Eye-Hand Interaction in XR through the Lens of Instrumental Interaction

Qiushi Zhou*
The Hong Kong University of Science and Technology (Guangzhou)

Ken Pfeuffer†
Aarhus University

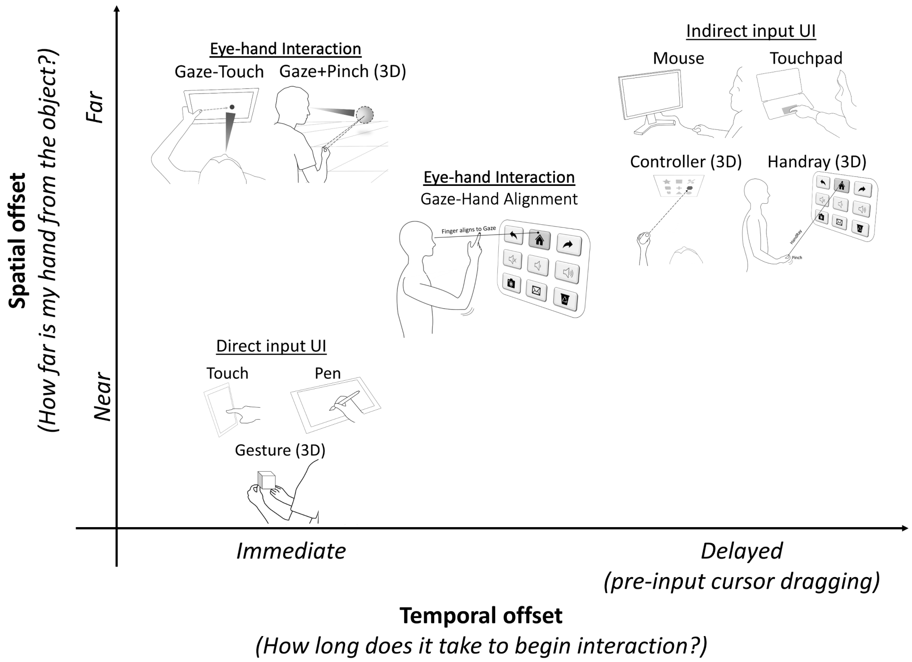

Figure 1: We analyse eye-hand interaction (EHI) in extended reality (XR) through the lens of instrumental interaction [1] while comparing different techniques regarding their properties as interaction instruments.

Abstract

Eye–hand interaction (EHI) has become a default input paradigm in extended reality (XR). Existing research largely evaluates these techniques in terms of performance and usability, as extensions of previously established WIMP techniques, whereas lacking theoretical and conceptual reflections. In this paper, we take a different approach and examine EHI through the theoretical lens of Instrumental Interaction. By conceptualising EHI as interaction instruments rather than mere user actions, we offer a novel perspective for understanding its mediating role between users and XR systems. We characterise representative EHI techniques using instrumental properties such as degree of indirection, integration, compatibility, and spatial and temporal offsets. This analysis highlights distinctive qualities of EHI, including its always-on activation, malleable indirection, and grounding in human perception–action loops, which enable fast and expressive interaction with both nearby and distant objects. Building on this perspective, we discuss design implications for XR interfaces and outline research directions that move beyond individual techniques toward the design of EHI as a foundational, XR-native interaction instrument.

Index Terms: Gaze, extended reality, instrumental interaction.

*e-mail: qiushizhou@hkust-gz.edu.cn
†e-mail: ken@cs.au.dk

1 Introduction

In extended reality (XR), Gaze&Pinch [23] has emerged as a dominant input technique set as available by default on major devices, including the Apple Vision Pro [1] and Samsung Galaxy XR [2]. While Gaze&Pinch has been adopted thanks to its ease to learn and perform for interacting with virtual UI elements and objects out-of-reach, these benefits do not belong exclusively to this single technique, but to a broader family of techniques that incorporate coordinated actions of the eyes and the hands. Eye-hand interaction (EHI) represents an emerging family of novel interaction techniques in XR that enable selection, manipulation, and other fundamental input operations. While some EHI techniques extend Gaze&Pinch by enabling novel capabilities, including multi-selection [11] and easier manipulation [19, 18], other works propose new principles including gaze-hand alignment (GHA) [17] and gaze&gesture [8] to explore novel approaches of coordinating eye-hand actions.

Despite its recent success in academic publication and in industry-adoption, EHI has yet to be understood as an explicit concept in human-computer interaction (HCI) beyond Gaze&Pinch. Currently, EHI has mostly been considered as user action, but rarely examined as a tool or instrument integrated with XR UI systems. For EHI to develop as a new interaction paradigm for XR, it needs to be conceptualised to **afford reflections on the design** of the interaction techniques following this approach. For instance, while it is hard to compare between each type of EHI techniques as they all have their own merits and limitations, it may be helpful to evaluate

[1]https://www.apple.com/apple-vision-pro/
[2]https://www.samsung.com/us/xr/galaxy-xr/galaxy-xr/

them as components of the UI system as a whole, and to analyse how they are fit for different tasks in different situated contexts. We believe that understanding EHI as an interaction instrument that is part of the XR "interface" provides a novel perspective for better understanding its pros and cons, and could shed light on its future research directions. In this work, we analyse EHI through the lens of instrumental interaction [1].

In common understandings of EHI, eye gaze is treated as an innate pointing modality and the hand as a confirmation or actuation modality, together emulating direct touch or mouse-based interaction in virtual environments. Though this understanding helps interaction research on modelling eye and hand input as human capabilities, it implicitly minimises the mediating role of interaction techniques using eye and hand that enable users to use XR systems. Instrumental Interaction [1] proposes that interaction is fundamentally mediated by *instruments*—entities [7] (e.g., a scroll bar) that affords human physical operation (e.g., through a mouse) in order to act upon domain objects (e.g., a document). The instrumental interaction model offers unique analytical perspectives for existing interaction techniques, including degree of indirection, integration, and compatibility. These perspectives offer novel opportunities to understand EHI in XR by placing it at the role of the mediating instrument between users and the object of interaction. For instance, an analysis of the degree of indirection of EHI regarding spatial and temporal offsets with reference to previously established interaction techniques would reveal the special traits of EHI in mitigating large spatial offset while enjoying small temporal offset, while enabling easy manipulation of distant objects (Figure 1).

In applying the lens of instrumental interaction to analysing EHI, our goal is not to establish formal definitions of it as an interaction instrument, or to come up with new standards for evaluating the performance or usability of existing techniques. Rather, we aim to characterise the design considerations and explore design opportunities of EHI techniques in XR in terms of instrumental properties, through providing a novel perspective that fosters fresh discussions. Based on our reading of EHI in XR as an always-on, malleable instrument grounded in human perception-action loops, we inform conceptual explorations of EHI and outline future directions [2].

2 Latest on EHI Research

As an early work that comprehensively explored and described the capacity of an EHI concept for supporting fundamental XR interaction including selection and manipulation, Gaze&Pinch enabled the **delivery of direct hand interaction** to anywhere in the 3D XR space **through gaze**, enabling easy and intuitive input [23]. This idea of switching between direct and indirect input modulated by gaze came from an earlier exploration of gaze+touch interaction on tablets [22]. The combination of gaze and hand benefits both from gaze's capability of instant access of distant targets, and from hand's capability of intuitively manipulate objects. Inspired by these early explorations, later works further investigated various implementations and evaluations of this concept, such as for manipulation [32], region selection [26], multi-selection [11], and evaluations of performance and gaze and pinch regarding late trigger and gesture complexity [21, 20].

Gaze&Pinch interaction has also been extended to more specific use contexts with more detailed investigations on its performance and novel capabilities. For instance, Lystbaek et al. explored novel affordances of bimanual manipulation using Gaze&Pinch [15]. Mikkelsen et al. investigated how Gaze&Pinch works for manipulation in XR comparing with existing input modalities including direct manipulation and handray [19], and evaluated how it can enable realistic hand behaviours in XR similar to that in physical manipulation [18].

Inspired by image plane techniques [25] that select and move distant virtual objects by occlusion with fingers, Lystbaek et al. proposed selection techniques based on the alignment of gaze and the index finger (GHA) [17]. This principle proposes a different approach from Gaze&Pinch that is based on coordinated eye-hand actions. GHA techniques have been further extended to text entry [16], and evaluated in a Fitts' Law study while being compared with other established selection techniques [30]. The GHA principle has also been combined with Gaze&Pinch in *SightWarp* that enabled easier manipulation of distant scene and objects through a virtual proxy initiated by the alignment of gaze and the pinching hand [13]. Other works explored novel variations of EHI techniques, including combining gaze and gestures following a touchpad metaphor [8], gaze and palm gesture based 3D plane manipulation [28], and palm-attached UI selection using gaze [24]. EHI techniques also include the combined use of gaze and physical tools held in the hand, such as Gaze+Pen based manipulation of control points in CAD contexts [29], translation of pen input on 2D tangible surface to 3D space through gaze [31], and teleportation in VR through gaze and stylus pointing for simultaneous specification of teleportation position and orientation [14].

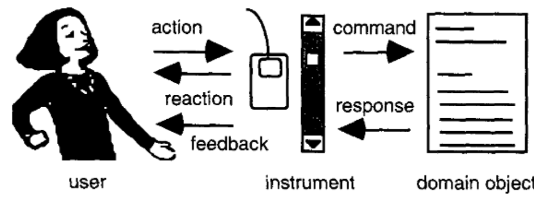

Figure 2: Interaction instrument mediating between a user and a domain object[1].

3 Eye-Hand Interaction as an Instrument for XR

Beaudouin-Lafon proposed instrumental interaction as an novel approach to understanding HCI and interaction design. Instead of seeing the UI as an object opposed to the user as subject, he distinguishes between *domain objects*, and *instruments*. Domain objects are the entities of interest in a task, and instruments mediate between the user and the domain object by translating user action to commands that are applied to the objects [1]. In this section, we describe EHI as instrumental interaction through mapping the components of each concept.

3.1 Instrument and Domain Object

If we consider EHI as the mediating layer between user and the XR interface system—that it translates all user intentions and actions to the UI elements and objects—we can view the EHI techniques as instruments, and the UI elements and objects as domain objects. In EHI-enabled XR systems:

- **Domain objects** include UI elements such as menus, buttons, windows, as well as manipulable virtual objects.
- **Instrument** is the coupled gaze+hand system that enables these commands.

In this structure, the domain objects basically include all interactive content in XR, and the instrument include all consciously and intentionally coordinated eye-hand actions that directly operate on the interactive contents. However, here, the EHI as instrument needs to be distinguished from the user "action" in the original definition of instrumental interaction (Figure 2). What conceptually serves as the instrument are the EHI techniques themselves, not the user action of performing those techniques. Therefore:

- **Action, reaction, and feedback** correspond to the user action of performing the EHI techniques, and the perception of their success or failure of their performance.

- **Commands** include selection, manipulation, navigation, and transformation that are applied to the domain objects and are issued from the techniques.

3.2 Physical and Logical Parts of Instruments

Instrumental Interaction further distinguishes between the *physical part* of an instrument, which the user manipulates, and its *logical part*, which represents the instrument's state within the system. Accordingly in EHI in XR, usually gaze works as a pointing mechanism while hand issues the command. However, in some cases, such as the GHA [17, 13] techniques, gaze and hand both serve as pointing modalities because pointing is achieved through the alignment of both pointers in the user's line of sight [25]. Thus in EHI:

- The movement of the eyes and of the hand constitutes the **physical** part of the instrument.

- The (invisible) "pointing cursor" constitutes the **logical** part, defining user intention.

In traditional pointing devices such as the mouse, the physical and logical parts are tightly coupled, meaning that the 2D hand movement on the desk is directly translated to 2D cursor movement on the display. In EHI, there is greater flexibility in the decoupling between the physical and the logical parts of the instrument. In Gaze&Pinch, gaze and hand are inherently decoupled while the hand is free from the pointing task, whereas in GHA, both gaze and hand are needed to collaboratively carry out the pointing task, in exchange for more precise selection and more explicit specification of intention. While the latter better emulates direct manipulation of physical objects, the former benefits from faster interaction with less physical effort [30].

3.3 Activating instruments

Activation is a central concern in Instrumental Interaction, as instruments typically incur either spatial costs (visible handles, widgets) or temporal costs (explicit activation actions). EHI exhibits a distinctive activation profile. Once eye and hand tracking are available, the instrument is effectively **always active**. Activation unfolds fast spatially, as gaze (or GHA) establishes a reference with an instant shift of attention, and temporally, as gaze and hand actions are easily coordinated thanks to innate human capabilities. Crucially, EHI introduces no screen real-estate cost, no attentional divider, and little explicit temporal overhead.

4 Instrumental Properties of EHI

In this section, we analyse EHI using instrumental properties, including degree of indirection, integration, and compatibility. We also discuss how meta-instrument relates to EHI in XR.

4.1 Degree of Indirection

Degree of indirection describes the distance between user action and its effect on domain objects and reflects the cost of mediation with the interaction instrument. An example of a low degree of indirection is selection handles next to objects on screen, and an example of high degree of indirection would be dialogue boxes that can be far away from the object it is associated with. Degree of indirection can be characterised with the spatial and temporal offsets. While the spatial offset is the distance on the screen between the logical part of the instrument and the object it operates on, the temporal offset is the time difference between the physical action on the instrument and the response of the object [1, 9].

4.1.1 Spatial Offset

Whereas a small spatial offset and a low degree of indirection is typically desired in instrumental interaction, things are different regarding EHI. The large spatial offset (distance between the hand and the target) is desirable in Gaze&Pinch for easy access to distant targets by gaze with less effort from hand [23]. Further, different techniques also require different placements of the hand. Whereas Gaze&Pinch fully decouples the hand position from selection, GHA techniques require the hand to overlap with the target along the line of sight of the user, hence demanding a small spatial offset on the image plane, regardless of the distance from the target in depth [17].

4.1.2 Temporal offset

Because eye movement typically leads and preceded hand movement towards the target in EHI [4], we may say that Gaze&Pinch and other techniques that use only gaze for pointing have a smaller temporal offset than GHA that needs coordinated eye-hand movement for pointing, while the extra step of moving the hand toward the target on the image place contributes to the temporal overhead. Though this temporal overhead is arguably mainly logical and only minimal regarding the cost of time, thanks to the innate human capabilities of coordinating eye and hand movement in everyday interaction with the environment [30].

In summary, we characterise the degree of indirection of EHI as:

- **Direct:** Direct hand touch and manipulation of virtual objects;

- **Indirect:** Gaze&Pinch, where gaze instantly points at the target and the hand confirms selection or performs manipulation away from the target [23]

- **Intermediate:** GHA, where hand needs to be directed to align with gaze over the target on the image plane but still being able to interact over a distance [17, 13]

Techniques such as GHA reduce spatial offset at the cost of an additional temporal step. However, this alignment step aligns with innate human perception-action loops, suggesting that not all temporal offsets are cognitively equivalent (Figure 1).

A distinctive property of EHI is that indirection is *malleable*. Users can dynamically shift between more direct and more indirect interaction through different ways of coordinating eye and hand movement, without explicit mode switching. Depending on gaze-object relations, gaze-hand alignment, and task context, the same hand action may result in different commands. These mode switches are not necessarily perceived by users as explicit system states but can emerge from human perceptual and motor behaviour. For instance, the natural action of bringing the hand to gaze or directing gaze to hand (i.e., simply "looking at the hand") during Gaze&Pinch interaction in SightWarp [13] enables a natural and implicit mode switching mechanism that allows fast and intuitive transitions between direct and indirect manipulation. This aligns with Instrumental Interaction's treatment of modes as properties of instruments, while shifting responsibility for mode management from the interface to the user's perceptual system [1].

4.2 Degree of Integration

The degree of integration measures the ratio between the number of degrees of freedom (DOF) provided by the logical part of the instrument and that captured by the input device [10, 1]. For instance, the mouse-scrollbar coupling has the degree of integration of 1/2 for using a 2D input device to control a 1D instrument. In EHI, the DOF is mainly associated with that of the 3D manipulation performed by the hand over the 3D movement of the distant target, because gaze as a pointing modality remains functionally similar across different techniques.

Our previous work investigated how manipulation tasks can be affected by DOF-Separation with the hand by mapping object translation and rotation separately to those of the wrist joint instead of the pinch point. We found that DOF-Separation generally improves manipulation performance of techniques that demands hand position or orientation to be coupled with the target, by allowing more freedom for the hand through decoupling translation from rotation. In that study, we found that Gaze&Pinch was not significantly affected by DOF-Separation because of the freedom of positioning and movement that is already enjoyed by the hand.

Similarly, we could postulate that GHA techniques may also benefit from DOF-Separation due to the tighter constraint of hand position comparing with Gaze&Pinch. Overall, we could say that Gaze&Pinch and GHA techniques have the same theoretical degree of integration as they both map 3D hand movement to that of the manipulating target. However, in practice GHA techniques have a lower degree of integration because the the manipulating hand can only initiate the input action from the line of sight of the user toward the target, limiting the expressivity of the hand for performing more natural grasping behaviours [18].

4.3 Degree of Compatibility

Compatibility measures how well an instrument's actions map to user intention for commands to be applied to the object (and its response) [1]. EHI for selection and manipulation in XR exhibits a high degree of compatibility because the hand movement is transported to the distant target through gaze, enabling an experience that is close to direct manipulation. The sense of looking at the target and performing a pinch or a grasp gesture mimics how users interact with physical objects while only breaking the spatial coupling between the hand and the virtual object. This is a unique benefit of EHI comparing with other indirect techniques such as handray, which requires an additional mediating layer/instrument of a ray/pointer/cursor that needs to be moved in the input space (Figure 1).

4.4 Meta-Instruments

Other than working as interaction instrument directly with the domain objects, EHI can also be understood as meta-instruments that operate other mediating instruments, such as controller [18], stylus [29, 14], and other tangible tools. These tools can augment stability, precision, or provide richer feedback for different tasks and in different contexts of use. For instance, 2D drawing on a physical surface can translate the stable pen action, that users are familiar with outside XR, to input on virtual objects in the 3D space through gaze [31]. Easy teleportation can be performed in virtual environments during 3D stylus input in CAD contexts by simply pointing to the desired location with the stylus while using gaze to specify orientation at the target location [14]. These examples illustrate further opportunities of using EHI as a meta-instrument to control other tools for XR input.

5 Discussion and Future Work

In this section, in light of the preceding description of EHI as an interaction instrument regarding its properties and characteristics, we discuss how the design of EHI techniques can benefit from such conceptual reflection, and explore how XR interface design can be re-thought while taking into consideration the unique features of EHI as an interaction instrument that is deeply integrated into the XR UI system.

5.1 From WIMP to EHI in XR

The original instrumental interaction paper analysed WIMP interfaces using Shneiderman's principles of direct manipulation [27] to motivate a new instrumental interaction model for better understanding and developing post-WIMP interfaces that match the principles of direct manipulation better. Here, we revisit those principles to illustrate how EHI can be further developed as another successor of WIMP interfaces for XR.

5.1.1 Continuous representations of objects of interest

Comparing with cursor-based input on desktop computers and with 3D input in XR mediated by cursors or handrays, EHI enjoys the benefit of minimum occlusion of the display space. Gaze can instantly locate the target of interest without any graphical mediation, while the hand can be quickly directed towards targets withou demanding explicit visual attention even in GHA techniques [17]. This merit of EHI enables users to utilise the entire 3D space around them as input space without needing any persistent instrument that has to be displayed, unlike in WIMP interfaces [1].

5.1.2 Physical actions on objects vs. complex syntax

EHI is inherently physical, as it is coordinated eye-hand actions toward target objects. In EHI, user action is tightly coupled with commands applied on domain objects, as gaze translates direct manipulation forward, without needing complex operations to manipulate remote objects. However, this is based on the assumption of using a simple gaze gesture [23] or a simple alignment of gaze and index finger [17]. Complex syntax is still possible in contexts where it is necessary, for instance, when arbitrary gesture combinations can be translated to complex command-series to the object.

Whereas it is more straightforward to design such interaction in WIMP interfaces because the involvement of keyboard input is arbitrary by nature, it is more tricky to design such gesture mappings in XR with EHI. For instance, with Gaze&Pinch, the mapping between the pinch gesture and the selection/acquisition command to the object is semantically intuitive to understand, whereas further mappings of more gestures to more output effects may not be. Future works need to carefully consider the design of complex gestures in EHI to ensure intuitive understanding by users and compatibility with existing gesture (e.g., pinch). For instance, out previous work explored multi-selection gestures to be integrated with Gaze&Pinch, and opted for those that exhibit input metaphors that may be familiar with users, such as sliding, rotating, and mode-switching using non-dominant-hand-pinching [11]. Overall, the simplicity of EHI is one of its unique traits to strive for in interaction design [6].

5.1.3 Fast, incremental and reversible operations with an immediately-apparent effect on the objects of interest

Gaze is often described to operate in the "speed of thought", for reflecting user's input intention fast. The intuitive mapping between 3D hand movement to the virtual objects being manipulated also helps the EHI operations to be easily incremental and reversible (e.g., continuous manipulation in multiple clutches [18]). The "immediately-apparent effect on the objects of interest", i.e., feedback, needs special attention when designing EHI XR techniques. User perception of success of gaze-pointing on targets replies on good design of visual feedback that can quickly respond to user eye movement in relation to the target in the input space [23, 11]. Future work should take the feedback design into important consideration when designing EHI techniques in XR. For instance, the visual feedback for an EHI operation can be designed to be only displayed on the target (outline-highlight, colour-change, or flashing), or at the gaze-pointed location (that is not necessarily occupied by any target), or shown as a visual effect in the entire field of view. The design choice needs to be made to be compatible with the use context, as that of XR extends far beyond WIMP interfaces.

Figure 3: Semantic pointing: a dialog box as seen on the screen (left) and as perceived in motor space (right). Enlarging the Save button in motor space facilitates its selection, while shrinking the Don't Save button in motor space makes it harder to select [5, 2].

5.1.4 Layered or spiral approach to learning

EHI techniques are typically easy to learn as they mimic real world eye-hand behaviour. After becoming familiar with Gaze&Pinch as the default interaction technique in many commercial XR systems today, users will be able to quickly learn other operations based on gaze pointing and indirect hand gesturing. However, for effective learning, the gestures or gesture combinations still need to be intuitive enough and compatible with natural eye-hand coordination as much as possible [4].

In sum, while most existing discussions on EHI frame it as a replacement for mouse or touch, we propose that it is an **XR-native interaction paradigm** because of the compatibility between the 3D embodied XR input-output space and the direct link of EHI to human perception-action loops that support interactions that are closer to direct manipulation [27]. The speed and the spatial freedom of performing input using EHI allows it to work as an excellent interaction instrument beyond the original definitions of degree of indirection, integration, and compatibility. EHI, as a more expressive instrument mediating human interaction with XR systems, opens new research directions focused on skill acquisition, mastery, and play. Designing for malleable indirection, implicit mode switching, and instrument awareness may enable richer forms of interaction that go beyond efficiency-oriented paradigms like WIMP [12, 2].

5.2 Rethinking XR Interaction Design with EHI

An important power of analysing EHI as an instrument for XR is the explicit placement of it at a role that mediates all human actions with the XR system. With EHI as an interaction instrument, we can imagine a re-design of current XR interfaces, which are still mostly based on WIMP principles with menu windows, cursors, menus, and buttons. To fully benefit from the merits of the EHI instrument, the XR interfaces can be restructured to better fit EHI techniques and the natural human capabilities they connect to. For instance, the CPN2000 project contributed "a graphical application with no menus, no dialog boxes, no title bars, no scrollbars and no notion of selection", yet demonstrably "more powerful and simpler to use than the earlier WIMP version" [3, 2]. The use of EHI as an instrument in XR provides an opportunity to more directly match the sensory-motor interface of humans to the interface of the XR system. This conscious awareness of the mediation of EHI should motivate future research on how to **design better EHI in XR beyond the limitations of the current forms of the interface**.

An example can be taken from Semantic Pointing (Figure 3) [5, 2] where buttons sizes adapt to more likely choices. Inspired by this example, future XR interface can rethink how the current UI elements easily afford gaze or not, and could explore design opportunities of utilising animation effects that attract gaze while invite indirect hand operations through different semantic mappings. In sum, the design of an entire XR interface, from its overall structure to details regarding button sizes and affordances for EHI techniques should all be considered with **EHI** in mind **as the fundamental mediating instrument between user and the XR system**.

If we are to foster the design of better interaction, we need not only sets of techniques with proven effectiveness, but also rules and principles for combining them without losing their advantages. Current research explorations are usually confined to the scope of single or a set of similar techniques. However, if we consider EHI in general as the mediating instrument between users and XR systems beyond any single techniques, we need to study **how different EHI techniques can be compatible with each other**. While Gaze&Pinch claimed early industry adoption, it is important to realise that it is only one (perhaps the most fundamental) design of many potential EHI techniques, whereas the real revolution of XR interaction is yet to come from.

6 Conclusion

This paper reframes eye–hand interaction in XR as an interaction instrument through the lens of Instrumental Interaction, rather than as a type of user actions or a collection of isolated input techniques. Through this lens, EHI emerges as a mediating instrument that shapes how users act on XR domain objects, revealing properties that are difficult to articulate when EHI is treated merely as user behaviour. Our analysis shows that EHI combines low temporal offset with flexible spatial indirection, supports high compatibility with user intention, and enables implicit mode-switching grounded in natural perception–action loops.

Viewing EHI as an always-active and malleable instrument also exposes design opportunities that go beyond optimising individual techniques such as Gaze&Pinch. While such techniques represent an important step toward effective XR input, they should be understood as early instantiations of a broader instrumental space. Future XR interfaces can benefit from explicitly designing for instrumental properties, such as adaptable indirection, integrated feedback, and compatibility across techniques, rather than replicating WIMP-style structures in three dimensions.

By foregrounding the mediating role of eye–hand interaction, this work invites **a shift in how XR interaction is conceptualised and designed in next stage of EHI**: from selecting the "best" technique for a task **toward shaping coherent EHI instrumental ecologies** that users can appropriate, master, and combine. Recognising **EHI as a fundamental interaction instrument** opens a path toward **XR-native interfaces** that better align with human sensorimotor capabilities and support richer, more expressive forms of interaction.

References

[1] M. Beaudouin-Lafon. Instrumental interaction: an interaction model for designing post-wimp user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, p. 446–453. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/332040.332473 1, 2, 3, 4

[2] M. Beaudouin-Lafon. Designing interaction, not interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, p. 15–22. Association for Computing Machinery, New York, NY, USA, 2004. doi: 10.1145/989863.989865 2, 5

[3] M. Beaudouin-Lafon and H. M. Lassen. The architecture and implementation of cpn2000, a post-wimp graphical application. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*, UIST '00, p. 181–190. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/354401.354761 5

[4] J. K. Bertrand and C. S. Chapman. Dynamics of eye-hand coordination are flexibly preserved in eye-cursor coordination during an online, digital, object interaction task. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3580866 3, 5

[5] R. Blanch, Y. Guiard, and M. Beaudouin-Lafon. Semantic pointing: improving target acquisition with control-display ratio adaptation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, p. 519–526. Association for Computing Machinery, New York, NY, USA, 2004. doi: 10.1145/985692.985758 5

[6] S. Bødker. It is still a star. 2007. 4

[7] S. Bodker. *Through the interface: A human activity approach to user interface design*. CRC Press, 2021. 2

[8] J. S. Esquivel, L. Zerbin, Q. Zhou, K. Pfeuffer, and H. Gellersen. Spatialpad – air-taps for interaction in xr. In *2025 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 40–45, 2025. doi: 10.1109/ISMAR-Adjunct68609.2025.00015 1, 2

[9] K. Hinckley. *Input technologies and techniques*, p. 151–168. L. Erlbaum Associates Inc., USA, 2002. 3

[10] R. J. K. Jacob, L. E. Sibert, D. C. McFarlane, and M. P. Mullen. Integrality and separability of input devices. *ACM Trans. Comput.-Hum. Interact.*, 1(1):3–26, Mar. 1994. doi: 10.1145/174630.174631 3

[11] J. Kim, S. Park, Q. Zhou, M. Gonzalez-Franco, J. Lee, and K. Pfeuffer. Pinchcatcher: Enabling multi-selection for gaze+pinch. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25. Association for Computing Machinery, New York, NY, USA, 2025. doi: 10.1145/3706598.3713530 1, 2, 4

[12] J. Li and P. O. Kristensson. On the benefits of sensorimotor regularities as design constraints for superpower interactions in mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 31(5):2568–2578, 2025. doi: 10.1109/TVCG.2025.3549876 5

[13] Y. Liu, T. Mikkelsen, Z. Liu, G. Tian, D. Mardanbegi, Q. Zhou, H. Gellersen, and K. Pfeuffer. At a glance to your fingertips: Enabling direct manipulation of distant objects through sightwarp. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, UIST '25. Association for Computing Machinery, New York, NY, USA, 2025. doi: 10.1145/3746059.3747653 2, 3

[14] Y. Liu, Q. Zhou, M. N. Lystb, A. Kehoe, M. Gutierrez, H. Gellersen, and K. Pfeuffer. Stylusport: Investigating teleportation using stylus in vr. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26. Association for Computing Machinery, New York, NY, USA, 2026. doi: 10.1145/3772318.3793420 2, 4

[15] M. N. Lystbæk, T. Mikkelsen, R. Krisztandl, E. J. Gonzalez, M. Gonzalez-Franco, H. Gellersen, and K. Pfeuffer. Hands-on, hands-off: Gaze-assisted bimanual 3d interaction. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, UIST '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3654777.3676331 2

[16] M. N. Lystbæk, K. Pfeuffer, J. E. S. Grønbæk, and H. Gellersen. Exploring gaze for assisting freehand selection-based text entry in ar. *Proc. ACM Hum.-Comput. Interact.*, 6(ETRA), May 2022. doi: 10.1145/3530882 2

[17] M. N. Lystbæk, P. Rosenberg, K. Pfeuffer, J. E. Grønbæk, and H. Gellersen. Gaze-hand alignment: Combining eye gaze and mid-air pointing for interacting with menus in augmented reality. *Proc. ACM Hum.-Comput. Interact.*, 6(ETRA), May 2022. doi: 10.1145/3530886 1, 2, 3, 4

[18] T. Mikkelsen, Q. Zhou, M. Gonzalez-Franco, H. Gellersen, and K. Pfeuffer. Preshaping hand behaviour for direct and indirect manipulation of 3d objects. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26. Association for Computing Machinery, New York, NY, USA, 2026. doi: 10.1145/3772318.3793424 1, 2, 4

[19] T. Mikkelsen, Q. Zhou, M. N. Lystbaek, Y. Liu, H. Gellersen, and K. Pfeuffer. DOF-Separation for 3D Manipulation in XR: Understanding Finger-Wrist Separation to Simultaneously Translate and Rotate Objects . In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1246–1256. IEEE Computer Society, Los Alamitos, CA, USA, Oct. 2025. doi: 10.1109/ISMAR67309.2025.00130 1, 2

[20] Y. Park, J. Kim, and I. Oakley. The impact of gaze and hand gesture complexity on gaze-pinch interaction performances. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '24, p. 622–626. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3675094.3678990 2

[21] Y. Park, J. Kim, and I. Oakley. Gazehandsync: Mitigating late-trigger errors for seamless gaze-hand interactions. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*, ETRA '25. Association for Computing Machinery, New York, NY, USA, 2025. doi: 10.1145/3715669.3723126 2

[22] K. Pfeuffer, J. Alexander, M. K. Chong, and H. Gellersen. Gaze-touch: combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, p. 509–518. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2642918.2647397 2

[23] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 99–108. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132180 1, 2, 3, 4

[24] K. Pfeuffer, J. Obernolte, F. Dietz, V. Mäkelä, L. Sidenmark, P. Manakhov, M. Pakanen, and F. Alt. Palmgazer: Unimanual eye-hand menus in augmented reality. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*, SUI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3607822.3614523 2

[25] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine. Image plane interaction techniques in 3d immersive environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, I3D '97, p. 39–ff. Association for Computing Machinery, New York, NY, USA, 1997. doi: 10.1145/253284.253303 2, 3

[26] R. Shi, Y. Wei, X. Qin, P. Hui, and H.-N. Liang. Exploring gaze-assisted and hand-based region selection in augmented reality. *Proc. ACM Hum.-Comput. Interact.*, 7(ETRA), May 2023. doi: 10.1145/3591129 2

[27] B. Shneiderman. Direct manipulation: A step beyond programming languages (abstract only). *SIGSOC Bull.*, 13(2–3):143, May 1981. doi: 10.1145/1015579.810991 4, 5

[28] U. Wagner, M. Albrecht, A. A. Jacobsen, H. Wang, H. Gellersen, and K. Pfeuffer. Gaze, wall, and racket: Combining gaze and hand-controlled plane for 3d selection in virtual reality. *Proc. ACM Hum.-Comput. Interact.*, 8(ISS), Oct. 2024. doi: 10.1145/3698134 2

[29] U. Wagner, J. Kim, Z. Wu, Q. Zhou, M. Romero, A. Iop, T. Feuchtner, and K. Pfeuffer. A study of multimodal pen + gaze interaction techniques for shape point translation in extended reality. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 454–464, 2025. doi: 10.1109/ISMAR67309.2025.00056 2, 4

[30] U. Wagner, M. N. Lystbæk, P. Manakhov, J. E. S. Grønbæk, K. Pfeuffer, and H. Gellersen. A fitts' law study of gaze-hand alignment for selection in 3d user interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581423 2, 3

[31] Z. Wu, U. Wagner, M. Romero, A. Iop, A. Matviienko, and K. Pfeuffer. Eyes on target, pen on table: Precise vr interaction with gaze and pen. In *2025 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 46–53, 2025. doi: 10.1109/ISMAR-Adjunct68609.2025.00016 2, 4

[32] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves. Gaze-supported 3d object manipulation in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445343 2