



Article

StARboard & TrACTOr: Actuated Tangibles in an Educational TAR Application

Emanuel Vonach , Christoph Schindler and Hannes Kaufmann

Institute of Visual Computing and Human-Centered Technology, TU Wien, 1040 Vienna, Austria;
e0828417@alumni.tuwien.ac.at (C.S.); hannes.kaufmann@tuwien.ac.at (H.K.)

* Correspondence: emmanuel.vonach@tuwien.ac.at

Abstract: We explore the potential of direct haptic interaction in a novel approach to Tangible Augmented Reality in an educational context. Employing our prototyping platform ACTO, we developed a tabletop Augmented Reality application StARboard for sailing students. In this personal viewpoint environment virtual objects, e.g., sailing ships, are physically represented by actuated micro robots. These align with virtual objects, allowing direct physical interaction with the scene. When a user tries to pick up a virtual ship, its physical robot counterpart is grabbed instead. We also developed a tracking solution TrACTOr, employing a depth sensor to allow tracking independent of the table surface. In this paper we present concept and development of StARboard and TrACTOr. We report results of our user study with 18 participants using our prototype. They show that direct haptic interaction in tabletop AR scores en-par with traditional mouse interaction on a desktop setup in usability (mean SUS = 86.7 vs. 82.9) and performance (mean RTLX = 15.0 vs. 14.8), while outperforming the mouse in factors related to learning like presence (mean 6.0 vs 3.1) and absorption (mean 5.4 vs. 4.2). It was also rated the most fun (13× vs. 0×) and most suitable for learning (9× vs. 4×).



Citation: Vonach, E.; Schindler, C.; Kaufmann, H. StARboard & TrACTOr: Actuated Tangibles in an Educational TAR Application. *Multimodal Technol. Interact.* **2021**, *5*, 6. <https://doi.org/10.3390/mti5020006>

Academic Editor: Christoph W. Borst
Received: 30 December 2020
Accepted: 5 February 2021
Published: 9 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Tangible Augmented Reality; Tangible User Interface; tabletop interaction; robots; actuation; haptics

1. Introduction

Tangible User Interfaces (TUI) match virtual data with a physical form for interaction, which brings numerous benefits, e.g., intuitive direct interaction and immediate feedback [1,2]. If the computer can also actively move the tangible objects, this is usually called Actuated TUI [3]. For a learning environment, the direct interaction and the involvement of multiple human senses is very beneficial and can improve the cognitive learning process [4,5]. Billingham et al. [6] suggested coupling the principles of TUI with Augmented Reality (AR), introducing the concept of Tangible Augmented Reality (TAR). Mapping physical object manipulations directly to virtual objects removes the extra step of mapping to logical functions, which can make interaction extremely intuitive and fast.

Nevertheless, previous works repeatedly presented results of a traditional desktop setup with mouse interaction outperforming TAR setups [7,8]. The interaction with a mouse is so familiar to us, that it is especially hard to compete with any new interaction technique. Upon a closer look however, most previous work in this area is based on desktop AR using fiducial markers as tangible objects and a separate display or handheld mobile device for output [9]. The resulting distance between the interaction and the perception space causes extra cognitive load for the user [10] and does not utilize the full potential of TAR. In order to fully integrate interaction and perception space, first-person immersive AR is favorable [11].

Billingham et al. [6] envision the physical objects in TAR as important as the virtual ones, providing true spatial registration. This allows seamless interaction with virtual objects just as with real objects, enabling two-handed manipulation with the user's own

hands, the same way as in the real world. First-person immersive AR with direct tangible object interaction (other than fiducial markers) has hardly been researched (see Related Work) and we believe there is a lot of untapped potential.

We propose a novel approach of employing actuated micro robots to provide haptic feedback in immersive AR and investigate the potential of TAR with active physical tangibles in an educational context (Figure 1). For that purpose, we developed an interactive AR learning system for sailing students StARboard, that teaches beginners to understand basic concepts of different bearings and sail trims in various wind situations. In order to provide direct tangible interaction tokens for input as well as output, we employed our previously developed ACTO system [12], that facilitates prototyping of actuated tangibles. For StARboard we equipped the 50 mm cubical micro robots with a button and a rotary knob on top. The sides of the robots as well as the mounted devices match their virtual counterparts in an AR overlay, e.g., a virtual ship, with a button and a sail beam on top. The user can directly interact with the virtual ship, for example pick it up or adjust the sail trim and will inevitably grab the physical ACTO instead. This direct interaction with real world objects is highly intuitive and efficient, and thereby facilitates learning and collaboration. Therefore, it allows us to evaluate the full potential of AR with active physical tangibles in an educational context.

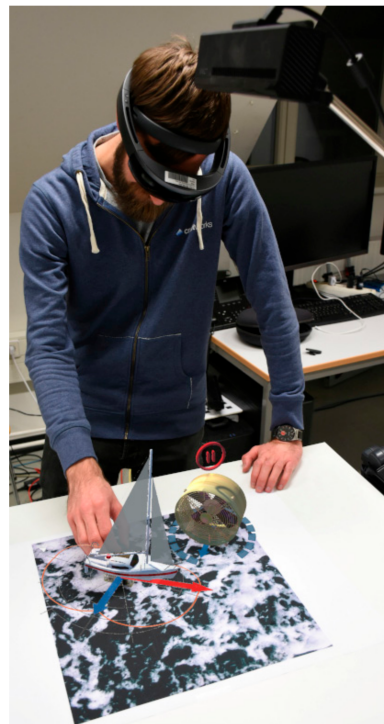


Figure 1. Actuated micro robots with virtual augmentations for direct haptic interaction.

In this work we present the concept and implementation of the educational AR tabletop application StARboard. Furthermore, we developed a tracking solution for the ACTO tangibles TrACTOr, which combines a depth sensor, fiducial markers and the Microsoft HoloLens feature based tracking to allow integration with an AR head-mounted display (HMD) and tracking independent of the table surface. We also report results of our conducted user study with 18 participants, comparing our setup with AR and actuated tangible feedback, with a common AR setup using HoloLens gestures and a traditional desktop setup with mouse input. We can show how our micro robot based TUI for immersive AR can rival mouse interaction in usability and performance in an educational context and even outperform it in factors related to learning like presence, absorption and fun.

Our goals and contributions are:

1. Investigation of employing multiple micro robots as tangibles for direct physical interaction in an immersive TAR environment.
2. Development of the StARboard tabletop application, a multi-user, personal viewpoint AR environment with direct haptic interaction for learning basic sailing concepts and terms.
3. Implementation of the TrACTOr tracking approach for surface independent tracking of actuated tangibles in this immersive tabletop AR setup.
4. A conducted user study with 18 participants, employing the developed application to explore the potential of our TAR approach in an educational context and in comparison to two more common interaction alternatives: AR with gesture interaction and a desktop setup with mouse input.

We structured the paper in a way that after discussion of related work we give an overview over our conceptual approach and system architecture. Next, we present concept and development of StARboard as well as TrACTOr. Then we report on procedure and results of our conducted user study and finally state our conclusions.

2. Related Work

Ever since Fitzmaurice et al. [1] and Ishii and Ulmer [13] established the concept of Tangible User Interfaces (TUI) over 20 years ago, a large and active research field evolved. TUIs allow us to build upon our everyday skills of physical object manipulation for the interaction with a computer since the tangible objects have familiar properties, constraints and affordances. That can bring very convincing properties, like ease of use, direct haptic interaction and multi-person collaboration with parallel user input [1]. For a broad overview over the research area, we recommend the work of Shaer and Hornecker [5].

2.1. Tangible Augmented Reality

Billinghamurst et al. [6] suggested applying the principles of TUI to Augmented Reality (AR), introducing Tangible Augmented Reality (TAR). Physical objects are mapped directly to virtual objects combining the benefits of direct manipulation with the unique display possibilities of AR, which should make interaction potentially extremely intuitive and fast. However, an overview over a wide range of publications in this field (e.g., Zhou et al. [9]) shows that TAR typically employs tracked fiducial markers as tangible objects and relies on hand gestures and marker covering for input. We think this approach does not exhaust the full potential of TAR. Dünser et al. [14] compared physical sliders and tracked paddles to mouse interaction in a desktop AR setup and report that while the accuracy is the same, mouse interaction is significantly faster. We speculate that although a tangible interface matching a task should improve user performance and satisfaction, test users will score generally very high with the common mouse, because despite its shortcomings (for example for 3D manipulation) we are extremely well-trained with the device in comparison to tracked markers that require an introduction and some getting used to. A recent examination by Bach et al. [7] yielded similar results. The authors hypothesized that a virtual environment matching the human capabilities should improve the understanding of 3D visualizations and compared the efficiency of a state-of-the-art TAR setup with Microsoft HoloLens and fiducial markers to a desktop mouse environment. In their study the desktop mouse environment generally also produced superior precision and speed, but they also observed tendencies that TAR performed a lot better for certain tasks. They also performed a smaller longer-term study and showed that the performance of the TAR setup improved with days of training compared to the mouse environment. In the paper the authors discuss that immersive AR should better match how humans perceive and interact with the real world regarding spatial proximity of perception and interaction, high degrees of freedom (DOF) of the tangible interaction and 3D perception. They suggest that the dynamic and interactive nature of AR would provide the best performance for tasks with focus on spatial interaction and perception as well as high DOF of interaction. We strengthen that assessment by exploring how active tangible interaction performs

in a TAR educational scenario and catch up on the lead the mouse has on alternative interaction techniques.

2.2. TAR & Education

The direct mapping between tangible input and digital output as well as the interactions the physical objects afford in a TAR setup, reduce the cognitive effort and complexity of problem solving because it allows the user to focus on the task without being distracted by the control mechanism. Verhaegh et al. [15] compared a tangible and a screen-based version of an educational game and observed that the tangible interface is both easier to learn and easier to use while on screen the learner has to cope with challenges posed by the interface first. TUIs can minimize unintended load caused by the interface and reduce the challenge mostly to the intended tasks. Schneider et al. [4] show that this effect makes TUIs very beneficial for educational applications and allows better task performance and higher learning gain with tangibility. They also describe that the tangibility promotes constructive behavior, e.g., playfulness is regarded as an important component of flow which can get learners so involved in a task that they will do it even at a greater cost. The directness and physical familiarity of the interaction is intuitive for novices, naturally supports collaboration and the external representation of a problem supports exploratory behaviors [16]. Furthermore, physical enacting educational content leads to stronger learning effects [17]. Radu and Antle [18] discuss the implications of this embodied learning in handheld AR. The educational content is spatially and temporally integrated with physical objects, thus the whole body is involved in learning as the user physically interacts with the learning content. This can lower the cognitive load, guide attention and allows risky simulations, offloading part of cognition into the physical world.

These qualities are very compelling, which is why a wide range of research projects investigated the employment of AR and in particular TAR in educational context. For an overview over AR in education and its challenges we refer to Akcayir and Akcayir [19] who recently published an extensive review of the field and found that most publications consistently reported enhanced learning but also that frequently users experienced usability issues or the published systems suffered from technical problems. Furthermore, the majority of TAR projects had in common, that they employed tracked fiducial markers for tangible interaction. A very recent example by Dass et al. [20] helps people to learn to code more easily with the help of AR. They compared first person AR with gestures with marker-based AR on a smartphone and a non-AR app on a tablet. Their results show that AR enhances beginners learning experience, but smartphone-based AR with markers turned out the fastest. While first person head-mounted AR was the most immersive, it was also the most difficult because it required to learn the use of the gestures. However, their results do not provide insights into first person AR with tangibles or how their system would perform compared to a desktop setup with mouse input. Shapira et al. [21] employ optically tracked physical toy blocks to represent virtual switches, batteries and other components in an educational Virtual Reality (VR) environment to experiment with basic electric circuitry. They point out that current limitations of other means for tactile interaction like gloves, ultrasonic blasts, actuated surfaces or gestures are too abstract or cumbersome while in their study children complete tasks a lot faster with the haptic feedback of their physical toys. They also report how the physical nature of the system allows everyone to join in (even without HMD) and the feeling of intuitive and safe interaction promotes exploration, discovery and collaborative play. Although their system provides an active virtual component as they simulate how the circuit components would interact, the physical objects are only passive. In this paper we explore the use of active tangible objects in an educational virtual environment.

2.3. Actuated Tangibles

As an alternative to fiducial markers often passive props are used as physical proxies for virtual objects in AR. These could be specific everyday objects or tools matched with a

virtual representation [22], currently available haptic properties or objects of the physical environment [23] or even arbitrary user selected objects defined as tangibles [24], but as Hettiarachchi and Wigdor [25] point out, you need to find suitable physical props to best represent every interactive object in the virtual environment or a custom set of props for each application, which can be very impractical or costly. In their proposed system they opportunistically find a set of physical objects in the current surrounding of the setup which best represent the required virtual objects. This can work very well, since researchers found out that while shape and size of the employed tangibles matter in terms of a more efficient interaction, more realism and easier manipulation with closer similarity [26], some mismatch is acceptable and it is mostly the basic shape that is important [27]. Nevertheless, the tangible objects are only passive and the system cannot actively produce a haptic output, which is why we suggest an approach with actuated tangibles. There are some researchers building up shape changing displays which can actively adjust the shape of the contact surface to provide the impression of differently shaped surface geometries [28,29]. However, they can only provide a limited display space at great constructional effort and do not support natural manipulations like picking up an object or rotating it. Other researchers employ robots to provide physical props in an immersive environment. Yamaguchi et al. [30] control a drone to position a sheet of paper as a contact surface in a large-scale VR setup. This allows a larger workspace but the possible contact forces are very low and positioning is not very precise. In an ongoing research project [31] we employ a robotic arm to actively present passive objects in an immersive infinite VR environment. While we believe this approach is very promising for large scale virtual environments, it does currently not support the natural object manipulations and experimentation we provide with our actuated micro robots in a desktop setup. Rosenfeld et al. [32] suggested to use two-wheeled motorized robots in their planar manipulator display. In this system the tangible robots could be actively positioned by the system, creating a bi-directional display with multi-handed and multi-user interaction on a tabletop. However, their platforms were only simple remote-controlled vehicles that did not provide more complex interactions or support an immersive AR setup. Kojima et al. [33] employed similar robot toys in a projected AR environment. They superimposed a virtual environment onto the real robots and the output of the virtual game was enacted by the actuated platforms to provide physicality to the projection-based game. In contrast to our approach the system only supported limited direct physical manipulations and was not fully immersive. Recently, Siu et al. [34] proposed to overcome the limitations of shape displays of being large, heavy and only providing a limited interaction area, by integrating an omnidirectional robot to create a mobile shape display for haptic interaction in an immersive tabletop VR environment. In this way their system allows 2D spatial manipulations and self-actuation of a shape display with lateral mobility. Although their display provides a more flexible shape, it is very bulky with over 200 mm³ and manipulations are limited to 2D manipulations. Our micro robots are only sized around 50 mm³ which allows the use of several devices at once and free 3D manipulations by the user. Zhao et al. [35] presented a system which employs passive objects assembled from magnetic blocks, which can then be positioned by active micro robots as interactive tangibles. Their prototype provides only limited actuation of the assembled objects and the possible spatial resolution of the magnetic blocks does not allow high fidelity interactions.

The ACTO system we use as foundation for our developed AR environment is a previously published platform for TUI prototyping [12]. Our **ACTuated Tangible Objects** (ACTO for short) are small cubical mobile robots intended as interaction tokens in a tabletop TUI. Their modular design allows easy adaptations of their physical configuration, and they already provide an actuation mechanism in form of a two-wheel motor module. Each one is compatible with the popular Arduino electronic platform and thus facilitates integration of arbitrary interface components like buttons or rotary encoders. The platform also includes a framework for Android smartphones, providing the basic communication and control

functionality. Please refer to the original publication for an in-depth description of the capabilities of the system [12].

Although the system also provides vision based tracking with a smartphone for a tabletop setup, the approach requires the use of a glass table, which is not ideal for state-of-the-art AR displays due to reflections, and also limits application possibilities to some extent. For that reason we extended the ACTO system with a surface independent tracking solution and tweaked the ACTOs to our needs, as we will describe below.

3. Materials and Methods

3.1. Concept and Architecture

In order to investigate actuated haptic feedback with micro robots in a TAR environment, we conceptualized an AR educational application we called StARboard. This application should help beginners in sailing to understand the basic terminology and correlations of bearings, sail trims and wind conditions. Moreover, it allowed us to explore the potential beneficial effects of such a TUI in an educational context. We hypothesized that the direct and intuitive nature of the tangible interactions would improve efficiency and success of the learning experience to a point, where it could outperform the traditional desktop interaction with mouse and keyboard.

The general concept is presented in Figure 2: A Microsoft HoloLens HMD displays virtual objects to the student in place of the ACTO robots. The ACTOs act as actuated interaction tokens which are fitted with a button, an infrared (IR) LED and a rotary encoder (Figure 3a).

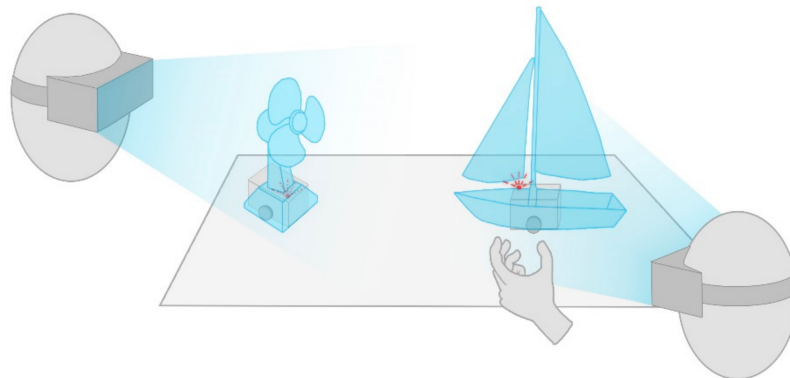


Figure 2. General concept of the StARboard application.

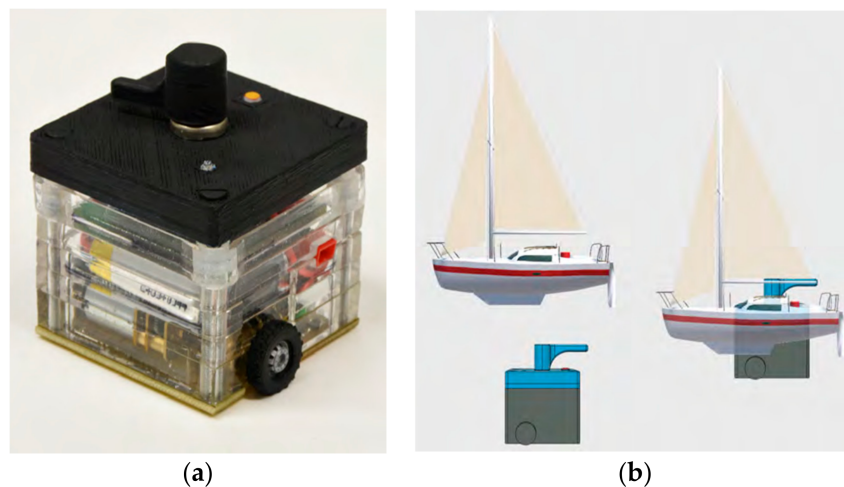


Figure 3. (a) ACTO fitted with button, rotary encoder and IR LED; (b) Virtual counterparts are aligned with the actuated physical interaction tokens.

The virtual overlay of a ship matches the position of an ACTO in such a way, that the interface components (i.e. button and rotary encoder) are aligned with their virtual counterparts (virtual button and beam of ship, see Figure 3b). The student can directly grab the virtual ship, and will physically pick up the ACTO instead, naturally rotating it or changing its location, direction etc. Similarly, the sail trim can be adjusted simply by pushing against the virtual sail beam, thereby moving the rotary encoder.

In Figure 4, we illustrate an overview over the system components and their relation. The positioning of the micro robots is controlled by the ACTO framework via radio frequency (RF), the communication with other components of the system is conducted over WiFi. The original ACTO prototyping platform was developed for Android, which is why the framework is running on a smartphone for our experimental setup. However, for a market-ready application, it could also be implemented for PC. It offers a graphical user interface (GUI) for feedback and diverse control functions. The ACTO robots manage their inner workings themselves, e.g., the motor control or the read-out of attached devices. They communicate any state change back to the ACTO framework.

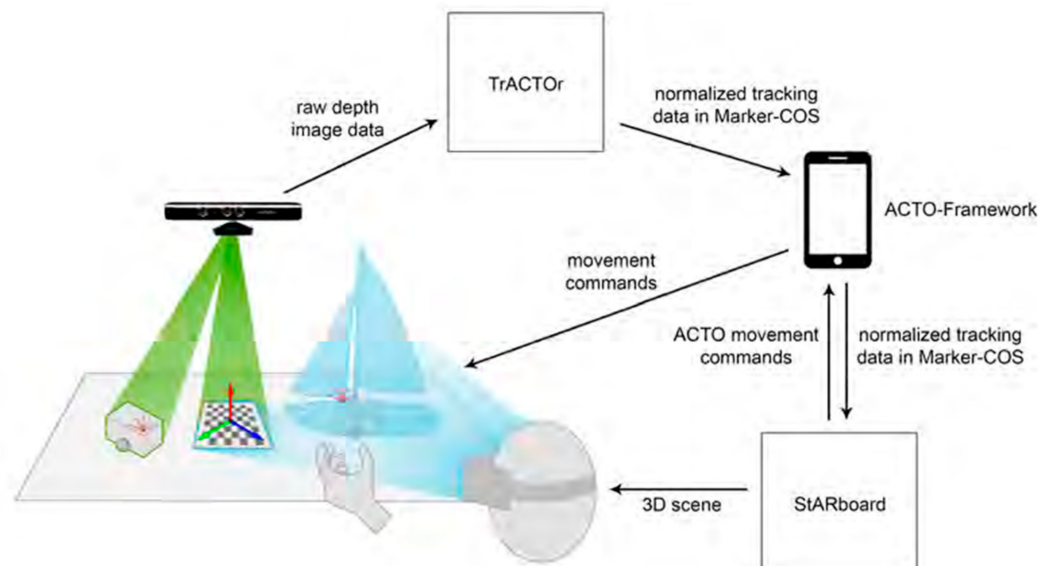


Figure 4. Overview of StARboard system components.

In the original setup, the tracking was performed from below a glass table, which is not ideal for our purposes. In order to be able to incorporate training material, e.g., sea charts or maps, the tracking has to be independent of the table surface. Furthermore, reflective or transparent surfaces like glass are known to cause problems for state-of-the-art AR glasses. For our concept fast and reliable tracking from above the table surface is important that allows accurate real-time alignment for a convincing virtual overlay and intuitive user interactions. For that reason, we developed a tracking system for tangibles based on a Microsoft Kinect depth sensor, the TrACTOr system. It provides positional tracking as well as the rotation around the vertical axis on any flat surface and performs identification of the individual ACTO robots. It runs on a PC and in addition to providing the ACTO tracking it also offers functionality for the management and calibration of the tracking area over a GUI. The acquired tracking data is sent via WiFi to the ACTO framework, from where it is distributed to all the involved software components.

In that way, the StARboard server running the AR application gets the latest tracking data. It manages the application logic of the educational environment and the synchronization of the virtual scene with the tangibles, e.g., which ACTOs have to move where for a maneuver demonstration. The virtual content is calculated on the server and streamed to the AR display clients, like the Microsoft HoloLens or a suitable smartphone, which manage their own viewpoint and render the scene.

In order to align the 3D spaces of the ACTO tracking system, the StARboard application and the AR displays, we use an optical tracking marker that can be recognized with the TrACTOr system as well as the HoloLens or current smartphones running the Vuforia Augmented Reality SDK. The fiducial marker has to be placed on top of the table surface during startup of the system in order to define a common origin for tracking. If the AR clients have their own tracking system integrated, for example in case of the SLAM tracking used by Microsoft HoloLens, it can be removed after the initialization. If individual AR clients, e.g., a smartphone, do not offer their own means of tracking, or the integrated tracking is prone to drift over time, the marker can also be left on the table, e.g., as a playing field, sea map, etc., to provide a fixed origin for repeated re-alignment.

3.2. Implementation—TrACTOr

In order to realize our TAR application we require real-time positional tracking and the rotation around the vertical axis as well as identification of the individual tangibles. The tracking system has to be robust against occlusions, since partially or entirely covered tangibles have to be expected due to the natural interaction with both hands. The tracking has to be fast and accurate to handle moving targets and real-time interaction as well as flexible to allow various setups, surfaces, etc. We also wanted the implementation not to be exclusively tailored to the StARboard environment, but it should integrate with the existing ACTO framework to allow reuse in diverse ACTO applications.

For these reasons, we decided to implement our tracking system based on a combination of feature-based object recognition and depth information. Similar approaches have been successfully employed before for tangible interaction with arbitrary objects (e.g., Funk et al. [24]). In our case, the unique shape of the ACTO robots allows us to use simple features and a lot of flexibility, for example regarding the ACTO equipment.

We employed the Microsoft Kinect 2 as a depth sensor (Figure 5). Based on practical experience, positioning it perpendicular to the center of the tracked space, in a distance of about 700 mm above the surface, proved to be recommendable. It provides an optimal trade-off between the size of the tracking area and the possible accuracy due to the field of view (FOV) of the depth sensor of 70° by 60°. In addition to the depth data, the Kinect 2 also provides a color and an IR image. In our tracking process we use the depth data for acquiring position and orientation and the IR image for identification. For image processing we rely on the open source computer vision library OpenCV and a cross-platform OpenCV wrapper for C# EmguCV to facilitate integration in the Unity 3D development engine.

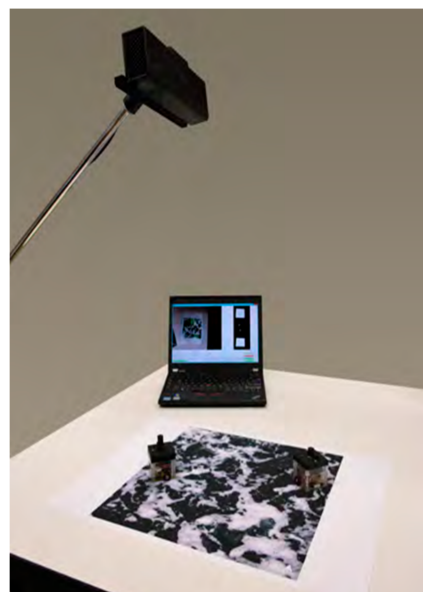


Figure 5. TrACTOr setup with Microsoft Kinect 2.

During the initialization step, an image marker placed on the table surface defines the origin and extent of the tracking area for the application, the playground. It is also used to align the coordinate system of other involved tracking systems, i.e. the AR displays. We normalize the tracking values between $+/-1$ from the edges of the marker inside the playground in order to unify all tracking values and the absolute distances are defined based on the physical size of the marker. The playground can be further delimited by defining an area of interest (AOI) with a GUI during startup to reduce any unnecessary processing overhead. We use the first 10 captured depth images after startup to create a background image of the empty scene, which makes the tracking process more stable against flickering, external light, misalignment and other noise. For that purpose, the tracking surface has to be clear of trackables during startup except for the image marker.

In Figure 6, we present the processing during the tracking phase: The current depth image (a) from the image stream is subtracted from the background image (b). The resulting foreground image contains only the objects of interest and is converted into a binary mask (c). The depth information is retained for the z-position. In the foreground image we search for shapes that could be potential ACTO candidates (d). For that purpose, we employ a Canny edge detector to find the edges in the image (e). In a first level, our algorithm analyzes the detected edges for contours with four vertices, that form rectangles with a certain minimum size and reasonable corner angles (f). The identified candidates could be either new ACTOs, or ones that already have been detected in previous frames. In a second level, we process the detected edges as individual lines. However, in this step we only analyze lines that are close to previously identified ACTO positions to reduce processing overhead. In that way, we can improve reliability and stability of the tracking, and also track partially occluded ACTOs.

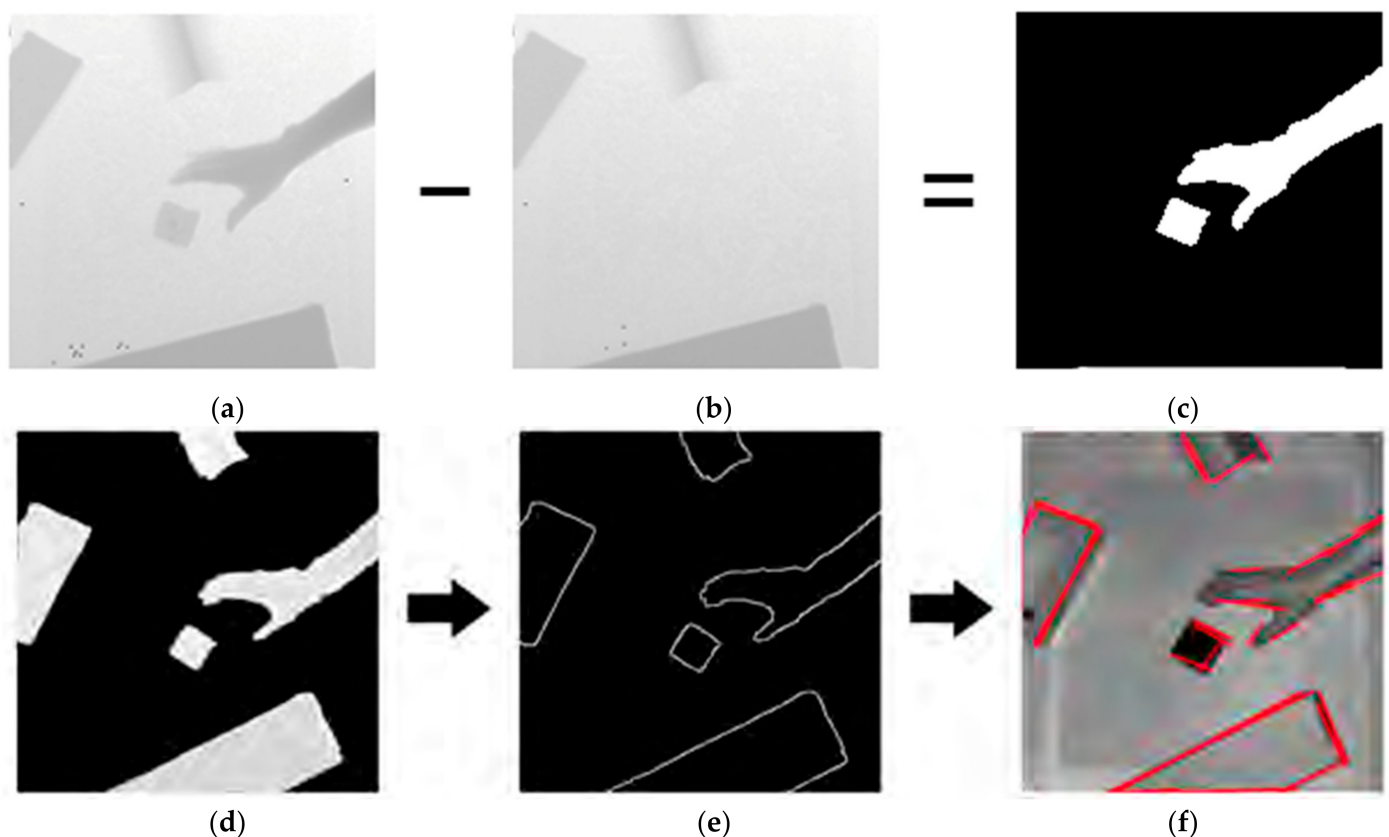


Figure 6. Image processing steps during tracking phase: (a) Current depth image; (b) background image; (c) binary foreground mask; (d) current binary image; (e) detected edges; (f) extracted edges of potential ACTO candidates.

In order to identify the actual ACTOs from the resulting list of candidates, the IR image is processed to find the bright spot caused by the mounted IR LED within the corresponding region of an ACTO candidate. Since the ACTOs are square, the tracked shape provides us with four possible orientations. For that reason, we mounted the IR LED close to a corner of the ACTO, which allows us to use the detected bright spot to derive the correct orientation.

For calculating the new position and orientation for a tracked ACTO, we employ a Kalman state estimator, which predicts the current pose by also including the tracking results from previous frames. This improves tracking quality by smoothing jitter and artifacts from the image sensor. It also allows us to identify and track an ACTO while the IR LED is not visible, by matching a detected rectangle to the pose of a previously identified ACTO. If no matching rectangle can be found in a reasonable distance to a previous ACTO pose, i.e. due to occlusion, the described line detection is applied. Based on information from previous frames, a single detected edge in the vicinity can still provide us with a valid pose. If no line is detected because an ACTO is completely covered, we still estimate its pose for up to one second, until it either reappears or it is assumed lost.

If an ACTO has not been tracked before (or was assumed lost), it has to be uniquely identified first. Using sequentially pulsing LEDs for tracking a TUI has been proven effective before [32]. For that reason, we employ the mounted IR LED to differentiate the individual ACTOs. The numerical unique ID of an ACTO is transmitted optically as a pattern over a number of frames. A message always starts with an initial bit 0, then the ID is coded as a series of 0 followed by a series of 1, and the message always ends with a 0. Only if the start and end bits are correct, we consider a transmitted message valid. Since we also use the LED to track the orientation of an ACTO, we inverted the coding (0 = on, 1 = off), so the LED is on most of the time. With a sampling rate of the Kinect sensor of 30 frames per second, in order to make sure the LED pulse is visible for more than half the sensor rate, a maximum of 15 bits per second could be transmitted optically with a single LED. However, there is a trade-off between bit-rate and error-rate, which is why we only use 10 bits per second to increase reliability. Depending on the length of an ID (related to the number of ACTOs in use), in that way an ID can be transmitted 2 to 5 times per second. In order to account for occlusions or sensor noise, an ID has to be detected twice before we consider an ACTO as identified, and only after the third time, the identification is fixed as reliable.

We extended the existing ACTO framework to accept tracking data over WiFi. In that way, the TrACTOr system streams the tracking data for each identified ACTO to the framework, while the ACTOs transmit the state of the mounted button and rotary encoder over RF. Finally, the ACTO framework shares all the data with the other connected components, i.e. the StARboard server and clients.

3.3. Implementation—StARboard

The StARboard application handles the visualization for the educational AR environment, but also the application logic. We employed the Unity 3D development engine for implementation, since it supports deployment on a variety of platforms, including the Microsoft HoloLens, Android smartphones and PC. We adopted a server client concept for our application. The application server is managing the virtual scene and the program logic. Tracking data and user input are received from the ACTO framework over WiFi and required movement commands are communicated back. All changes in the virtual scene are streamed to one or multiple clients simultaneously, for example AR capable devices like the Microsoft HoloLens, a state-of-the-art smartphone or a PC. They are implemented as thin clients, responsible for managing their own viewpoint on the scene and displaying it.

For the AR overlay we modeled two different ships, a classical trireme and a modern sailboat, as well as a tabletop fan representing the wind source (Figure 7). The virtual models and their physical counterparts, the ACTOs, are equipped with a matching 3D printed

button and a sail beam or knob for a rotary encoder. For reducing possible reflections of IR light, the top of the ACTOs was painted with matte black acrylic color. The virtual models are scaled in order to align well with the sides of the trackables. Any physical change in input, for example grabbing and rotating a ship, is immediately reflected in the visualization as well as the physical behavior of the ACTO robot in order to promote exploratory behavior and experimentation. For an accurate physical feedback, we calculate direction and speed for moving the ACTO robots with respect to the set bearing and sail trim, wind direction and speed and characteristics of ship and sail type. However, in order to simplify the physical model for our application, we disregard other factors like payload, heeling, currents, etc.

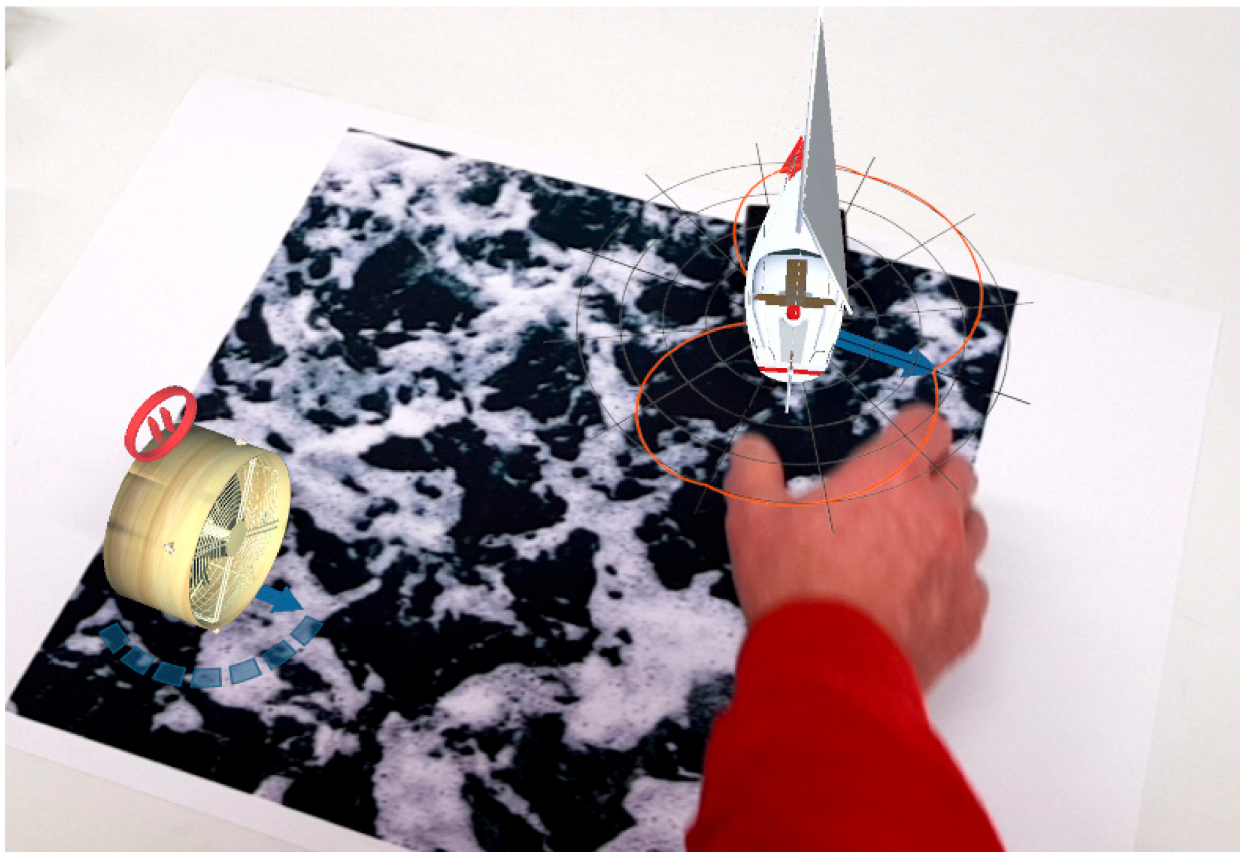


Figure 7. Modern sailboat and fan overlay in StARboard application.

The two ships offer the same interaction modalities, i.e. positioning and rotating, pushing the sail beam and pressing the button. The button switches between different diagrams we implemented to visualize the different characteristics of the ships. A wind diagram shows the true wind direction and speed represented by a blue arrow (Figure 8a). A second red or green arrow (depending on the current mode) presents the resulting driving direction and speed of the ship. This diagram supports the experimentation with different sail trims under different wind conditions and helps to explore the concept of true and apparent wind. Apparent wind is the sum of the wind caused by the ships own motion plus the velocity of the true wind.

Another implemented visualization are polar diagrams (Figure 8b). These diagrams characterize each ship type by illustrating the maximum possible speed in relation to the bearing and the wind conditions. An orange outline around the ship indicates the possible speed per bearing under the current wind speed and direction, assuming ideal sail trim. This makes differences between ship types evident, since for example a classical trireme supports different bearings than a modern yacht.

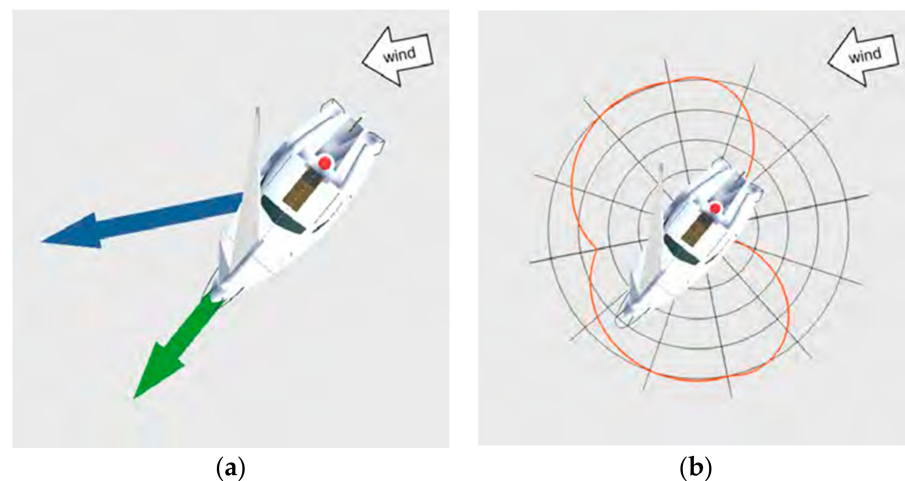


Figure 8. (a) Wind diagram showing true wind as blue arrow and resulting driving direction as green arrow; (b) Polar diagram visualizing.

The tabletop fan allows direct control of wind speed with a rotary knob on top and direction of the wind by positioning it as desired (Figure 9). The speed of the rotating blades as well as a blue arrow and an indicator visualize the current settings. Pressing the button on top allows switching between different play modes. In planning mode, the user can experiment with different bearings or sail trim, for example in combination with the different diagrams, without the ship immediately executing the settings. Switching to live mode with the button, makes the ship move according to the set direction, sail trim and wind conditions. The third mode is the maneuver mode, which is used to demonstrate more complex maneuvers by moving the ship in a preplanned sequence. A verbal audio explanation in English or German language explains the maneuver and points out the important steps, for example a change in sail trim at a certain point. For generating the presented audio tracks as well as facilitating the integration of future additions, we employed the free text-to-speech engine text2mp3. The fourth implemented mode is the test mode, which performs a preplanned movement only up to a crucial point. There the user has to take over and complete the maneuver by setting the required bearing and sail trim. After each test, an audio comment gives feedback about the achieved performance. For a complete example scenario of using the application, please refer to Section 3.5. User Study Design & Procedure.



Figure 9. Fan model representing wind speed and direction.

For evaluation, we also implemented all possible interactions with the application without tangibles. We took particular care to avoid preferring one interaction modality over the other and all possible interactions should be easy and intuitive with any of the modalities.

In case of the Microsoft HoloLens we used the available gesture control and gaze direction. Gazing at either the virtual ship or the fan, moving the extended index finger down towards the thumb and back up (tap gesture) has the same functionality as pressing the hardware button on the corresponding tangibles. In order to rotate a virtual object, the user has to gaze at the desired object (either ship hull, sail or fan), perform the tap gesture but hold the index finger in the downward position (tap and hold) and then move the hand sideways. For the fan, in addition to rotating, moving the hands up and down changes the fan speed. Changing the position of the entire virtual object requires gazing at the object and performing the tap and hold gesture for a second without moving (long tap). After that, the object can be moved on the table surface following the gaze direction and finally placed by ending the gesture. We decided to always stick to the table surface and not allow lifting the object in order to avoid confusion.

For interaction in a desktop setup without AR display, we implemented a classic mouse interaction as well. A left click on the ship or fan corresponds to pressing the hardware button on top of the tangibles. Drag and drop with the right mouse button rotates the selected object (ship hull, sail or fan), while drag and drop with the left mouse button positions the entire object in the scene. The wind speed can be adjusted by scrolling the mouse wheel on top of the fan. In order to account for the free viewing angle into the scene in an AR setup, in the desktop setup the user can rotate the entire scene with the arrow keys on the keyboard.

3.4. Technical Evaluation

We analyzed the most important characteristics of our system from a technical viewpoint. Our test setup consisted of a 4 GHz Intel i7 4790K CPU, 32 GB RAM and an Nvidia GeForce GTX 980Ti running the TrACTOr tracking system. In order to provide us with a clear picture of the individual system components, we employed a separate PC with a 3.4 GHz Intel i7 6800K CPU, 32 GB RAM and an Nvidia Titan Xp GPU for the StARboard server. The ACTO framework was running on a Samsung Galaxy S4 with 2 GB RAM. As an AR client, the Microsoft HoloLens was used. The Microsoft Kinect 2 was mounted 720 mm above the tracking surface. On the surface, we placed a 390 mm by 390 mm image marker in the center and defined a tracking area of 500 mm by 550 mm as a region of interest.

Regarding the frame rate of our TrACTOr system, it is necessary to provide the tracking data for the ACTO robots with more than 24 fps, to permit a matching real time AR overlay that is convincing. The processing effort and therefore the time until the tracking data for a frame is available, increases with each handled ACTO. Since the frame rate of our employed sensor, the Microsoft Kinect 2, is 30 fps, this rate can be considered a bottleneck. We wanted an estimate on how many ACTOs the tracking system can handle, without a negative impact on the available 30 fps. For that purpose, we measured the mean processing time for 1000 frames from receiving a depth frame in TrACTOr to receiving the resulting tracking data at ACTO framework for 0 to 3 ACTOs in a representative scenario (see Table 1). Showing a nearly linear increase in processing time in our measurements, we can roughly estimate a potential of tracking up to 26 ACTOs simultaneously, before dropping below 30 fps.

Table 1. Mean processing time for TrACTOr (1000 frames in representative scenario).

	0 ACTOs	1 ACTO	2 ACTOs	3 ACTOs
Processing time	3.80 ms (SD 0.89)	4.98 ms (SD 1.37)	5.87 ms (SD 0.72)	7.10 ms (SD 1.04)

The identification process for an ACTO over the sequentially pulsing IR LED introduces a delay, if it was not tracked in previous frames. We measured the time for an ACTO sending the first IR pulse of an ID until being recognized and considered reliable by the TrACTOr system. The resulting mean for 10 measurements each was for 1 ACTO 420 ms (SD 35 ms) and for 3 ACTOs 650 ms (SD 26 ms). Measuring the total time from receiving a depth frame in TrACTOr for a previously identified ACTO to the rendered 3D model on the AR client showed a mean overall latency of 110 ms (SD 42 ms) for 14 measurements.

In order to evaluate the accuracy of the provided tracking data, we tracked 2 ACTOs mounted on a measuring stick with a known distance of 130 mm to each other. We measured 10 different positions in the tracking space over 100 frames each and calculated the deviations from the real distance. The resulting mean deviation was 1.46 mm (SD 0.95 mm). For the jitter, we measured a single stationary ACTO at 10 different positions and rotations in the tracking space over 1000 frames each. The results showed a mean variance in x: 0.46 mm (SD 0.42 mm), in y: 0.48 mm (SD 0.42 mm) and in z: 4.71 mm (SD 2.92 mm). The mean rotational variance was 0.70° (SD 0.54°). We repeated the same procedure with two edges occluded and the resulting mean variance was in x: 1.38 mm (SD 1.22 mm), in y: 1.51 mm (SD 1.25 mm), in z: 6.36 mm (SD 4.38 mm) and rotational: 5.10° (SD 3.10°).

This performance can be considered in relation to the maximum resolution of the depth sensor of 512 px by 424 px and the distance to the tracking surface of 720 mm, which results in about 2 mm per pixel. The results for accuracy and jitter for x and y are well below this resolution due to the line detection and filtering executed in our algorithm. The possible maximum tracking surface in this distance is 1000 mm by 830 mm. At the maximum specified Kinect 2 range of 4.5 m the tracking surface would allow a theoretical area of 6.3 m by 5.2 m, but the resolution would degrade to 12.3 mm per pixel.

3.5. User Study Design & Procedure

We conducted a user study in order to make a statement about the effects of direct haptic interaction with AR content on the learning experience and success in an educational application. Since a user should focus on the learning task rather than on the interface, we also emphasized observing the ease of use and intuitiveness of our approach. Finally, we also wanted to know, how enjoyable using our system is, since having fun also has a positive effect on learning. A total of 18 volunteers (five female, 13 male), aged between 22 and 57 years (mean 31.5, SD 8.4), participated in our study. On a scale of 1 to 5 (1 not at all—5 very much), their previous experience with the Microsoft HoloLens was very low (mean 1.9, SD 1.1) while AR experience was intermediate (mean 2.9, SD 1.4). Prior knowledge about sailing was reported rather low (mean 1.8, SD 1.1).

In order to isolate the effect of the haptic feedback in AR, we compared our approach of direct haptic feedback provided by micro robots (condition AR-H) with common gesture interaction without any haptic feedback (condition AR-G). As a third contender, we included the especially well known and familiar classic interaction with mouse on a desktop system (condition D-M). We implemented all the interaction modalities with great care to be as intuitive and comfortable as possible, with the same range of functionality, in order to allow an objective analysis. During the experiment, every participant solved 2 tasks under each condition, one with a bearing and one with a maneuver, to a total of 6 tasks. In order to account for possible learning effects, the order of conditions as well as the order and assignment of tasks were counterbalanced among the participants, so all combinations were tested an equal number of times. During the experiment, the participant was standing in front of the interactive table surface or sitting at the same table in front of a notebook in case of condition D-M. We also recorded the sessions on video, captured the display output and the experimenter took notes of noteworthy incidents. For consistency, the experimenter followed a written test protocol through the whole procedure.

After a general introduction, the experimenter explained the purpose of the study and the participant filled out a short questionnaire with demographic questions and a general consent. For the AR conditions, we included the simple Visually Induced Motion

Sickness (VIMS) questionnaire by Hwang et al. [36] to quantify motion sickness. Then the StARboard application as well as the elements of the test setup were explained, followed by the functionality and visualizations for the first interaction modality, i.e. the available gestures for AR-G or the micro robots for AR-H. The participant could get acquainted with the interaction modality at hand during 4 introductory tasks, covering the basic application mechanics (select, rotate, button and play mode). For the actually rated tasks, the interactive objects (ship and fan) positioned themselves at starting locations. A voice overlay explained a sail trim and a maneuver (i.e. tack, jibe or shoot up), while the ship was moving along a certain path to demonstrate the principle. As a test, the user was asked to set the correct sail trim for a given course. For the maneuver, the system would drive up to a crucial point with a change in bearing, at which the user had to change the sail trim correctly to complete the maneuver. We measured how many tries were required to successfully finish the maneuver. We defined a correct sail trim when providing at least 60% of the maximum possible speed for the situation.

After completing the tasks for an interaction modality, the participant filled out a short post-questionnaire. It presented a post-test VIMS question (in case of the AR conditions) and a System Usability Scale (SUS) by Brooke [37]. Furthermore, we included 4 scales (frustration level, mental demand, performance and temporal demand) from the NASA Task Load Index (TLX) by Hart [38], in this form often referred to as Raw TLX (RTLX). We also added the four items to measure Absorption as a core component of Flow proposed in the Flow Short Scale (FSS) by Rheinberg et al. [39], two items adapted from the Slater-Usch-Steed questionnaire by Usch et al. [40] to measure Presence and two open questions inspired by the Critical Incident Technique (CIT) by Flanagan [41]. We arranged all questions to avoid associated items to follow upon each other directly. The same procedure was repeated for the other interaction modalities. At the end, the participants had to take a short test about the concepts learned, consisting of nine questions (three per condition) in a randomized order. Finally, they could state their opinion comparing all three interaction modalities and give some final thoughts in an open question. Altogether, an average test session lasted for 1 hour 16 minutes.

4. Results and Discussion

In this chapter we present the results from the experiment, employing a Bonferroni corrected alpha level of significance to account for the number of pairwise comparisons.

For the VIMS questionnaire, administered before and after each AR condition, the participants had to rate their current physical condition regarding symptoms of motion sickness. The employed scale ranged from 0 (none) to 4 (very Severe), emphasized through corresponding pictograms. The results represent the change in participants self-assessment between before and after each test condition. With one exception, none of the participants reported a significant change in symptoms (AR-H: mean -0.17 , SD 0.51 ; AR-G: mean -0.11 , SD 0.58). A single test person started with a rating of 0, which changed considerably to 2 (moderate) after the AR-G condition and further to 3 (severe) after the AR-H condition. In view of the ratings from the other participants, we consider this as an individual case of high sensitivity for motion sickness. A Wilcoxon Signed-Ranks test supports that the difference between the two AR conditions is not significant (mean ranks AR-H 2.00, AR-G 2.00, $Z = -0.58$, $p = 1.000$) and therefore does not indicate any systemic issues with our system.

The SUS questionnaire allowed the participants to rate the perceived usability under the different conditions with 10 items, on a scale between 1 (strongly disagree) and 5 (strongly agree) for each item. Following the indicated scoring scheme [37], the mean usability ratings for the different conditions were AR-H 86.7 (SD 11.3), AR-G 59.6 (SD 26.8) and D-M 82.9 (SD 15.2) (Figure 10).

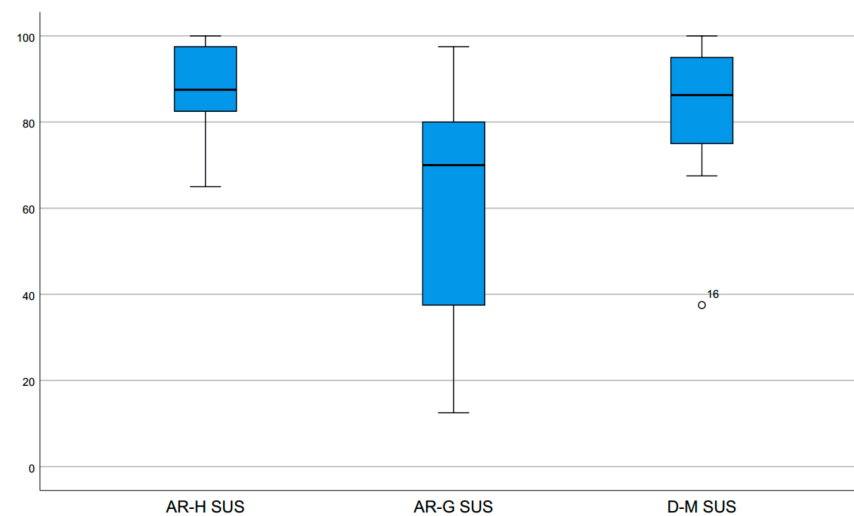


Figure 10. Mean usability ratings (SUS) for the tested conditions.

Since the SUS does not represent a linear scale, the results have to be interpreted in comparison to other systems. According to Sauro [42], anything below 68 is considered below average. Thus, AR-H can be described as “best imaginable” and scores in the top 4% in a percentile range. The ratings for AR-G can be interpreted as “OK” and lie well below average in the lower 34%. The condition D-M is rated “excellent”, with a score in the top 10% percentile range. A Friedman test indicates a significant difference in ratings (mean ranks AR-H 2.44, AR-G 1.31, D-M 2.25, $\chi^2(2) = 14.15$, $p = 0.000$). However, a Wilcoxon Signed-Ranks test indicates that AR-H (mean rank 10.06) and D-M (mean rank 7.81, $Z = -0.67$, $p = 0.523$) are not rated significantly different, while AR-H (mean rank 9.47) compared to AR-G (mean rank 1.50, $Z = -3.55$, $p = 0.000$) as well as D-M (mean rank 9.54) compared to AR-G (mean rank 4.00, $Z = -2.90$, $p = 0.002$) are rated significantly higher.

Although the SUS scores of AR-G are lower than for the other conditions, there is also a high standard deviation. We believe, this could be attributed to the gesture interaction, which is hard to use intuitively after only a short introduction if the test person is not familiar with it. Similarly, the high scores of D-M could be related to the fact, that users are well accustomed to the interaction with a computer mouse.

The four RTLX questions frustration level, mental demand, performance and temporal demand had to be answered on a scale from 1 (low/poor) to 18 (high/good) in order to measure the perceived workload under the different conditions. We analyzed the results as a composite (RTLX), but also the individual component subscales to get an indication of the nature of the differences between the conditions. The results for RTLX are shown in Figure 11, all details are presented in Table 2.

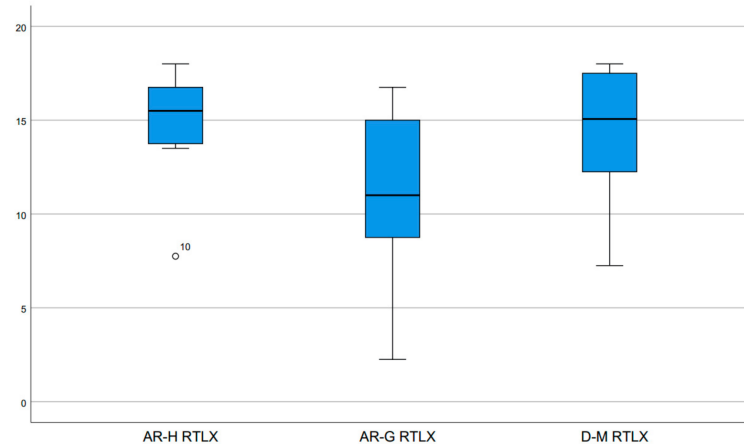


Figure 11. Mean Raw Task Load ratings (RTLX) for the tested conditions.

Table 2. Detailed results of RTLX and the individual TLX scales.

RTLX (Lower is better)				
mean	AR-H 3.99 (SD 2.34)	AR-G 8.08 (SD 4.29)	D-M 4.22 (SD 3.01)	
Friedman mean ranks	AR-H 1.67	AR-G 2.78	D-M 1.56	$\chi^2(2) = 16.44, p = 0.000$
Wilcoxon mean ranks	AR-H 8.00	*	D-M 11.86	$Z = -0.11, p = 0.924$
	AR-H 10.00	<<	AR-G 9.47	$Z = -3.29, p = 0.000$
	D-M 4.00	<<	AR-G 10.60	$Z = -3.20, p = 0.001$
Frustration Level (Lower is better)				
mean	AR-H 2.72 (SD 2.11)	AR-G 9.03 (SD 5.30)	D-M 4.03 (SD 3.57)	
Friedman mean ranks	AR-H 1.53	AR-G 2.67	D-M 1.81	$\chi^2(2) = 13.85, p = 0.001$
Wilcoxon mean ranks	AR-H 6.50	*	D-M 8.06	$Z = -1.26, p = 0.224$
	AR-H 1.75	<<	AR-G 9.97	$Z = -3.46, p = 0.000$
	D-M 4.67	<<	AR-G 9.93	$Z = -2.96, p = 0.002$
Mental Demand (Lower is better)				
mean	AR-H 5.56 (SD 4.21)	AR-G 8.81 (SD 4.52)	D-M 5.17 (SD 3.87)	
Friedman mean ranks	AR-H 1.83	AR-G 2.64	D-M 1.53	$\chi^2(2) = 13.56, p = 0.001$
Wilcoxon mean ranks	AR-H 6.95	*	D-M 10.10	$Z = -0.54, p = 0.611$
	AR-H 6.00	<<	AR-G 8.31	$Z = -2.73, p = 0.004$
	D-M 6.00	<<	AR-G 8.86	$Z = -2.90, p = 0.002$
Performance (Higher is better)				
mean	AR-H 15.58 (SD 3.09)	AR-G 12.36 (SD 4.61)	D-M 15.67 (SD 2.85)	
Friedman mean ranks	AR-H 2.36	AR-G 1.42	D-M 2.22	$\chi^2(2) = 11.42, p = 0.002$
Wilcoxon mean ranks	AR-H 6.00	*	D-M 7.50	$Z = -0.71, p = 0.503$
	AR-H 8.00	*	AR-G 8.00	$Z = -2.05, p = 0.040$
	D-M 8.68	>>	AR-G 7.25	$Z = -2.78, p = 0.003$
Temporal Demand (Higher is better)				
mean	AR-H 14.75 (SD 3.45)	AR-G 11.14 (SD 5.20)	D-M 14.67 (SD 4.24)	
Friedman mean ranks	AR-H 2.19	AR-G 1.58	D-M 2.22	$\chi^2(2) = 5.73, p = 0.060^*$
Wilcoxon mean ranks	AR-H 7.86	*	D-M 7.14	$Z = -0.16, p = 0.888$
	AR-H 9.23	*	AR-G 4.63	$Z = -2.36, p = 0.016$
	D-M 8.09	*	AR-G 5.33	$Z = -2.30, p = 0.018$

* not significant

Friedman tests indicated significant differences for all questions, except for temporal demand. Wilcoxon Signed-Ranks tests revealed, that AR-G was perceived significantly worse for all scales than both AR-H and D-M, only the difference between AR-G and AR-H for performance did not meet the Bonferroni corrected p -value required for significance of $p = 0.017$. The high mental demand and frustration level for AR-G could indicate that a lot of mental capacity was occupied with interaction and therefore was not available for the actual learning. However, the standard deviation for frustration level is very high, which would suggest that the range of how well users were able to handle the interaction in this condition was more diverse than with the other conditions. The AR-H condition scored en-par with D-M for all questions, the differences between the two conditions were not significant. The low perceived mental demand for D-M seems to confirm how well established this interaction technique is among users.

For measuring the experienced Presence, the feeling of actually being in the virtual environment, we adapted two items from the Slater-Usch-Steed questionnaire [40]: The

users had to rate their sense of the ship “being there” between 1 (Not at all) and 7 (Very much) and how they perceived the ship between 1 (Images that I saw) and 7 (Something that I interacted with). The mean reported presence ratings presented in Figure 12 are for the different conditions AR-H 6.00 (SD 1.15), AR-G 5.64 (SD 1.19) and D-M 3.11 (SD 1.68). A Friedman test indicated significant differences in ratings (mean ranks AR-H 2.56, AR-G 2.28, D-M 1.17, $\chi^2(2) = 20.90$, $p = 0.000$). A Wilcoxon Signed-Ranks test confirms that in both AR conditions, AR-H (mean rank 9.50) compared to D-M (mean rank 1.00) ($Z = -3.58$, $p = 0.000$) as well as AR-G (mean rank 9.50, $Z = -3.58$, $p = 0.000$) compared to D-M (mean rank 1.00), users felt significantly more present in the environment than during the interaction with a mouse pointer on a display. However, in this experiment the direct haptic interaction in AR-H condition (mean rank 7.90) did not reflect in significantly higher presence ratings compared to AR-G (mean rank 8.20, $Z = -1.09$, $p = 0.296$).

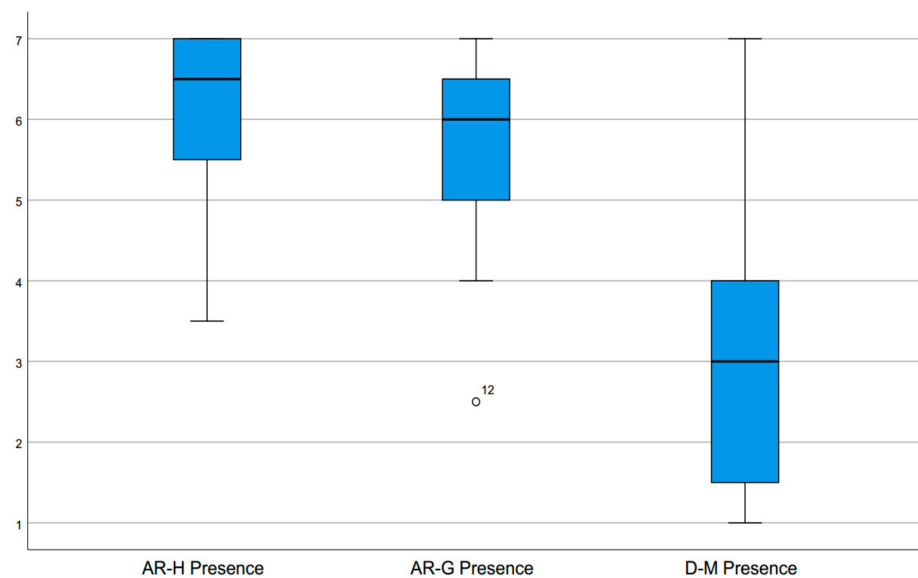


Figure 12. Mean Presence ratings for tested conditions.

The four items to measure Absorption by activity as proposed in the Flow Short Scale had to be answered on a scale from 1 (not at all) to 7 (very much). The mean ratings (Figure 13) of AR-H 5.43 (SD 0.92), AR-G 5.31 (SD 1.08) and D-M 4.22 (SD 1.36) were significantly different according to the performed Friedman test (mean ranks AR-H 2.39, AR-G 2.39, D-M 1.22, $\chi^2(2) = 19.60$, $p = 0.000$). A Wilcoxon Signed-Ranks test shows that both AR conditions, AR-H (mean rank 9.00) compared to D-M (mean rank 1.00) ($Z = -3.47$, $p = 0.000$) and AR-G (mean rank 8.97) compared to D-M (mean rank 1.50, $Z = -3.49$, $p = 0.000$), showed significantly higher absorption ratings. If learners are more absorbed in what they are doing in the AR conditions, this would also indicate, that they would potentially be willing to invest more time and effort into studying compared to a desktop scenario. The difference between AR-H (mean rank 7.67) and AR-G (mean rank 5.33, $Z = -0.55$, $p = 0.607$) was not significant, which suggests that haptic interaction is not the significant factor for absorption.

In order to analyze the achievable learning success with the different conditions, we compared how many tries it took the participants to find the correct solution for the individual tasks. The differences between the conditions for task 1 AR-H (1.67, SD 0.59), AR-G (1.89, SD 0.32) and D-M (1.89, SD 0.32) were found not to be significant with a Friedman test (mean ranks AR-H 1.83, AR-G 2.08, D-M 2.08, $\chi^2(2) = 2.00$, $p = 0.531$). The same applies for task 2 AR-H (1.78, SD 0.43), AR-G (1.67, SD 0.49) and D-M (1.78, SD 0.43, mean ranks AR-H 2.06, AR-G 1.89, D-M 2.06, $\chi^2(2) = 0.67$, $p = 0.817$). The results from the concluding quiz with 9 questions about the learned concepts AR-H (2.11, SD 0.96), AR-G (2.06, SD 1.06) and D-M (2.22, SD 1.00) also did not show a significant difference

in a Friedman test (mean ranks AR-H 2.00, AR-G 1.92, D-M 2.08, $\chi^2(2) = 0.35$, $p = 0.855$). We conclude, that for significant results the experiment would have to be conducted with more difficult tasks and a larger number of participants.

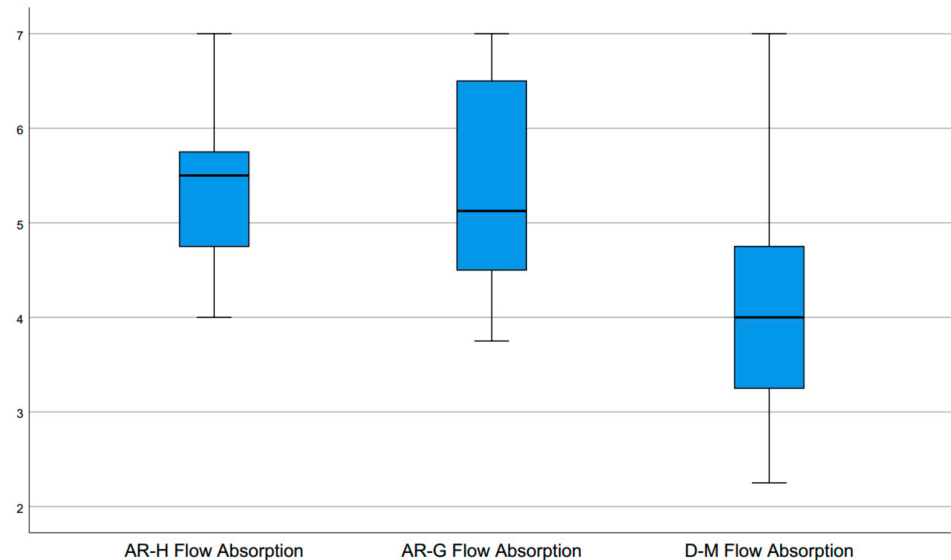


Figure 13. Mean Absorption by Activity ratings of tested conditions.

In two open questions, participants were asked to report the most positive and the most negative incidents experienced during the experiments. D-M (13× positive, 1× negative) was positively mentioned as easy to use and familiar due to common work with desktop PCs. AR-H (9× positive, 1× negative) was positively described as especially intuitive and easy to use and the good overview when looking around the scene was pointed out. AR-G (3× positive, 11× negative) was reported as difficult to use and having problems with gesture interaction and it was stated that one had to concentrate on interaction so much, that focusing on the task was impeded. The level of immersion was stated positively for AR-H (4×) and AR-G (3×), and negatively for D-M (3×). For AR-H tracking jitter when occluding the robots was mentioned negatively (8×). For AR-H (4×) as well as AR-G (3×) the FOV was negatively characterized as too small. The examiners also repeatedly observed users stepping back from the table, out of arms reach of the robots, to get an overview of the scene.

Stating their opinion on all interaction modalities in direct comparison, participants described AR-H (13×) and AR-G (3×) as the most fun to use. D-M was not mentioned in this context (2× no tendency). AR-H (9×) was described as the most useful for learning, followed by D-M (4×) and AR-G (2×) (3× no tendency). Ranking the interaction modalities, AR-H was preferred over D-M (9×), D-M over AR-H (5×) and AR-H equal to D-M (4×). AR-G was ranked last (12×) or not mentioned (6×).

5. Conclusions and Future Work

We have presented the concept and implementation of the educational AR tabletop application StARboard, as well as a tracking solution for the ACTO tangibles TrACTOR. Our approach, combining HoloLens positional head tracking, optical depth and rectangle tracking with IR ID transmission for the tangibles and fiducial marker tracking to register the environment, proved appropriate to track multiple ACTOs in our educational scenario and allowed us further exploration of the potential of TAR in this context.

In the conducted user study with 18 participants, we could compare our prototype with actuated tangible feedback in AR to a common AR setup using HoloLens gestures and a traditional desktop PC with mouse input in an educational context. Our results show that both AR conditions outperform the desktop scenario in factors positively related

to learning, like absorption by activity and presence. The low scores and high standard deviation apparent in SUS and frustration level for the common AR setup, as well as corresponding statements in the open questions, seem to be related to the HoloLens gestures. These are hard to use intuitively after only a short introduction for an untrained test person but could have room for improvement with more learning time and a better FOV of future hardware.

Direct interaction and intuitive handling provided by our micro robot-based TAR allowed scores to rival traditional mouse interaction in its strong qualities performance and usability. Our prototype was also rated the most fun and the most suitable for learning by the participants, which contributes more positive effects to the learning experience, confirming the potential of the presented approach to provide a great learning environment.

In the future we expect hardware to evolve further, e.g., providing AR displays with a larger FOV. The occasional jitter of occluded tangibles in the TAR setup should be resolvable with additional filtering during occlusion handling in future improvements of our setup. We would like to conduct an experiment with a larger group of participants and more difficult tasks with a future version of our prototype, because we think that we can show the potential to outperform the common mouse and desktop setup significantly in performance and learning gain. We are also interested in extending our environment to approach other educational topics as well as explore the benefits of our proposed interaction technology outside an educational context, e.g., in a traditional desktop PC work setting, to challenge the long-term contender mouse and desktop in its most solidified application fields.

Looking even further we also believe a combination of our approach with a shape changing surface and actively changing configuration of the micro robots would warrant continuing investigations. Moreover, we see a lot of potential in collaborative settings with several users and multiple different AR displays, supporting our vision of an actuated TAR based on micro robots as highly versatile and engaging interactive platform.

Author Contributions: Conceptualization, E.V. and C.S.; Methodology, E.V. and C.S.; Software, E.V. and C.S.; Validation, E.V. and C.S.; Formal Analysis, E.V. and C.S.; Investigation, E.V. and C.S.; Resources, E.V., C.S. and H.K.; Data Curation, E.V. and C.S.; Writing—Original Draft Preparation, E.V. and C.S.; Writing—Review & Editing, E.V. and H.K.; Visualization, E.V. and C.S.; Supervision, E.V. and H.K.; Project Administration, E.V. and H.K.; Funding Acquisition, E.V. and H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and the TU Wien Informatics Code of Ethics (<https://informatics.tuwien.ac.at/code-of-ethics/> (accessed on 30 December 2020)).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy reasons.

Acknowledgments: We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. We also want to thank the students of HTBLVA Spengergasse who sparked the basic idea for a sailing application during the work for the original ACTO project 2014 with INNOC—Österreichische Gesellschaft für innovative Computerwissenschaften as well as the volunteers who participated in our experiment.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fitzmaurice, G.W.; Ishii, H.; Buxton, W.A.S. Bricks: Laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI '95*; ACM Press: New York, New York, USA, 1995; pp. 442–449.
2. Ishii, H. Tangible bits: Beyond pixels. In *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction—TEI '08*, Bonn, Germany, 18–20 February 2008; pp. xv–xxv.

3. Poupyrev, I.; Nashida, T.; Okabe, M. Actuation and Tangible User Interfaces: The Vaucanson Duck, Robots, and Shape Displays. In Proceedings of the 1st International Conference on Tangible and Embedded Interaction—TEI '07, Baton Rouge, LA, USA, 15–17 February 2007; ACM Press: New York, NY, USA, 2007; pp. 205–212.
4. Schneider, B.; Jermann, P.; Zufferey, G.; Dillenbourg, P. Benefits of a Tangible Interface for Collaborative Learning and Interaction. *IEEE Trans. Learn. Technol.* **2011**, *4*, 222–232. [[CrossRef](#)]
5. Shaer, O.; Hornecker, E. *Tangible User Interfaces: Past, Present, and Future Directions*; Now Publishers Inc.: Norwell, MA, USA, 2010; Volume 3, pp. 1–137. [[CrossRef](#)]
6. Billinghurst, M.; Kato, H.; Poupyrev, I. Tangible augmented reality. *ACM Siggraph Asia* **2008**, *7*, 1–10.
7. Bach, B.; Sicat, R.; Beyer, J.; Cordeil, M.; Pfister, H. The Hologram in My Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality? *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 457–467. [[CrossRef](#)] [[PubMed](#)]
8. Besançon, L.; Issartel, P.; Ammi, M.; Isenberg, T. Mouse, Tactile, and Tangible Input for 3D Manipulation. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems—CHI' 17, Denver, CO, USA, 6–11 May 2017.
9. Zhou, F.; Dun, H.B.L.; Billinghurst, M. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In Proceedings of the 7th IEEE International Symposium on Mixed and Augmented Reality 2008—ISMAR 2008, Cambridge, UK, 15–18 September 2008.
10. Sharlin, E.; Watson, B.; Kitamura, Y.; Kishino, F.; Itoh, Y. On tangible user interfaces, humans and spatiality. *Pers. Ubiquitous Comput.* **2004**. [[CrossRef](#)]
11. Steptoe, W.; Julier, S.; Steed, A. Presence and discernability in conventional and non-photorealistic immersive augmented reality. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality—Science and Technology 2014—ISMAR 2014, Munich, Germany, 10–12 September 2014.
12. Vonach, E.; Gerstweiler, G.; Kaufmann, H. ACTO: A Modular Actuated Tangible User Interface Object. In Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces—ITS '14, Dresden, Germany, 16–19 November 2014; ACM Press: New York, NY, USA, 2014; pp. 259–268.
13. Ishii, H.; Ullmer, B. Tangible bits: Towards seamless interfaces between people, bits and atoms. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI' 97, Atlanta, GA, USA, 22–27 March 1997; ACM Press: New York, NY, USA, 1997; pp. 234–241.
14. Dünser, A.; Looser, J.; Grasset, R.; Seichter, H.; Billinghurst, M. Evaluation of Tangible User Interfaces for Desktop AR. In Proceedings of the 2010 International Symposium on Ubiquitous Virtual Reality—ISUVR 2010, Gwangju, Korea, 7–10 July 2010; pp. 36–39.
15. Verhaegh, J.; Fontijn, W.; Jacobs, A. On the benefits of tangible interfaces for educational games. In Proceedings of the 2nd IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning—DIGITEL '08, Banff, BC, Canada, 17–19 November 2008; pp. 141–145.
16. Marshall, P.; Price, S.; Rogers, Y. Conceptualising tangibles to support learning. In Proceeding of the 2003 Conference on Interaction Design and Children—IDC '03, Preston, UK, 1–3 July 2003.
17. Goldin-Meadow, S.; Wagner, S.M. How our hands help us learn. *Trends Cogn. Sci.* **2005**, *9*, 234–241. [[CrossRef](#)] [[PubMed](#)]
18. Radu, I.; Antle, A. Embodied learning mechanics and their relationship to usability of handheld augmented reality. In Proceedings of the 2017 IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual and Augmented Reality, KELVAR 2017, Los Angeles, CA, USA, 19 March 2017.
19. Akcayir, M.; Akcayir, G. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educ. Res. Rev.* **2016**, *20*, 1–11. [[CrossRef](#)]
20. Dass, N.; Kim, J.; Ford, S.; Agarwal, S.; Chau, D.H. (Polo) Augmenting Coding. In Proceedings of the Sixth International Symposium of Chinese CHI—Chinese CHI '18, Montreal, QC, Canada, 21–26 April 2018; ACM Press: New York, NY, USA, 2018; pp. 156–159.
21. Shapira, L.; Amores, J.; Benavides, X. TactileVR: Integrating Physical Toys into Learn and Play Virtual Reality Experiences. In Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality—ISMAR 2016, Merida, Mexico, 19–23 September 2016.
22. Hinckley, K.; Pausch, R.; Goble, J.C.; Kassell, N.F. Passive real-world interface props for neurosurgical visualization. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence—CHI '94, Boston, MA, USA, 24–28 April 1994; ACM Press: New York, NY, USA, 1994; Volume 30, pp. 452–458.
23. Henderson, S.J.; Feiner, S. Opportunistic controls. In Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology—VRST '08, Bordeaux, France, 27–29 October 2008.
24. Funk, M.; Korn, O.; Schmidt, A. An augmented workplace for enabling user-defined tangibles. In Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems—CHI EA '14, Toronto, ON, Canada, 26 April–1 May 2014.
25. Hettiarachchi, A.; Wigdor, D. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In Proceeding of the SIGCHI Conference on Human Factors in Computing Systems—CHI '16, San Jose, CA, USA, 7–12 May 2016.

26. Kwon, E.; Kim, G.J.; Lee, S. Effects of sizes and shapes of props in tangible augmented reality. In Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality, Orlando, FL, USA, 19–22 October 2009.
27. Simeone, A.L.; Velloso, E.; Gellersen, H. Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences. In Proceedings of the ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 10–15 April 2015.
28. Hirota, K.; Hirose, M. Simulation and presentation of curved surface in virtual reality environment through surface display. Proceedings of IEEE Virtual Reality Annual International Symposium '95, Research Triangle Park, NC, USA, 11–15 March 1995; pp. 211–216.
29. Iwata, H.; Yano, H.; Nakaizumi, F.; Kawamura, R. Project FEELEX. In *Proceedings of the 28th annual conference on Computer graphics and Interactive Techniques—SIGGRAPH '01*; ACM Press: New York, NY, USA, 2001; pp. 469–476.
30. Yamaguchi, K.; Kato, G.; Kuroda, Y.; Kiyokawa, K.; Takemura, H. A Non-grounded and Encountered-type Haptic Display Using a Drone. In *Proceedings of the 2016 Symposium on Spatial User Interaction—SUI '16*; ACM Press: New York, NY, USA, 2016; pp. 43–46.
31. Vonach, E.; Gatterer, C.; Kaufmann, H. VRRobot: Robot actuated props in an infinite virtual environment. In Proceedings of the IEEE Virtual Reality 2017, Los Angeles, CA, USA, 18–22 March 2017; 2017; pp. 74–83.
32. Rosenfeld, D.; Zawadzki, M.; Sudol, J.; Perlin, K. Physical objects as bidirectional user interface elements. *IEEE Comput. Graph. Appl.* **2004**, *24*, 44–49. [[CrossRef](#)] [[PubMed](#)]
33. Kojima, M.; Sugimoto, M.; Nakamura, A.; Tomita, M.; Inami, M.; Nii, H. Augmented Coliseum: An Augmented Game Environment with Small Vehicles. In Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems—TABLETOP '06, Adelaide, Australia, 5–7 January 2006; Volume 2006, pp. 3–8.
34. Siu, A.F.; Gonzalez, E.J.; Yuan, S.; Ginsberg, J.B.; Follmer, S. shapeShift: 2D Spatial Manipulation and Self-Actuation of Tabletop Shape Displays for Tangible and Haptic Interaction. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems—CHI '18, Montreal, QC, Canada, 21–26 April 2018.
35. Zhao, Y.; Kim, L.H.; Wang, Y.; Le Goc, M.; Follmer, S. Robotic assembly of haptic proxy objects for tangible interaction and virtual reality. In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces—ISS 2017, Atlanta, GA, USA, 10–15 April 2017.
36. Hwang, A.D.; Deng, H.; Gao, Z.; Peli, E. Quantifying visually induced motion sickness (VIMS) during stereoscopic 3D viewing using temporal VIMS rating. *Electron. Imaging* **2018**, *2018*, 1–9. [[CrossRef](#)]
37. Brooke, J. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* **1996**, 189–194.
38. Hart, S.G. NASA-task load index (NASA-TLX) 20 years later. In *Proceedings of the Human Factors and Ergonomics Society*; Sage: Los Angeles, CA, USA, 2006.
39. Rheinberg, F.; Vollmeyer, R.; Engeser, S. Die Erfassung des Flow-Erlebens [The Assessment of Flow Experience]. In *Diagnostik von Selbstkonzept, Lernmotivation und Selbstregulation [Diagnosis of Motivation and Self-Concept]*; Hogrefe-Verlag: Göttingen, Germany, 2003; ISBN 3-8017-1674-0.
40. Usoh, M.; Catena, E.; Arman, S.; Slater, M. Using Presence Questionnaires in Reality. *Presence Teleoperators Virtual Environ.* **2000**, *9*, 497–503. [[CrossRef](#)]
41. Flanagan, J.C. The critical incident technique. *Psychol. Bull.* **1954**, *51*, 327–358. [[CrossRef](#)] [[PubMed](#)]
42. Sauro, J. 5 Ways to Interpret a SUS Score. Available online: <https://measuringu.com/interpret-sus-score/> (accessed on 26 May 2020).