Touchless Interaction-Novel Chances and Challenges

René de la Barré, Paul Chojecki, Ulrich Leiner, Lothar Mühlbach, and Detlef Ruschin

Fraunhofer Institute for Communication Technologies,
Heinrich-Hertz-Institut, Einsteinufer 37
10587 Berlin, Germany
{Rene.de_la_Barre,Paul.Chojecki,Ulrich.Leiner,Lothar.Muehlbach, Detlef.Ruschin}@hhi.fraunhofer.de

Abstract. Touchless or empty-handed gestural input has received considerable attention during the last years because of such benefits as removing the burden of physical contact with an interactive system and making the interaction pleasurable. What is often overlooked is that those special forms of touchless interaction which employ genuine gestures – defined as movements that have a meaning – are associated with the danger of suffering from the same drawbacks as command based interfaces do, which have been widely abandoned in favor of direct manipulation interfaces. Touchless direct manipulation, however, is about to reach maturity in certain application fields. In our paper we try to point out why and under which conditions this is going to happen, and how we are working to optimize the interfaces through user tests.

Keywords: interaction, touchless, direct manipulation, gestures, hand tracking, user experience.

1 Touchless Interaction through Hand Movements

The topic of this paper is touchless interaction between humans and artificial systems, which takes place with the help of new interaction devices. Interaction devices are the units that can either execute or sense interactive behavior. For example in gaze controlled computing both the human eyes, for which the relevant behavior is their line-of-sight change, and the gaze tracker employed to monitor the eyes' movements, are interaction devices. Interaction is said to be touchless if it can take place without mechanical contact between the human and any part of the artificial system, which means that for example the interaction with a Wii through its wireless controller is not touchless. Touchless interaction can be multimodal, in which case the interactive behavior produces simultaneous events in the visual modality (color, form, or position change), in the auditory modality (speech, sounds) or in the olfactory modality (odors). In the following, however, we will be concentrating on form and position changes of the user's hands as relevant events in touchless interaction.

Such hand movements should not be equated with gestures. The proper classification of manual behavior into gesture and non-gesture helps to distinguish between two types of touchless interaction, which correspond with the well-known dichotomy between direct and indirect manipulation introduced by Shneiderman [1].

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2009, LNCS 5611, pp. 161–169, 2009.

According to a distinction made by Cadoz [2], hand movements can be of three different kinds, which the author had termed *semiotic*, *ergotic*, and *epistemic* (quoted after Mulder [3]). The semiotic kind of hand movements is used to convey meaning, the ergotic kind is used to change the state of the physical environment, and the epistemic kind is performed to feel the environment and thus to gain knowledge.

As far as gestures are concerned, a consensus seems to be emerging about what kind of behavior should count as gesture, which Kendon [4] formulated in the following way: *"'Gesture' we suggest, then, is a label for actions that have the features of manifest deliberate expressiveness"*. This means that a gesture is a body movement which is being performed with the perceivable intention to express something. According to this definition hand gestures are clearly a subclass of semiotic hand movements. Hence those movements are ruled out as gestures which are either performed to serve some practical goal, i.e. the ergotic movements, or to explore the world, i.e. the epistemic movements. Another important subclass of semiotic movements, which is, however, not of interest here, consists of the involuntary movements that are considered meaningful by observers, like trembling as a symptom of strain.

2 Interaction Styles

2.1 Direct Manipulation

Shneiderman in his already mentioned article tried to identify a common denominator of several then new and exciting user interfaces, which did not rely on written command input. From diverse examples ranging between screen-based text editors, interactive maps, and video games he concluded that this common denominator was what he termed direct manipulation. Through mentioning an article by Nelson about "the design of virtuality" [5] Shneiderman made clear that the alleged directness is often only based on "a representation of reality that can be manipulated". One of his examples for such a representation was the dragging of a document icon onto a printer icon to produce a hardcopy of the document which that icon represented.

The notion of directness may be misleading, because it only refers to the subjective impression gained from the interaction. Direct manipulation resembles what in engineering is called control, where a manipulated variable influences a controlled variable, and the controlled variable is fed back to provide information that can be used to readjust the manipulated variable. Due to the plasticity of human sensory-motor coordination an impression of directness occurs not only when the turning of a steering wheel causes car wheels to turn but also when the hand moves a mouse or a pen and thus causes a screen cursor to move. Furthermore in the discussed computer-oriented examples the user manipulates only signs, even though an illusion of influencing some kind of represented reality may occur.

The main point we would like to make from this review of direct manipulation is that the human behavior in this mode is always either ergotic or epistemic. Even if only signs are manipulated the interactive behavior serves to change some perceivable state but not to express something. An example for epistemic movements is the head motion performed to look around objects. Such head motions can be used to directly manipulate an interactive imaging system, which reacts to the motion by producing a perspective change of the rendered the image.

2.2 Indirect Manipulation

The true opposite of direct manipulation is not the use of command lines as Shneiderman suggested, but rather what could be termed *communication based interaction*. The reason for this claim is that the basic distinction which underlies Shneiderman's dichotomy is the one between "do it yourself" and "have the machine do it" approaches, as the title of one of his later articles ("Beyond intelligent machines: Just Do It!") [6] seems to confirm. However, commands can also be issued nonverbally through pointing and clicking. To use f.i. a "delete" button is certainly a command, and it is difficult to see why Shneiderman classified such actions as direct manipulation. Furthermore many of the messages exchanged during indirect manipulation are not commands in a strict sense. This is especially true for messages like "error", which are issued by the interactive system and which in most cases are only informative. However, it is also debatable whether the human input consists of nothing but commands. Compare for example the y/n feedback that many applications request.

Therefore indirect manipulation in principle comprises any type of interactive behavior that can be interpreted as a speech act in the sense of Searle [7], i.e. as an assertion, declaration, question, promise and so on, amongst which commands are just one example. Because speech acts are the minimal units of communication, indirect interaction could be properly termed communication based.

3 Hand Movement Types and Interaction Styles

As was explained above direct manipulation is generally performed through ergotic or epistemic movements. This remains true when the interaction happens in a touchless way. Imagine for example a hypothetical kind of Squash gaming application, in which a ball is shown to move and bounce from a virtual wall via computer animation and the players in front of the screen are acting as if they were hitting a real ball with their open palms. Because the palms are tracked through computer vision, information about their movements can be used to influence the animation of the ball, and thus the players can "directly" manipulate the ball.

By contrast we can consider a kind of hand movements that contain a meaning to trigger a response from a system. An example could be the formation of a thumbs-up hand configuration to operate a light switch. With that movement a human can communicate to a system that he wants it to switch the light on. Given the above definition, such a deliberately produced semiotic movement is a gesture. Therefore indirect, or communication based, manipulation correspond with gesture based interactions, whereas all direct manipulation interfaces correspond with interactions based on non-gestures.

This leads to an important conclusion. As interface designers, when we are faced with the decision whether we should incorporate gestures into the available repertoire of interactive behavior, we should be aware that any of the classical arguments against indirect manipulation interfaces also hold concerning gesture based interfaces. One of these arguments is the necessity to memorize arbitrary communication rules. This concerns even gestures that are considered to be highly plausible, because most of the words used for written commands are also highly plausible and can nevertheless be confused (c.f. "purge", "remove", or "delete" to destroy a digital entity).

Features of (verbal or non-verbal) speech acts such as intuitivity, plausibility, familiarity are usually more helpful for recognition than for recall.

4 The Naturalness of Hand-Based Interaction

Sometimes it is claimed that gesture based interaction is highly natural [8], [9] and should be therefore preferred. With a loose definition of gesture, which does not distinguish true gesture from meaningless hand movements, this claim might be justifiable. However, for true gestures the naturalness claim points to a visionary far future. The idea behind it is that of the computer as artificial human, which in its most extreme form envisages the computer to interpret facial expressions, looking behavior, spoken language, gestures and so on. With such artificial humans around the real humans would no more need to adapt themselves to the machines. Unfortunately with the current state of the art special knowledge and skills are as yet necessary for successful human-machine interaction, for which gestural type is a good example.

Individuals without disabilities use gestures almost exclusively in combination with speech. The co-speech gestures are syntactically and semantically integrated with the verbal utterance (c.f. McNeill [10]) in a way that is not yet completely understood, and their interpretation is virtually impossible, if the accompanying words are unknown. Furthermore most of the gestures are uncoded, i.e. their meaning can only be inferred and not be looked up in a lexicon. In real human machine interaction this complex way of communicating must be replaced with the use of primitive artificial sign languages, about which Cassel [11] wrote, "*I don't believe that everyday human users have any more experience with, or natural affinity for, a 'gestural language' than they have with DOS command*". An illustration of such a "gestural language" is the pen-based gesture inventory built into Microsoft's Windows Vista SDK [12], because the same gestures could also be employed in a touchless interface. In this inventory for example the "insert" command can be evoked by drawing a triangle.

5 Reasons for Touchlessness

The previous discussion relates to touch-based as well as to touchless interaction forms, but what are the special benefits of touchless user interfaces?

The conditions under which touchless or "free-form" (c.f. Saffer [13]) interactions seem to have advantages as compared with touch based interfaces include:

- Sterile environments or clean rooms: Usually, surfaces that are touched have to be made aseptic after being used. As an example, when using touch screens in operating rooms for controlling medical devices, the screens usually are covered with transparent plastic films, which have to be changed after each operation. When using touchless interaction technologies, effort, time and money can be saved.
- Vandalism-prone environments: When using cameras or similar devices for sensing fingers, hands, eyes etc. from a certain distance, it is possible to place the display device and the sensor at vandalism-proof places, e.g. behind glass walls. Among other things, this can be a useful solution for information systems at public places or for interactive store window displays.

- Co-located shared use: Interactive systems that are jointly used by a group of persons and therefore employ more distantly located large displays, f.i. in a class or lecture room.
- Placing and moving objects in three dimensions: If a task requires to move a (real or virtual) object in all three dimensions it will be easier or more convenient to use a touchless user interface, provided that it has the capability to scan a finger or hand position in x, y, and z. Such operations can be necessary - among other things - in construction tasks.
- Eventually contact-free interaction can be more suitable than contact-based interaction in cases of fleeting use, i.e. when there is extremely little time to seek, grasp and comprehend a system's input devices. These systems can be screenless, such as doors and lights in public spaces, conveyor transport systems, service robots etc.

Apart from the above examples that have been chosen on rather rational grounds there will exist cases in which the main advantage of touchless manipulation resides in the emotional sphere (c.f. [14]) or at the "visceral level", as Norman calls it [15].

Even if a touchless interactive system does not work easier, faster, more error-proof etc. than a touch-based counterpart people may nevertheless like it more and experience more fun with it. How that could be exactly explained has not yet been sufficiently investigated. Our hypothesis is that the main factor is a sense of wizard power users might gain by controlling a scene from a distance without having to touch it, an interaction style which might be called "direct manipulation by mystic means".

6 Research at Fraunhofer HHI

6.1 Interaction Concepts

For the cases in which touchless interaction is beneficial we have been developing solutions at the Fraunhofer HHI, and we have been restricting ourselves to direct manipulation interfaces as much as possible. However, in systems that incorporate more than one function, a means must be provided to switch between these functions, and it is a proven practice to employ menus or command buttons for this purpose. This means to mix up elements of direct and communication based interaction in the same user interface. The main reason why this mixture works so well is the utilization of the *menu principle*, through which the possibility of arbitrary input is replaced with a limited choice from available items.

A touchless nonverbal method to choose from a menu consists of two components, namely an addressing mechanism and an action signal, which are equivalent to pointing and clicking, respectively, in mouse based interactions. With our implementation of a touchless addressing mechanism the user can move a fingertip through the air in a predefined virtual frame and thus control a screen cursor. The virtual frame is defined through calibration of a vision based hand tracker, which can be located either above or below the hand. In our experience this frame is most conveniently placed in a fronto-parallel plane in the usual human gesture space, which is located in chest height.

To trigger the action it is necessary to define a suitable gesture. For the sake of robustness we aimed at using a simple behavior of a single finger and thus to circumvent the recognition of hand configurations. After some trial and error we were left with two candidates, amongst which we needed to decide on an experimental basis, namely

1. staying on a point, i.e. the finger or the cursor must be held over an interactive element for a minimum dwell time to start an action associated with that element (Fig. 1 left part), and
2. movement across a z-threshold, i.e. positioning the cursor over an element and then moving the finger over some distance towards the display (Fig. 1 right part). This resembles the pushing of a button in the air.

In our experiment with 20 subjects we used an interactive element in the form of a graphically represented button, and we investigated five dependant variables that are suitable to indicate the relative usability of the two candidate gestures to activate the button. Those dependent variables were

1. the learning/exploration time till the subjects felt to master the gesture subsequent to a verbal instruction
2. the precision of addressing the button during its activation
3. the minimal button size the subjects felt to be acceptable for each gesture
4. the subjective satisfaction with each kind of gesture
5. the task completion time for addressing and activating a button.

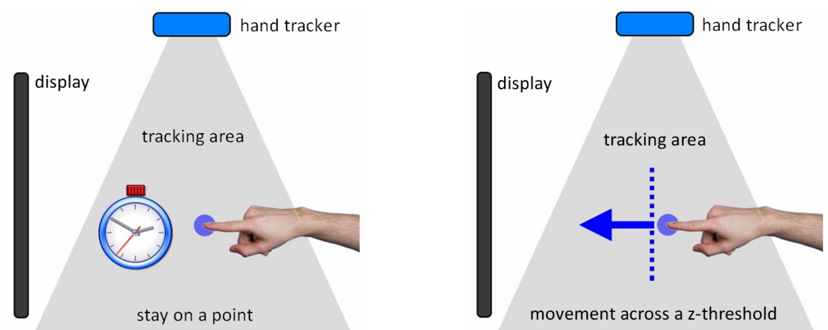

Fig. 1. Illustration of button activation by staying on a point (left) and movement across a z-threshold (right)

We found that there was a significant advantage (statistically and otherwise) for the staying on a point variant in the first four variables (Fig. 2), with task completion time, once the gesture had been learned, being the same.

As a result of these experiments we now normally employ staying on a point as an action signal in our touchless systems. We are also using the described single finger based interaction for direct manipulation. Thus touchless dragging, zooming, rotating and painting are easily possible. However, our research is ongoing, because we feel that in future comparisons slightly different z-movements might score better as an activation signal or as a grabbing action.

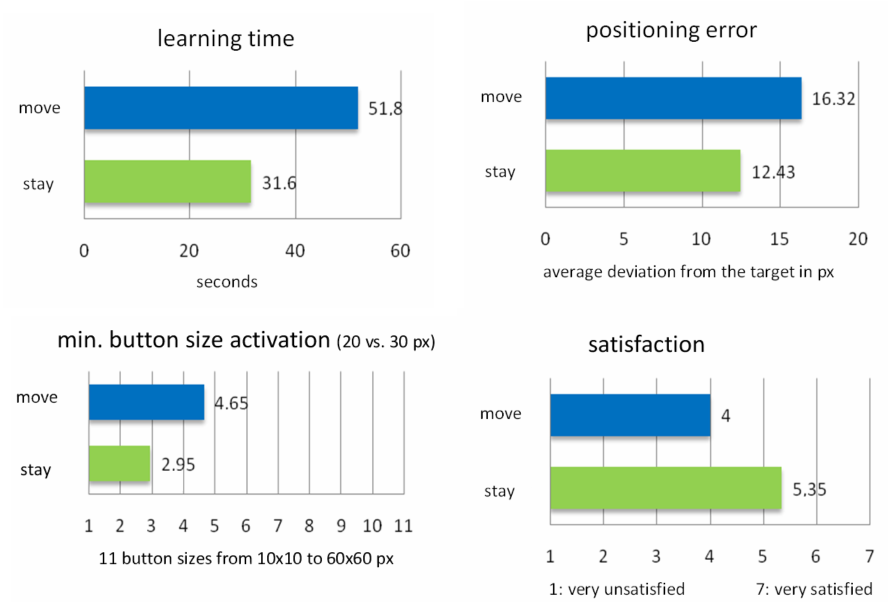

Fig. 2. Result of the gesture usability experiment

6.2 Application Scopes

Based on the mentioned scenarios that are advantageous for touchless interaction, Fraunhofer HHI has been developing various prototypes. They are used both as demonstrators on fairs or exhibitions and as testbeds for usability studies and further technology concepts [16]. Commercially available appliances include

Fig. 3. The iPoint Explorer, a touchless information system

- Information systems for public places like railway stations, town halls, museums etc. (Fig. 3)
- Presentation system for up to 120 data modules handling images, movies, slide shows or 3D objects
- Historic "Pong"-game revised for 2D and 3D contact-free playing
- Medical information system for patient data, to be used in operating theaters
- Various 3D user interfaces for autostereoscopic displays, which are also long term area of HHI's competence.

6.3 Input Devices for Hand-Based Interaction

Our technique to sense and interpret hand configurations and movements is vision-based and uses a stereo infrared-camera with active lighting. With that technique it is possible to detect and track all ten fingertips of a user (Fig. 4) in real-time with an image frequency of 50 Hz and high resolution in three spatial dimensions.

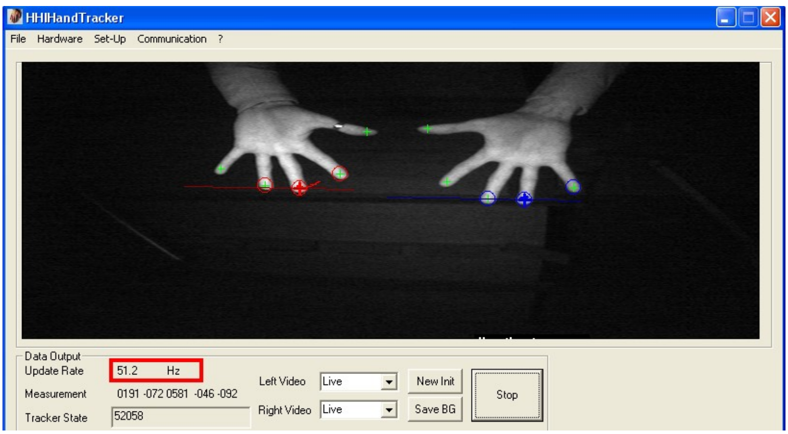

Fig. 4. Administrative user interface of Fraunhofer HHI hand tracker. Control image shows 10 detected fingertips

A number of measures like discarding static backgrounds have been contributing to the tracker's robust performance. Potential problems are the limited range of the active lighting and interfering objects in the camera's field of view that could be mistaken for fingers.

Several alternatives to our approach have been suggested in the literature or are already available on the market. Those include ultrasound position measurement devices, electrical field sensing devices, and time-of-flight depth cameras. The first two of these solutions cannot be used to recognize forms, and currently it is impossible to realize multitouch interaction with them. Therefore both are only suitable as finger mice, with three degrees of freedom, though. Electrical field sensing requires the placement of antennae near the user's hand, and some implementations of this principle require the user to stand on a metal ground plate. The time-of-flight principle is promising, but commercially available devices seem as yet to suffer from poor resolution, and the price tag of higher resolving units that are on the development roadmaps can be expected to be considerably higher as that of stereo cameras. Furthermore it remains to be seen how their performance is affected by objects in the field of view, the depth of which is mismeasured due to their problematic light reflection properties (either strong scattering or strong reflection).

Altogether the emerging techniques will soon offer useful, usable and fascinating forms of touchless interaction, but it is still necessary to carefully compare them with each other in terms of accuracy, reliability, and costs.

References

1. Shneiderman, B.: Direct manipulation: A step beyond programming languages. IEEE Computer 16(8), 57–69 (1983)
2. Cadoz, C.: Les réalités virtuelles. Dominos, Flammarion (1994)
3. Mulder, A.: Hand gestures for hci. Technical report, Hand Centered Studies of Human Movement Project, Simon Fraser University (1996)
4. Kendon, A.: Gesture: Visible Action as Utterance. Cambridge University Press, New York (2004)
5. Nelson, T.: Interactive systems and the design of virtuality. Creative Computing 6(11), 56–62 (1980)
6. Shneiderman, B.: Beyond intelligent machines: Just Do It! IEEE Software 10(1), 100–103 (1993)
7. Searle, J.: Speech Acts. Cambridge University Press, New York (1969)
8. de la Barré, R., Pastoor, S.: Conomis, Ch., Przewozny, D., Renault, S., Stachel, O., Duckstein, B., Schenke, K.: Natural Interaction in a desktop Mixed Reality environment. In: 6th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 2005 (2005)
9. Varona, J., Jaume-i-Capó, A., Gonzàlez, J., Perales, F.J.: Toward natural interaction through visual recognition of body gestures in real-time. Interacting with Computers 21(1), 3–10 (2008)
10. McNeill, D.: Gesture and thought. University of Chicago Press, Chicago (2005)
11. Cassel, J.: A Frameworkfor gesture generation and interpretation. In: Cipolla, R., Pentland, A. (eds.) Computer Vision for Human-Machine Interaction, pp. 191–215. Cambridge University Press, New York (1998)
12. Application Gestures and Semantic Behavior (Windows), `http://msdn.microsoft.com/en-us/library/ms704830(VS.85).aspx`
13. Saffer, D.: Designing Gestural Interfaces. O'Reilly, Sebastopol (2008)
14. Norman, D.A.: Emotional Design: Why We Love (Or Hate) Everyday Things. Basic Books, New York (2004)
15. Hassenzahl, M.: Hedonic, emotional, and experiential perspectives on product quality. In: Ghaoui, C. (ed.) Encyclopedia of Human Computer Interaction, pp. 266–272. Idea Group, Hershey (2006)
16. Heinrich-Hertz-Institut - Department Overview, `http://www.hhi.de/en/departments/interactive-media-human-factors/department-overview/`