Evaluation of Static and Dynamic Freehand Gestures in Device Control

Anne Köpsel (anne.koepsel@uni-ulm.de)
Department of General Psychology, Ulm University
Ulm, Germany

Anke Huckauf (anke.huckauf@uni-ulm.de)
Department of General Psychology, Ulm University
Ulm, Germany

Abstract

Increasingly, devices are controlled by gestures. Hence, questions concerning the usability of gestures arise. Most of the used gestures are promoted as 'intuitive', suggesting that training can be avoided. In the present work, we suggest certain criteria, namely recognition, learnability and executability. In addition, we present a paradigm of how to evaluate gestures with respect to these criteria. The empirical tests are based on the proposal of universal interaction design, i.e., the recommendation that gestures should be as independent from a certain task setting as possible. Preliminary data suggest that the proposed way of evaluating gesture systems as well as single gestures effectively results in a catalogue of criteria which can be weighted in their importance depending on the current task set.

Keywords: Gesture interaction; Augmented Reality; Gesture recognition

Introduction

Nowadays, controlling devices by gestures has become more and more common. Smartphones are used with multi-touch displays, Microsoft's Kinect uses full-body gestures and spoken commands instead of holding a game controller and new developed 3D motion–capturing devices (e.g., Leap Motion Controller), even recognizes the difference between fingers, thumbs and even pencils simultaneously. However, in respective evaluation studies, the usability of certain gestures is estimated according to a special scenario. In terms of a universal design of human-machine-interaction, one might wish to have one gesture system fitting for all or, at least, most of the cases. Until now, studies comparing advantages and disadvantages of various gesture systems are still missing. The aim of the present study is to do a first step into this field by examining systems of already existing manual gestures.

In the literature on input devices, applications are often promoted as based on 'intuitive' gestures (e.g., Bhuiyan, & Picking, 2009; Höysniemi, Hämäläinen, Turkki & Rouvi, 2005; Riener, Rossbory & Ferscha, 2011). But what does 'intuitive' mean when controlling an artificial device via gestures? Henseler (2011) described intuitive action as something a person can use, formerly learned action patterns for. These patterns are based on mental models; which are representative of our knowledge. In other words, in order to call something like gestures intuitive, the user's mental model has to be synchronized with the designer's one. But, as Wobbrock et al. pointed out, gestures created by designers are "not necessarily reflective of user behavior" (Wobbrock, Ringel, Morris & Wilson, 2009, p.1083). Hence, the question arises about the most relevant characteristics and criteria with regard to the usability of control gestures.

Gestures differ in various aspects. In sign language, these are the place of articulation, movement, handshape and orientation (e.g., Brentari, 2005). Various gesture systems in the literature differ unsystematically with respect to several of these features. In order to get an idea of the importance of certain characteristics, gestures of various systems must be compared. In the study at issue here, these were the following: American Sign Language (ASL, American Sign Language University, 2013) is one example of an oral language system which has already evolved for several decades. EdgeWrite (Wobbrock & Myers, 2005), which was chosen as an example for a gesture system based on written language, which is, of course, still less common. EdgeWrite has been shown to function highly accurate and stable and is suitable also for people with motor impairments (Wobbrock, Aung, Myers & LoPresti, 2005). More common gestures are wave or wipe gestures which are used with game consoles (e.g., Microsoft's Kinect) and which can be seen as the free hand version of the related touch gestures, used for controlling tablet computers and smartphones (e.g., Villamor, Willis & Wroblewski, 2010). Even if the common gesture systems are dynamic, we integrated also a static pointing system. At least for automatic gesture recognition, static gestures have many advantages and thus might become faster and better suited as controlling system. All of these gesture systems did not differ with respect to the place of articulation (Figure 1).

In terms of usability, the effectiveness and the efficiency, as well as the satisfaction while using and the acceptance of a system are important criteria to evaluate (e.g., Cuomo & Bowen, 1994).

In addition, the better the gestures will fit the mental models of as many members of the respective target group as possible, the easier the gestures should be learned and remembered. According to this, a proper gesture system should be easy and fast to learn and easily remembered, because “the user must know the set of gestures that the system recognizes” (Baudel & Beaudouin-Lafon, 1993, p.30). So, optimal gestures seem to be those which actually have no need to be learned, so they can be recognized and kept. Hence, our testing battery included gesture recognition, gesture learning, as well as gesture execution tasks.

In order to examine these usability criteria for the gesture systems, we developed a test battery, examining the most relevant characteristics of some gesture systems. We aimed at investigating free hand gesture systems, which might fit for various application scenarios, that is, which are maximally independent of any specific task set. Therefore, we investigated different systems in isolation, independent of a special control task. Thus, there was no input device presented. Such a scenario might, however, in fact function for a certain application area, as of controlling an augmented reality device with Head-Mounted-Displays via gestures (e.g., Huckauf, Urbina, Böckelmann, Schega, Mecke, Grubert, Doil, & Tümler, 2011; Saxen, Rashid, Al-Hamadi, Adler, Kernchen, & Mecke, 2012) or an oversize projection screen or in-car applications while driving. But importantly, the gestures themselves are suitable for many more application scenarios.

In summary, in order to assess the usability of these four different gesture systems, they were tested in terms of easiness of recognition without any previous knowledge, the time they need to be learned, whether they are kept and how good and fast they can be executed.

Methods

For evaluating the selected criteria, the gestures were adjusted to be executed in space without any reference surface and for just one hand which makes them usable also in environments, where the other hand is needed (e.g. driving a car).

Eight meanings or commands were chosen and therefore a gesture for each meaning of the four systems (Table 1). These eight meanings can be regarded to serve as basic commands for controlling any application. For each of these commands a video sequence was created (32 videos in total). In these short videos a gesture was presented from an ego perspective point of view (Figure 1). This allows participants to directly asses the gestures without the need of mentally rotating the gestures as becomes necessary when observing another person.

Table 1: commands/meanings used.

Gesture Meaning	
On / Off	Forward
Help	Backward
Zoom In	Up
Zoom Out	Down

Every gesture in the video started and ended with a fist, with exception of the static pointing gestures which have no movement. The fist served as start and end signal for the participants. The videos last three seconds in mean. The gestures were presented in front of a grey wall with four red dots marking the corners. These dots marked the area for the gestures in all systems and were especially used for the EdgeWrite gestures.

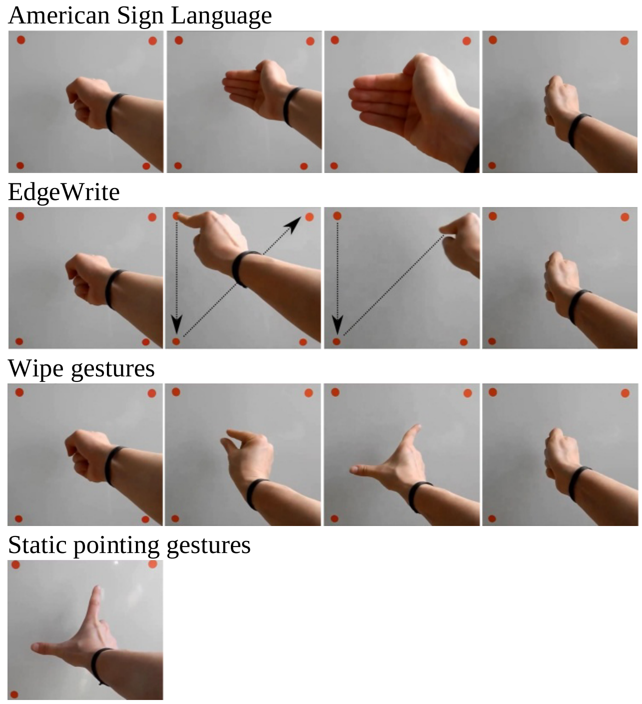

Figure 1: Video sequence in four pictures (left to right) for each of the four systems. The depict gesture is Zoom In. Every gesture starts and ends with a fist, except the static pointing gestures where no movement was shown. For ASL it was a pulled gesture, for EdgeWrite the way between the passed points is marked.

The stimuli were presented on a BenQ SensEye TFT screen. Key presses and reaction latencies were locked.
Every participant was tested with two of the gesture systems. They completed four experimental blocks for each gesture system. One session lasted about 45 min in mean.

Block 1 was a recognition task; in which the eight videos of one gesture system were presented to the participants in random order. For each video the participants had to choose which of the eight possible meanings, presented below the video, was demonstrated in the displayed gesture. Every video was presented twice (16 trials). Identification rate was measured.

In Block 2, after recognition, the same videos were presented while their respective meanings were presented with the videos. The participants were instructed to learn about the meaning of the gestures and how to execute them. During learning, each gesture was presented five times (40 trials). The duration was self-paced, that is, participants started each presentation of a new video by pressing a key.

Block 3 was a repetition of the recognition block (Block 1). This time identification rate and reaction time were measured.

In Block 4, the execution of gestures was examined. While pressing a key (spacebar), the meaning of one gesture was written on the screen. The participant's task was to release the key as soon as they can start executing the respective gesture. Every gesture meaning was presented once (8 trials). Identification rate and response latencies were measured. Between Block 3 and Block 4 the participants had to answer a short questionnaire. We report the preliminary results of twelve individuals (1 Male; mean age = 22 y).

Results and Discussion

The purpose of our investigation was finding a way to examine the usability of gestures and providing some first results demonstrating the potential of such an account.

The preliminary results presented for each block were submitted to separate analysis of variance (ANOVAs). During Block 1, before learning, best identification performance was found for the static pointing system (mean M = 73 %; standard error SE = 6%) followed by the wipe system (M = 71%; SE = 11%), the ASL system (M = 47%, SE = 12%) and finally the EdgeWrite system (M = 26%, SE = 8%). These differences are significant ($F[3, 20] = 5.622$; $p < .01$) and they show that static pointing gestures seem to be easier to understand, even without learning. These gestures fit to our knowledge about the normal use of arrows and the typical direction for going backwards, like used in web browsers to navigate. They might thus be termed intuitive for recognition.

During Block 2, participants had as much time as they needed. The times from the beginning until the end of the learning block are given in Figure 2. The learning times did not differ between the systems ($F[3, 20] = 2.688$; $p = .075$). Nevertheless, learning time was shortest for the static pointing system. This strengthens the easiness of these simple gestures and might be expected when considering the high identification rates during Block 1.

In Block 3, after learning, the wipe system produced the lowest effective performance with an identification rate of 96% (SE = 3%; static pointing: 99% [SE = 1%], ASL: 98% [SE = 2 %]; EdgeWrite: 99% [SE = 1%]). There were no differences between them ($F[3, 20] = 0.518$, $p > .68$). The respective reaction times for recognition (Block 3) after learning for the four gesture systems collapsed over meanings are depicted in Figure 3. These differences are significant ($F[3,20] = 7.896$; $p < .01$). As can be seen, the easiness of recognizing static pointing gestures is also supported by the analysis of latencies.

In Block 4, participants were asked to execute the previously learned gestures. Here, key-releasing latency (i.e., the time while pressing the spacebar, until starting performing the gesture) and the portion of correctly executed gestures were measured. It took participants the shortest times to correctly execute EdgeWrite gestures (M = 1.38 s; SE = 0.32 s), followed by the static pointing gestures (M = 1.76 s; SE = 0.29 s). Wipe (M = 2.13 s; SE = 0.28 s) as well as ASL (M = 2.2 s; SE = 0.32 s) required more time than the others. However, these differences did not reach significance ($F[3, 20] = 1.506$, $p = .24$).

The same pattern holds for the portion of correct executions. With 96% (SE = 4%) it was best for EdgeWrite, followed by the static pointing gestures with 94% (SE = 2%) and the wipe gestures with 92% (SE = 5%). It was worst for ASL with 83% (SE = 4%). Again, the differences between correctly executed gestures were not significant ($F(3, 20) = 1.708$, $p = .20$).

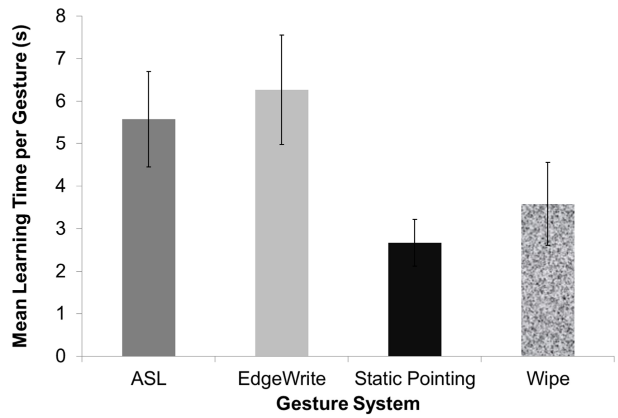

Figure 2: Mean learning time per gesture of the four different gesture systems (Error bars correspond to ±1 standard error of the mean)

Based on the current state of analysis, we would recommend static pointing gestures for device control. Static pointing gestures are simple, they seem perfectly fitted to the participant's mental model, and might thus be called intuitive. But they have a disadvantage, they have no beginning or ending. The participants made a real fast movement during execution, without keeping the gesture still, which would make recognition by a system extremely hard. Also, wipe gestures performed very well; however, eleven of our participants use a smartphone in daily lives, so one cannot exclude that they are learned well instead of being intuitive.

EdgeWrite produced good results after learning but it was not preferred by the participants. They described it as being illogical showing that most of the executed movements were regarded as unrelated to the gestures meaning. Here, it would be of interest whether an explicit explanation would suit to overcome this deficit. ASL was participant-rated similar to the wipe gestures as “good to use”. This fits to the fact that these two systems were related in the movements to each other more than to the other systems. According to these preliminary results, we gave every gesture system points from one to three for the different criteria (Table 2).

The data suggest systematic differences between the gesture systems. Before inferring about the general usability of a certain gesture system, one has to carefully investigate its single gestures to see how the single meanings and gestures may contribute to the overall effects. In order to analyse this, still more data have to be collected.

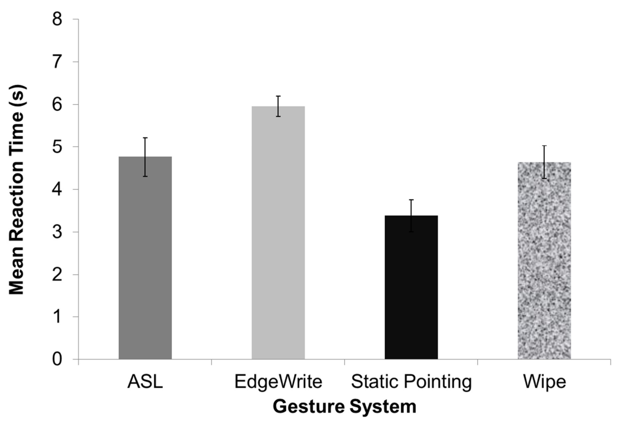

Figure 3: Reaction time for recognition (Block 3) for the four gesture systems in mean (Error bars correspond to ±1 standard error of the mean)

Based on a larger sample and a further detailed analysis, the described methods can be regarded as a powerful tool for the investigation of the usability of controlling gestures. In future research we plan to extend the criteria and investigate the usability of gesture systems for various applications and systems.

Table 2: Rating of the four gesture systems related to the criteria

	ASL	EdgeWrite	Static Pointing	Wipe
Recognition	+	++	+++	+++
Learning	+	+	+++	+++
Execution	++	+++	++	++

Acknowledgments

The work reported is performed within the Transregional Collaborative Research Centre SFB/TRR 62 “Companion-Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG). Transfer project (T01).

References

American Sign Language University (2013, February 11). Retrieved from lifeprint.com

Baudel, T. and Beaudouin-Lafon, M. (1993): Charade: Remote control of objects using free-hand gestures. *Communications of the ACM 36* (7), 28-35.

Bhuiyan, M. and Picking, R. (2009):Gesture Control User Interface, what have we done and what's next?. *Proceedings of the fifth collaborative research symposium on security, E-learning, Internet and Networking (SEIN-2009),* University of Plymouth, Darmstadt, pp. 59.

Brainard, D.H., *The Psychophysics Toolbox.* Spatial Vision 1997. 10: p. 4.

Brentari, Diane (2005). Representing handshapes in sign languages using morphological templates. *In Leuninger and Happ*, 145–177.

Cuomo, D.L. & Bowen, C.D. (1994): Understanding usability issues addressed by three user-system interface evaluation techniques. *Interacting with Computers,* Vol.6, No.1, pp. 86-108.

Henseler, W. (2011). Von GUI zu NUI: Die nächste Generation des User-Interface-Designs. *Konturen,* p. 60-67.

Höysniemi, J., Hämäläinen, P., Turkki, L., Rouvi, T. (2005). Children's intuitive gestures in vision-based action games. *Communications of the ACM - Interaction design and children*, Volume 48 (1), 44-50.

Huckauf, A. Urbina, M., Böckelmann, I., Schega, L., Mecke, R., Grubert. J., Doil, F., & Tümler, J. (2011). Perceptual issues in optical see-through displays. ACM *Applied Perception of Graphics and Visualization (APGV),* 41-48.

Riener, A., Rossbory, M., Ferscha, A. (2011): Natural DVI based on intuitive hand gestures: *Workshop Proceedings: User Experience in Cars, INTERACT 2011,* September 5, Lisbon/Portugal, Springer LNCS, 63-67.

Saxen, F., Rashid, O., Al-Hamadi, A., Adler, S., Kernchen, A., & Mecke, R. (2012) Image-Based Gesture Recognition for User Interaction with Mobile Companion- based Assistance Systems. *In Proceedings of the 4th International Conference of Soft Computing and Pattern Recognition (SoCPaR 2012),* pages 200-204.

Villamor, C., Willis, D. and Wroblewski, L. (2010) Touch Gesture Reference Guide.

Wobbrock, J.O., Aung, H.H., Myers, B.A. and LoPresti, E.F. (2005). Integrated text entry from power wheelchairs. *Journal of Behaviour and Information Technology* 24 (3). London: Taylor and Francis, pp. 187-203.

Wobbrock, J.O. and Myers, B.A. (2005). EdgeWrite: A new text entry technique designed for stability. *Proceedings of the RESNA 28th Annual Conference (RESNA '05).* Atlanta, Georgia (June 23-27, 2005). Arlington, Virginia: RESNA Press.

Wobbrock, J. O., Ringel Morris, M., Wilson, A.D. (2009) User-defined Gestures for Surface Computing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09).* Boston, USA (April 4-9, 2009). ACM New York, NY, USA, 1083-1092.