RESEARCH-ARTICLE

Can't (Midas) touch this: An Evaluation of Clutching Methods for Opportunistic Tangible Interfaces on Interactive Tabletops

MAXIMILIAN THOMAS KILGER, University of Regensburg, Regensburg, Bayern, Germany

ELISABETH PHILIPP, University of Regensburg, Regensburg, Bayern, Germany

VITUS MAIERHÖFER, University of Regensburg, Regensburg, Bayern, Germany

Open Access Support provided by:

University of Regensburg

Published: 31 August 2025

MuC '25: Mensch und Computer 2025
August 31 - September 3, 2025
Chemnitz, Germany

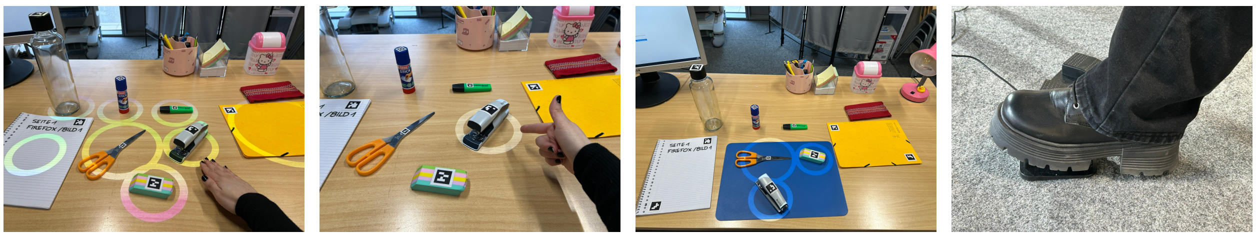
MuC '25: Proceedings of the Mensch und Computer 2025 (August 2025)
https://doi.org/10.1145/3743049.3743055
ISBN: 9798400715822

Can't (Midas) touch this: An Evaluation of Clutching Methods for Opportunistic Tangible Interfaces on Interactive Tabletops

Maximilian Thomas Kilger
University of Regensburg
Regensburg, Germany
maximilian-thomas.kilger@stud.uni-regensburg.de

Elisabeth Philipp
Universität Regensburg
Regensburg, Germany
elisabeth.philipp@stud.uni-regensburg.de

Vitus Maierhöfer
University of Regensburg
Regensburg, Germany
vitus.maierhoefer@ur.de



(a) Activation Gesture (b) Address Gesture (c) Active Zone (d) Pedal

Figure 1: Four clutching methods compared in our study. Performing each of these interactions puts everyday objects in an *active* state (indicated by white rings), allowing for them to be used as tangibles for tangible interaction.

Abstract

Opportunistically employing everyday objects as tangibles provides a promising avenue for wider adoption of Tangible User Interfaces. However, the potential double purpose of these objects both as tangibles and for their originally intended purpose makes opportunistic tangible interaction susceptible to the Midas Touch effect. Explicit clutching mechanisms that control the activation and deactivation of these tangibles offer a solution to this problem, but the optimal choice of clutching method remains unclear in the context of opportunistic tangibles on interactive tabletops. This paper investigates suitable clutching mechanisms through a controlled user study (n=24), comparing four different clutching methods: Activation- and Address Gestures, Active Zones, and Pedal clutches. We analyze quantitative measures of user experience as well as qualitative feedback, identifying advantages and disadvantages and singling out Address- and Activation gestures as points of reference for future work.

CCS Concepts

• **Human-centered computing** → *Mixed / augmented reality.*

Keywords

Tangible Interaction, Opportunistic Tangible Interaction, Midas Touch, Interactive Tabletops

MuC '25, Chemnitz, Germany

ACM ISBN 979-8-4007-1582-2/25/08
https://doi.org/10.1145/3743049.3743055

ACM Reference Format:
Maximilian Thomas Kilger, Elisabeth Philipp, and Vitus Maierhöfer. 2025. Can't (Midas) touch this: An Evaluation of Clutching Methods for Opportunistic Tangible Interfaces on Interactive Tabletops. In *Mensch und Computer 2025 (MuC '25), August 31–September 03, 2025, Chemnitz, Germany.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3743049.3743055

1 Introduction

As central hubs for activities such as working, dining, playing and socializing, tables provide both an ideal space for the digital augmentation of everyday activities and a large, visible surface for displaying and contextualizing digital information. In the push to integrate digital user interfaces more seamlessly into the physical world, interactive tabletops have emerged as one of the focal points of Spatial Augmented Reality [51, 59].

While there are multiple ways of facilitating interaction on an interactive table, such as touch- and stylus input, [38, 39], *Tangible Interaction* in particular allows for haptic feedback and a greater utilization of our inherent motoric and spatial awareness skills [2, 17, 53]. But Tangible User Interfaces (TUIs) generally require internal or external hardware for tracking and detection, such as RFID chips and sensors [10, 43] or sensing tablets [44]. Thus, TUIs are often reliant on the availability of dedicated, technically sophisticated tangibles, tailored to specific contexts of use. This presents a considerable barrier towards the adoption of tangible interaction in everyday life.

Therefore, the replacement of dedicated tangibles with existing everyday objects has been proposed as a necessary step [27]. Within the last 15 years, the field of *Opportunistic Tangible Interaction* (OTI) [25] has emerged as a sub-field of Tangible Interaction, placing the emphasis on the use and re-use of features with useful affordances that already exist within the context of use. These features may be surfaces [9], edges [22] or everyday objects [7, 11, 19, 34].

The original definition of Opportunistic Tangible User Interfaces (OTUIs) [25] limits them to affordances and features that are currently unused, as their original purpose is entirely replaced when repurposed for interaction. For example, if the button of a ballpoint pen is used as an input device (as described by Corsten et al. [11]), the user may be unwilling to sacrifice its primary function as a writing tool or, conversely, its use as a controller.

While the simplest solution would be to allow the pen to function for both writing and button-pressing as needed, this introduces a new challenge: the OTUI must distinguish between deliberate input and unintentional actions stemming from the object‘s original use. Failure to do so causes the *Midas Touch problem* [32]: Actions using an objects primary function are misinterpreted as input to the computing system, leading to unintended effects. In the pen scenario, for example, pressing the button to extend the ballpoint tip would also trigger the mapped digital function, even if just in preparation for writing. This failure to accurately interpret user input leads to a loss of control.

There are several ways to mitigate the Midas Touch effect. One approach is to infer user intent from context [14, 50]. For example, if a pen is held tip-down over paper, an autonomous system could assume the user intends to write and classify their actions as non-input.

Detecting a user's true intention, however, is practically impossible to do reliably from outside observation alone, and thus prone to error. If an interaction intent is missed (false-negative), the interface will not respond to input. If an intent is wrongly detected (false-positive), the Midas Touch effect occurs. Therefore, it may be more practical to let users themselves explicitly indicate *when* and *how* they want to interact with an opportunistic tangible as an input device. The simplest way to achieve this is through a *clutch* [32] — a deliberate action that toggles user input on and off.

Clutching actions can take a variety of forms. Previous work proposes a broad spectrum of interactions as clutches, ranging from gaze-activated clutches [31] over gesture control [58] to voice commands [12]. Since no single clutching method has emerged as a general optimal choice, the "best" clutching method is highly dependent on the context of use.

Several studies, for example, compared clutching interactions for mid-air gesture interaction in a variety of different contexts [12, 23, 49]. But no research has examined clutching methods specifically for Opportunistic Tangible Interaction. Likewise, no study has investigated them for their use on interactive tabletops. The question as to which clutching method is suitable for this area of application in particular remains unanswered.

Our work addresses this research gap by evaluating user experience and preferences of different clutching methods for Opportunistic Tangible Interaction on interactive tabletops in a controlled user study. We compared four different clutching methods: *Activation Gesture*, *Address Gesture*, *Active Zone* and *Pedal* (see Section 3.2). While none of these four could clearly distinguish itself as the optimal choice, based on the individual advantages and disadvantages of these clutching methods, we identify the two gesture-based clutching methods as a recommended general starting point for future work.

2 Related Work

2.1 Origins of Tangible Interaction

The graphical user interface is the prevalent paradigm for human-computer-interfaces. However, it lacks the spaciality and tactility that characterize our interaction with the physical world. Therefore, the concept of tangible interaction [30] was developed to connect the physical and digital worlds more directly.

The first major push for physical elements in human-computer interaction were Fitzmaurice et al. [17]'s *Bricks*, a system that used small wood blocks as surrogates for handles in a GUI. These *"Graspable User Interfaces"* were claimed to leverage inherently human motor- and spacial reasoning skills and to provide a more direct, space-multiplexed form of input, with the goal of "offer[ing] a seamless blend between the physical and virtual worlds" [17].

The concept of Graspable User Interfaces was generalized and expanded by Ishii and Ullmer [30] to produce the concept of *Tangible Bits*. With Tangible Bits, the authors propose a deep and seamless connection between digital and physical objects, and an extension of the human-computer interface to the entire user environment [30].

At its most basic level, this concept can be realized by coupling digital information to a physical artifact [55, 56], which is then commonly called a *tangible*. Physical interaction with this tangible, like placing, moving, flipping or rotating it, is then translated into interaction with the digital information it represents [28].

2.2 Interactive Tabletops

Desks are a natural point of interaction between physical items and the digital realm, and as such a place where we need to exchange information between the two spaces. We eat, work, play and socialize at tables, and as a well-defined, flat space, tables are ideal for displaying the digital information that can support all those activities and more.

The *DigitalDesk* [63] represents the first major effort to integrate the digital world of the virtual workstation with the surface of the desk itself. Using a top-mounted projector-camera-system, it was able to detect interaction and project information on the tabletop. Another, more conceptual interactive tabletop system called the *MetaDESK* was presented by Ullmer and Ishii [54] to demonstrate key concepts of Tangible Interaction, with electromagnetically tracked tangibles serving as a stand-in for the icons, menus, handles, and controls of the graphical user interface. A successor project, *SenseTable* [44], utilized specialized pucks on commercial sensing tablets as tangibles.

Throughout the early 2000s, research on tangible interaction for interactive tabletops tended towards producing specialized technology that expanded the capabilities of tangibles [4, 13, 43]. Like *SenseTable*, these often relied on specialized sensors and environments, limiting the adaptability of Tangible User Interfaces [27]. Even systems utilizing everyday objects required the objects to be prepared with RFID-tags before use [41, 43].

2.3 Opportunistic Tangible Interaction

Henderson and Feiner [25] introduced the concept of *Opportunistic Controls* to describe a "tangible user interface [...] that leverages naturally occurring, tactilely interesting, and otherwise unused affordances" [25]. Features with "unused affordances" could, theoretically, be anything, from machine parts [24, 25], walls and surfaces [8, 9], or edges [22] to unaugmented everyday objects [7, 11, 16, 65]. An often-cited example for a suitable affordance is the rotatability of a round coffee mug, which allows for its use as a makeshift dial [15, 19, 47].

Opportunistic Tangible User Interfaces are often designed to be created spontaneously and used temporarily [11, 20, 60, 61]. They emphasize the role of the user as a co-designer of the interface [5, 34, 62], giving them control over important design aspects like the selection and function of tangibles, allowing for highly customized and personalized interface solutions. All of these aspects — spontaneous creation, temporary use, user authoring and non-reliance on pre-fabricated technology — give OTUIs an extraordinary potential for flexibility.

Tabletops — interactive or not — present a prime opportunity for the use of interactive tangibles [7, 19, 27, 61]. The primary function of tables is the placement of objects on them, providing a perfect interaction space for tangible interaction. Also, tables are often cluttered with objects that are not constantly in use [1, 7], which provide a useful selection of naturally occurring, tactically interesting affordances to be leveraged by an OTUI.

While many objects may have affordances suitable for OTUIs, they still have other purposes that can interfere with the requirement of being "otherwise unused" [25]. OTUI systems tend to substitute a given object's original purpose with its role as an input device [7, 11, 16, 19, 26, 65]. However, previous work also recognizes the potential need to utilize a tangible's substituted purpose [7, 11, 19, 65]. They propose mechanisms to explicitly put an object into this state of substituted purpose, like mid-air gestures [11, 65] or voice recognition [16], but also to reverse the substitution process, such as a time-to-live mechanism [7] or a two-step gesture-based "unlinking" process [19]. These interactions generally serve as long-term bookends at the beginning or end of the interaction with the tangible, and de-emphasize fast transitions for short-term interaction needs.

2.4 The Midas Touch problem

The Midas Touch problem describes the phenomenon of unintentional user input triggered by unrelated actions of the user. The term (as well as the first solutions for it) originates in the field of eye tracking [32], where the precise and intentional issuing of commands is nearly impossible due to the passive, unconscious nature of eye movements. Dwell time [57] has emerged as a simple and common solution for the Midas Touch problem in eye tracking specifically that uses natural fixation patterns in human gaze. However, it requires a trade-off between precision and interaction time [45], so alternatives and enhancements to dwell time have been studied as well [18, 35, 40, 64]. Among these is the *clutch* [31], a deliberate action by the user which can turn eye tracking input on or off [32].

Mid-air gesture interaction is also prone to the Midas Touch problem. While various solutions exist, such as gaze-based prediction [50], none match the simplicity and ubiquity of dwell time in eye tracking. A plethora of clutching interactions have been proposed as a solution as well . For example, Vogiatzidakis and Koutsabasis [58] introduced a preliminary *address gesture* to be performed with or before a control gesture to specify the IoT device it controlled. Surale et al. [52] compared gestural clutches for drawing tasks in Virtual Reality. Rivière et al. [49] evaluated pedals and embedded buttons as clutching interactions for a non-opportunistic TUI. Similarly, Heinrich et al. [23] compared hand gestures, foot gestures, and voice commands as clutching methods in a medical context. Cronin et al. [12] also studied clutching interactions for medical imaging systems, utilizing voice commands, activation gestures, gaze recognition and an active zone (a marked area in space, outside which gestures would not be recognized) as clutches. The results of these evaluation studies are mixed. Rivière et al. [49] argued for the use of foot pedals based on non-significant results. Heinrich et al. [23] identified voice activation as a clear favorite among users, but this was partially attributed to the poor accuracy of their gesture recognition system. Cronin et al. [12] found several pros and cons for voice-, gaze- and gesture-based clutches, arguing for active zones as the ideal clutching methods in the operating room.

In conclusion, clutching interactions are a viable solution to the Midas Touch problem and can be applied across different interaction paradigms, e.g. eye-tracking and mid-air gesture interaction. The factors that make one clutching method preferable over others, however, can change depending on the context in which the interaction paradigms are used.

2.5 Summary and Research Gap

While the Midas Touch problem is well-understood and widely studied in the areas of eye tracking and mid-air gesture interaction [12, 31], awareness and understanding of the topic are still lacking in the field of Opportunistic Tangible Interaction. OTUIs are generally designed under the assumption that opportunistic tangibles are utilized exclusively for either Human-Computer Interaction or their original purpose and give the user no short-term way of switching between the two modalities.

Previous work suggests that clutching interactions are a viable solution to this problem [12, 18, 23, 31, 49]. However, the suitability of individual interaction techniques depends on the context of use [12, 23, 49]. The use of clutching interactions in the context of opportunistic tangible interaction on interactive tabletops is poorly understood, with future interaction designs lacking a point of reference for clutching interactions.

Our work addresses this gap by comparing different clutching interactions in a controlled study. Specifically, we aim to address the following questions:

(1) How do different clutching methods compare in measures of user experience?
(2) How do different clutching methods compare in terms of user preference?

3 Method

To answer the questions posed in Section 2.5, we selected four different clutching methods and compared them in a controlled user study. Where possible, we adapted clutching methods from other fields to Opportunistic Tangible Interaction on an Interactive Tabletop.

As our focus was on human factors in clutching interaction design, we deliberately avoided a technical evaluation of different clutching methods. This minimized the impact of implementation-specific factors such as latency or detection accuracy, which could otherwise affect usability. Instead, we used a Wizard-of-Oz approach to simulate clutching interactions where appropriate.

Below, we outline the study design, the specific clutching methods tested, and the evaluation process.

3.1 Design

We conducted a controlled user study using a within-subjects design. Our main independent variable was CLUTCHING METHOD, with TASK as a control variable. We measured the dependent variables user experience, task load and clutching frequency, and we recorded qualitative data on participants' preferences, assessments and experiences.

In the study, participants used four different clutching methods to activate objects on the table, turning them into tangibles, or to deactivate them to restore their original non-interactive state (depicted in Figure 1). We selected *Activation Gesture*, *Address Gesture*, *Active Zone* and *Pedal* as our four conditions for the main independent variable CLUTCHING METHOD (detailed in Section 3.2).

To evaluate these interactions, we designed two tasks, a text task and a search task. Each participant completed four runs, with one task per run under a different clutching condition. The tasks alternated to ensure balanced exposure. We tracked the task completed during a trial as a control variable to identify potential differences between the text and search tasks.

To assess cognitive load, we employed the NASA Task Load Index (TLX) [21], examining whether specific clutching methods increased or reduced task load. Additionally, we used the User Experience Questionnaire (UEQ) [36] to assess more subjective factors of usability, appeal and perceived efficiency . These quantitative measures provided a multi-faceted comparison of different clutching methods with respect to abstract concepts such as workload or usability.

In addition to quantitative data, we conducted structured interviews with the participants to gather qualitative insights. These interviews helped us understand participant behavior, identify underlying causes, and capture participants' impressions and preferences.

Before finalizing the study design, we conducted a pretest with three students. The aim was to evaluate the overall study procedure, identify potential usability issues, and refine task instructions and interaction flow. Based on the feedback, we made minor adjustments to improve clarity and feasibility of our instructions for participants in the main study. Additionally, we tweaked the artificial latency of the *Pedal* condition to match the reaction time of the wizard. The students that participated in the pretest did not participate in the main study.

3.2 Conditions

In our study, we compared four different clutching methods. Objects designated for tangible interaction could be in an *active* or *inactive* state; with interaction with the objects only triggering user input in the *active* state. While the core interaction with the objects remained the same, the methods differed in how activation and deactivation were triggered, varying in required input and user effort.

In the following section, we describe each method in detail.

3.2.1 Activation Gesture. With the Activation Gesture method (Figure 1a), all available objects become interactive when the user performs a predefined gesture [11]. To deactivate them, the user performs the same gesture again. We selected a double tap gesture on the table with a flat hand (analogous to the proposed "Play"-gesture in [29]), since this gesture is unlikely to be performed accidentally and is easy to track [12]. It provides a balance between ease of use and ergonomics. Alternative gestures, such as knocking on the table, may cause discomfort over time, especially in prolonged work sessions. Additionally, the flat-hand tap minimizes unintended activations and remains practical in real-world tabletops environments.

3.2.2 Address Gesture. To contrast the activation gesture, we investigated a clutching method that allows users to activate one specific object at a time, similar to [58]. With the Address Gesture (Figure 1b), users activate an object by performing a predefined gesture directed at it. We selected a pointing gesture with the index finger [19] as it allows for precise selection and intuitive interaction [29]. Multiple objects can be activated simultaneously, and repeating the gesture on an active object deactivates it.

We considered alternative gestures, such as snapping fingers near the object or flicking it. However, test runs revealed that these gestures were physically tiring and uncomfortable over time, making them less practical for extended use. In contrast, pointing is a natural and intuitive selection method that aligns well with interaction patterns in both digital and physical environments.

3.2.3 Active Zone. The Active Zone method (Figure 1c) activates an object when it is placed within a predefined area on the table [12]. Only inside this zone, interaction with the tangible triggers input events (similar to [3]). Removing the object from the Active Zone restores its original non-interactive state. In our study, we used a desk pad 41×30 cm in size to demarcate the Active Zone. As mentioned by Cronin et al. [12], this method can be "advantageous because explicit actions are not required by the user, potentially reducing the cognitive load [...]". Since objects only become tangibles within the Active Zone, activation is restricted to a specific area on the desk. This constraint helps prevent accidental activations and unintended interactions outside the designated workspace.

3.2.4 Pedal Clutch. We also considered foot gestures, especially the foot pedal that was used in several previous studies [23, 49]. For the Pedal Clutch, we used a *PC Sensor*[1] mechanical foot pedal positioned under the table (See Figure 1d). The device featured three pedals, all configured to perform the same function. Pressing any pedal once activated all objects on the table, while pressing it again deactivated them. We deliberately opted for a toggle mechanism instead of requiring users to keep the pedal pressed. Holding it down for extended periods could lead to discomfort and impracticality, particularly during longer tasks. Additionally, a continuous press could increase the risk of unintended errors, such as accidental deactivation due to fatigue or inconsistent pressure.

[1] https://pcsensor.cn/Product-details1?product_id=1019

The Pedal was the only clutch whose activation was detected automatically, rather than by a human operator, during the experiment. Test runs revealed a bias in favor of the pedal, as it was unaffected by the latency introduced by human reaction time for other clutching methods. To ensure comparability, we artificially added a latency of ~0.5 s (averaged) to the pedal activation, which simulated the reaction time of the wizard. The delay was randomized using a Gaussian distribution ($\mu = 0.5\,s, \sigma = 0.08\,s$).

3.3 Apparatus

Our study was conducted at an interactive desk outfitted with a top-mounted projector and a camera. The desk was decorated with common desk utensils, such as a pencil holder, a lamp and a desk trash can. The desk also featured a monitor as the output device for the tasks, as well as a tablet displaying the instructions for the tasks. We tried to keep the setup of the table close to a common, everyday private desk to provide a naturalistic environment for our study.

Participants had to complete tasks with a rudimentary Tangible User Interface consisting of everyday objects. Interactions were tracked through a mix of automatic detection and human observation. Objects were tagged with Apriltags [42], which allowed for the robust detection of their rotation and position through the camera [6]. Additionally, the user's hands were tracked [37]. Several opportunistic controls determined the state of a tangible with this information and derived a discrete or continuous value, such as the position of a finger on a tagged surface, or the rotation angle of a marker as a value between 0 and 1. Controls could be *activated* and *deactivated*. Their current value could be queried by other processes, but was only returned in the *active* state to prevent processing of the input in the *inactive* state. These values controlled or triggered common digital functionality — simulating a scroll wheel, adjusting the zoom level, or triggering a variety of keyboard shortcuts.[2]

Clutching and declutching interactions were detected by a researcher through observation of the participant and a camera feed from the tracking camera. The activation and deactivation of tangibles, were triggered manually with the help of a *Novation Launchpad mini*. The functions of non-automatically tracked tangibles was also controlled through the Launchpad. These functions followed the same activeness/inactiveness rules as automatic controls and were equally affected by clutching interactions. The *active/inactive* status of each tangible control was indicated to the wizard through LEDs embedded in the Launchpad's buttons, and to the participant by a small circle projected around tangibles that were currently active (similar to [19]).

3.3.1 Design of the User Interface. To ensure that our experimental setup closely resembled a regular desk environment, we utilized a dataset by Achmann-Denkler [1] with images of 86 individual desks in an everyday context.

All images were annotated with the objects that were visibly on the desks, which were then ordered by total frequency over the full dataset. Excluding technical devices like smartphones, monitors and keyboards, we then analyzed all objects which occurred in more than two images according to the taxonomies of gestures and affordances presented by Becker et al. [5] and Gong et al. [20]. This step yielded a set of suitable gestures and interactions for each object. We then selected objects as tangibles for our interface based on afforded gestures and interaction techniques, their frequency, their potential usefulness in experimental tasks and the possibility of tagging the objects with fiducial markers.

The final mapping of tangibles to interaction techniques and functions was chosen in accordance with the requirements of the tasks (see 3.4). Where the tasks required simple interaction techniques that followed a trigger-response schema, we tried to align the digital function of an object with the user's mental model of their physical use. For example, erasers are generally used for the reversal of mistakes, which is why we assigned it the undo-action and an interaction technique that simulates the use of an eraser. With more involved interactions, such as adjusting the zoom level of an image, we put more emphasis on form factors and interaction techniques afforded by them, rather than similarities of purpose.

Some of the interaction techniques selected only produced a momentary trigger or switched between two binary states, such as pressing down a stapler or performing a gesture with an eraser. We opted to simulate these interaction techniques with a Wizard-of-Oz approach, as doing so would be simpler than implementing a robust detection of these interaction techniques. More involved interaction techniques, such as dialing a continuous value with a rotating object, required more precision than discernible with the naked eye and were detected automatically. We implemented three types of controls: A dial which computed a value from the z-angle of an object (used for the bottle), a touchpad which returned the position of a fingertip within certain bounds around a marker (used for the folder), and a state selector detecting whether a corresponding marker was visible or invisible (used for the notepad). All other interactions with objects were detected by eye and realized through button presses.

3.3.2 Devices. For our study, we used a PC equipped with an *Intel i9* CPU, 64 GB RAM and a *Nvidia GeForce RTX 2060* graphics card, as well as a *ViewSonic LS 700* 4k projector with a resolution of 3840 × 2160 pixels and a frame rate of 60 Hz. The camera we used was a *Logitech BRIO 500* webcam with a resolution of 4096 × 2160 pixels.

3.4 Task

Related work employs a wide variety of tasks for user studies on Opportunistic Tangible Interaction. These Tasks are usually short and limited in scope, aiming to evaluate individual interaction techniques instead of their continuous use over a larger time frame [7, 25, 33, 60, 65]. As a result, there was no suitable task design to be readily repurposed for our study.

We therefore aimed to extend tasks from previous work into longer (~10 minutes) activities that required mixed use of common tangibles for appropriate physical and digital functions. Switching between physical and digital modalities of use could either be mandated at specific points during the task (as in [7, 12, 50]) or left to the participant, trading control over the frequency of clutching interactions for naturalism.

We chose to use one task for each of these two paradigms in our study. The first task was a text editing task in the applications

[2]The code for our experimental setup is available under:
https://github.com/MaximilianKilger/Cant-Midas-Touch-this-Experimental-Setup.git

object	mode of interaction	function in task 1	function in task 2
glass water bottle	rotating the bottle (clockwise and counterclockwise)	scrolling the document (up and down)	zooming in and out of the image
flat paper folder	2D finger sliding	-	moving viewport / frame
notepad	turning pages to reveal/conceal markers	switching between windows	switching between pictures
highlighter	Dragging highlighter on table (in a drawing motion)	selecting all text	marking a spot in the image with a red circle
scissors	opening/ closing scissor blades	cutting text (to clipboard)	-
gluestick	dragging on table (in a drawing motion)	paste from clipboard	-
stapler	moving down lever	create comment in Google Docs	-
eraser	rubbing on table	undo action / revert creation of comment	-

Table 1: Mapping of actions and interaction modalities to objects in task 1 and 2.

Google Docs and *Windows Notepad*, where participants were guided by a 16-step instruction manual, which took them 11 minutes and 18 seconds on average. The steps were grouped into blocks of digital tasks requiring the tangibles, such as copying a text section with the highlighter and the scissors or reading a text while scrolling with the bottle, and physical tasks, such as cutting out cards and gluing them together or answering questions about the text on the notepad. At each transition between these blocks, participants were intended to use the clutch. We did not explicitly tell our participants when to clutch, as we wanted them to engage with our clutch deliberately and not as a mere reaction to a prompt. The task had two variations with different texts, cards, and questions, and a slightly different order of steps. We adapted the idea of interacting with text and answering questions from [7] and [12].

The second task was a "Where's Waldo"-style search task, in which participants navigated through a slideshow of hidden object pictures using tangibles and located specific people. The target individuals' faces were displayed on profile pages placed in the notepad, alongside three activities. When a participant found a matching person, they marked them in the slideshow with a red circle, highlighted the activity they performed, removed the corresponding page from the notepad and filed it in a folder. Each task had three images, with two people to look for in each picture. There were two task variations with different images and people to search for. Apart from an initial tutorial, participants were free to clutch when they needed to, with the physical interactions ensuring that declutching was required whenever a person was found. We adapted the idea from simple slideshow tasks in the literature [5, 11, 65], but added the "Where's Waldo"-element to increase the time that participants were actively engaging with the slideshow. On average, this task occupied participants for 17 minutes and 47 seconds.

3.5 Procedure

The experiment was conducted by two researchers. One assumed the role of *investigator*, guiding the participant through the experiment, explaining the tasks and conditions, preparing questionnaires and conducting the interview. The other assumed the role of *wizard*, observing the participant and triggering clutching interactions and keyboard shortcuts when the participant was performing certain pre-defined actions. This gave the illusion that the system responded to user input automatically (see Figure 2).

Figure 2: Study setup: A participant (foreground) completes a task using an everyday object as a tangible. The wizard (background) observes and manually triggers the effects of the participant's actions.

Participants were greeted and seated at the table, gave their informed consent and filled out a demographic questionnaire. After this, the investigator explained the purpose and procedure of the experiment and the working principle of the interface. Then, the participant completed four trials of the experiment. Each trial saw the participant completing one of the two tasks, then filling out the User Experience Questionnaire and the NASA Task Load Index. Before filling out the questionnaire, the investigator instructed the participant to focus specifically on their interaction with the clutching method used during the current trial, rather than on the task itself or the overall system and prototype. This was intended to mitigate the impact of other factors on participants' answers, providing a clearer understanding of their preferences.

The task type was alternated between trials, so that no subsequent trials contained the same task. The task for the beginning trial was also alternated between participants to prevent sequence effects. While the functionality of tangibles was only dependent on task type, the clutching method changed with every trial. The order of clutching methods was balanced using a Latin square design.

After four trials were completed and participants had experienced all clutching methods, the investigator conducted a short, semi-structured interview. Participants answered which clutching method they considered the most efficient, as well as the clutching method they considered the worst. If necessary, the investigator inquired about the reason for their judgement. If a participant had gone long stretches of time during a trial without using the clutch, the investigator inquired into the cause of this, too. Additionally, participants were asked about their experience of using the OTUI as a whole, if they would consider using an interface like this in their daily lives, and whether they felt that the table setup was realistic. These last questions were included mainly to verify our apparatus. Finally, participants were asked for any thoughts or feedback they still had about the experiment.

3.6 Participants

We recruited 24 participants through a combination of university mailing lists and personal contacts (13f, 11m). The median age was 23 (μ=25.41, std=9.80), with 21 participants within the 20-25 age range. 21 participants were students, and all but 2 participants had a background in Media Informatics. 19 Participants possessed a general university entry qualification as their highest level of education, and 4 participants sported an academic degree. Participants studying Media Informatics were compensated with course credits, if required, and all participants received snacks as a thank-you gift.

4 Results

In this section, we present the results of our analysis. This chapter is split into the analysis of quantitative data (i.e. the two questionnaires *UEQ* and *NASA-TLX* alongside *Clutching Frequency*) and qualitative data (derived from interviews and recordings).

4.1 Quantitative Data

User Experience Questionnaire. We analyzed the User Experience Questionnaire using the standard analysis tool [3]. The UEQ consists of the six subscales *Attractiveness*, *Perspicuity*, *Efficiency*, *Dependability*, *Stimulation* and *Novelty*. The mean values (alongside standard deviation) for each subscale are reported in Table 2.

UEQ subscale	Act. Gesture		Add. Gesture		Act. Zone		Pedal	
	μ	std	μ	std	μ	std	μ	std
Attractiveness	1.55	0.82	1.24	1.43	0.94	1.27	1.23	1.09
Perspicuity	2.14	0.60	1.76	1.01	1.42	1.22	2.06	0.71
Efficiency	1.32	0.90	0.83	1.37	0.67	1.14	1.34	0.88
Dependability	1.10	1.08	0.89	1.22	0.94	1.05	1.34	0.74
Stimulation	1.58	1.04	1.70	1.28	1.40	1.26	1.14	1.34
Novelty	1.32	1.41	1.69	1.15	1.48	1.34	1.02	1.46

Table 2: UEQ subscales per clutching method.

As a Shapiro-Wilk test revealed that all of the subscale scores were non-parametric (except for *Efficiency*), we conducted Friedman tests for each subscale and a repeated-measures ANOVA for *Efficiency*, whose results are reported in Table 4.

	χ^2	p
Attractiveness	4.085	0.252
Perspicuity	3.014	0.389
Dependability	4.533	0.209
Stimulation	7.828	**0.050**
Novelty	8.826	**0.032**
	F	p
Efficiency	2.469	0.069

Table 4: Statistical Analysis of UEQ subscales: Test statistics and p-values for the Friedman-Test (Attractiveness to Novelty) and ANOVA (Efficiency).

As the tests indicated a significant effect on *Stimulation* (Kendall's W=0.109) and *Novelty* (Kendall's W=0.123), we conducted pairwise Wilcoxon signed-rank post-hoc tests with the Bonferroni-Holm-correction for those subscales. While the tests with corrected p-values failed to reveal any significant interactions (all p>0.05), tests with the Bonferroni-Holm-correction omitted suggest differences between Activation Gesture and Pedal (Z=49.0, $p_{uncorrected}$=0.035, Pearson's r=10.002) or Address Gesture and Pedal (Z=62.5, $p_{uncorrected}$=0.037, r=12.758) for *Stimulation*, and between Address Gesture and Pedal (Z=43.0, $p_{uncorrected}$=0.020, r=8.777) for *Novelty*.

[3]https://www.ueq-online.org/Material/Data_Analysis_Tools.zip

NASA-TLX. The NASA TLX questionnaire consists of six subscales *Mental Demand*, *Physical Demand*, *Temporal Demand*, *Performance*, *Effort* and *Frustration*. The means and standard deviations are reported in Table 3.

TLX subscale	Act. Gesture		Add. Gesture		Act. Zone		Pedal	
	μ	std	μ	std	μ	std	μ	std
Mental Demand	38.75	26.26	38.13	22.4	50.42	26.37	44.58	24.93
Physical Demand	28.75	21.33	33.33	25.78	45.21	31.67	29.58	23.03
Temporal Demand	26.25	16.1	31.67	19.37	29.79	17.48	31.46	21.64
Performance	29.38	25.76	30.83	22.3	35.42	25.96	27.5	17.19
Effort	38.12	27.77	38.33	22.0	49.38	27.08	32.5	18.0
Frustration	31.46	22.48	33.96	27.03	46.46	25.69	34.38	22.81

Table 3: NASA TLX subscales per clutching method.

Since none of the subscales were found to be parametric by a Shapiro-Wilk test, we conducted a Friedman test for each subscale. The results are reported in Table 5.

	χ^2	p
Ment. Demand	4.473	0.215
Phys. Demand	7.918	**0.048**
Temp. Demand	0.607	0.895
Performance	4.076	0.253
Effort	7.409	0.06
Frustration	5.365	0.147

Table 5: Statistical Analysis of NASA TLX subscales: Test statistics and p-values for the Friedman-Test.

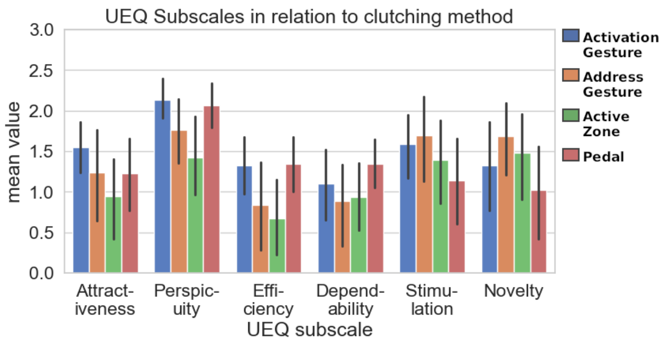


Figure 3: UEQ subscales in relation to clutching method (with 95%-confidence interval).

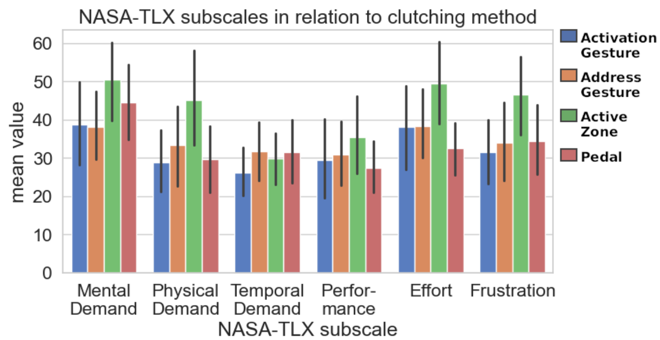


Figure 4: NASA-TLX subscales in relation to clutching method (with 95%-confidence interval).

As the Friedman test indicated a significant effect for the condition *Physical Demand* (Kendall's W =0.0745), we conducted Bonferroni-Holm-corrected pairwise Wilcoxon signed-rank post-hoc tests for this subscale. We found significant effects between Active Zone and Activation Gesture (Z=16.0, p-value=0.024, r=2.309) and Active Zone and Pedal (Z=30.5, p-value=0.025, r= 4.402). Regarding the variable *Task*, all of the subscale scores turned out to be non-parametric as evidenced by the Shapiro-Wilk test. Therefore, we conducted a Wilcoxon signed-rank test that indicated significant differences for *Effort* (Z=358.5, p-value=0.046, r=51.745) between the two tasks (see Figure 4).

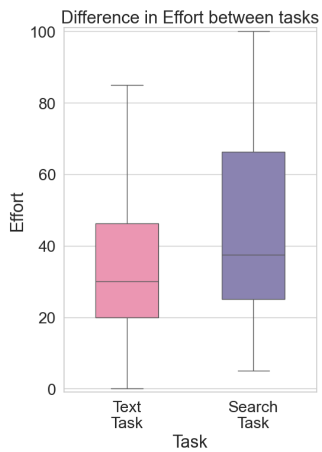


Figure 5: NASA TLX *Effort* subscale in relation to task type.

	Pedal		Active Zone		Address Gesture	
	stat	p	stat	p	stat	p
Activation Gesture	-0.457	0.246	-1.599	**<0.001**	-0.933	**0.002**
Address Gesture	0.458	0.267	-0.666	0.049		
Active Zone	1.124	**<0.001**				

Table 6: Results of the Tukey-HSD test for clutches per minute.

Clutching Frequency. We extracted the total number of clutching interactions and task completion time per trial from the recorded videos. From this, we computed the average number of clutching interactions per minute, which was 0.8586 for the Activation Gesture (std = 0.6605), 1.792 for the Address Gesture (std = 0.6577), 2.458 for the Active Zone (std = 0.870) and 1.334 for the Pedal (std = 1.061), as seen in Figure 6. Raw task completion time was discarded from our analysis, as a Mann-Whitney-U test revealed considerable sequence effects of TASK ORDER (U=1898.5, p=9e−13), masking all effects of CLUTCHING METHOD.

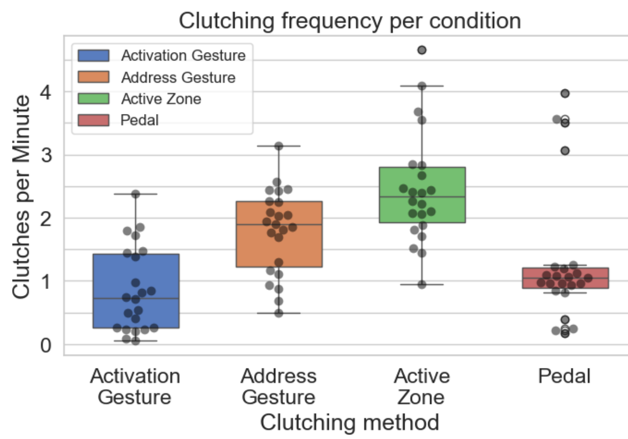


Figure 6: Average clutching interactions per minute in relation to clutching method.

We conducted a three-way ANOVA to find the effect of TASK TYPE, TASK ORDER and CLUTCHING METHOD on CLUTCHES PER MINUTE. The ANOVA revealed a significant effect for TASK ORDER (F=6.915, p=0.0102) and CLUTCHING METHOD (F=15.884, p<0.001), but not TASK TYPE (F=1.242, p=0.2684). As a post-hoc-test, we conducted Tukey's HSD test for all levels of CLUTCHING METHOD, the results of which are reported in Table 6.

4.2 Qualitative Data

Beyond the quantitative analysis, we conducted semi-structured interviews to explore user experiences, preferences, and the reasoning behind their assessments of the clutching methods. Additionally, we used video recordings to make observations about participant behavior.

We analyzed the interviews by performing initial transcription of the recordings using an LLM[4]-powered tool [46]. We double-checked and corrected these transcripts using our notes taken during the interview and the original recordings. To identify recurring themes and gain insight into user behavior and preferences, we focused on clustering related answers and interpreted them based on frequency. As many participants gave similar answers to the same questions, we classified answers based on similarity and determined the most frequent answers per question from those categories.

4.2.1 Interviews. Participants were asked to identify their "most efficient" and "least favorite" clutching methods (multiple answers allowed), providing valuable insights for both qualitative and quantitative analysis of participant preferences. *Address Gesture* was named most often as the most efficient clutching method, while *Active Zone* and *Pedal* were most commonly named as the least preferred methods (see Figure 7).

When giving reasons for their choice, participants often mentioned the ability to choose the activated object directly, but did so significantly more often for the Address Gesture (9 times) than for the Active Zone (2). Those partial to the Active Zone often underscored that interacting with an object already entailed bringing it towards oneself and into the Active Zone (3), making it a natural way of clutching and declutching. Participants favoring the Activation Gesture explained their preference mostly with its hedonic qualities, such as the quick (5) and easy (5) manner in which it was performed, while those favoring the Pedal liked that the ability to clutch while simultaneously doing something else on the table (3).

[4] using OpenAI Whisper 0.28.0 [48] as a backend.

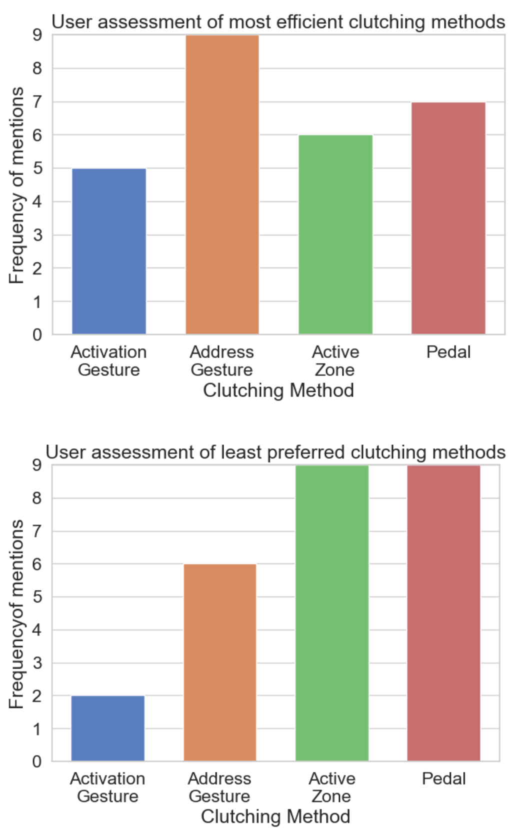


Figure 7: Participants' assessment of most efficient (above) and least preferred (below) clutching method

When discussing their dislike for Pedal and Activation Gesture, participants reported occurrences of the Midas Touch effect when all objects were active (10) — especially when moving the hand over the folder, where touch and hover states could not be distinguished due to limitations of the hand tracking. With regard to the pedal, participants also disliked having to coordinate their feet (5), especially when they could not see the pedal they were supposed to hit (3). With the Address gesture, participants criticized the experienced latency (4), as well as having to perform the gesture as an extra action (4) — the latter criticism also being directed against the Activation Gesture (2). The most points of criticisms were met by the Active Zone: Participants often considered it too small (4) and cluttered (3), and disliked that it required constantly shifting objects in and out (3), that it restricted the space they had for non-digital interactions (3), and the mental toll of keeping track of its location and the objects currently inside or outside of it (2).

When asked whether they found this type of control useful or a hindrance, six out 24 of participants stated that they considered it useful. Eight people stated that traditional input methods, such as mouse and keyboard, were significantly faster and more efficient. These assessments primarily referred to the control of specific tangibles, such as the folder or the bottle, that have no connection to the function they control, or to individual clutching methods, rather than the overall concept. Certain interactions worked better for specific objects — marking text was seen as practical, while using scissors for cutting or a bottle for scrolling was perceived as less intuitive. Task complexity also influenced participants' perceptions. During Task 2, participants experienced frustration, describing the interaction as slow and inconvenient.

When asked whether they could imagine using this method of control in daily life, nine participants stated they would use it. Six other participants expressed uncertainty, stating that this system requires time to adapt before using it in daily life. They also stated that the usefulness would depend on how the concept develops over time. Nine participants stated that they would not use the system regularly. While many did not see potential in its use for desk work, they imagined potential applications in gaming (3), educational contexts (2) or for accessibility purposes (1). The desk environment was largely perceived as authentic, with 15 participants stating that the study setup felt realistic.

Of the ten participants who forgot to declutch for prolonged periods during the tasks, six stated as their reason that they had simply forgotten to do it. Explanations for this were varied: Some were too focused on the task, one person didn't think unwanted effects could occur, and one blamed the position of the pedal under the table where they couldn't see it. One person also found that covering markers with their thumb was a faster alternative to clutching.

Participants also mentioned technical issues, such as the false-positive recognition of markers on the notepad leading to unintentional switching between pictures and applications. Also, using the Address Gesture when objects were close together sometimes resulted in the wrong object being activated.

4.2.2 Observations. While using the Pedal or the Activation Gesture, some participants failed to deactivate the tangibles for a long time, sometimes throughout the entire task. Many of these participants started properly deactivating the tangibles after experiencing the Midas Touch effect for the first time.

For Task 2 in particular, participants tended to place the main tangibles (folder, bottle and notepad) in a specific arrangement throughout the task and kept it that way, even across conditions. This behavior was especially common for the *Active Zone* condition.

During physical tasks in the *Active Zone*-condition, some participants would place and move items in quite inconvenient ways to keep them outside the Active Zone. Even though participants could move the mat that delimited the Active Zone, most did not.

Also during Task 2, participants would often try to use the eraser to undo a mark they created using the highlighter, even though this interaction was not implemented and not mentioned in their instructions. When prompted to select text in a TXT-file on the computer, a few individual participants erroneously tried to select printed text on the notepad with the highlighter.

5 Discussion

Our analysis revealed significant differences in user experience and workload across the four clutching methods. The UEQ questionnaire showed notable effects for *Stimulation* and *Novelty*, with the Pedal method tending towards lower ratings compared to Activation and Address Gestures. In contrast, Efficiency did not show significant differences between conditions.

The NASA-TLX questionnaire showed a significant effect for Physical Demand, where the Active Zone was more *physically demanding* compared to all other methods. Furthermore, task type influenced the *Effort* subscale, suggesting variations in cognitive and physical strain between the two tasks.

The analysis of clutching interactions per minute revealed significant differences, with the Active Zone method leading to the highest interaction frequency. The Address Gesture also resulted in a higher clutching rate than the Activation Gesture and Pedal method.

Participants also showed a diversity of preferences towards different clutching methods, highlighting different advantages and disadvantages for each method. Their feedback, alongside observations made while revising the recordings, paint a picture of the individual strengths and weaknesses of different clutching methods.

Below, we will discuss the implications of these results.

5.1 Comparison of clutching methods

Although the Activation Gesture was rarely mentioned as the most efficient clutching method, few people seemed to dislike it. While some participants liked that all objects were activated and deactivated at once, more participants disliked that behavior. The main advantage of the Activation Gesture seems to lie in its hedonic qualities, as evidenced by interview answers and the good scores on some scales of the UEQ, such as Efficiency and Attractiveness. This reaffirms our decision for using the double-tap as a clutching gesture. From our perspective, these results characterize the Activation Gesture as a neutral, baseline method for clutching.

The Address Gesture was considered the most efficient clutching method by one third of all participants, mostly due to the ability to select the active tangible separately. While this quality seemed to be the most well-liked aspect of the clutching method, it also requires more clutching interactions to activate all necessary tangibles than other methods, such as the Activation Gesture. The pointing gesture itself was hard to anticipate and the identification of the selected object was sometimes ambiguous. These two factors slowed down the response time of the wizard and lead to a perception of the Address Gesture as sluggish and imprecise.

While the Active Zone was only rated as the third most efficient gesture, it offered a unique advantage as it fit right into the natural movement sequence of using tangibles. Participants noted that they intuitively placed the objects they wanted to interact with within reach, eliminating the need for additional activation gestures. However, it also required participants to constantly reposition objects when using them for physical tasks, especially as the size of the zone (41×30cm) wasn't big enough for several objects at once. This was reflected in an elevated *Physical Demand* in the NASA TLX. For our research, we chose a smaller desk pad to ensure that it would fit with other items on the table. A larger pad might have left no space to complete physical tasks with the tangibles elsewhere. We found that with the Active Zone, even minor factors like placement and size involve a trade-off between the physical and digital use of tangibles.

The Pedal was considered the second most efficient method, mostly because it allowed participants to activate objects without using their hands, allowing them to stay focused on other tasks. Many participants valued its straightforwardness, low required effort, and the ability to trigger all objects simultaneously. However, despite its efficiency, the Pedal was also among the least favored methods alongside the Active Zone. UEQ data suggests that it lacks the fun factor and novelty of other clutching methods. Many participants found coordinating their feet challenging, especially since the pedal was often out of sight under the table. Additionally, the Midas Touch effect could still occur when all objects were active simultaneously, which some found disruptive. While it required less physical effort than other methods like the Active Zone or Address Gesture, it still demanded an extra action, making it feel less intuitive. This contrast highlights how efficiency alone does not necessarily translate to user preference, as the Pedal's usability depended on individual comfort with foot coordination.

Our results do not point to a single clutching method as definitively the most efficient. Instead, each method comes with its own trade-offs, making efficiency dependent on user preferences and context. However, the findings suggest that Activation and Address Gesture may be the more promising approaches. The Activation Gesture was perceived as a neutral method with few strong criticisms, while the Address Gesture was rated as the most efficient by a third of participants. In contrast, the Active Zone and the Pedal, despite their efficiency in some aspects, were the least preferred overall. This indicates that future implementations of Opportunistic Tangible User Interfaces should prioritize the strengths of Activation and Address Gestures while addressing their usability challenges. Although Pedal and Active Zone might still be a viable solution to the Midas Touch problem in some contexts, their limitations will need to be considered and carefully designed around.

5.2 General observations on clutching and opportunistic tangibles

Providing users with a clutching mechanism does not automatically solve the Midas Touch problem. The responsibility for deactivating unnecessary tangibles remains with the user. Very few participants declutched tangibles right after they finished using them, and as such, our trials often featured several objects being active at a time. Unused tangibles (especially the folder / trackpad) could therefore still trigger the Midas Touch effect. Given the novelty of our system and the cognitive demands of the task, it appears that integrating clutching into the workflow required additional effort from our participants. Like them, some users may find it challenging to consistently remember to use the clutching mechanism.

These effects seemed to be more pronounced with clutching methods that activated and deactivated all objects at once, namely the Activation Gesture and the Pedal. With clutching methods that can select the tangible to be activated, participants seemed to clutch more carefully. While this selectivity reduces the risk of the Midas Touch effect, it also increases the overall interaction load, as these methods led to a much higher number of clutching interactions. Participants noted this trade-off especially when needing to activate several tangibles in quick succession. Thus, clutching frequency appears to be a critical design consideration for clutching methods. Some of the potential for interfering interactions across tangibles may be mitigated through more precise detection methods, although we still anticipate a trade-off here which future work will have to consider.

In terms of the interface itself, our participants usually needed some time to get used to the novel interaction method. While interacting with everyday objects seemed to be fun and enjoyable to our participants, it is no replacement for traditional input devices. This is in line with the approaches of previous work, who conceptualized Opportunistic Tangible User Interfaces as an auxiliary technology, placing the emphasis on its availability and flexibility rather than optimal usability [16, 20].

5.3 Limitations

While we strove to design our experiment in accordance with previous work, a number of our design decisions might limit the generalizability of our study results. Our apparatus combined automatic detection with a Wizard-of-Oz design, making it difficult to assess the impact of both human and machine error on our results. A fully automatic detection of clutching interactions, with different ways of observing input and resolving ambiguities, might shift user preferences towards different methods of clutching than those suggested by our results. Additionally, the idiosyncrasies of the selected tangibles, their associated interaction techniques and user's mental model of them also influenced our study in ways that are hard to characterize. Several users found individual interaction techniques unnatural, indicating flaws in our interpretation of an OTUI that limit comparability to other OTI systems.

Another limitation of our study is the sampling of participants. The majority of participants were young adults, which were primarily current or former students of Media Informatics. As a result, our sample might be more technically knowledgeable and more enthusiastic about novel interaction methods compared to the general population. Our recruitment method allowed us to find a relatively sufficient number of motivated participants within limited time, at the cost of representativeness compared to the general population.

Since participants were aware of the experimenter's presence, we acknowledge that observer effects may have subtly influenced their behavior. Social acceptability of clutching techniques was not the focus of this study but may be relevant for future research in public or shared workspaces.

Additionally, we only explored a limited number of potential clutching methods. The most notable method not included in our study is definitely voice activation, as explored by Cronin et al. [12]. Furthermore, we were unable to confirm whether the gestures selected for the *Activation-* and *Address Gestures* were the optimal choices.

Another design decision which might present a limitation is the dual-task-design of our experiment. Previous work studied clutching interactions for specific use cases [12, 23, 49], which stands in contrast to our use-case-agnostic investigation on clutching interactions. The same goes for previous OTUI studies, which did not focus on transitions between virtual and physical interactions [7, 60, 65]. While we grounded our task design in related work, we needed to extend previous tasks to allow frequent clutching with a variety of tangibles. The use of two kinds of tasks made it possible to study both guided and unguided clutching interactions. It provided variety between different trials and allowed us to "hedge our bets" on our two unproven task designs, but limited comparability of within-subject results within our study and with previous and future work.

Finally, the number of clutching interactions in our experiment varied widely between conditions. While trials with very low clutching frequency may be regarded as an indicator of low acceptance for a clutching method, it limits the validity of self-reported usability metrics and the comparability of individual trials. Our decision against explicitly prompting participants to clutch aimed to increase ecological validity, but involved a trade-off with the validity of our measurements.

6 Conclusion

Clutching interactions can be a useful way of preventing the Midas Touch effect. We compared four different clutching methods for an Opportunistic Tangible User Interface in the context of an interactive tabletop: *Activation Gesture*, *Address Gesture*, *Active Zone* and *Pedal*. To this end, we conducted a user study with two tasks, in which participants completed two different tasks with all four clutching methods. Although no single method emerged as the universally most efficient or most favored, our results provided valuable insights into the strengths and weaknesses of each approach.

6.1 Primary Findings

To answer our research questions, we analyzed how different clutching methods compare in terms of user experience and user preference between various clutching methods. The Activation Gesture emerged as a neutral option, while the Address Gesture was considered the most efficient, followed by the Pedal and Active Zone method. The Active Zone was polarizing, with some participants appreciating its natural synergy with object placement habits while others found its spatial limitations restrictive. The Pedal, although rated the second most efficient method, faced criticism for its reliance on hand-foot coordination and the Midas Touch effect, which detracted from its overall usability.

In terms of user preference, the Activation and Address Gesture were the most favored methods. The Activation Gesture was seen as a neutral, easy-to-use method and broadly acceptable, as it provided simple and predictable interaction with minimal criticism. The Address Gesture, despite its complexity and issues with accuracy, was appreciated for its ability to select specific objects, making it the most preferred method for its efficiency. In contrast, the Active Zone was criticized for requiring constant object repositioning resulting in a higher work load. The Pedal was disliked due to its reliance on foot coordination and tendency to activate all objects at once, leading to a less intuitive experience.

Overall, our study did not identify a single clutching method as definitively the most efficient. Each method has advantages and disadvantages that can be used depending on the context of use. However, the individual characteristics of clutching methods emerging during the study lead us to consider the Address- and Activation Gesture to be more broadly suitable for Opportunistic Tangible Interaction.

6.2 Impact and future work

Our work sheds light on possible solutions for the Midas Touch problem in Opportunistic Tangible Interactions. Our insights into different clutching methods can help designers of future OTUI systems implement clutches that suit the specific context of use. Future work in this field may compare our findings with the evaluation of other explicit clutching methods, such as voice clutches, as well as implicit clutching methods, such as gaze direction. Our results suggest that future work should build on the Activation Gesture and Address Gesture, as they were either more widely preferred methods or lacked major drawbacks. The Activation Gesture provides a neutral, reliable baseline, while the Address Gesture allows for precise activation but requires robust automatic recognition and target detection to make up for its shortcomings in our Wizard-of-Oz setup. Future research could focus on enhancing these strengths and addressing their limitations to create more intuitive and efficient clutching interactions.

As an initial, general evaluation of clutching methods, we left the context of use beyond a generic interactive tabletop intentionally vague. However, we also found context-sensitive aspects such as the location of the Pedal and the form factors of the Active Zone to have a profound impact on our results, which underlines how understanding context is vital for understanding clutching methods in general. While out of the scope of our work, the influence of context factors like social acceptability, different use-case-scenarios or the interplay with other forms of interaction on clutching methods remains an important field of study.

As our evaluation was exploratory in nature and relied in part on a Wizard-of-Oz setup, we recognize the limitations of our current metrics and study design in capturing the full complexity of clutching interactions. Future work should include setups with automatic recognition systems to reduce bias and improve measurement reliability. Additionally, future studies could incorporate more diverse and context-sensitive evaluation metrics — such as long-term usage patterns, error rates, or social acceptability — to complement subjective assessments and better reflect real-world usability.

We also developed two tasks to specifically study switching behavior between human-computer-interaction and physical modalities of tangibles. To our knowledge, this is a relatively novel requirement for task design in HCI, and we are curious to see how future work will validate and expand on our task designs.

While the capabilities of Opportunistic Tangible User Interfaces are constantly being improved, practical issues like the Midas Touch effect remain a non-technological barrier to the wide-spread adoption of OTUIs. Our work contributes to the understanding of how issues like this can be solved, potentially aiding in the marriage of capability and practicality that might one day make Opportunistic Tangible Interaction a part of our everyday lives.

References

[1] Michael Achmann-Denkler. 2022. Objects on tables in everyday use: A dataset. (2022). (unpublished).

[2] Alissa N. Antle, Milena Droumeva, and Daniel Ha. 2009. Hands on What?: Comparing Children's Mouse-based and Tangible-based Interaction. In *Proceedings of the 8th International Conference on Interaction Design and Children (IDC '09)*. ACM, New York, NY, USA, 80–88. doi:10.1145/1551788.1551803

[3] Thomas Baudel and Michel Beaudouin-Lafon. 1993. Charade: remote control of objects using free-hand gestures. *Commun. ACM* 36, 7 (1993), 28–35.

[4] Patrick Baudisch, Torsten Becker, and Frederik Rudeck. 2010. Lumino: tangible blocks for tabletop computers based on glass fiber bundles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1165–1174.

[5] Vincent Becker, Sandro Kalbermatter, Simon Mayer, and Gábor Sörös. 2019. Tailored Controls: Creating Personalized Tangible User Interfaces from Paper. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces (ISS '19)*. Association for Computing Machinery, New York, NY, USA, 289–301. doi:10.1145/3343055.3359700

[6] G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

[7] Kai-Yin Cheng, Rong-Hao Liang, Bing-Yu Chen, Rung-Huei Laing, and Sy-Yen Kuo. 2010. iCon: utilizing everyday objects as additional, auxiliary and instant tabletop controllers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1155–1164. doi:10.1145/1753326.1753499

[8] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. 2021. SemanticAdapt: Optimization-based Adaptation of Mixed Reality Layouts Leveraging Virtual-Physical Semantic Connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 282–297. doi:10.1145/3472749.3474750

[9] Yi Fei Cheng, Christoph Gebhardt, and Christian Holz. 2023. Interactionadapt: Interaction-driven workspace adaptation for situated virtual reality environments. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[10] Keywon Chung, Michael Shilman, Chris Merrill, and Hiroshi Ishii. 2010. OnObject: gestural play with tagged everyday objects. In *Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology (UIST '10)*. Association for Computing Machinery, New York, NY, USA, 379–380. doi:10.1145/1866218.1866229

[11] Christian Corsten, Ignacio Avellino, Max Möllers, and Jan Borchers. 2013. Instant user interfaces: repurposing everyday objects as input devices. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces (ITS '13)*. Association for Computing Machinery, New York, NY, USA, 71–80. doi:10.1145/2512349.2512799

[12] Sean Cronin, Euan Freeman, and Gavin Doherty. 2022. Investigating clutching interactions for touchless medical imaging systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–14.

[13] Peter Dalsgaard and Kim Halskov. 2012. Tangible 3D tabletops: combining tangible tabletop interaction and 3D projection. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*. 109–118.

[14] Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. 2021. Towards gaze-based prediction of the intent to interact in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications (ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3448018.3458008

[15] Adam Drogemuller, James Walsh, Ross T. Smith, Matt Adcock, and Bruce H Thomas. 2021. Turning everyday objects into passive tangible controllers. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21)*. Association for Computing Machinery, New York, NY, USA, 1–4. doi:10.1145/3430524.3442460

[16] Ruofei Du, Alex Olwal, Mathieu Le Goc, Shengzhi Wu, Danhang Tang, Yinda Zhang, Jun Zhang, David Joseph Tan, Federico Tombari, and David Kim. 2022. Opportunistic Interfaces for Augmented Reality: Transforming Everyday Objects into Tangible 6DoF Interfaces Using Ad hoc UI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–4. doi:10.1145/3491101.3519911

[17] George W. Fitzmaurice, Hiroshi Ishii, and William A. S. Buxton. 1995. Bricks: Laying the Foundations for Graspable User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 442–449. doi:10.1145/223904.223964

[18] Alexandre Freitas, Diego Santos, Rodrigo Lima, Carlos Gustavo Santos, and Bianchi Meiguins. 2023. Pactolo Bar: An Approach to Mitigate the Midas Touch Problem in Non-Conventional Interaction. *Sensors* 23, 4 (Jan. 2023), 2110. doi:10.3390/s23042110 Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[19] Markus Funk, Oliver Korn, and Albrecht Schmidt. 2014. An augmented workplace for enabling user-defined tangibles. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 1285–1290. doi:10.1145/2559206.2581142

[20] Weilun Gong, Stephanie Santosa, Tovi Grossman, Michael Glueck, Daniel Clarke, and Frances Lai. 2023. Affordance-Based and User-Defined Gestures for Spatial Tangible Interaction. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1500–1514. doi:10.1145/3563657.3596032

[21] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[22] Fengming He, Xiyun Hu, Jingyu Shi, Xun Qian, Tianyi Wang, and Karthik Ramani. 2023. Ubi Edge: Authoring Edge-Based Opportunistic Tangible User Interfaces in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3544548.3580704

[23] Florian Heinrich, Kai Bornemann, Laureen Polenz, Kai Lawonn, and Christian Hansen. 2023. Clutch & Grasp: Activation gestures and grip styles for device-based interaction in medical spatial augmented reality. *International Journal of Human-Computer Studies* 180 (Dec. 2023), 103117. doi:10.1016/j.ijhcs.2023.103117

[24] Steven Henderson and Steven Feiner. 2010. Opportunistic Tangible User Interfaces for Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 16, 1 (Jan. 2010), 4–16. doi:10.1109/TVCG.2009.91 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[25] Steven J. Henderson and Steven Feiner. 2008. Opportunistic controls: leveraging natural affordances as tangible user interfaces for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology (VRST '08)*. Association for Computing Machinery, New York, NY, USA, 211–218. doi:10.1145/1450579.1450625

[26] Anuruddha Hettiarachchi and Daniel Wigdor. 2016. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1957–1967. doi:10.1145/2858036.2858134

[27] Clemens Holzmann and Andreas Hader. 2010. Towards tabletop interaction with everyday artifacts via pressure imaging. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction (TEI '10)*. Association for Computing Machinery, New York, NY, USA, 77–84. doi:10.1145/1709886.1709901

[28] Eva Hornecker and Jacob Buur. 2006. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. Association for Computing Machinery, New York, NY, USA, 437–446. doi:10.1145/1124772.1124838

[29] Masoumehsadat Hosseini, Tjado Ihmels, Ziqian Chen, Marion Koelle, Heiko Müller, and Susanne Boll. 2023. Towards a Consensus Gesture Set: A Survey of Mid-Air Gestures in HCI for Maximized Agreement Across Domains. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–24. doi:10.1145/3544548.3581420

[30] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. ACM Press, Atlanta, Georgia, United States, 234–241. doi:10.1145/258549.258715 ZSCC: 0004869.

[31] Howell Istance, Richard Bates, Aulikki Hyrskykari, and Stephen Vickers. 2008. Snap clutch, a moded approach to solving the Midas touch problem. In *Proceedings of the 2008 symposium on Eye tracking research & applications (ETRA '08)*. Association for Computing Machinery, New York, NY, USA, 221–228. doi:10.1145/1344471.1344523

[32] Robert JK Jacob. 1995. Eye tracking in advanced interface design. *Virtual environments and advanced interface design* 258, 288 (1995), 2.

[33] Rahul Jain, Jingyu Shi, Runlin Duan, Zhengzhe Zhu, Xun Qian, and Karthik Ramani. 2023. Ubi-TOUCH: Ubiquitous Tangible Object Utilization through Consistent Hand-object interaction in Augmented Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3586183.3606793

[34] Chang-Min Kim and Tek-Jin Nam. 2022. Exploration on Everyday Objects as an IoT Control Interface. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*. 1654–1668.

[35] Rick Kjeldsen and Jacob Hartman. 2001. Design issues for vision-based computer interaction systems. In *Proceedings of the 2001 workshop on Perceptive user interfaces (PUI '01)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/971478.971511

[36] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*. Springer, 63–76.

[37] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019).

[38] Vitus Maierhöfer, Andreas Schmid, and Raphael Wimmer. 2024. TipTrack: Precise, Low-Latency, Robust Optical Pen Tracking on Arbitrary Surfaces Using an IR-Emitting Pen Tip. In *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–13.

[39] Nicolai Marquardt, Ricardo Jota, Saul Greenberg, and Joaquim A Jorge. 2011. The continuous interaction space: interaction techniques unifying touch and gesture on and above a digital surface. In *Human-Computer Interaction–INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part III 13*. Springer, 461–476.

[40] Pallavi Mohan, Wooi Boon Goh, Chi-Wing Fu, and Sai-Kit Yeung. 2018. DualGaze: Addressing the Midas Touch Problem in Gaze Mediated VR Interaction. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 79–84. doi:10.1109/ISMAR-Adjunct.2018.00039

[41] Elena Mugellini, Omar Abou Khaled, Stéphane Pierroz, Stefano Carrino, and Houda Chabbi Drissi. 2009. Generic Framework for Transforming Everyday Objects into Interactive Surfaces. In *Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction*, Julie A. Jacko (Ed.). Springer, Berlin, Heidelberg, 473–482. doi:10.1007/978-3-642-02580-8_52

[42] Edwin Olson. 2011. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE international conference on robotics and automation*. IEEE, 3400–3407.

[43] Alex Olwal and Andrew D Wilson. 2008. SurfaceFusion: unobtrusive tracking of everyday objects in tangible user interfaces. In *Proceedings of graphics interface 2008*. 235–242.

[44] James Patten, Hiroshi Ishii, Jim Hines, and Gian Pangaro. 2001. Sensetable: a wireless object tracking platform for tangible user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 253–260.

[45] Yesaya Tommy Paulus and Gerard Bastiaan Remijn. 2021. Usability of various dwell times for eye-gaze-based object selection with eye tracking. *Displays* 67 (2021), 101997.

[46] Jennifer Philipp, Kristina Barufke, and Janine Keller. 2024. Transcription Tool: A software assignment. (2024). (unpublished).

[47] Henning Pohl and Michael Rohs. 2014. Around-device devices: my coffee mug is a volume dial. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 81–90. doi:10.1145/2628363.2628401

[48] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*. PMLR, 28492–28518.

[49] Guillaume Rivière, Nadine Couture, and Patrick Reuter. 2010. The activation of modality in virtual objects assembly. *Journal on Multimodal User Interfaces* 3, 3 (April 2010), 189–196. doi:10.1007/s12193-010-0038-0

[50] Julia Schwarz, Charles Claudius Marais, Tommer Leyvand, Scott E. Hudson, and Jennifer Mankoff. 2014. Combining body pose, gaze, and gesture to determine intention to interact in vision-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3443–3452. doi:10.1145/2556288.2556989
[51] Jürgen Steimle, Mohammadreza Khalilbeigi, Max Mühlhäuser, and James D. Hollan. 2010. Physical and Digital Media Usage Patterns on Interactive Tabletop Surfaces. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM, New York, NY, USA, 167–176. doi:10.1145/1936652.1936685
[52] Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019. Experimental analysis of barehand mid-air mode-switching techniques in virtual reality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.
[53] Philip Tuddenham, David Kirk, and Shahram Izadi. 2010. Graspables revisited: multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, Atlanta, Georgia, USA, 2223. doi:10.1145/1753326.1753662
[54] Brygg Ullmer and Hiroshi Ishii. 1997. The metaDESK: models and prototypes for tangible user interfaces. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*. 223–232.
[55] B. Ullmer and H. Ishii. 2000. Emerging frameworks for tangible user interfaces. *IBM Systems Journal* 39, 3.4 (2000), 915–931. doi:10.1147/sj.393.0915 Conference Name: IBM Systems Journal.
[56] Brygg Anders Ullmer. 2002. *Tangible interfaces for manipulating aggregates of digital information*. Ph. D. Dissertation. Citeseer.
[57] Boris Velichkovsky, Andreas Sprenger, and Pieter Unema. 1997. Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem". In *Human-Computer Interaction INTERACT'97: IFIP TC13 International Conference on Human-Computer Interaction, 14th–18th July 1997, Sydney, Australia*. Springer, 509–516.
[58] Panagiotis Vogiatzidakis and Panayiotis Koutsabasis. 2021. Mid-air gestures for manipulation of multiple targets in the physical space: comparing the usability of two interaction models. In *CHI Greece 2021: 1st International Conference of the ACM Greek SIGCHI Chapter (CHI Greece 2021)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3489410.3489425
[59] Stephen Voida, Matthew Tobiasz, Julie Stromer, Petra Isenberg, and Sheelagh Carpendale. 2009. Getting Practical with Interactive Tabletop Displays: Designing for Dense Data, "Fat Fingers," Diverse Interactions, and Face-to-face Collaboration. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '09)*. ACM, New York, NY, USA, 109–116. doi:10.1145/1731903.1731926
[60] James A Walsh. 2013. Tangible Agile Mapping: Ad-hoc Tangible User Interaction Definition. *User Interfaces* 139 (2013).
[61] James A. Walsh, Stewart von Itzstein, and Bruce H. Thomas. 2014. Ephemeral interaction using everyday objects. In *Proceedings of the Fifteenth Australasian User Interface Conference - Volume 150 (AUIC '14, Vol. 150)*. Australian Computer Society, Inc., AUS, 29–37.
[62] Tianyi Wang, Xun Qian, Fengming He, Xiyun Hu, Ke Huo, Yuanzhi Cao, and Karthik Ramani. 2020. CAPturAR: An Augmented Reality Tool for Authoring Human-Involved Context-Aware Applications. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. Association for Computing Machinery, New York, NY, USA, 328–341. doi:10.1145/3379337.3415815
[63] Pierre Wellner. 1993. Interacting with paper on the DigitalDesk. *Commun. ACM* 36, 7 (1993), 87–96.
[64] Niu Ya-Feng, He Jia-Xin, and Liu Jin. 2024. Magilock: a reliable control triggering method in multi-channel eye-control systems. *Frontiers in Human Neuroscience* 18 (March 2024). doi:10.3389/fnhum.2024.1365838 Publisher: Frontiers.
[65] Qian Zhou, Sarah Sykes, Sidney Fels, and Kenrick Kin. 2020. Gripmarks: Using Hand Grips to Transform In-Hand Objects into Mixed Reality Input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3313831.3376313