and investigated; discussing their respective benefits and limitations in order to assess whether certain gestural types are more suited towards particular tasks. A range of underlying technologies commonly utilized in gesture recognition systems are then similarly categorized and investigated. The major (current) areas of application of gesture recognition are then outlined; their motivation and viability as a replacement for existing interaction paradigms is evaluated by combining the limitations outlined in the previous two sections. Finally the overall naturalness of systems is evaluated in order to determine whether they successfully operate as a NUI.

Gesture sets have been effectively designed for a range of areas of application. The major limitation of current gestural systems is the need to design for digital interactions which have no physical counterpart. Not only does this require extra work for developers, but greatly increases the learning curve for applications which would otherwise be simple to perform with traditional peripherals.

Gesture recognition technology is rapidly improving, however there is still a fundamental tradeoff between usability and accuracy. This is most clearly illustrated in the split between markerless and marker based recognition technology, which emphasize the aforementioned attributes respectively. Future improvements in image recognition and depth sensing technology could render marker based gesture recognition redundant. For now however, it is still the recommended form of gesture recognition technology where dexterity is pivotal to success.

The gesture recognition systems reviewed in the previous sections were as diverse in their ranges of application as they were in their success. Although the gesture based interaction paradigm can be applied to almost all areas of digital interaction, it is obvious that it is limited in its potential areas of *effective* utilization. Even with the continual improvements in gesture recognitions underlying technology, until an effective generic gesture-based interaction paradigm is developed that can be used across all existing areas of digital interaction, traditional peripherals will continue to be more suitable to certain areas of application. Furthermore, even if such a system was to be developed; the level of complexity required of it to facilitate the vast range of virtual actions with no physical counterparts would likely be so high that it could not be considered a NUI.

REFERENCES

1. Boussemart, Y., Rioux, F., Rudzicz, F., Wozniewski, M. and Cooperstock, J.R. 2004. A framework for 3D visualisation and manipulation in an immersive space using an untethered bimanual gestural interface. In *Proc. of the ACM symposium on Virtual reality software and technology* (VRST '04), 162-165. DOI=10.1145/1077534.1077566 http://doi.acm.org.ezproxy.auckland.ac.nz/10.1145/1077534.1077566
2. Freeman, D., Vennelakanti, R., Madhvanath, S. 2012. Freehand pose-based Gestural Interaction: Studies and implications for interface design. In *Proc. of Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on*, 1,6, 27-29. DOI=10.1109/IHCI.2012.6481816 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6481816&isnumber=6481768
3. Goth, G. 2011. Brave NUI world. *Commun. ACM* 54, 14-16. DOI=10.1145/2043174.2043181 http://doi.acm.org/10.1145/2043174.2043181

4. Khan, M., Ahamed, S.I., Rahman, M., Ji-Jiang Yang. 2012. Gesthaar: An accelerometer-based gesture recognition method and its application in NUI driven pervasive healthcare. In *Proc. Emerging Signal Processing Applications (ESPA), 2012 IEEE International Conference on*,163,166. DOI=10.1109/ESPA.2012.6152471 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6152471&isnumber=6152429

5. Liu W. 2010. Natural user interface- next mainstream product user interface. In *Proc. of Computer-Aided Industrial Design & Conceptual Design (CAIDCD), 2010 IEEE 11th International Conference on*, 203,205. DOI=10.1109/CAIDCD.2010.5681374 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5681374&isnumber=5681221

6. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O. and Mackay, W. 2011. Mid-air pan-and-zoom on wall-sized displays. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11), 177-186. DOI=10.1145/1978942.1978969 http://doi.acm.org/10.1145/1978942.1978969

7. Nishino, H., Nariman, D., Utsumiya, K., Korida, K. 1998. Making 3D objects through bimanual actions. In *Proc. of Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, 11-14. DOI=10.1109/ICSMC.1998.726623 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=726623&isnumber=15672

8. Shiratuddin, M.F. and Wong, K.W. 2012. Game design considerations when using non-touch based natural user interface. In *Transactions on Edutainment VIII*, Zhigeng Pan, Adrian David Cheok, Wolfgang Müller, Maiga Chang, and Mingmin Zhang (Eds.). Springer-Verlag, Berlin, Heidelberg, 35-45. http://dl.acm.org/citation.cfm?id=2363273.2363278&coll=DL&dl=GUIDE
Note: This paper appears to no longer be available through the UoA library proxy (despite that being where I obtained it from). Nor is the PDF available anymore on the ACM website.

9. Zafrulla Z., Brashear H., Starner T., Hamilton H., and Peter Presti. 2011. American sign language recognition with the kinect. In *Proc. of the 13th international conference on multimodal interfaces* (ICMI '11), 279-286. DOI=10.1145/2070481.2070532 http://doi.acm.org/10.1145/2070481.2070532

10. Zigelbaum, J., Browning, A., Leithinger D., Bau O., and Hiroshi Ishii. 2010. g-stalt: a chirocentric, spatiotemporal, and telekinetic gestural interface. In *Proc.of the fourth international conference on Tangible, embedded, and embodied interaction* (TEI '10), 261-264. DOI=10.1145/1709886.1709939 http://doi.acm.org/10.1145/1709886.1709939