

Gesture Based User Interface Experience – Accessibility, Evolution and Challenges

William Vida

User Experience Evolution

Today, the keyboard and mouse provide the majority of user input for computer systems. But there are other methods of user input including hand or other body gestures, eye tracking and console controllers [1]. Microsoft's Kinect for the Xbox 360 is an example of a gesture-based system. The Kinect was controller free and allowed users to make hand and arm movements which were used as inputs instead of a controller for certain video games [14].

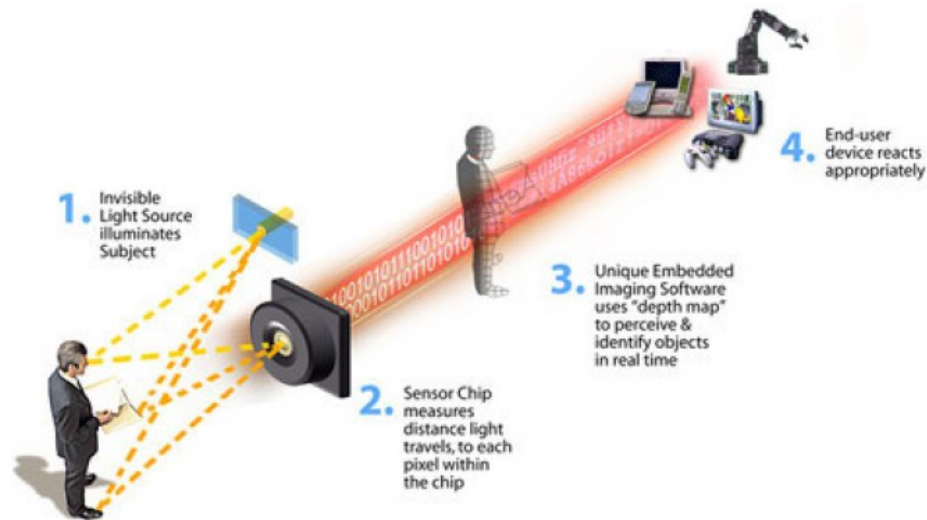


Figure 1: Depth perception using an infrared camera.

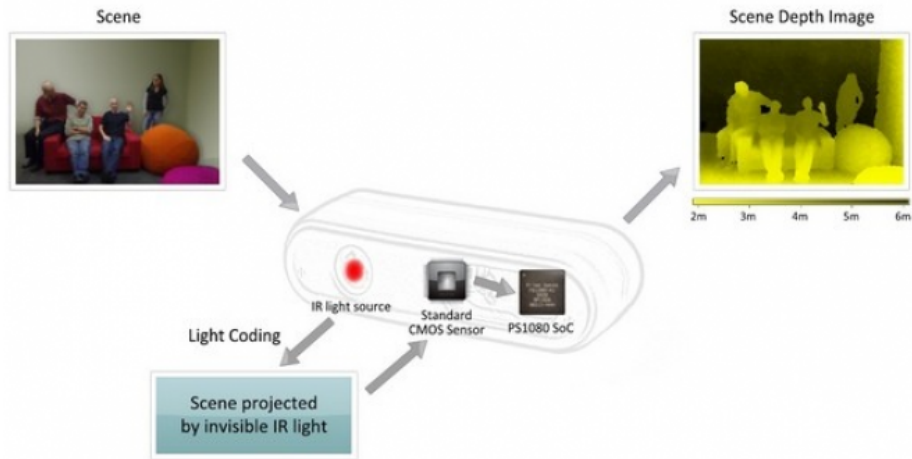


Figure 2: Transfer of information from the camera to the TV.

The mouse was created in the early 1960s at the Stanford Research Institute by Douglas Engelbart. The mouse was originally meant to be a cheap replacement for light pens. Engelbart showed the uses of the mouse in a movie in 1968. Many of the uses shown in the movie are still used today. The mouse commercially appeared in products such as the Apple Lisa and the Xerox Star in the 1980s [7].



Figure 3: The first mouse.

The first pen-based computer was the Rand Tablet which was developed by the Advanced Research Projects Agency (ARPA). The first trainable gesture recogniser was invented by Warren Teitelman in 1964. Tom Ellis used the Rand Tablet to create an early demo of gesture recognition called Graphic Input Language (GRAIL). It was very customary for light pen-based systems to provide gesture detection. Since the 1970s, gesture recognition has been used

in industrial CAD systems [7].

In recent years, gesture-based interaction has become an essential part of modern interfaces. This can be seen across multiple areas in our way of life including workplaces and entertainment. Gestures can include touch screen gestures and free form or in-air gestures. Touch screen interaction is done by touching a screen directly with one's fingers. While in air gestures are done by making a certain gesture with one's hand and/or fingers in front of a camera or a number of sensors. But, to use mid-air gestures a dedicated input device must be used which can include gloves or controllers, or the gestures executed must be done within a certain range of depth-sensing sensors or infrared sensors [4].

Gesture-based input has advantages over normal input devices which can include naturalness and the ease of learning. Gesture-based input would preferably allow users to concentrate on the current task and not on the input device. Users can feel an increased sense of power and control when the computer reacts to gestures that are perceived to be natural. Users can also feel more personal identity and pleasure when using a system that reacts to their gestures [14].

Gesture-based input can help the elderly and those with disabilities. Due to changes in demographics, in the coming future, we will see more elderly people and less young people. This could mean that mouse and keyboard will be replaced by gesture-based applications for elderly people and also those with disabilities [1].

One area of gaming that has been on the rise in recent years is virtual reality. Virtual reality involves using a headset device strapped to one's face, motion controllers and of course a monitor and usually a powerful computer, to get a superior virtual reality gaming experience, in order to play. Virtual reality can add more immersion and intensity to one's gaming experience than a normal controller or keyboard and mouse can. Virtual reality gaming can "enhance[] overall satisfaction, enjoyment, engrossment, creativity, sound, and graphics quality". Steam, which is a digital video game marketplace, 16% of video games available for sale there supported a virtual reality headset in 2016. And 83% of those video games required a virtual reality headset in order to play [12].



Figure 4: The Oculus Rift headset and the Oculus Touch motion controllers.

Gestures as a Communication Tool

Hand gestures are just part of non-verbal communication channels. Gestures can be complex in part due to the fact that similar movements are used to send different types of messages. Gestures can be iconic or arbitrary, they can refer to objects or go hand in hand with the words of the speaker or they can be directed at themselves or to other people around the speaker. Speech is usually accompanied by changes to bodily movements in particular to the hands and face of the speaker. Those two have been found to be very highly linked with speech even down to the level of each syllable. An explanation for this proposes that language "evolved from gestures-first pointing at objects and animals, then representing animals by imitating their movements, and objects by the movements used to make or use them" [3].

The frequency of gestures used along with speech can depend on the culture one person is in. For example, Italians are said to be a people where hand gestures are used very often and use them more than the English which is considered a low gesture usage people. But it must be said that it is unclear if Italians communicate more information by using hand gestures. Observed cultural differences in Italian and Jewish migrants to the United States showed different gestures by them. Italian migrants used more "kinetographs (showing a bodily action) and pictographs (drawing a picture)". Italian gesturing can also include sign language. British-Americans do not make their hand gestures as openly as Italian-Americans. British-American children are taught not to make gestures as it is seen as rude. There are individual differences when it comes to the capability to send and receive nonverbal signals but these differences have been shown to not be related. A study found that the ability to identify voice-tone emotions has a correlation with intelligence and with the ability to vocally express feelings [3].



Figure 5: Some international hand gestures.

Gestures used during verbal communication have been found to have an information value that can be utilised by both speakers and listeners. For example, those who are also language learners use gestures as a means to provide verbal help when they interact with native listeners. An experiment was conducted where listeners were asked to repeat a story told to them by a speaker whose gestures did not always match the corresponding speech. The results showed that "information in the gestural channel is retained by the listeners and integrated with information from the spoken channel into what can be thought of as an intermodal cognitive representation of meaning" [5].

Face to face interaction is where most gestures take place. The listener in this case usually looks at the face of the speaker. Assuming this is followed, then gestures made would have to be taken in through one's peripheral vision.

Research and analysis of gesture recognition, particularly in face-to-face conversation, have rarely had sufficient control over the visual perception of the listener but mostly relied on video recordings from the speaker’s point of view [5].

Studies have shown that while speech is immersed in noise, gestures may increase language understanding, have a beneficial effect on future memory of the language information and may even serve as a disambiguation cue when speech is unclear or includes an uncertain term. A 2010 study, proposed the integrated-systems hypothesis which states the usage of gesture and speech is mutual and compulsory. This means that gestures are always taken into consideration and one cannot circumvent using them in a speech [8].

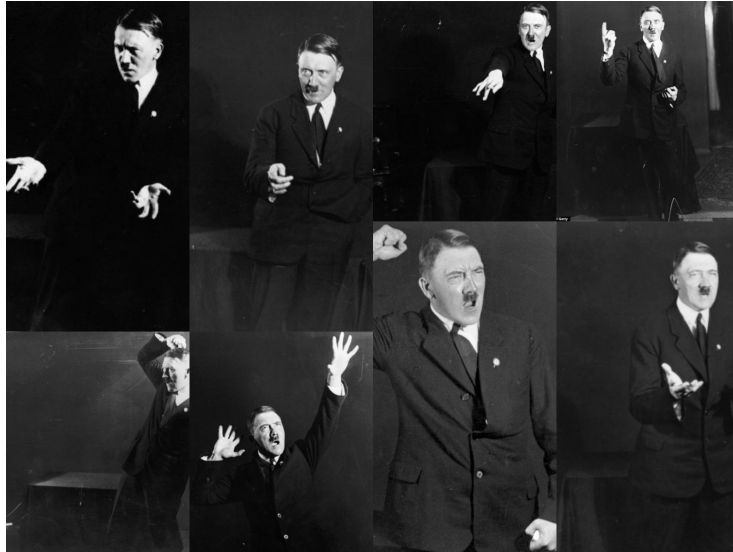


Figure 6: Adolf Hitler practising his hand gestures for his speeches. Reportedly taken in 1925.

A 2004 study explored the effects of gestures on language comprehension. The study recorded event-related potentials (ERPs) which are "electrophysiological neural responses time-locked to a stimulus [6]", elicited by spoken words said at the same time with gestures that either agreed with the words spoken or disagreed with the words spoken. Stimuli were made by recording an actor while the actor was gesturing to either a tall, thin glass and a short, wide dish. The actor said one of the following speech tokens, tall and thin (the glass) and short and wide (the dish). The matching condition involved the speech made by the actor matching with both the object described in the speech as well as the same dimension (the actor said the tall and gestured the height of the tall, thin glass). The mismatch involved the speech token matching to one object while the gesture matched to the other. When there was no gesture condition

the speech was presented alone [16]. The results showed that:

Early effects of gesture congruency (between 100 and 352 ms), with mismatching and complementary stimuli eliciting relative to other conditions larger P1 and P2 components, which reflect auditory sensory processing. N400-like effects were also observed at bilateral temporal electrode sites, with mismatch trials eliciting more negative ERPs than all other conditions around 450 ms post-stimulus. These findings suggest that gesture congruency affects both early sensory as well as higher order semantic processing of words [16].

Challenges for Design of Applications

For many human-computer interaction systems, hand gesture movement seems to be the most natural and appropriate as it is similar to how humans interact with each other. One suggestion is to use a virtual mouse. There is a sizeable market for hand gestures systems and there remain challenges for the development of hand gesture recognition systems. A virtual mouse can replace the use of a physical keyboard and mouse but this still has problems that need to be sorted out. Hand gesture recognition is a key issue that needs to be tackled when designing a virtual mouse device. A very effective form of equipment for recording hand gesture movement is through electromechanical or magnetic sensing devices (data gloves). This process uses sensors attached to a glove that converts finger movements into electrical signals to recognise hand gestures. Data gloves that are used with hand gestures have been used in sign language analysis and training and to control robotic arms. Unfortunately, data gloves have many drawbacks. These include that due to their complexity their costs remain high. The size of data gloves can also cause issues. This makes it difficult to use data gloves for day to day activities. Vision-based gesturing provides a very hopeful alternative to data gloves as it is non-contact and feels more natural than data gloves. Vision-based gesturing is based on computer vision algorithms. Algorithms are used for hand segmentation based on a background model. Colour-based algorithm systems are used to distinguish the area of the object of interest. Hand gesture identification helps increase precision when analysing a vast amount of hand gestures [13].

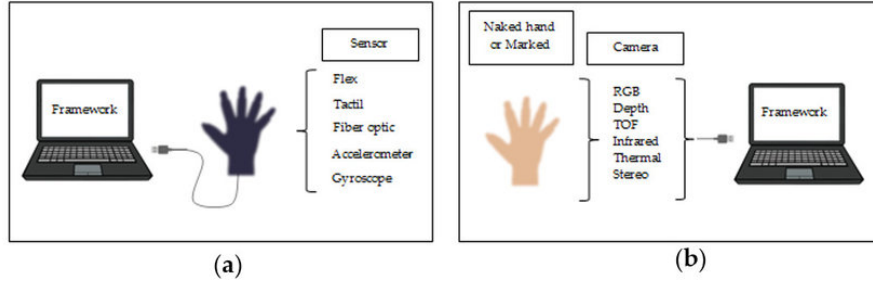


Figure 7: Two different techniques for hand gestures showing a glove sensor and mid-air hand gestures.

It can be difficult to achieve a high hand gesture recognition accuracy when having complex backgrounds, changes to the external lighting and shadows of the hand gestures. It was suggested to use RGB cameras to detect hands but it was hard to use in certain lighting conditions and for different skin colours. It was better to use depth cameras to differentiate hands in front of a camera. A study found that when hand gestures contained a background, a human face and a human body, the accuracy for detecting hand gestures was 97.1% with a simple background while the accuracy dropped to 85.3% when a complex background was in place. Tests were done using a Kinect to show that if there is a larger object than a hand then the gestures made will be inaccurately analysed by the Kinect. A deep learning neural network was made which analysed a large set of computer-generated 3D hand images with some real camera images as part of the training data set. The testing and training data sets for both had 24 classes of hand gestures and the hand gesture accuracy was 77.08% [17].

Study gaps can be easy to find in the field of hand gesture recognition systems as research studies tend to focus on computer applications, sign language, and interacting with a 3D model in a simulated environment. However, numerous scientific papers are concerned with strengthening the system for hand gesture recognition or designing new algorithms rather than applying a realistic application with respect to healthcare. A big challenge researchers face is to design a functional system that overcomes the most common problems with fewer constraints and produces precise and consistent performance [9].

Many of the hand gesture systems that are suggested can be placed into two different categories of computer vision techniques. The first one is to use image processing methods by either using Open-NI libraries or OpenCV libraries and other resources if necessary to provide real-time interaction that recognises time consumption due to real-time processing. This has drawbacks which can include "backdrop problems, lighting changes, distance restrictions, and multi-object or multi-gear issues". The second category involves using gestures contained in data sets to match the gestures made by the user where significantly more complicated patterns need more advanced algorithms. Artificial intelligence and deep learning methods can be used to recognise a vast number of gestures

in real-time by analysing gestures in a data set containing specific postures or gestures. This method does have issues which can include overlooking some gestures due to the accuracy contrast of the classification algorithm [9].

When designing hand gesture systems, the designers must take into account the differences in hand gestures in different nations and different cultures. Certain hand gestures can have different meanings around the world. A gesture that is positive or neutral in one country might be considered offensive or insulting in another one even if they were unintentional. People who are not from the countries of the designers of hand gesture systems might move their hands in a different way than the way the system was designed for which could reduce the precision of the hand gesture recognition system. A study was done using 54 Lao alphabets while four individuals did 540 gestures. The camera had difficulty in keeping up with the speed of the hand gestures and it had an accuracy rate of 79% [11].

Challenges for Implementation

Hand gesture user interfaces should be able to perform gesture recognition at high accuracy in real-time. Those systems should be able to detect hands, track their movement and the gestures they make. Tracking describes the process of following the user's hands frame by frame. Detecting hands can be difficult as there can be a high variation in one's hand shape and size, skin colour and the lighting condition the user is in. Tracking issues can happen if there are tight spaces, a cluttered environment and quick hand movements. When deploying a hand gesture system it can be difficult for everyday environments. The camera sensor and lens, the background and the differences of the user can make rolling out the product difficult [15].

The recognition of hand gestures includes complex processes such as motion modelling, motion analysis, pattern recognition and machine learning. The environmental background of the user such as background lighting and the speed of the movement affects the tracking and the accuracy of the prediction. Anticipated problems that can arise include "temporal variance, spatial complexity, movement epenthesis, repeatability and connectivity as well as multiple attributes such as change of orientation and region of gesture carried out". Criteria to judge gesture recognition systems include "scalability, robustness, real-time performance and user-independent" [2].

Virtual reality has potential but currently, it has some drawbacks and a big one includes cybersickness. Symptoms of cybersickness can include "headaches, eyestrain, stomach awareness, and nausea, sometimes severe enough to result in vomiting". Cybersickness is similar to motion sickness but the main difference is that motion sickness occurs when "actual physical bodily motion does not match what is perceived visually". A lot of research has been done to find the causes and to avoid cybersickness when using virtual reality equipment. There was very little difference between the first generation of the Oculus Rift DK1 and its successor the DK2 when it came to cybersickness [10]. The market for

virtual reality is quite small compared to other forms of gaming such as console and PC gaming. This means that there is a lack of games for virtual reality and a lack of games that are of good quality. There are a few reasons for this, one is because that virtual reality is a new technology and it takes more time for developers to develop and test these games. Another is that virtual reality equipment is expensive. Virtual reality headsets can cost more than \$500 and one would also need a powerful computer to run these games. And one would, of course, need space to play the games as well [12].

Conclusions

Hand gestures are used very often when people talk to each other or when giving a speech. They can help make one person a better speaker and allows them to connect to whom you are talking to better. Audiences feel more connected if a speaker is making hand gestures than one who is inactive.

The ways for using computers have evolved but the mouse and keyboard remain the dominate methods used for human interaction with computers to this day even after a few decades after their invention. Phones have evolved from needing a number pad to being completely touch screen in less than 15 years. There have been attempts to bring hand gestures into video games with the Kinect being the most famous but even though it seemed to be revolutionary at the time, the Kinect was arguably a failure for Microsoft. Virtual reality video games remain an expensive alternative to the usual keyboard and mouse for gaming but a problem with virtual reality is that they lack dedicated games as most of the PC gaming market is with keyboard and mouse. This is not to say that there never will be a market for hand gestures interfaces. They seem to be more useful in other fields but maybe as technology advanced they might make another rise in gaming. Hand gesture technology is advancing quickly and maybe within a few decades or less they will be commonplace in our daily lives. Hand gesture technology has advantages such as it being controller/mouse and keyboard free and no set up is required other than a camera. But as mentioned it may be a long time until they are mainstream.

References

- [1] Moniruzzaman Bhuiyan and Rich Picking. Gesture-controlled user interfaces, what have we done and what's next. In *Proceedings of the fifth collaborative research symposium on security, e-learning, internet and net-working (SEIN 2009)*, Darmstadt, Germany, pages 25–29. Citeseer, 2009.
- [2] Ming Jin Cheok, Zaid Omar, and Mohamed Hisham Jaward. A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics*, 10(1):131–153, 2019.

- [3] Jean Ann Graham and Michael Argyle. A cross-cultural study of the communication of extra-verbal meaning by gestures(1). *International Journal of Psychology*, 10(1):57–67, 1975.
- [4] Lisa Graichen, Matthias Graichen, and Josef F. Krems. Evaluation of gesture-based in-vehicle interaction: User experience and the potential to reduce driver distraction. *Human Factors*, 61(5):774–792, 2019. PMID: 30694705.
- [5] Marianne Gullberg and Kenneth Holmqvist. Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics & Cognition*, 7(1):35–63, 1999.
- [6] Sean McWeeny and Elizabeth S. Norton. Understanding event-related potentials (erps) in clinical and basic language and communication disorders research: a tutorial. *International Journal of Language & Communication Disorders*, 55(4):445–457, 2020.
- [7] Brad A. Myers. A brief history of human-computer interaction technology. *Interactions*, 5(2):44–54, March 1998.
- [8] Christian Obermeier, Thomas Dolk, and Thomas C. Gunter. The benefit of gestures during communication: Evidence from hearing and hearing-impaired individuals. *Cortex*, 48(7):857 – 870, 2012. Language and the Motor System.
- [9] Munir Oudah, Ali Al-Naji, and Javaan Chahl. Hand gesture recognition based on computer vision: A review of techniques. *Journal of Imaging*, 6(8), 2020.
- [10] Daniel M Shafer, Corey P Carbonara, and Michael F Korpi. Factors affecting enjoyment of virtual reality games: a comparison involving consumer-grade virtual reality technology. *Games for health journal*, 8(1):15–23, 2019.
- [11] Vaidyanath Areyur Shanthakumar, Chao Peng, Jeffrey Hansberger, Lizhou Cao, Sarah Meacham, and Victoria Blakely. Design and evaluation of a hand gesture recognition approach for real-time interactions. *Multimedia Tools and Applications*, pages 1–24, 2020.
- [12] William J Shelstad, Dustin C Smith, and Barbara S Chaparro. Gaming on the rift: How virtual reality affects game user satisfaction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 61, pages 2072–2076. SAGE Publications Sage CA: Los Angeles, CA, 2017.
- [13] Tsung-Han Tsai, Chih-Chi Huang, and Kung-Long Zhang. Design of hand gesture recognition system for human-computer interaction. *Multimedia Tools and Applications*, 79(9):5989–6007, 2020.

- [14] Maurice H. P. H. van Beurden, Wijnand A. Ijsselstein, and Yvonne A. W. de Kort. User experience of gesture based interfaces: A comparison with traditional interaction methods on pragmatic and hedonic qualities. In Eleni Efthimiou, Georgios Kouroupetroglou, and Stavroula-Evita Fotinea, editors, *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, pages 36–47, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- [15] Juan Pablo Wachs, Mathias Kölsch, Helman Stern, and Yael Edan. Vision-based hand-gesture applications. *Commun. ACM*, 54(2):60–71, February 2011.
- [16] Ying Choon Wu and Seana Coulson. How iconic gestures enhance communication: An erp study. *Brain and Language*, 101(3):234 – 245, 2007. Gesture, Brain, and Language.
- [17] Mais Yasen and Shaidah Jusoh. A systematic review on hand gesture recognition techniques, challenges and applications. *PeerJ Computer Science*, 5:e218, 2019.