HAL open science

Better Than You Think: Head Gestures for Mid Air Input

Katrin Plaumann, Jan Ehlers, Florian Geiselhart, Gabriel Yuras, Anke Huckauf, Enrico Rukzio

To cite this version:

Katrin Plaumann, Jan Ehlers, Florian Geiselhart, Gabriel Yuras, Anke Huckauf, et al.. Better Than You Think: Head Gestures for Mid Air Input. 15th Human-Computer Interaction (INTERACT), Sep 2015, Bamberg, Germany. pp.526-533, 10.1007/978-3-319-22698-9_36 . hal-01609427

HAL Id: hal-01609427

https://inria.hal.science/hal-01609427

Submitted on 3 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Better than you Think: Head Gestures for Mid Air Input

Katrin Plaumann[1], Jan Ehlers[2], Florian Geiselhart[1], Gabriel Yuras[2], Anke Huckauf[2], Enrico Rukzio[1]

[1]Institute of Media Informatics, Ulm University, Ulm, Germany
[2]Institute of Psychology and Education, Ulm University, Ulm, Germany
{firstname}.{lastname}@uni-ulm.de

Abstract.
This paper presents a systematical comparison of pointing gestures in the context of controlling home appliances in smart homes. The pointing gestures were conducted with head, hand, arm and a computer mouse serving as baseline. To the best of our knowledge, we are the first to report on such a systematical comparison of the mentioned modalities. Our results indicate that although being overall slower and more inaccurate than hand and arm gestures, head gestures are more suitable for mid air input than previous research indicated. We show that disadvantages like slowness and inaccuracy can be compensated by a larger target size. In addition, head gestures have the largest learning effect. Considering our results and the possibilities head gestures would provide in daily life, we recommend thinking of head gestures as a feasible input modality besides hand and arm gestures.

Keywords. Pointing gestures; smart home; head gestures; comparative study

1 Introduction

In recent years, smart homes have gained more and more attention from the research community. Due to the increased accessibility of low cost sensors like the Microsoft Kinect and other depth cameras, the vision of a ubiquitous gestural interface for smart homes increasingly becomes a reality. Accordingly, mid air interaction has been a big topic in recent research [1, 4, 6, 10, 12, 13, 14, 16, 19, 20, 23].

However, all the aforementioned projects have in common that they focus on gestures performed with the complete or at least part of the arm. In contrast, little research has been done to analyse the feasibility of head gestures as an appropriate alternative. The few research projects covering head gestures are mostly motivated by impairments of users whereby head-based interaction is the only possible form of mid air interaction [9, 11]. Other projects simply focused on the detection of head gestures [8]. However, little is known about the performance of head gestures compared to gestures performed with arm or hand.

We consider this an issue for smart homes interaction research, since head gestures could be useful in many scenarios, including working, cooking and leisure activities. For example, imagine typing a letter on your PC at late afternoon. As it turns dark, you would like to turn the light on. Instead of moving your hand away from the keyboard, performing a gesture, and moving your hand back to the keyboard, all needs to be done is moving your head in the direction of the lamp. In this scenario, it is not even necessary to interrupt your writing. The same goes for adjusting the heat of the oven while chopping vegetables, and turning the music off while reading a book: using head gestures, it could still be possible to control home appliances with no (or just a very short) interruption of the main task. Appliances could be controlled directly or through a graphical interface. Graphical interfaces could be displayed either on an ambient display or projected near the device.

To find out more about the performance of head movements in comparison to hand and arm gestures, we designed and conducted a study, which is, to the best of our knowledge, the first to give insights on the relative performance of head, hand and arm gestures, compared to mouse input as a baseline. We consulted pointing gestures due to their simplicity; such movements are not bound to a certain context and occur frequently in our daily communication. In addition, pointing gestures can be used for a variety of functions, such as turning on and off any device, selecting devices and things, and interacting with graphical interfaces. So far, most research comparing different mid air input techniques has focused on comparing movements of parts of the arm [2, 3, 5, 20].

Our main contributions in this paper therefore are (1) a quantitative evaluation of pointing performance across different modalities and (2) a relative comparison showing the differences between pointing modalities.

2 User study

2.1 Participants

We invited 37 participants (six male; mean age: 22 (SE: 0.58)) to take part in our user study. All participants were students of the local universities psychology department; two of them were left-handed. All participants received course credits for their participantion.

2.2 Study design

The participants' task was to point at several targets. The configuration of the targets is shown in figure 1. The layout was designed in consideration of the recommendations by Soukoreff and MacKenzie [22]. As depicted, 17 targets were arranged. The centre target is further referred to as home. The remaining 16 circles result from the combination of the independent variables direction (up, down, left and right), size (small and large), and distance to home (short and long). The diameter of the small targets was 3.6 cm and 11 cm for the large targets. The shorter distance was 18 cm from the centre of home to the centres of the targets-to-reach and 54 cm for the longer

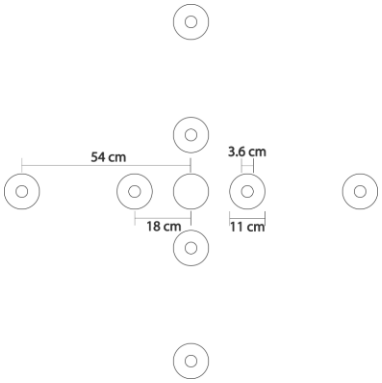
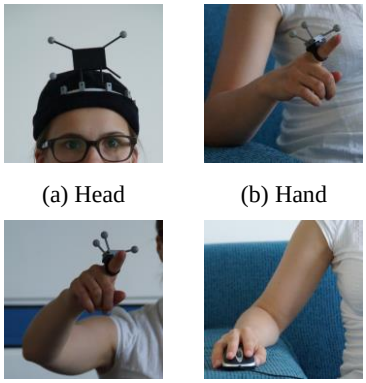


Fig. 1. Target configuration: 16 targets, two sizes and distances to the centre target ("home") as well as four directions. At each position, both small and large targets are depicted.



(a) Head (b) Hand

(c) Arm (d) Mouse

Fig. 2. The four input modalities head hand, arm, and mouse.

distance. The resulting indexes of difficulty according to [22] were 1.4, 2.6, and 4 bits. Pointing tasks were performed in a seated position, central to and with 2.5m distances to the projected targets.

The participants' task was to select the currently presented target in less than three seconds by moving a cursor from home (middle) to the currently visible target. Thereby a dwell time of 0.6 seconds had to be completed for successful selection (the dwell time was included in the maximum selection time). As a proper selection mechanism is needed to avoid the Midas Touch problem, we used dwelling. Dwelling does not need a second input modality like buttons or voice commands, and according to [21], is a proper selection technique for mid air interaction especially concerning users' perceived quality and selection accuracy. The order of target appearance was randomized; a complete round was composed of all 16 targets, whereby only home and the respective target circle were visible. For each modality, every participant passed 10 rounds. A target circle appeared when the cursor rested on home and disappeared after the feedback was given. If a target was successfully selected, it was marked green and a success tone was played. Otherwise, the target was marked red and an unpleasant no-success tone resounded.

Using this set-up, we compared the four input modalities head, hand, arm and mouse as depicted in figure 2. Head pointing was achieved by moving the head. To point with their hand, participants placed their elbow on an armrest. Arm pointing was performed out of the shoulder joint with the complete arm outstretched. As surface for the mouse, an armrest was used. Every participant had to use all four input modalities for all target sizes, distances and directions, resulting in a 4 X 2 X 2 X 4 factorial within subject design. The order of the modalities was counterbalanced. As dependent variables we measured "time to succeed": the duration it took to successfully select an envisaged target, starting when the target appeared and ending when the target

was successfully selected (only successful selections were used); and "success rate": the amount of successfully selected targets among all presented targets.

2.3 Procedure

After explaining the background, purpose and course of the study, participants signed an informed consent. Every participant was given time to acquaint with all input modalities. When the participants felt ready, we started with the first modality as described above. When the modalities changed, participants where given as much time as they needed to familiarise with the new input technique. During the course of the study, participants could take breaks at any time.

2.4 Gesture recognition

To recognise the head, arm and hand movements, we used the marker based optical tracking system OptiTrack [18] because of its accuracy. Participants wore either a cap or a finger ring (depending on the tested input modality); both were equipped with retro-reflecting markers, as can be seen in figure 2. The markers were tracked with eight cameras, places around the sitting position of the participants. The system was calibrated as suggested by the manufacturer, resulting in millimeter accuracy. For each wearable, a rigid body was defined. To calculate the coordinates of the cursor participants had to move, ray casting was implemented. Therefor, the position of the particular rigid body and its orientation were retrieved using the NatNetSDK [17]. The position of the rigid body served as the starting point of the ray, while the rigid bodies' orientation defined the direction of the ray. To avoid natural hand tremor to disturb the cursor position, the 1€-Filter [7] was implemented. The intercept was set to 0.6 while beta was set to 0.04.

2.5 Results

Overall results. Figure 3 shows the resulting "time to succeed" and "success rate" for each modality. As can be seen, mouse was the fastest input modality (m = 1524.6ms, SE = 17.45), followed by arm (m = 1790.0ms, SE = 21.93), hand (m = 1795.9ms, SE = 26.93) and head (m = 2042.2ms, SE = 17.46). The same pattern can be observed for the success rate, with mouse being the most successful (m = 99.6%, SE = 0.24) followed by arm (m = 91.4%, SE = 0.91), hand (m = 87.9%, SE = 2.03) and head (m = 76.7%, SE = 1.75).

Due to inhomogeneity of variances, a Friedman test was used to compare time to success and success rates of the four input modalities. The Friedman test revealed that both "time to succeed" and "success rate" differ significantly ($X^2(3) = 95.23$, $p < 0.01$ and $X^2(3) = 84.34$, $p < 0.01$ respectively). Regarding the time to succeed, a pairwise comparison of all modalities with t-tests (normality was assured by Shapiro-Wilks tests) showed that except for hand and arm, the differences between all modalities are significant ($p < 0.001$). To pairwise compare the success rate, Wilkoxon-Sign-Rank

tests were used since the data was not normally distributed. As with the time to succeed, the only not significant ($p < 0.01$) difference was between hand and arm.

Results according to distance from centre. As expected, targets within the short distance from the centre were faster and more accurately selected than the targets within the longer distance. This holds true for all input modalities. Figure 4 shows the mean time to succeed and the success rate for targets with short and targets with long distance to the centre.

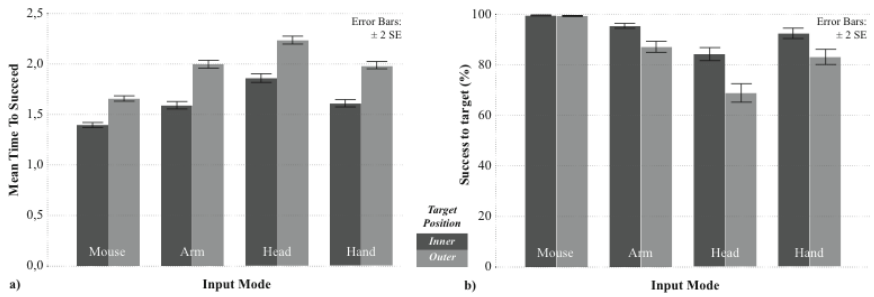


Fig. 3. Mean time to succeed (a)) and success rate (b)) for targets with short (inner) and long (outer) distance to home.

Results according to target size. Figure 4 shows the mean times to succeed and the mean success rates. For the mean times to succeed, the same pattern as described above could be found, with mouse being the fastest modality followed by hand, arm and head for both target sizes. The main difference is that larger targets could be selected faster than smaller ones, as expected. The same is true for the success rates for the smaller targets. The success rates for the larger targets, however, deviate from this pattern in two ways. First, all success rates are rather high with above 96%. Second, although a Friedmans test showed that the input modalities differ significantly ($X^2(3) = 82.85$, $p < 0.001$), there is no significant difference between the success rate for head gestures and the success rate for hand gestures, as pairwise comparisons with Wilcoxon Sign Rank tests showed.

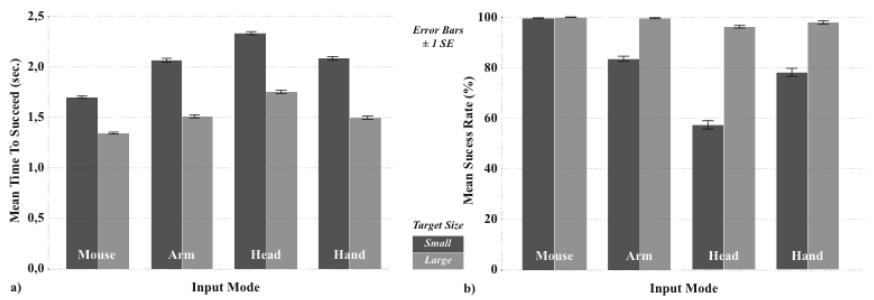


Fig. 4. Mean time to succeed (a)) and success rate (b)) for large and small targets.

Figure 4 also shows that the reaction of ponderous effectors (like head) to an increased index of difficulty is stronger, as proposed in [15].

Learning effects. Figure 5 shows the success rate for each of the ten rounds for all target sizes, large targets and small targets. As can be seen, the improvement is especially high for head gestures.

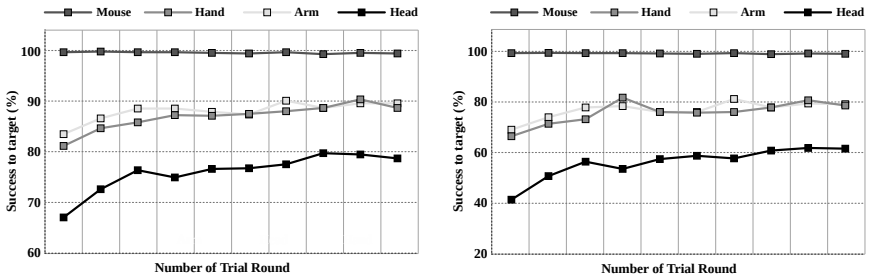


Fig. 5. Accuracy for all four modalities for every round for all target sizes (a)), and small targets (b).

During the ten rounds of our study trial, participants improved their success rate in head pointing from 66.9% in the first round to 78.6% in the last round. For small targets, this increase was even higher, from 41.4% to 61.7%. A Friedmans test revealed that the overall differences among the pointing modalities are significant ($p < 0.001$). A pairwise comparison with Wilcoxon Signed Rank tests showed that the differences between mouse and head, hand and arm respectively were significant ($p < 0.05$).

3 Discussion

A key finding of the presented study is the performance of head-based pointing gestures. Although being the slowest and most error-prone pointing modality (compared to the overall results), the performance difference to selection gestures based on head, hand or arm movements is smaller than expected (time to succeed: 268.2ms and 253.7ms respectively, success rate: 14.4% and 11.2% respectively). Furthermore, head pointing can be comparably successful, if an appropriate target size is available. Despite being labeled "large" and "small" the current target sizes are in fact rather small. With 11 cm in diameter, the large circles are comparable to the size of most modern light switches. Our results showed that at this size, head pointing can be compared to hand pointing and hast a relatively high success rate. It should also be considered that mid air interaction is predominantly used over certain distances; consequently, active selection fields in smart home environments have to feature an appropriate extent to ensure smooth and trouble-free interaction.

We also observed learning effects among all mid air interaction modalities, with head-based pointing featuring the largest performance increase. It is obvious that new input techniques (in this case head pointing) lead to difficulties, particularly at the

beginning and that subjects will initially perform worse compared to more usual movements (e.g. mouse input). However, the current learning results indicate, that practicing head-based pointing behaviour brings out the capacity for quick and precise interaction beyond the conventional motor pathways for pointing. This should not only open the possibility for alternative input channels but may enable simultaneous working steps and bypasses the necessity of interrupting parallel tasks.

4 Conclusion

This paper presented a comparative study of pointing gestures performed with head, hand, arm and mouse pointing gestures. We systematically compared all three modalities in a pointing task with regard to "time to succeed" and "success rate". The intresults indicated that head-based pointing gestures can perform almost equally well compared to arm or hand movements; always assuming that targets feature an appropriate size. Furthermore, target selection performance over an extended period exhibited a steeper learning curve for head-based movements in comparison to the alternative input modalities.

Considering the presented results and the current state of research regarding head gestures, we strongly recommend thinking of head gestures as a feasible alternative or at least additional way to interact in future smart home applications.

Acknowledgements

This work was conducted within the project "Gaze- and Gesture-Based Assistive Systems for Users with Special Need" funded by the BMBF and was supported by the Transregional Collaborative Research Centre SFB/TRR 62 Companion Technology for Cognitive Technical Systems funded by the DFG.

References

1. Anastasiou, D., Jian, C., and Zhekova, D.: Speech and Gesture Interaction in an Ambient Assisted Living Lab. In SMIAE 2012, pp. 18-27, 2012.Association for Computational Linguistics.
2. Balakrishnan, R., and MacKenzie, I. S.: Performance differences in the fingers, wrist, and forearm in computer input control. In CHI 1997, pp. 303-310. ACM, 1997.
3. Boyle, J. B., and Shea C. H.: Wrist and arm movements of varying difficulties. In Acta psychologica 137, no. 3 2011, pp. 382-396. 2011
4. Brown, J. N., Kaufmann, B., Huber, F. J., Pirolt, K. H., and Hitz, M.: "… Language in Their Very Gesture" First Steps towards Calm Smart Home Input. In HCI-KDD, pp. 256-264. Springer 2013.
5. Cao, X., Villar, N., and Izadi, S.: Comparing user performance with single-finger, whole-hand, and hybrid pointing devices. In CHI 2010, pp. 1643-1646, 2010. ACM.

6. Caon,M., Yue, Y., Tscherrig, J., Mugellini, E., Abou Kahled, O.: Context-aware 3d gesture interaction based on multiple kinects. In AMBIENT 2011, pp. 7-12, 2011
7. Casiez,G., Roussel, N., and Vogel, D.: 1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems. In CHI 2012, pp. 2527-2530, 2012. ACM.
8. Galanakis,G., Katsifarakis, P., Zabulis, X., and Adami, I.: Recognition of Simple Head Gestures Based on Head Pose Estimation Analysis. In AMBIENT 2014, pp. 88-96, 2014.
9. Gray, J. O., Jia, P., Hu, H. H., Lu, T., and Yuan, K.: Head gesture recognition for hands-free control of an intelligent wheelchair. In Industrial Robot: An International Journal, 34(1), pp. 60-68, 2007.
10. Grosse-Puppendahl, T., Beck, S., Wilbers, D., Zeiß, S., von Wilmsdorff, J., and Kuijper, A.: Ambient Gesture-Recognizing Surfaces with Visual Feedback. In N. Streitz and P. Markopoulos (Eds.), Distributed, Ambient, and Pervasive Interactions, pp. 97-108, 2014. Springer.
11. Hirsch, M., Cheng, J., Reiss, A., Sundholm, M., Lukowicz, P., and Amft, O.: Hands-free Gesture Control with a Capacitive Textile Neckband. In ISWC 2014, pp. 55-58, 2014. ACM.
12. Kim,D., and Kim, D.: An intelligent smart home control using body gestures. In ICHIT 2006, pp. 439-446,2006. IEEE.
13. Kim, H. J., Jeong, K. H., Kim, S. K., and Han, T. D.: Ambient wall: Smart wall display interface which can be controlled by simple gesture for smart home. In SIGGRAPH Asia 2011 Sketches, p. 1, 2011. ACM.
14. Kosunen, I., Jylha, A., Ahmed, I., An, C., Chech, L., Gamberini, L., Cavazza, L., Jacucci, G.: Comparing Eye and Gesture Pointing to Drag Items on Large Screens. In ITS 2013, pp. 425–428, 2013. ACM.
15. Langolf, G. D., Chaffin, D. B., and Foulke, J. A.: An investigation of Fitts‘ Law using a wide range of movement amplitudes. Journal of Motor Behavior, 8, pp. 113-128, 1976.
16. Lin, C.-Y., and Lin, Y.-B.: Projection-Based User Interface for Smart Home Environments. In COMPSACW 2013, pp. 546-549, 2013. IEEE.
17. NaturalPoint Inc.: NatNet SDK. https://www.optitrack.com/products/natnet-sdk/, 2015. (Last Access: 2015/03/24)
18. NaturalPoint Inc.: OptiTrack. https://www.optitrack.com/, 2015. (Last Access: 2015/03/24)
19. Neßelrath, R., Lu, C., Schulz, C. H., Frey, J., and Alexandersson, J.: A Gesture Based System for Context–Sensitive Interaction with Smart Homes. In Ambient Assisted Living pp. 209-219, 2011. Springer.
20. Sambrooks, L., and Wilkinson, B.: Comparison of Gestural, Touch, and Mouse Interaction with Fitts’ Law. In OzCHI 2013, pp. 119–122, 2013. ACM.
21. Schwaller, M., & Lalanne, D. : Pointing in the Air: Measuring the Effect of Hand Selection Strategies on Performance and Effort. In SouthCHI 2013, pp. 732–747, 2013. Springer.
22. Soukoreff, R. W., and MacKenzie, I. S. : Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts’ law research in HCI. International Journal of Human-Computer Studies, 61(6), pp. 751–789, 2004.
23. Vatavu, R.-D.: A Comparative Study of User-defined Handheld vs. Freehand Gestures for Home Entertainment Environments. J. Ambient Intell. Smart Environ., 5(2), pp. 187–211, 2013.