Gesture-Based User Interfaces for Public Spaces

Andreas Kratky

USC School of Cinematic Arts, Interactive Media Division,
900 West 34th Street, SCA 201, Los Angeles, CA 90089-2211
akratky@cinema.usc.edu

Abstract. Gesture-based approaches are becoming an increasingly popular technique in Human Computer Interaction. Recent developments in the hardware field have made it more affordable and more reliable to use gesture-based interfaces and they are becoming more of a standard way for human users to interact with computers. Most of the research has been investigating the usage of gestures in personal and limited access situations. But gesture interfaces are promising great benefits to usage scenarios in public spaces or general access environments. This paper will summarize and evaluate the particular aspects of using gesture-based interfaces in application contexts in public and semi-public spaces.

Keywords: Gesture-based Interfaces, Alternative reality environments, Human Computer Interaction, Wearable interfaces, Proxemic interaction

1 Introduction

The field of Human Computer Interaction has seen a paradigm shift towards so called "natural interfaces" in the recent years moving away from the standard of keyboard and mouse-based interfaces towards alternative interaction approaches. The term "natural" in this context refers to interfaces that replace the input of commands into a command line interface or the usage of interface devices such as pointing devices and keyboards by communication means that are more close to the ways human beings communicate in general and thus feel more natural [1], [2]. A common area to look for naturalness is interpersonal communication and the means people use to communicate with each other. The ideal for natural communication is direct, device-free communication, which means speech and gestures, while device-mediated forms of communication are regarded as less natural [3], [4]. Reliable speech recognition is a complicated problem stemming from the difficulty of mapping the sound phenomenon to linguistic elements and the extraction of meaning from those elements. This problem field has been researched for more than 50 years and it still remains difficult to resolve ambiguities in spoken language [5]. Compared to the advances made in speech recognition-based interface technologies, gesture processing has seen a number of recent breakthroughs that made gesture-based interfaces significantly more powerful and popular. The successful launch of the *Kinect* controller was just one of the inventions that started a widespread attention towards gesture interfaces. The *Kinect* game controller is a computer-vision-based controller that allows to capture

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2011, LNCS 6766, pp. 564–572, 2011.

and interpret body movements and gestures from users with a depth camera enabling gesture recognition with a degree of reliability and precision that has been so far not been achieved with solely computer-vision-based approaches. Integrated into a game console, this controller became a successful mass-product making gesture recognition a widespread and affordable technology. Since its launch the *Kinect* was sold 8 million times [6] and it was compared to the legendary gestural interface that made its appearance in the 2002 movie "Minority Report" [7]. Before the *Kinect* another gesture-oriented game controller had a very successful release: The *Wii Remote* was released in 2005 and sold 6.29 million only in the year 2007 [8]. Other popular devices were the *iPhone* and *iPad* with their gesture-oriented touch screen controls allowing to "flick" tough pages, "pinch to zoom" etc. Compared to the *Kinect* the other interface systems do not operate device free – the user has to hold or touch the controller – but they are part of a strong movement towards so called natural human interfaces, which Steve Ballmer, the CEO of Microsoft characterized as a key part of the fifth revolution in computing technology in his 2008 keynote of the consumer electronics show CeBIT 2008 [9].

With this strong emphasis on gesture interfaces as a successful approach to natural user interfaces, this paper will look at the implications and possibilities of implementing gesture interfaces in public spaces and in particular in museum or gallery spaces, which we consider semi-public spaces. If we understand public space as a space that is accessible to everybody without any restriction such as access control, fees etc. we would normally think of public streets, squares and the like. In the scope of this text we will include places that are accessible to a large public without requirements such as the membership in a specific group, race or similar criteria. We will consider this valid for public museums, shopping areas etc. While this definition will not be able to qualify public space from a sociological point of view, it allows us to analyze certain usage patterns that are addressed by human computer interaction. What we will exclude from this study are non-geographic public spaces that solely exist in a form mediated through interfaces such as shared online spaces.

Human computer interaction that takes place in public spaces can benefit significantly from gestural interfaces, specifically because they allow unencumbered, device-free interaction. In this regard gesture oriented interface technologies seem more suitable than other techniques [10]. It is it promising to investigate them for several reasons:

1. As a key technology of natural user interfaces they are suitable to implement intuitive interface solutions that are easy to understand and do not require complex skills to be used successfully.
2. An increased familiarity of the audience thanks to those recent successes mentioned above will support the ease of use.
3. There is a potential to reduce maintenance cost due to less or no mechanical parts exposed to the public.
4. With the mass-market introduction of gesture interface technologies the price of deployment is significantly reduced, which formerly was prohibitively high [11].
5. Gesture interfaces – in particular computer-vision-based implementations – allow for seamless engagement: The user can walk into the active range of

the interface and start interacting before he even realizes and from there start to explore gradually deeper levels of engagement.

The paper will review existing implementations of gesture-based interfaces in public and semi-public spaces and examine their benefits and problems compared to other implementations. It will close with a set of research agenda and evaluation criteria to improve the performance and accessibility of future implementations.

2 Evaluation Criteria

The design of interface technologies for the public space has a number of specific requirements that differ from other task-oriented applications in private usage scenarios. Most interface design criteria are formulated for single-user, non-public settings supported by interface devices such as keyboard, mouse, or even *Wii Remote*. It is in the nature of the control-device that it can only be handled by one user at a time. And even though gesture interfaces lend themselves inherently to more open interaction scenarios, a lot of research still focuses on single-user applications and non-public environments [12], [13]. Some research also targets multi-user systems, in particular with the depth camera technology that is part of the *Kinect* controller, which was explicitly developed to have multi-user capabilities and thus allow for collaborative gaming experiences [14]. This research has generated several guidelines for successful interface design in the domain of gesture-based interfaces as exemplary outlined by Nielsen et al. [15]:

- Easy to perform and remember
- Intuitive
- Metaphorically and iconically logical towards functionality
- Ergonomic; not physically stressing when used often

These criteria hold true also for situations in public space, but need to be extended in order to address some of the specific requirements. In public space it is essential for interfaces to have a steep learning curve. The development of engagement with an application or experience (we will refer to the task that is experienced by the user through the gesture interface as *application*, even though this term implies a task orientation, which might be questionable terminology in an educational or entertainment oriented experience) is different than in most task-oriented applications experienced in non-public environments. While in the latter users generally approach the application with the intention to engage and use it to solve the task at hand (such as write a text with a word processing application or play a game on a game console), in public environments that engagement is often spontaneous as the user encounters the application. Between engagement and non-engagement may be only a very short time of examining the application. In this sense it is vital that the first approach already enables the user to successfully engage into controlling the application and require an easy interface approach that can be grasped very quickly to not deter the user from engaging. Over the time of engagement – once the user made the decision to stay with the application for some time – a certain amount of practice and newly learned skills will establish. In general, though, the learning curve should be steep and then plateau with only little new information learned in terms of operating the application [16]. This kind of phase-staggered engagement can be supported by an appropriate design of the interface complexity allowing the user to ease-into the experience, through the aesthetic appearance that is attractive and accessible to the user, by making the application responsive so that it can sense and respond to the presence of a user, as well as through an appropriate spatial set-up. The spatial characteristics should make the user feel comfortable to spend time in this space, it should be easy to find or enter, and it should make it easy for the user to identify potential tools and positions etc [17]. This can mean different things depending on the application and where it is set up. It is important to mark the space that belongs to the application, i.e. the area that is tracked and thus registering input versus the areas that are not. We often see installations that use projected images set up in "black boxes" in order to keep ambient light controlled and improve the quality of a projection. The light traps that are used with those installations force the user to walk through a dark corridor and constitute a high threshold potentially inhibiting engagement with the application. The user has to go through this corridor and decide to engage before even knowing much about the application at the interior.

Several studies (i.e. Nielsen et al. as mentioned above) have investigated aspects of appropriateness of gestures, such as the relationship of the gesture and the meaning that it stands for (arbitrary versus iconic relationship etc.). A particular criterion arising in public spaces is appropriateness whether a user feels comfortable enacting a certain gesture. As the users are potentially seen by other users, they should not feel embarrassed or uncomfortable carrying out the gestures. Also, gestures can be disruptive in a space shared with other users [18].

3 Example Interfaces

In a short review of some interface systems we will examine exemplary implementations in respect to how they address the above-mentioned criteria. The list of implementations will at the same time revisit several historic steps, but given the scope of the text, it does not claim to be exhaustive. Only applications that use device-free interaction are included.

3.1 Responsive Environments

One of the early implementation of an application that responds to the presence of a user in order to engage him into an interaction is the so-called "responsive environment" *Glowflow*, created by the American artist Myron Krüger. *Glowflow* originated in 1969 as the first application of a series of works that respond to the presence of a user and change their states accordingly. The interaction is carried out by entering the environment deeper and exploring it. In the installation the user steps on sensitive mats that communicate his location to the controlling computer, which then triggers changes in the illumination of the room and the sound created in real-time by a synthesizer. A successor to this work was the installation *Metaplay*, which used instead of the phosphorescent tube lighting of *Glowflow* a video projection of the user in conjunction with a superimposed computer-generated image [19]. Krüger describes as his main focus in these environments as "the only aesthetic concern is the quality of interaction" [19]. Both installations are examples of responsive environments that sense the user and respond to his presence to invite him and draw him into further interaction. The level of complexity of triggering these first state changes of the application system is extremely easy and is done without the user even noticing it. Only based on the response from the system he realizes that he already interacts and is motivated to find out more. We can call such an interface approach "explorative" as the user is invited to explore the system and through this kind of engagement the system reveals its content as well as its interaction methodology.

The notion of a responsive environment has been taken up in later applications. We find a similar approach in a much more recent application by the French artist Marie Sester called *Access* [20]. In the *Access* installation users walk into the range of an overhead computer-vision camera and get tracked by the system. As they get tracked, a spotlight follows them around in the space along with a focused sound projection. The system is construed again in such a way that the interaction starts without the user noticing it, simply by walking into the range of the tracking system the user starts to interact and then realizes the response of the system. The tracking system receives input also from remote users through a web interface where they have access to the imagery captured by the tracking system and are able to select the target to track.

The responsive environments from Krüger were all installed in a gallery space – a semi-public space. Due to the intensive technological preparation of the space it is almost impossible to show these applications in an openly accessible space. Pressure sensors have to be mounted on the floor, the video capturing requires a neutral background, and the interaction system can only present meaningful feedback to a limited amount of users in the system because in case there are too many users the readability of the feedback becomes difficult and state changes are hard to attribute to specific interactions. The *Access* system is more suitable to be presented in an open access public space and is referred to as an installation traveling from "one public space to another" [20], as it provides a seamless entry and exit from the experience and the bounds of the sensitive area are marked by the spotlight.

The type of application described with the help of the above mentioned examples provides a good solution towards the question of seamless development of user engagement with the application and provides a certain possibility to explore the application through continued interaction. While the learning requirements for the user to be able to successfully control this type of application are small, the range of possible interactions is limited to triggering changes in illumination of installed light sources or attracting the spotlight.

3.2 User Representation

A different approach has been followed in a type of application that provides a more complex representation of the user and his interactions with the system. A later application developed by Myron Krüger in 1970 is the *Videoplace* system. Here the image of the user is captured by a camera and shown as a silhouette in a projection [21]. The silhouette image is combined with superimposed computer graphics images of small objects as well as a representation of remote users. The silhouette image and the superimposed images of small objects etc. interact, the objects for example can bounce off of the contour line of the user's silhouette. The system reflects a representation of the user like a mirror and there is an easily understandable mapping of interactions to state changes of the system. With this very direct representation of the user and his influence on the application it is possible to realize a potentially more nuanced set of interactions while preserving the ease of engagement and the possibility to explore the application's functionality through continued engagement.

This approach has also been picked up in a range of applications installed in gallery settings as well as shopping malls and other semi-public spaces. An example for a later implementation is the installation *Bubbles* by Wolfgang Münch and Kiyoshi Furukawa made in 2005, where users see their silhouette interacting with virtual soap bubbles that bounce off of the silhouette outline [22]. Similar installations can be found in science centers (e.g. TELUS World of Science, Vancouver, Canada) or shopping malls (e.g. Glendale Galleria, Glendale, CA, USA). The ability of this type of applications involving a representation of the user in form of a reflection or a shadow to convey more complex and layered content has been observed in several test cases described by Snibbe and Raffle [17].

Most of these applications have been installed in semi public-spaces, as they require a clean projection surface that provides enough contrast and visibility of the projected image, mounting, power, and protection for the tracking equipment and some way to channel users in and out and control the amount of users at a time.

3.3 Ubiquitous Gesture Interfaces

From the previous examples it becomes clear that it is a challenge to deploy gesture-based interfaces in public space for technical reasons as outlined above. Also it seems that interaction sequences in public spaces are significantly different, shorter, and less complex than we know it from other task oriented interfaces in non-public spaces. A lot of the research that spurred the development of technologies like *Kinect* or *Wii Remote* comes from computer gaming. The focus here is on individualized long-term interaction of a comparatively high degree of complexity [23], [24], [25]. There are few examples of gesture interfaces deployed in public space that provide a complex sequence of interactions. Besides the large scale tracking interface of *Access* which is suitable for the public space we can observe the opposite movement towards small mobile mechanisms that provide gesture interface capabilities for the public space. One example is the *Traces* application that consists of a rugged integrated display and tracking device that is "left" in public space so that random passers-by can pick it up and use it. The device captures gestures and colors created by a user and responds by outputting color and motion on a flexible display. The device is intended for "transient public spaces" and presents an instance of a movement towards portable, ubiquitous gestural interfaces [26].

Another application along similar lines is the *PyGmI* interface consisting of an entirely wearable system that integrates a gesture tracker and a micro projector to generate the output of the system [27]. The tracker allows the user to communicate with the system through gestures carried out in proximity to his body. It is thus a system for individuals and not encouraging any multi-user scenarios as the tracker is targeting only the proximity around the user carrying the system.

4 Conclusion and Future Work

The examples discussed in this paper indicate that even though device-free gesture-based interfaces for the public space have great promises there are few implementations of them. The current mainstay of this interface technology is in semi-public and private spaces. We can observe two tendencies in the deployment of gestural interfaces in shared public environments: One tends to make the entire environment responsive with sensing systems covering a larger space, which can be entered, explored and exited by users in a seamless way. Examples for this are mentioned above and further comprise implementations of intelligent spaces [28] and inhabitable interfaces [29].

Several technical implications are responsible for the current distribution of gesture interfaces and the recent development in more reliable tracking methods such as the depth camera technology suggest that future development will break up the current separation between public and semi-public environments for gestural interfaces.

The discussed examples further suggest that users do not tend to engage into interaction of high complexity and spend less time with applications in public space compared to private spaces, while the semi-public spaces are probably in the middle-ground between public and private. These issues of the complexity of applications can potentially be addressed by a meaningful staggering of different levels of engagement. User who do not want to engage simply pass by, others just have a fast glimpse at the application, while those who are ready to engage in a more extended fashion can spend time and proceed to more complex layers of the application. The notion of *proxemic* interactions [30], where this staggering of engagement is based on the distance between the user and the system is a promising approach for future applications.

References

1. Moeslund, T.B., Störring, M., Granum, E.: A Natural Interface to a Virtual Environment through Computer Vision-Estimated Pointing Gestures. In: Wachsmuth, I., Sowa, T. (eds.) GW 2001. LNCS (LNAI), vol. 2298, pp. 59–63. Springer, Heidelberg (2002)
2. Weiyuan, L.: Natural User Interface- Next Mainstream Product User Interface. In: Computer Aided Design & Conceptual Design (CAIDCD), vol. 1, pp. 203–205 (2010)
3. Schapira, E., Sharma, R.: Experimental Evaluation of Vision and Speech based Multimodal Interfaces. In: PUI 2001: Proceedings of the 2001 Workshop on Perceptive User Interfaces, pp. 1–9 (2001)
4. Zhao, J., Chen, T.: An Approach to Dynamic Gesture Recognition for Real-Time Interaction. In: Wang, H., et al. (eds.) The Sixth ISNN. AISC, vol. 56, pp. 369–377 (2009)
5. Diehl, R.L., Lotto, A.J., Holt, L.L.: Speech Recognition. Annual Review of Psychology 55, 149–179 (2004)
6. Thorsen, T.: Xbox 360 sells 50 million, Kinect ships 8 million. In: Gamespot at CES 2011 (2011), http://ces.gamespot.com/story/6285921/xbox-360-sells-50-million-kinect-8-million?tag=top_stories;title;2 (retrieved March 3, 2011)
7. Mollina, B.: Review: Microsoft Kinect marks big leap for motion gaming. In: USA Today (2010), http://content.usatoday.com/communities/gamehunters/post/201

0/11/review-microsoft-kinect-marks-big-leap-for-motion-gaming/1 (retrieved March 03, 2011)
8. Boyer, B.: 2007 U.S. Game Industry Growth Up 43% to $17.9 Billion. In: Gamasutra (2008), http://www.gamasutra.com/php-bin/news_index.php?story=17006 (retrieved March 3, 2011)
9. Ballmer, S.: Keynote CeBIT 2008 (2008), http://www.microsoft.com/presspass/exec/steve/2008/03-03cebit.mspx (retrieved March 3, 2011)
10. Cabral, M.C., Morimoto, C.H., Zuffo, M.K.: On the Usability of Gesture Interfaces in Virtual Reality Environments. In: CLIHC 2005 Proceedings of the 2005 Latin American Conference on Human Computer Interaction, pp. 100–108 (2005)
11. Dadgostar, F., Sarrafzadeh, A., Messom, C.: Multi-layered Hand and Face Tracking for Real-Time Gesture Recognition. In: Köppen, M., Kasabov, N., Coghill, G. (eds.) ICONIP 2008. LNCS, vol. 5506, pp. 587–594. Springer, Heidelberg (2009)
12. Wu, M., Balakrishnan, R.: Multi-Finger and Whole Hand Gestural Interaction Techniques for Multi-User Tabletop Displays. In: UIST 2003 Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology, pp. 193–202 (2003)
13. Villar, N., et al.: Mouse 2.0: multi-touch meets the mouse. In: UIST 2009 Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, pp. 33–42 (2009)
14. Wilson, A.D., Benko, H.: Combining Multiple Depth Cameras and Projectors for Interactions On, Above, and Between Surfaces. In: UIST 2010 Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, pp. 273–282 (2010)
15. Nielsen, M., Störring, M., Moeslund, T.B., Granum, E.: A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 409–420. Springer, Heidelberg (2004)
16. Ritter, F.E., Schooler, L.J.: The Learning Curve. In: International Encyclopedia of the Social and Behavioral Sciences, pp. 8602–8605. Pergamon, Amsterdam (2002)
17. Snibbe, S.S., Raffle, H.S.: Social Immersive Media – Pursuing Best Practices for Multi-user Interactive Camera/projector Exhibits. In: CHI 2009 Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 1447–1456 (2009)
18. Rico, J., Brewster, S.: Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability. In: CHI 2010 Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 887–896 (2010)
19. Krüger, M.W.: Responsive Environments. In: AFIPS 1977 Proceedings of the National Computer Conference, June 13-16, pp. 423–433 (1977)
20. Sester, M.: Access, http://www.accessproject.net/ (Retrieved March 3, 2011)
21. Krüger, M.W.: Videoplace – an Artificial Reality. In: CHI 1985 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 35–40 (1985)
22. Münch, W., Furukawa, K.: Bubbles, http://empac.rpi.edu/events/2005/bubbles.html (retrieved March 3, 2011)
23. Bleiweiss, A., et al.: Enhanced Interactive Gaming by Blending Full-Body Tracking and Gesture Animation. In: SA 2010 SIGGRAPH Asia Sketches (2010)
24. Yin, T.K., et al.: Integrating A Gesture Interface to a Commercial Online Dance Game. In: ICME 2009, IEEE International conference on Multimedia and Expo., pp. 1688–1691 (2009)
25. Park, H.S., Jung, D.J., Kim, H.J.: Vision-Based Game Interface Using Human Gesture. In: Chang, L.-W., Lie, W.-N. (eds.) PSIVT 2006. LNCS, vol. 4319, pp. 662–671. Springer, Heidelberg (2006)

26. Downie, M., Shirvanee, L.: Trace: Embedded Adaptive Devices in Public Spaces (2008), `http://web.media.mit.edu/~anjchang/ti01/trace.pdf` (retrieved on March 3, 2011)
27. Schwaller, M., Lalane, D., Khaled, O.A.: PyGmI – Creation and Evaluation of a Portable Gestural Interface. In: NordCHI 2010 Proceedings of the 6th Nordic Conference on Human-Computer Interaction, pp. 773–776 (2010)
28. Eng, K., et al.: Ada – Intelligent Space: An Artificial Creature for the Swiss Expo.02. In: ICRA 2003 IEEE International Conference on Robotics and Automation, vol. 3, pp. 4154–4159 (2003)
29. Huang, J., Waldvogel, M.: The Swisshouse: An Inhabitable Interface for Connecting Nations. In: DIS 2004 Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 195–204 (2004)
30. Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., Wang, M.: Proxemic Interactions: The New Ubicomp? Interactions 18(1), 42–50 (2011)