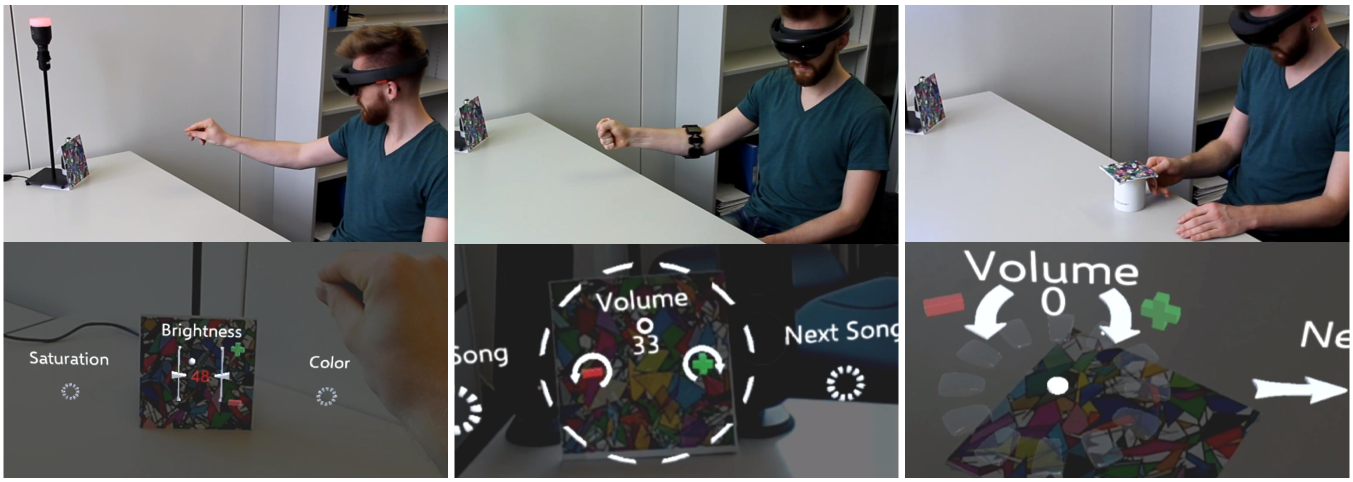
Investigating Universal Appliance Control through Wearable Augmented Reality

Vincent Becker
vincent.becker@inf.ethz.ch
Department of Computer Science,
ETH Zurich, Switzerland

Felix Rauchenstein
rafelix@ethz.ch
Department of Computer Science,
ETH Zurich, Switzerland

Gábor Sörös
gabor.soros@inf.ethz.ch
Department of Computer Science,
ETH Zurich, Switzerland



(1) (2) (3)

Figure 1: We present a concept and a prototype system for universal control of networked smart appliances. The appliances externalize their user interfaces to networked wearable computers. More specifically, a head-mounted display renders automatically generated digital user interface primitives overlaying the physical objects. We compare three ways of manipulating the primitives: (1) a pinching gesture, (2) other natural gestures like arm rotations, and (3) utilizing tangible physical objects at hand. The images in the top row show the outside view while the images in the bottom row show the AR view at the same moment in time during the different modes of interaction.

ABSTRACT

The number of interconnected devices around us is constantly growing. However, it may become challenging to control all these devices when control interfaces are distributed over mechanical elements, apps, and configuration webpages. We investigate interaction methods for smart devices in augmented reality. The physical objects are augmented with interaction widgets, which are generated on demand and represent the connected devices along with their adjustable parameters. For example, a loudspeaker can be overlaid with a controller widget for its volume. We explore three ways of manipulating the virtual widgets: (a) in-air finger pinching and sliding, (b) whole arm gestures rotating and waving, (c) incorporating physical objects in the surrounding and mapping their movements to the interaction primitives. We compare these methods in a user study with 25 participants and find significant differences in the preference of the users, the speed of executing commands, and the granularity of the type of control.

AH2019, March 11–12, 2019, Reims, France

ACM ISBN 978-1-4503-6547-5/19/03…$15.00
https://doi.org/10.1145/3311823.3311853

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; **Graphical user interfaces**; **Gestural input**; **Mobile devices**.

KEYWORDS

Tangible User Interfaces; Augmented Reality; Ubiquitous Computing; Smart Object; Wearable Computing

ACM Reference Format:
Vincent Becker, Felix Rauchenstein, and Gábor Sörös. 2019. Investigating Universal Appliance Control through Wearable Augmented Reality. In *Augmented Human International Conference 2019 (AH2019), March 11–12, 2019, Reims, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3311823.3311853

1 INTRODUCTION

New ways of interaction between humans and appliances have gained significant interest with the emergence of networked smart appliances. Since the times appliances could be operated only by hard-wired knobs and buttons, we have witnessed the inventions of a number of convenience features. Remote controls were introduced in the 1950s, allowing to comfortably control a single appliance from a distance, and since then we reached the point that home systems allow almost universal device control via smartphone apps or even via speech and gesture recognition systems. This process can be described by the term *user interface externalization*. The physical user interface of an appliance is *externalized* from the appliance to the remote control, to a companion smartphone app, or in an abstract sense to the speech or gesture command language.

As appliances in our homes are becoming smarter with improved sensing, actuation, and computing capabilities as well as connectivity to other devices and cloud services [18], there are new opportunities for more advanced interaction and also for externalizing the appliances' user interface to other devices. Besides the ability to control things remotely, further advantages of such "outsourced" user interfaces are smaller cost, overcoming the lack of display space, and simpler UI feature updates via software.

We present the concept of externalizing user interfaces of appliances to personal wearable computers. More generally, we show a concept how a network of smart things worn on our body allows universal interaction with a network of smart things in our environment. We employ visual representations of device characteristics and actions in augmented reality (AR) and allow the user to interact with the devices through this AR interface. For example, a digital color lamp may be represented by its brightness and color values, which can be manipulated via one-dimensional sliders. As we make use of a head-mounted AR display, all representations appear in the user's ego-centric view, i.e. allow the direct interaction with an appliance, without an intermediary remote control device. Furthermore, we can create one unified interface for all devices and, in particular, the devices themselves are not required to have a display or exhibit any kind of direct input and output.

This concept holds great potential for the future, especially as AR headsets will become smaller, and we believe it is necessary to study the user's behavior in such an environment as well as the different possible interaction modalities. For this reason, we built a working prototype of the described concept with a Microsoft HoloLens for displaying the visual AR representation, and various techniques for input. Devices are recognized via visual markers, and the user can select a specific smart thing with his/her gaze, i.e. by looking at them. Then, the supported interaction primitives (knobs, sliders, etc.) overlay the devices in the AR view. In the spatially registered AR view, the visual content describes the appliance's capabilities and the interaction primitives are represented depending on how they can be manipulated.

For actuating the smart objects, i.e. providing input to modify the primitives and thereby the device parameters, we design three different possibilities which we enable by employing different wearable computers, e.g. a Thalmic Lab's Myo or the HoloLens itself: first, using a simple pinching gesture to adjust virtual sliders and click on buttons (Figure 11), second, using a larger gesture set recognized by an electromyography armband (Figure 12), and third, by displaying the representations on a physical object, which allows to control an appliance by moving or rotating the physical object (Figure 13). This can be seen as an extreme form of externalization, which even allows to embed "dumb", unconnected objects in the space of smart, connected appliances. We compare the three possibilities as interaction scenarios in a study with 25 participants within a small test environment as shown in Figure 2.

Figure 2: The setup of connected smart devices used for the study consisting of a lamp, a sound system, and a phone. The mug used in scenario 3 is also shown. In our prototype, all have an individual marker attached for visual recognition.

Beyond the implementation of the concept, the main contribution of this paper is the thorough evaluation of users' behavior while controlling devices in an AR environment which may support future interaction design for AR control interfaces or device control interfaces in general. Furthermore, it goes beyond previous research by also providing the comparison of different ways of interaction in the AR space.

2 RELATED WORK

One way of externalization deployed nowadays is the use of handheld devices such as smartphones or tablets to display information and controls of other devices. This is also a common solution for many commercial products. Interesting in the scope of our work are approaches which also facilitate the selection of a device. Several researchers explored this by combining device recognition and device control. Mostly, devices are recognized visually from camera images [6, 20], but also other modalities such as visual codes [14] or electro-magnetic signatures are used [26]. Some works additionally take the user's location, i.e. the proximity to the devices, into account for the selection process [6, 16]. Subsequently, user interfaces on the handhelds provide visual control elements to actuate the devices. While these works approach the same problem as we do, they require the user to interact with the handheld, i.e. a proxy device in between. In contrast, we provide a direct interaction interface in the AR space observed from the user's ego-centric perspective.

Improving on the works mentioned above, several researchers employ visual AR on handhelds for device control [9, 13]. This allows the user to interact with the visual counterpart of the real device in the camera image after the device has been recognized and not only with a simple UI representation consisting of buttons and sliders. Huo et al. [11] enhanced this method in *Scenariot* by locating the devices through ultra-wide-band RF units so the application always knows the position of the devices even if they are outside of the camera's view. Nevertheless, the user still requires a handheld and merely interacts with a visual representation of the device. In our approach, the user can directly interact in the mixed reality space "on the device" itself without having to look back and forth from the remote control to the device. *PICOntrol* [23] is also based on a handheld device, but uses an incorporated projector to display a graphical interface on the controllable appliances and employs visible light communication to transmit control information. Actions are selected using buttons on the handheld and by the transmission of the corresponding light patterns. Thereby, the visual UI elements augment the real device and not a virtual counterpart, nonetheless, the user still has to hold a handheld device. Research on configuring connections between smart devices includes the *RealityEditor* by Heun et al. [9] and a magic lens proposed by Mayer et al. [19], however, these also require a handheld mediator.

Finally, only few researchers have used head-mounted displays for interaction in the AR space. Sorgalla et al. [24] recently proposed a system called *ARGI*, which uses the HoloLens to augment the user's view with interaction widgets placed on controllable devices. While the basic concept is similar to ours, the interaction possibilities do not go beyond the standard pinching gestures provided by the HoloLens. Going further, we create three different ways of interacting with the devices and conduct a user study to investigate the use of the AR system and the interaction methods. Another system, *Ubii* [17] uses a Google Glass to represent digital devices such as computers or smartphones and the files they contain in the glass' view and supports two main operations through simple pinching and dragging gestures: the transfer of files between two devices or to a device such as a printer and the pairing of devices. *Ubii* is different from our system by only targeting digital devices and supporting only two specific operations for these and not allowing device control in a general sense. *AmbiGaze* [25] employs a wearable eye tracker to allow device interaction using a target selection method called pursuits, which correlates eye movements to the movements of targets displayed on the device. By following the target, the user can select certain actions. However, this only works if the device has a display to show targets or exhibits mechanical movement itself. All other devices require a stationary projector displaying the targets, i.e. in comparison to our concept it requires an additional augmentation of the devices.

Further interesting in the scope of this paper are works which externalize interfaces to physical elements, which were not deliberately designed to act as a controller. One possibility to do this is to reuse existing controls, e.g. use the switch of a light to play or pause a sound system [4, 10]. On the other hand, one could use simple physical objects available in the environment, as we intend to do. This was already proposed by Ishii et al. [12] more than two decades ago. Recently, Pohl et al. [22] explored the space of everyday objects potentially useful for control and found that there is both a diverse set of objects available and potential use for them. Henderson et al. [8] proposed so-called *Opportunistic Controls*, a system to support the use of tangible elements, e.g. bolts of a machine, for input when using a head-mounted AR display. They found that the opportunistic controls support faster completion times than a baseline technique, which uses simple virtual buttons in AR. Several previous approaches allowing to use simple objects as control elements instrument a space which is observed by cameras from above or below [1, 3, 5, 7]. The cameras can recognize the movement of the objects within the scene. Some also use projectors to display information on the objects. The concept of these works is very similar to one of our approaches. We, however, build a system which only requires a HoloLens instead of an instrumented workspace, hence, our approach is mobile and potentially unbounded.

3 INTERACTION SCENARIOS

Our general interaction paradigm is to visually augment an appliance with a representation of its parameters and the way to apply modifications to them (see the bottom row of Figure 1). We recognize the devices with an egocentric camera and by their individual visual markers. The user sees the point of his/her gaze in AR, which is approximated by the head pose, and can select an appliance by looking at it, more precisely at its marker. This procedure is the same for all three scenarios. However, each scenario adopts a different type of user input to change the parameters of a connected appliance and thus also incorporates a different state visualization of the appliance. Our intention was to find out how users would interact in the AR space, hence we only choose scenarios within this space. Since visual AR is able to show interaction elements, we allow a direct interaction with these elements. We want to compare this to the use of natural gestures. Finally, we also include a tangible object. We do not include a control device like a connected knob, as this would defy the idea of a wearable system that can always be with us. We furthermore do not include speech commands, as we use visual elements to represent the appliances, which provide the ability to not only show state information but also allow their manipulation by moving and clicking them. In terms of this representation, it appears more reasonable to manipulate the elements and their underlying states using gestures than speech. In the following, we describe the three scenarios in more detail. A video showing all scenarios is included in the supplementary material.

Scenario 1. In the first scenario, we aim at a simple mode of control, composed of simple gestures and simple widgets, which allows for a direct interaction with the devices, i.e. the user will also look at the devices while interacting because the widgets are placed on the devices themselves. We use the predefined HoloLens gesture, pinching thumb and forefinger as depicted in Figure 3, as the only option for user input. The gesture is recognized by the HoloLens based on visual recognition. It can be used to click on an AR element by simply tapping as well as dragging an element by holding the closed form of the gesture as shown in Figure 11. This enables us to use virtual sliders and buttons as AR representations augmenting the real appliance, which the user can directly interact with. For example, we show a volume slider above a sound system. The user can select the volume level by looking at it and then employ the pinch gesture to hold and drag the slider and alter the volume level.

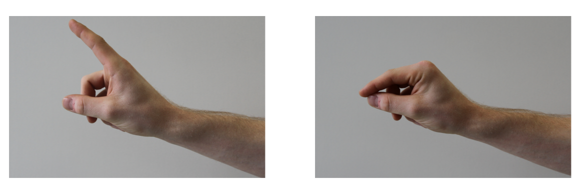


Figure 3: Left: The beginning of the pinching gesture. Right: The gesture while holding a control element, e.g. a slider.

Scenario 2. Our second scenario is also based on the direct manipulation of the AR controls by employing gestures. However, we allow for potentially more freedom in the way the gestural interaction takes place by utilizing a broader and more natural gesture set. We chose an electromyography (EMG) armband, the Thalmic Labs Myo (cf. Figure 4), for this purpose. The advantage of EMG over other forms of gesture tracking is that the arm and hand can be in any pose and does not have to be in the field of view of any camera. We chose the Myo because it recognizes gestures user-independently, i.e. without having to train the recognition per user. The Myo can recognize five predefined, user-independent gestures: Waving one's hand to the inside (*wave in*) or to the outside (*wave out*), making a fist (*fist*), spreading one's fingers (*spread*), and double tapping thumb and middle finger (*double tap*). Furthermore, it contains motion sensors to measure acceleration and rotation around all axes. We use the *fist* gesture to activate the parameter control. The user can then set values by rotating his/her arm. We also use the waving gestures for discrete input. Our representation informs the user, which gesture he/she may use to perform which action. For example, Figure 5 shows the AR view instructing the user that he/she can either wave to the right to play the next song or make a fist to set the volume. Upon performing the fist gesture, a continuous value setting appears as shown in Figure 12, telling the user that a clockwise rotation of the arm increases the volume level and a counter-clockwise rotation decreases it.

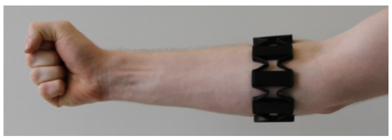


Figure 4: A user wearing the Myo and currently holding a control by making a fist and adjusting it by rotating the arm.

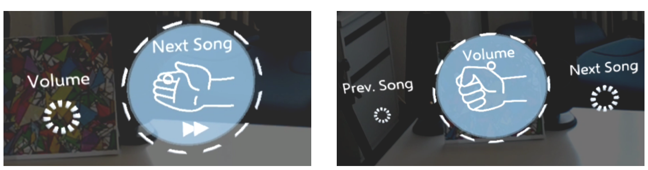


Figure 5: An AR representation augmenting a sound system, informing the user about possible gestures and corresponding actions.

Scenario 3. In the third scenario, we explore the idea of using objects in the user's environment to control devices. These potentially simple, unconnected objects can be moved or rotated by the user. The movement is registered by our system and thereby allows input by simply manipulating the object, now providing a tangible user interface. In comparison to the two other scenarios which allow the direct control of the devices, the object becomes a remote or proxy control which additionally allows us to evaluate the influence of this characteristic on the participants' perception of the system. For our study, we use a simple tea mug, to which we attached a visual marker to be tracked with the HoloLens' camera as depicted in Figure 6. The mug has a circular shape and naturally can be rotated to e.g. change the volume of a sound system. Additionally, it can be moved in several directions which might be used for other triggers. As we do for the Myo gestures in scenario 2, we inform the user about the possible interactions through AR visualizations, i.e. which movement of an object will cause which effect. Moreover, we show the current state of the appliance in AR overlayed on the mug. Figure 13 shows a user adjusting the volume of a loudspeaker by rotating the mug. Note that in order to track the movement of the mug, its marker has to be in the field of view of the HoloLens camera, which is a limitation as a user might rather want to look at the appliance he/she is currently controlling than at the mug. This is a general limitation, as either the object has to be tracked by an external camera, or contain motion sensors itself.

Figure 6: The tea mug with a visual marker for recognition used in the third scenario.

4 IMPLEMENTATION

The main component of our system is a Microsoft HoloLens, which tracks the three appliances and the mug used in scenario 3 by detecting visual markers attached to them. We chose the HoloLens because at the time of the study, it was the only high-quality HMD which fulfilled our requirements of having a see-through display and robust and accurate inside-out tracking to place the interaction elements in space. For the recognition and tracking, we use the Vuforia SDK[1]. All the visual AR representations are implemented in Unity. In our prototype, the primitives are fixed for the specific devices, however, they could be described in an abstract form in an XML document, downloaded from a web server on demand and parsed in order to automatically create the representation of the appliance as Mayer et al. have shown [21]. The HoloLens controls all the appliances by sending commands via a WiFi network. For the purpose of exchanging state information and commands, we implemented a simple message protocol based on JSON. To simplify the control of the appliances, we use an Android smartphone (a Nexus 5X running Android 8.1), which receives the commands from the HoloLens and forwards them to the devices. The smartphone itself also acts as sound system by being connected to speakers (the smartphone is not visible to the participants). The smart lamp, produced by LIFX[2], can be directly controlled by sending UDP packages over the WiFi network. The smartphone is also used to receive the gestures from the Myo armband via BLE in scenario 2, which are in turn forwarded to the HoloLens. As we only carry out mockup calls, the telephone merely has to be recognized, but not controlled here. A schematic diagram of all the devices and their information flows is given in Figure 7. A desirable property of the prototype is that it only utilizes the head-worn computer and smart

[1] www.vuforia.com
[2] www.lifx.com

things connected over a WiFi network, therefore it is mobile and not confined to a restricted space.

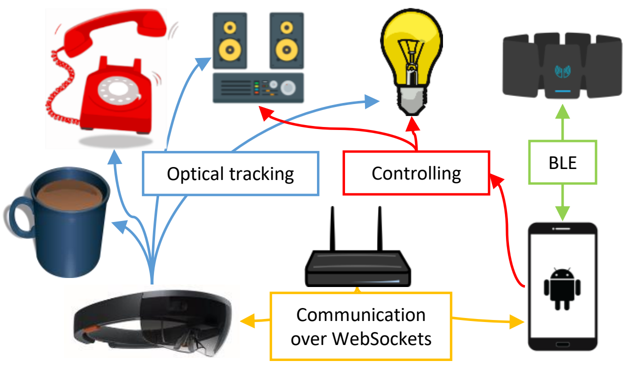


Figure 7: The HoloLens tracks the appliances and the mug (in scenario 3). It communicates to the appliances and the Android smartphone via a WiFi network. The Myo is connected to the smartphone via BLE.

5 STUDY DESIGN

We compare the three interaction methods in an appliance environment consisting of a smart lamp, a sound system, and a landline telephone in an office (cf. Figure 2). The lamp can change its color in the HSV color space, i.e. the adjustable parameters are the hue color angle, the saturation, and the brightness. It can be turned off by setting the brightness to zero. For the sound system, we preselected a set of three songs. The user can switch to the previous or the next song in the list and adjust the volume level. Similar to the lamp, the sound system can be turned off by setting the volume to zero. When selecting the telephone, the user can select a contact from a predefined list of five imaginary contacts and then start a call. Instead of actually calling someone, we provide visual feedback in AR by a telephone icon, that the call is in progress. During a call, the user has the option to hang up, which is also confirmed by visual feedback. This means the telephone is the only appliance which is actually not connected and actuated, but only allows a mock-up call, which shown in AR. Note that in our prototype implementation the setup is fixed to these appliances for this experiment; nevertheless our concepts generalize to arbitrary appliances, which may even be located in different places.

At the beginning, each participant signed a consent form, and was instructed about the purpose of the study by the investigator. Each participant carried out a fixed series of appliance interactions in each of the three scenarios. The interactions series were embedded into a story told by the investigator and included the following steps in the given order:

(1) Changing the light color to green, setting the brightness to 30% and the saturation to 80%.
(2) Turning on the sound system, setting the volume to 20, and choosing either the next or previous song.
(3) Turning off the music.
(4) Making a phone call (initiating the call).
(5) Terminating the call.
(6) Resuming the music player, switching to another track. Then the participants should change to another chair standing at a different position to the setup (this is particularly interesting in scenario 3 because they have to take the mug with them).
(7) Increasing the volume of the music. Afterwards,
(8) They sit back at the table and stop the music player.
(9) Turning the brightness of the lamp to 50%.
(10) Switching off the lamp.

Every participant carried out 10 interactions per scenario, i.e. 30 in total. The order of the interaction within each scenario was the same, the order of the scenarios was random for each participant in order to avoid a learning bias. Before each scenario, the participants were briefly instructed on the new type of interaction. After finishing with each scenario, the participants filled in a questionnaire on the scenario containing the following questions, each to be answered on a five-point Likert scale from "disagree" to "agree" with a score ranging from -2 to 2 with neutral score at 0:

Q1 The way how to control devices was easy to understand.
Q2 It was fast.
Q3 It was appealing.
Q4 It was not physically demanding.
Q5 It was not mentally demanding.
Q6 I felt comfortable.
Q7 I felt that I was in control of the devices.
Q8 I can imagine to use this method in my daily life.

Additionally, the participants were given a User Experience Questionnaire (UEQ) [15][3] score sheet in order to obtain a standardized measure of the user experience. The 26 items have a seven-point score ranging from -3 to 3. The output of the UEQ analysis are scores in the following six dimensions: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty.

After having performed all three scenarios the participants fill in another questionnaire with pairwise comparisons (also allowing for a neutral decision) on which of the scenarios they preferred with regard to the following questions (i.e. there are three pairwise comparisons per question):

P1 Which method did you generally prefer?
P2 Which method was faster to use?
P3 Which method was easier to understand?
P4 Which gave you the feeling of the most control?
P5 Which was more fine-grained?

Besides, we measured the task completion time from the moment the investigator instructed the participant to perform a task until it was fulfilled. One entire session took around 45 minutes.

Participant Population. We recruited five participants for the pilot study (average age 24.0 with a standard deviation of 1.9, ranging from 22 to 26, two females) and 25 further participants (average age 29.3 with a standard deviation of 9.7, ranging from 21 to 52, six females) for the actual study. 36% of the participants had previous experience with VR (virtual reality) and 16% had previous experience with AR applications. The participants took part on a voluntary basis without compensation.

[3]http://www.ueq-online.org/

6 RESULTS

We carry out a statistical analysis on the four measures: the questionnaire, the UEQ, the pairwise comparison questionnaire, and the task completion times. For all the four measures, the normality condition for applying an ANOVA test is violated (significant Shapiro-Wilk's test and visible from QQ-plots), hence we employ the Friedman test. If the Friedman test finds significant differences, we carry out post-hoc tests employing a Bonferroni-corrected Wilcoxon signed-rank test to reveal where the differences lie. In general, we use a significance level of 0.05, in the cases using Bonferroni-correction 0.017, respectively. We only include the data of the 25 participants from the actual study and not the additional five from the pilot study. For the sake of shorter notations, we use the abbreviations S1, S2, and S3 for scenario 1, 2, and 3, respectively.

Questionnaire Results. The questionnaire, filled in after each scenario, asked the participants to answer the questions Q1 to Q8 in Section 3 on a five point Likert scale ranging from -2 to 2. The average results for the 25 participants are displayed in Figure 8. The mean scores for all the questions are positive. For all questions except Q5, S1 obtains the highest average scores. S2 mostly achieves higher scores than S3, however the only significant differences we find are between S1 and S2 (Q2 $p < 0.005$, Q3 $p < 0.015$, Q7 $p < 0.01$), and S1 and S3 (Q2 $p < 0.005$, Q3 $p < 0.005$, Q8 $p < 0.017$). The results indicate that on average the participants have a more positive impression of S1 compared to the others. Especially, for the subjective impression of speed (Q2), the scores are more than twice as high for S1 than for S2 and S3.

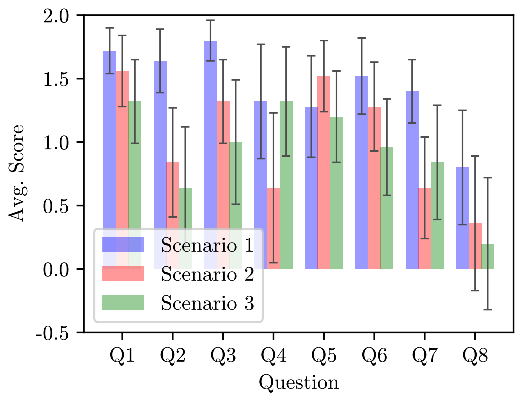


Figure 8: Average results of the Likert items in the questionnaire for all three scenarios with 95% confidence intervals.

UEQ results. The results of the UEQ are six scales, calculated as averages from the items on the UEQ sheet, i.e. we can calculate a score for each scale for each participant and scenario. The average scores over all participants for these six are shown in Figure 9. Again, all the results are positive. The results in the UEQ generally support the findings of the questionnaire mentioned above. All the average scores are positive and again, apart from the scale Novelty, S1 has the highest average scores. For the first four scales we find significant differences between S1 and S2 (Attr. $p < 0.005$, Eff. $p < 0.005$, Dep. $p < 0.005$), and S1 and S3 (Attr. $p < 0.005$, Per. $p < 0.005$, Eff. $p < 0.005$, Dep. $p < 0.001$), only for Perspicuity the difference between S1 and S2 is insignificant ($p = 0.03$). Between S2 and S3 we find no significant differences. For Stimulation and Novelty, the means are similar (as can be seen from the plot), and we cannot find any significant differences.

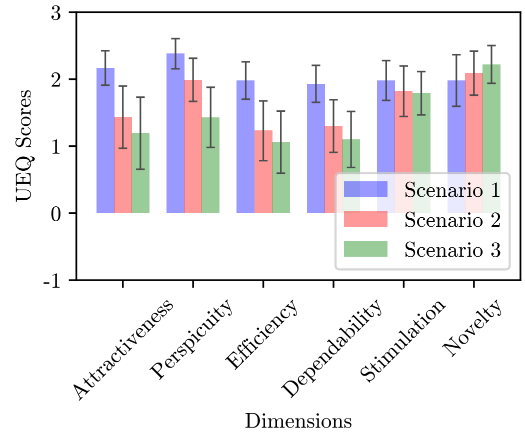


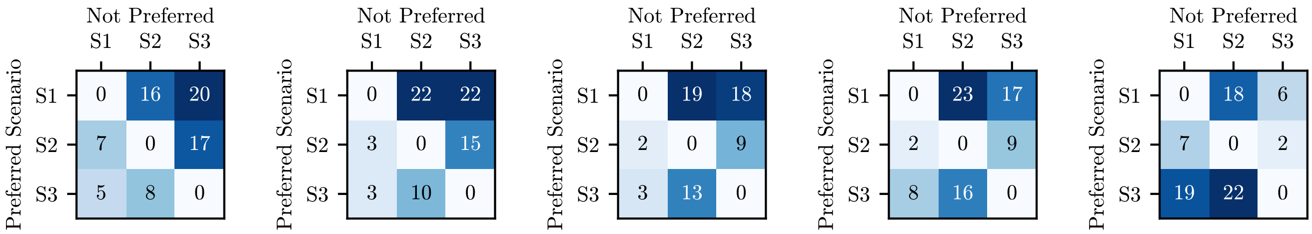
Figure 9: The average results of the UEQ, segmented in the six dimensions, for all three scenarios. The error bars show the 95% confidence intervals.

Pairwise comparisons. The results for the pairwise comparisons are shown in Figure 10 as the sums over the participants' preferences per question item. Over all question items and participants, there were no cyclic dependencies (such as S1 being preferred over S2, S2 over S3, and S3 over S1), i.e. it was possible to transform the pairwise answers into rankings for each question item. For these, we can carry out a Friedman test and post-hoc tests as above. As for the other measures, S1 is mostly preferred over S2 and S3, which is also supported by the statistical tests (S1 - S2: P2, P3, P4 significant with $p < 0.001$; S1 - S3: P1 $p < 0.005$, P2 $p < 0.001$, P3 $p < 0.005$, P4 $p < 0.017$, P5 $p < 0.01$). Only for P5, S3 was chosen more often, i.e. the majority believe that S3 offers a more fine-grained control. Between S2 and S3, there were no statistically significant differences, apart from the aforementioned P5 (S2 - S3: $p < 0.001$).

Task completion times. We calculate the mean task completion time per participant and scenario. Since task 6 and 8 included changing from one spot to the other which constitutes a rather undefined period of time difficult to compare, we omit those tasks. The averages over all participants are given in Table 1. The average overall task completion time for S2 is nearly twice as high as for the fastest, S1. For S3, it is 50% higher than for S1. The differences in the overall mean completion times between all scenarios are significant (S1 - S2: $p < 0.001$, S1 - S3: $p < 0.001$, S2 - S3: $p < 0.01$). Interestingly, the order S1 < S3 < S2 holds in each task.

Table 1: Mean completion times for each measured task over all participants for each scenario.

	T1	T2	T3	T4	T5	T7	T9	T10	all
S1	31.8	9.0	21.2	3.5	3.6	1.2	8.4	1.6	10.0
S2	51.5	17.0	42.4	4.8	5.1	5.6	20.2	6.1	19.1
S3	41.1	12.3	32.6	4.0	4.9	5.4	17.6	4.2	15.3



(a) P1 **(b) P2** **(c) P3** **(d) P4** **(e) P5**

Figure 10: The results for each pairwise comparison item summed over all participants. The plots show how often which scenario was preferred over which other scenario.

7 DISCUSSION

As the positive scores in the questionnaire (e.g. Q3) and Attractiveness in the UEQ show, our participants liked our system in general and were excited to use it (high scores for Stimulation), most were even amazed at this new way of interaction. We received many comments, such as "It was a great experience", "So cool, this is amazing" and "I am mind-blown". Besides the general appeal, all participants found the system to be innovative.

General advantages of our approach which apply to all three scenarios are that it makes appliances easily accessible, as the user only has to look at them in order to control them, and shares the benefit of remote controls by enabling a interaction from a distance. Furthermore, our interaction approach was easy to learn over all scenarios, as the scores for Q1 and Perspicuity indicate. We believe this to be the result of the possibility to show interaction instructions in a first-person view, e.g. which way the user should rotate his/her arm to increase a level. This makes the control very intuitive, as none of the interactions has to be remembered. Moreover, as the AR representation is locally collocated with the appliance (except for S3), there is no need to switch one's gaze between a remote control and the controlled device as it is the case for standard remote controls. This also makes it easy to understand which appliance is currently affected by changes. Several participants stated that "it is easy to get used to".

When comparing the three scenarios, we find a strong preference for S1, indicated by the questionnaire, the UEQ, and especially by the strong overall votes for S1 in the pairwise comparisons. Appliance control was also significantly faster on average than in the other scenarios. This objective measure aligns with the subjective impression of higher speed as shown in Q2 and Efficiency. Many participants expressed their liking of S1 because of its simplicity compared to the others, which also made it easier to understand: "Easy to understand", "Intuitive", "worked like a charm", "more intuitive and faster", "good that you only have to learn a single gesture", "I don't have to think too much", "really intuitive, although I'm not a geek". One important reason we believe for this scenario being so popular is that it provided the most direct interaction, because one is virtually moving sliders and clicking buttons on the appliance itself. One participant described the reason for his preference for S1 accordingly as "the feeling of the direct interaction". The results for Q7 and P4 also show a significantly higher score for the feeling of being in control of the devices, supporting our assumption.

The only major shortcoming of S1 was the difficulty to adjust the values in a fine-grained manner. Participants stated it was easy to set an appliance parameter to the minimum or maximum, but rather difficult to set it to an exact intermediate value. This is also exhibited by the preference of S3 over S1 in the pairwise comparison item P5 on which scenario offered the most fine-grained control.

S2, using the Myo for gesture recognition, received a higher score than S3 for many of the questionnaire items, however only the difference in the scores for P5 were significant, but then in favor of S3. What participants valued, similarly to S1, was the direct interaction with the appliances. Moreover, some preferred the use of natural gestures, such as waving to the left or right, over the clicking on buttons and dragging sliders in S1: "The coarse-grained gestures in S2 are better than the fine-grained gesture in S1". Furthermore, a participant mentioned this form of gesture input is better because one does not have to raise the hand to be in the field of the camera, but can be anywhere. Another participant even said that he thinks this would be faster than S1, because one does not have to raise the arm. One interesting aspect is that the subjective impression of speed does not match the objective task completion times. There are slightly higher scores for the question items on the speed for S2 than for S3 (but not significant), although the task completion times show that S2 is significantly slower than S3.

In comparison to the two other scenarios, there was a strong conceptual difference in S3. Instead of a direct gestural control, S3 offered a tangible user interface through the movement of a mug. Participants found this scenario very innovative and interesting, but at the same time not as suitable for appliance control: "more innovative than S1, but also more cumbersome" or "very cool, but liked much less than S1". The participants strongly prefer the form of direct control that S1 and S2 offer, as mentioned above. Some also expressed their dislike of the necessity to frequently change the viewpoint back and forth from the appliances to the mug. In our implementation, the mug is only tracked visually by the HoloLens, which requires it to be in the field of view of the HoloLens' cameras. This means the user has to be looking at the controlling objects to be able to perform actions and cannot look at the controlled appliance at the same time, which was very unnatural for many participants. Several participants said that looking down is uncomfortable when wearing the relatively heavy HoloLens. This limitation could be solved by visually tracking the object with an external camera, however this defies the goal of a mobile application.

On the other hand, one strong advantage S3 has over the other scenarios, is that it offers haptic feedback, as one participant expressed for example "the haptic feedback is great". This goes hand in hand with the fact that S3 was on average preferred for setting the parameter values in a fine-grained manner, as shown by the results for P5. The participants showed strong interest in this scenario and understood our intention to potentially use any object as it "would be great if this worked with all objects" and one would not have to take it with oneself.

In summary, we found a strong preference for the simplest way of control, S1, using only a single gesture together with comprehensible visual representations. According to our participants, this offered the most direct form of control. A similar scheme of interaction was given in S2, however, with more gestures. This appeared to be significantly slower and more physically demanding. S3 affords a novel interaction paradigm, which provides tangible feedback through the used object. The participants certainly found this interesting, however, our evaluation shows that the more direct form control is preferred, which does not use an interaction proxy. Nevertheless, the haptic experience of the interaction allows for a fine-grained adjustment.

A main conclusion we can draw when contemplating all scenarios is that an AR representation offers the benefit of allowing for a direct control of appliances without requiring any proxy controller device in between. We believe this is also helpful advice for the design of potential future human interfaces of appliances.

7.1 Limitations

The general positive attitude towards our approach is despite of the HoloLens being relatively bulky and heavy, and furthermore having a little field of view. These constraints currently are a limitation to the practical use of our approach and are reflected in the relatively low scores for the question, whether participants could imagine to use our system in their daily life (Q8). Nevertheless, we believe there to be more practical AR devices in the future with a wider spread among the general public. We envision an AR system, which is built into a person's glasses, ubiquitously available, making our approach also practical for the daily life.

A further limitation of our work is that the prototype system, including the interfaces displayed, was fixed to three devices and to a single office. The concept itself is applicable to many more devices and a much larger space. As mentioned before, we designed the interfaces in a way that they could be used in a generalized form, which could be automatically generated from abstract device descriptions as shown in [21]. Especially the extension to a wider space with more rooms would be an interesting continuation because then also proximity-based interactions as in [11, 16] could be explored. In this sense, another limitation is that often the devices were within reaching distance for the user. This raises the question whether the widget augmentation and the use of gestures is necessary at all. The main reason for this is that the recognition of the markers does not work well at a large distance without significantly increasing the size of the marker, hence it was limited in our study. This limitation could be mitigated by implementing object recognition based on the appearance of the devices (which would eliminate the necessity of having to use markers over all), or using additional technology such as infrared emitters and receivers to transmit IDs between devices and the HoloLens (augmented with an IR sensor) which works across a range of several meters [2, 4]. However, the implementation of these more complex approaches is out of the scope of this paper. Nevertheless, in some situations the appliances were clearly out of reach for the participants, especially after moving to another chair. Furthermore, the participants never questioned the interaction design in spite of the proximity of the appliances. In any case, we believe that the widget augmentation and the interactions we present here still have benefits compared to interfaces on the appliances themselves, in particular because the appliance are not required to have an interface of their own.

Another limiting factor could be the fact that 15 out of 25 participants had no previous experience with VR or AR applications which might bias these participants in their general assessment of the interaction system by being amazed by the AR experience in general. Nonetheless, also among the participants with previous VR and AR experience most were positively surprised by the way our AR approach supported device interaction and also explicitly stated this in their comments. In fact, there were little differences in the attractiveness scores between the two groups for example (only S1 having a significantly higher score for the inexperienced group). In the comparison of the scenarios there should be no bias, as none of the participants knew the system beforehand.

8 CONCLUSION

We presented an interaction paradigm in which we externalize the user interface of appliances into visual representations in AR, using a head-mounted display. These representations provide an overview of the properties and capabilities of every appliance through the user's first-person view and also allow appliance control. We implemented three different ways of interaction: (1) using simple gestures to adjust virtual sliders and click virtual buttons, (2) employing more gestures recognized by an EMG armband, and (3) extending the representation to an object, in our case a tea mug, which can be utilized for control by moving it and thereby provides a tangible interface. We conducted a user study in an environment consisting of three appliances and our evaluation shows that the participants prefer the direct control and simplicity featured by the first scenario, which is also significantly faster than the other two options. Although the use of the object may seem the most intriguing interaction possibility, it introduces the object as a proxy control. This lets the user interact with the object, rather than with the appliance, which was less preferred by the participants.

We believe that our work demonstrates the potential of universal device control in augmented reality. In the future, this work can be further extended with markerless object detection, and scaled up to a larger space not confined to a single room. One can further imagine composite user interfaces in augmented reality, which can merge capabilities and allow simultaneous control of independent devices. Also, we are further investigating extensions of the third scenario, in which the proxy objects themselves are also smart, and can understand more about the user's intentions.

ACKNOWLEDGMENTS

We thank all participants who volunteered for our study for their time, effort, and valuable feedback.

REFERENCES

[1] D. Avrahami, J. Wobbrock, and S. Izadi. 2011. Portico: Tangible Interaction on and Around a Tablet. In *UIST '11*.
[2] T. Bartindale, D. Jackson, K. Ladha, S. Mellor, P. Olivier, and P. Wright. 2014. RedTag: Automatic Content Metadata Capture for Cameras. In *TVX '14*.
[3] B. Berghmans, A. Faes, M. Kaminski, and K. Todi. 2016. Household Survival: Immersive Room-Sized Gaming Using Everyday Objects As Weapons. In *CHI EA '16*.
[4] R. Boldu, H. Zhang, J. Cortés, S. Muthukumarana, and S. Nanayakkara. 2017. In-Sight: A Systematic Approach to Create Dynamic Human-controller-interactions. In *AH '17*.
[5] K. Cheng, R. Liang, B. Chen, R. Laing, and S. Kuo. 2010. iCon: Utilizing Everyday Objects As Additional, Auxiliary and Instant Tabletop Controllers. In *CHI '10*.
[6] A. de Freitas, M. Nebeling, X. Chen, J. Yang, K. Ranithangam, A. Kirupa, and A. Dey. 2016. Snap-To-It: A User-Inspired Platform for Opportunistic Device Interactions. In *CHI '16*.
[7] M. Funk, O. Korn, and A. Schmidt. 2014. An Augmented Workplace for Enabling User-defined Tangibles. In *CHI EA '14*.
[8] S. Henderson and S. Feiner. 2008. Opportunistic Controls: Leveraging Natural Affordances As Tangible User Interfaces for Augmented Reality. In *VRST '08*.
[9] V. Heun, J. Hobin, and P. Maes. 2013. Reality Editor: Programming Smarter Objects. In *UbiComp '13 Adjunct*.
[10] V. Heun, S. Kasahara, and P. Maes. 2013. Smarter Objects: Using AR Technology to Program Physical Objects and Their Interactions. In *CHI EA '13*.
[11] K. Huo, Y. Cao, S. Yoon, Z. Xu, G. Chen, and K. Ramani. 2018. Scenariot: Spatially Mapping Smart Things Within Augmented Reality Scenes. In *CHI '18*.
[12] H. Ishii and B. Ullmer. 1997. Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms. In *CHI '97*.
[13] S. Kasahara, R. Niiyama, V. Heun, and H. Ishii. 2013. exTouch: Spatially-aware Embodied Manipulation of Actuated Objects Mediated by Augmented Reality. In *TEI '13*.
[14] B. Kriesten, C. Mertes, R. Tünnermann, and T. Hermann. 2010. Unobtrusively Controlling and Linking Information and Services in Smart Environments. In *NordiCHI '10*.
[15] B. Laugwitz, T. Held, and M. Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *USAB '08*.
[16] D. Ledo, S. Greenberg, N. Marquardt, and S. Boring. 2015. Proxemic-Aware Controls: Designing Remote Controls for Ubiquitous Computing Ecologies. In *MobileHCI '15*.
[17] S. Lin, H. Cheng, W. Li, Z. Huang, P. Hui, and C. Peylo. 2017. Ubii: Physical World Interaction Through Augmented Reality. *IEEE Trans. on Mobile Computing* (2017).
[18] F. Mattern. 2003. From Smart Devices to Smart Everyday Objects. In *SOC '03*.
[19] S. Mayer, Y. Hassan, and G. Sörös. [n. d.]. A Magic Lens for Revealing Device Interactions in Smart Environments. In *SA '14*.
[20] S. Mayer and G. Sörös. 2014. User Interface Beaming - Seamless Interaction with Smart Things using Personal Wearable Computers. In *BSN '14*.
[21] S. Mayer, A. Tschofen, A. Dey, and F. Mattern. 2014. User Interfaces for Smart Things – A Generative Approach with Semantic Interaction Descriptions. *ACM Trans. Comput.-Hum. Interact.* (2014).
[22] H. Pohl and M. Rohs. 2014. Around-device Devices: My Coffee Mug is a Volume Dial. In *MobileHCI '14*.
[23] D. Schmidt, D. Molyneaux, and X. Cao. 2012. PICOntrol: Using a Handheld Projector for Direct Control of Physical Devices Through Visible Light. In *UIST '12*.
[24] J. Sorgalla, J. Fleck, and S. Sachweh. 2018. ARGI: Augmented Reality for Gesture-based Interaction in Variable Smart Environments. In *VISIGRAPP '18*.
[25] E. Velloso, M. Wirth, C. Weichel, A. Esteves, and H. Gellersen. 2016. AmbiGaze: Direct Control of Ambient Devices by Gaze. In *DIS '16*.
[26] R. Xiao, G. Laput, Y. Zhang, and C. Harrison. 2017. Deus EM Machina: On-Touch Contextual Functionality for Smart IoT Appliances. In *CHI '17*.