# EyePointing: A Gaze-Based Selection Technique

**Robin Schweigert**
University of Stuttgart
Stuttgart, Germany
robin.schweigert@gmail.com

**Valentin Schwind**
University of Regensburg
Regensburg, Germany
valentin.schwind@acm.org

**Sven Mayer**
Carnegie Mellon University
Pittsburgh, USA
info@sven-mayer.com

## ABSTRACT

Interacting with objects from a distance is not only challenging in the real world but also a common problem in virtual reality (VR). One issue concerns the distinction between attention for exploration and attention for selection – also known as the Midas-touch problem. Researchers proposed numerous approaches to overcome that challenge using additional devices, gaze input cascaded pointing, and using eye blinks to select the remote object. While techniques such as MAGIC pointing still require additional input for confirming a selection using eye gaze and, thus, forces the user to perform unnatural behavior, there is still no solution enabling a truly natural and unobtrusive device free interaction for selection. In this paper, we propose *EyePointing*: a technique which combines the MAGIC pointing technique and the referential mid-air pointing gesture to selecting objects in a distance. While the eye gaze is used for referencing the object, the pointing gesture is used as a trigger.

## CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Pointing**; **Gestural input**.

## KEYWORDS

MAGIC pointing; mid-air pointing; ray casting; eye tracking; selection technique.

**ACM Reference Format:**
Robin Schweigert, Valentin Schwind, and Sven Mayer. 2019. EyePointing: A Gaze-Based Selection Technique. In *Mensch und Computer 2019 (MuC '19), September 8–11, 2019, Hamburg, Germany.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3340764.3344897

*MuC '19, September 8–11, 2019, Hamburg, Germany*

ACM ISBN 978-1-4503-7198-8/19/09…$15.00
https://doi.org/10.1145/3340764.3344897

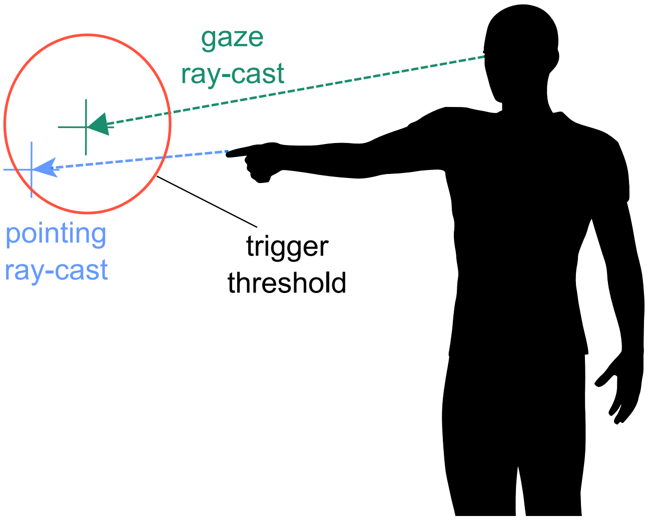

**Figure 1: The proposed interaction concept. While eye tracking is used for estimating the target position, the pointing ray cast is used to trigger a selection.**

## 1 INTRODUCTION AND BACKGROUND

Already in 1980, Bolt [2] proposed to replace the mouse to interact with objects on large screens from a distance using mid-air gestures. In the future, we envision human skeleton tracking will move rom on-body marker based systems (e.g., OptiTrack, Vicon) to room scale pervasive tracking systems (e.g., Kinect skeleton tracking) [7, 8, 23]. Moreover, we not only see tracking systems that are pervasive but we also envision eye tracking devices. While high accuracy eye trackers have been stationary for a long time, they became mobile in the last years while still relying on infrared cameras. However, eye tracking is becoming pervasive in the environment and even using standard RGB cameras [3, 9, 11, 28]. Thus, wearing body-worn eye trackers will not be necessary anymore. In the following, we provide a literature review of the two parts leading up to a novel interaction technique called *EyePointing*. This method combines skeleton and eye tracking to provide users with a natural interaction method allowing the user to interact, select, and reference objects in the distance; thus, overcoming the Midas-touch problem.

### Mid-Air Pointing

In contrast to relative ray casting techniques, where users are directing a cursor bound to their relative movements, absolute mid-air pointing ray casting methods [26] enable an interaction which does not rely on explicit visual feedback such as a crosshair or mouse cursor. Thus, we focus on absolute ray casting techniques.

Argelaguet et al. [1] distinguish between *eye-rooted* and *hand-rooted* techniques. For *eye-rooted* techniques, researchers presented a set of viable options: head ray cast (HRC) [19], and eye-finger ray cast (EFRC) [20]. However, today EFRC actually uses the "Cyclops Eye", which is the position between the eyes as root [12, 15]. On the other hand, *hand-rooted* techniques use the hand as the origin for the ray [17, 18]. Here, two techniques are most common: index finger ray cast (IFRC) [4], and forearm ray cast (FRC) [19]. These ray-casting techniques support the user to point on distant objects, however, do not enable selecting these objects. Therefore, Vogel and Balakrishnan [26] present *AirTap* and *ThumbTrigger* both with sound and visual feedback to counteract the lost physical feedback when using a mouse to select objects. Mayer et al. [15, 16] showed that the accuracy is still limited even with high precision motion tracking. However, they presented a correction model to increase accuracy. Moreover, Mayer et al. [15] showed that presenting feedback can be beneficial for accuracy, however, it increases the task completion time (TCT).

### MAGIC Pointing

The concept of MAGIC pointing was first presented by Zhai et al. [27]. This technique enables the user to rapidly overcome large distances on a screen by fixating the target and pressing a dedicated trigger to activate a cursor warp. For small and precise selections, the user can still use traditional mouse input after a cursor warp.

Drewes and Schmidt [6] proposed using a touch-sensitive mouse for triggering cursor warps. On the other hand, Vertegaal [25] substituted the trigger with dwell time which increases TCT and lowers accuracy. Zhai et al. [27] showed MAGIC pointing can reduce the required effort and TCT. Dickie et al. [5] showed that for pure screen switching MAGIC pointing was 110% faster than the mouse. In contrast, Lischke et al. [14] showed that on a 165″ large high-resolution display (LHRD) the TCT is worse than for the mouse. To enrich MAGIC pointing various projects enhanced the interaction such as Turner et al. [24] who added rotation support during drag and drop operations using a tablet. Later, Kytö et al. [13] combined MAGIC pointing with manual input to further improve eye pointing target accuracy for precise interaction in augmented reality (AR).

### Midas-Touch Problem

Midas-touch Problem references the Greek mythology in which King Midas turned anything into gold he touched. This prevented him from touching anything anymore and in the legend, he starved to death. This problem can be transferred to human-computer interaction for all channels which are always active. As the problem arises to distinguish between natural movement and actual interactions. For instance, gaze movements are happening constantly, but not always as an input for a system [10]. Thus, is often problem is referred to as the Midas-touch problem.

### Summary

While mid-air pointing offers a natural way to point at objects in the distance, it lacks the natural possibility to confirm a selection to complete the input interaction. As this would require the user to always carry a selection trigger (e.g., button) with them; however, this is not natural nor convenient. Moreover, we see that MAGIC pointing suffers the same issues. The pointing technique of using eye gaze is pervasive, however, misses a natural trigger for selection. Thus, the proposed approaches still all have the Midas-touch problem or none natural interaction. Thus, in the following, we propose *EyePointing* to overcome the Midas-touch problem when selecting objects in the distance.

## 2 EYEPOINTING

We propose *EyePointing* to overcome the Midas-touch problem allowing the user to reference objects in a distance in a natural way. *EyePointing* combines the gaze direction and natural human pointing gesture to interact with objects from the distance. The interaction itself is a two-step process. In the first step, the target object needs to be referenced by the gaze of the user. Afterwards, the user needs to perform a pointing gesture triggering the action if the pointing gesture is within the *trigger threshold*. Potentially, any action associated with the object itself can be triggered. For example, in a smart home environment, it would be possible to increase and decrease the temperature by looking on predefined areas near a radiator and then point at to for selection. In addition, voice commands could potentially enrich the interaction, to trigger a variety of actions associated with the object. The whole interaction process is presented in Figure 1.

As gaze direction, we propose using the 3D gaze vector provided by any modern eye tracker, in the following labelled as gaze ray cast (GRC). For the human pointing gesture literature has shown a number of ways to ray cast the humans' body posture to a distant object. However, Mayer et al. [15] showed that eye-finger ray cast (EFRC) is the most accurate. In the following, we will use EFRC as the default pointing ray-casting technique; however *EyePointing* is not limited to EFRC, other ray casting technique such as index finger ray cast (IFRC) will work as well with an adjusted trigger threshold.

As any pointing gesture would, therefore, trigger an action, we use a *trigger threshold*, see Figure 1. Therefore, whenever the *threshold* is larger than the distance between GRC and EFRC an action is triggered. Thus, this *threshold* reduces false positive interactions.

We see *EyePointing* as a major advantage over previously presented solutions as it enables natural controller-free interaction with objects in a distance. Moreover, the use of eye gaze as interaction will not fatigue the eye muscles more than with traditional mouse pointing [6].

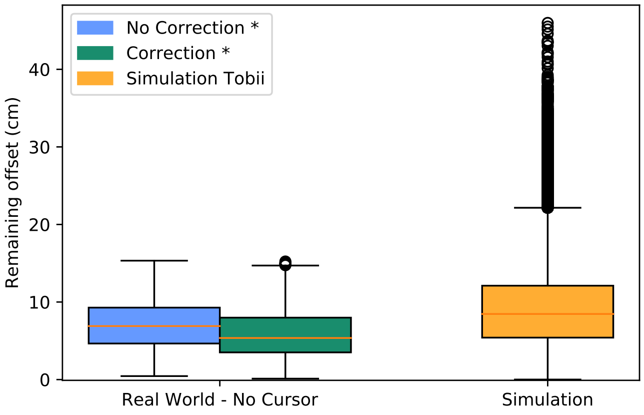

**Figure 2: Remaining offsets for simulation for the Tobii X3-120 ($\mu$ = .4°, $\sigma$ = .24°) compared to pointing in a real world study presented in a box plot. With the comparison to the results by Mayer et al. [15] who implemented a pointing correction model for mid-air pointing.**

## 3 USE CASES

Our proposed combination of MAGIC pointing and mid-air pointing has a wide range of possible use cases. *EyePointing* will open a whole new way to interact naturally and controller-free.

### Smart Homes

With voice assistants like Alexa and Google Home voice interaction made its way into homes. The latest versions can even make use of displays in the home to present content. However, when turning on a light using a voice assistant referencing a specific light bulb with a bulb specific reference label. Memorizing all lights in the room is already hard but with an increasing number of? smart objects such as smart blinds, this is becoming progressively complex. Here, *EyePointing* offers a way to enable a truly natural multimodal interaction.

### Large high-resolution display (LHRD)

In the last years, screens have got bigger. Today, single screens upwards of 80 inches are available to consumers featuring up to 8K. Stitching multiple screens together offer a large interaction space. Thus, the media room envisioned by Bolt [2] already in 1980 can become reality soon. However, interacting on such a large display with the mouse and keyboard is cumbersome and often not practical. *EyePointing* can replace today's cumbersome interaction and offer a fast alternative to reference and interact with content on the screen. We see *EyePointing* especially as the perfect solution for short and rapid interactions, such as in meetings and exhibitions where it is not feasible to equip everyone with the appropriate hardware tools to interact with the LHRD. Thus, *EyePointing* also promotes collaborative work in meetings.

### AR and virtual reality (VR)

VR gear, like the HTC Vive and the Oculus Rift, is providing a high-quality VR experience and with the Oculus Go VR experience becomes more portable and more affordable than ever before. Moreover, while some already work without an external tracking system, they all still rely on a controller to interact with the virtual environment. As Mayer et al. [15] showed that mid-air pointing performance in VR and the real world is similar, we envision *EyePointing* as a mean to interact with virtual content when using VR headsets.

This is not only true for VR headsets but also for full body cave automatic virtual environment (CAVE) setups like the one presented by Fender et al. [7]. CAVEs have the unique advantage over VR headsets that the user can see their own body without rendering a virtual body. We argue that this will make *EyePointing* interaction more natural and immersive.

The first AR devices such as the MOVERIO BT-100 and 200 had a handheld touch controller as the main mean of input. Later versions such as the Google Glas were using mainly voice control but also had a multi-touch gesture input control element mounted to the side of the glasses temples. Finally, the HoloLense can be operated without an external selection device, relying purely on gestures; however, it is still being shipped with a controller. Here, *EyePointing* can replace today's controller input or replace the HoloLenses' mid-air tap gesture.

## 4 SIMULATION

We were not able to conduct a user study of the whole interaction technique since today's hardware is not accurate enough but in an effort to quantify the theoretical accuracy of the *EyePointing* as an interaction technique using today's technology we conducted a simulation using industry-grade hardware specifications.

**Table 1: Remaining Offsets for pointing and eye tracking in cm at $2m$ distance.**

| | Input | M | SD |
|---|---|---|---|
| EFRC | w/o correction | 7.1 | 3.3 |
| | w/ correction | 6.2 | 3.6 |
| GRC | $\mu = .4^\circ$, $\sigma = .24^\circ$ | 9.2 | 5. |
| | $\mu = 4.2^\circ$, $\sigma = 6.1^\circ$ | 28.9 | 16.2 |

Previous work has shown that EFRC is the most accurate mid-air pointing technique [15, 16]. However, when using *EyePointing* the ray-cast of the gaze direction is used to determine the selection in sight. Thus, determining the gaze direction needs to be as accurate as possible. To understand how well today's eye trackers are suitable for distant object selection and if they outperform presented ray casting techniques, we simulate *EyePointing* interaction. Our simulation is based on the study design by Mayer et al. [15]. While they asked their users to point on targets in a $2m$ distance, our simulated participants looked at targets in $2m$ distance.

We modelled body sizes for 50% female and 50% male with a Gaussian distribution according to the anthropometric data [21]. We then simulated GRC for each simulated participant and intersection with a grid of targets on a plane $2m$ in front of the person. In total, we simulated 10.000 GRC for each target summing up to a total of 350.000 simulated GRCs. To simulate the GRC we again use a Gaussian distribution with different $\mu$'s and $\sigma$'s reported by state-of-the-art eye tracking reports. For this purpose, the perfect ray between the simulated participant's eye and the target is rotated both up/down and left/right randomly according to this Gaussian distribution. When assuming accuracy and precision as reported by Tobii for their X3-120 ($\mu = .4^\circ$, $\sigma = .24^\circ$) we get an offset of $M = 9.2$ ($SD = 5.0$), see Figure 2 and Table 1. Using reported measurement values by Schüssel et al. [22] ($\mu = 4.2^\circ$, $\sigma = 6.1^\circ$) we even have an offset of $M = 28.9$ ($SD$ = 16.2) for SMI Glasses 2.0.

Our simulation, shows that today's eye-tracking technology can operate in a similar range as a high-precision motion tracking system [15]. However, this also requires equipping the user with makers, a truly device free system would need to rely on less accurate tracking techniques, e.g. depth-camera skeleton tracking. Therefore, the combination of eye-tracking in the environment [11] and depth-camera tracking is a viable step toward a truly natural interaction.

The 350.000 GRCs represent the same number of participants each pointing once on the target. In contrast to Mayer et al. [15, 16] we could not average over multiple samples per participant, resulting in a wider spread of the data, c.f. Figure 2.

## 5 CONCLUSION

In this work, we proposed a new interaction technique to overcome the Midas-touch problem when interacting with objects in a distance. Here, we proposed using the eye gaze for object referencing and the pointing gesture with the eye-finger ray cast as a trigger for natural interaction. We showed various use cases such as in VR where this technique can be deployed. Moreover, we showed while today's eye tracking still suffers accuracy, it potentially enables users to naturally interact with their surrounding without using an additional device such as a mouse or unnatural behaviour.

While in this paper we proposed and showed the feasibility of using *EyePointing* as a way to naturally interact with objects in the distance, there are a number of open questions which we aim to address in the future. As a next step, we plan to determine the *trigger threshold*. Moreover, in the future, we want to implement the system which is capable of *EyePointing* input. We plan to study different setups such as smart home environment and *EyePointing* as a controller replacement for AR and VR. Finally, we want to investigate the performance when displaying feedback to help the user during selection time as proposed by Vogel and Balakrishnan [26].

## ACKNOWLEDGEMENTS

This work was financially supported by the German Research Foundation (DFG) within Cluster of Excellence in Simulation Technology (EXC 310/2) at the University of Stuttgart.

## REFERENCES

[1] Ferran Argelaguet, Carlos Andujar, and Ramon Trueba. 2008. Overcoming Eye-hand Visibility Mismatch in 3D Pointing Selection. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology (VRST '08)*. ACM, New York, NY, USA. https://doi.org/10.1145/1450579.1450588

[2] Richard A. Bolt. 1980. Put-that-there: Voice and Gesture at the Graphics Interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '80)*. ACM, New York, NY, USA. https://doi.org/10.1145/800250.807503

[3] Dong-Chan Cho and Whoi-Yul Kim. 2013. Long-Range Gaze Tracking System for Large Movements. *IEEE Transactions on Biomedical Engineering* (Dec 2013). https://doi.org/10.1109/TBME.2013.2266413

[4] Andrea Corradini and Philip R. Cohen. 2002. Multimodal speech-gesture interface for handfree painting on a virtual paper using partial recurrent neural networks as gesture recognizer. https://doi.org/10.1109/IJCNN.2002.1007499

[5] Connor Dickie, Jamie Hart, Roel Vertegaal, and Alex Eiser. 2006. LookPoint: An Evaluation of Eye Input for Hands-free Switching of Input Devices Between Multiple Computers. In *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments (OZCHI '06)*. ACM, New York, NY, USA. https://doi.org/10.1145/1228175.1228198

[6] Heiko Drewes and Albrecht Schmidt. 2009. The MAGIC Touch: Combining MAGIC-Pointing with a Touch-Sensitive Mouse. In *Human-Computer Interaction (INTERACT '09)*. Springer, Berlin, Heidelberg.

[7] Andreas Fender, Philipp Herholz, Marc Alexa, and Jörg Müller. 2018. OptiSpace: Automated Placement of Interactive 3D Projection Mapping Content. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 269. https://doi.org/10.1145/3173574.3173843

[8] Juergen Gall, Carsten Stoll, Edilson de Aguiar, Christian Theobalt, Bodo Rosenhahn, and Hans-Peter Seidel. 2009. Motion capture using joint skeleton tracking and surface estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*. https://doi.org/10.1109/CVPR.2009.5206755

[9] Craig Hennessey and Jacob Fiset. 2012. Long Range Eye Tracking: Bringing Eye Tracking into the Living Room. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA. https://doi.org/10.1145/2168556.2168608

[10] Robert JK Jacob. 1995. Eye tracking in advanced interface design. *Virtual environments and advanced interface design* (1995).

[11] Mohamed Khamis, Axel Hoesl, Alexander Klimczak, Martin Reiss, Florian Alt, and Andreas Bulling. 2017. EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA. https://doi.org/10.1145/3126594.3126630

[12] Alfred Kranstedt, Andy Lücking, Thies Pfeiffer, Hannes Rieser, and Marc Staudacher. 2006. Measuring and Reconstructing Pointing in Visual Contexts. In *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue (brandial'06)*. Universitätsverlag Potsdam, Potsdam.

[13] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 81. https://doi.org/10.1145/3173574.3173655

[14] Lars Lischke, Valentin Schwind, Kai Friedrich, Albrecht Schmidt, and Niels Henze. 2016. MAGIC-Pointing on Large High-Resolution Displays. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA. https://doi.org/10.1145/2851581.2892479

[15] Sven Mayer, Valentin Schwind, Robin Schweigert, and Niels Henze. 2018. The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 653. https://doi.org/10.1145/3173574.3174227

[16] Sven Mayer, Katrin Wolf, Stefan Schneegass, and Niels Henze. 2015. Modeling Distant Pointing for Compensating Systematic Displacements. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA. https://doi.org/10.1145/2702123.2702332

[17] Mark R. Mine. 1995. *Virtual Environment Interaction Techniques*. Technical Report. University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[18] Mark R. Mine, Frederick P. Brooks Jr., and Carlo H. Sequin. 1997. Moving Objects in Space: Exploiting Proprioception in Virtual-environment Interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA. https://doi.org/10.1145/258734.258747

[19] Kai Nickel and Rainer Stiefelhagen. 2003. Pointing Gesture Recognition Based on 3D-tracking of Face, Hands and Head Orientation. In *Proceedings of the 5th International Conference on Multimodal Interfaces (ICMI '03)*. ACM, New York, NY, USA. https://doi.org/10.1145/958432.958460

[20] Jeffrey S. Pierce, Andrew S. Forsberg, Matthew J. Conway, Seung Hong, Robert C. Zeleznik, and Mark R. Mine. 1997. Image Plane Interaction Techniques in 3D Immersive Environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics (I3D '97)*. ACM, New York, NY, USA. https://doi.org/10.1145/253284.253303

[21] A Poston. 2000. Human engineering design data digest. *Washington, DC: Department of Defense Human Factors Engineering Technical Advisory Group* (2000).

[22] Felix Schüssel, Johannes Bäurle, Simon Kotzka, Michael Weber, Ferdinand Pittino, and Anke Huckauf. 2016. Design and Evaluation of a Gaze Tracking System for Free-space Interaction. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA. https://doi.org/10.1145/2968219.2968336

[23] Loren Arthur Schwarz, Artashes Mkhitaryan, Diana Mateus, and Nassir Navab. 2012. Human Skeleton Tracking from Depth Data using Geodesic Distances and Optical Flow. *Image and Vision Computing* (2012). https://doi.org/10.1016/j.imavis.2011.12.001

[24] Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA. https://doi.org/10.1145/2702123.2702355

[25] Roel Vertegaal. 2008. A Fitts Law Comparison of Eye Tracking and Manual Input in the Selection of Visual Targets. In *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI '08)*. ACM, New York, NY, USA. https://doi.org/10.1145/1452392.1452443

[26] Daniel Vogel and Ravin Balakrishnan. 2005. Distant Freehand Pointing and Clicking on Very Large, High Resolution Displays. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST '05)*. ACM, New York, NY, USA. https://doi.org/10.1145/1095034.1095041

[27] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA. https://doi.org/10.1145/302979.303053

[28] Xucong Zhang, Yusuke Sugano, and Andreas Bulling. 2019. Evaluation of Appearance-Based Methods and Implications for Gaze-Based Applications. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. https://doi.org/10.1145/3290605.3300646