Augmented Reality and Older Adults: A Comparison of Prompting Types

Thomas J. Williams
tjw44@bath.ac.uk
University of Bath
Bath, United Kingdom

Simon L. Jones
s.l.jones@bath.ac.uk
University of Bath
Bath, United Kingdom

Christof Lutteroth
c.lutteroth@bath.ac.uk
University of Bath
Bath, United Kingdom

Elies Dekoninck
e.a.dekoninck@bath.ac.uk
University of Bath
Bath, United Kingdom

Hazel C. Boyd
hazelboyd@designability.org.uk
Designability
Bath, United Kingdom

ABSTRACT

Older adults can benefit from technologies that help them to complete everyday tasks. However, they are an often-under-represented population in augmented reality (AR) research. We present the results of a study in which people aged 50 years or older were asked to perform actions by interpreting visual AR prompts in a lab setting. Our results show that users were less successful at completing actions when using ARROW and HIGHLIGHT augmentations than when using ghosted OBJECT or GHOSTHAND augmentations. We found that user confidence in performing actions varied according to action and augmentation type. Users preferred combined AUDIO+TEXT prompts (our control condition) overall, but the GHOSTHAND was the most preferred visual prompt. We discuss reasons for these differences and provide insight for developers of AR content for older adults. Our work provides the first comparative study of AR with older adults in a non-industrial context.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; *User studies*; Laboratory experiments.

KEYWORDS

augmented reality, prompting tools, older adults, task prompting

ACM Reference Format:
Thomas J. Williams, Simon L. Jones, Christof Lutteroth, Elies Dekoninck, and Hazel C. Boyd. 2021. Augmented Reality and Older Adults: A Comparison of Prompting Types. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3411764.3445476

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445476

1 INTRODUCTION

Trends in augmented reality (AR) and pervasive computing suggest that AR is set to become integrated into daily living in the near future [14, 34]. Already there have been various successful applications of AR seen in specific settings, including medicine [28] and cultural heritage tourism [33]. AR has also been successfully applied to task prompting tools in industrial assembly and maintenance contexts [35], with evidence showing that the use of AR can lead to fewer unsolved errors compared with other types of support [13], during the completion of novel assembly tasks. As AR becomes more prevalent in society, it is becoming increasingly important to determine how and to what extent this technology might be used in everyday contexts.

An area in which task prompting systems have been widely used is that of assistive technologies (AT) to support an aging population to complete activities of daily living (ADLs). For example, existing (non-AR) task prompting tools have been shown to be effective at improving the independence of people living with memory difficulties [8, 19, 30]. In general, technology can be used in a variety of settings to help older adults with their changing needs in later life [16], hence it is appropriate to explore the potential applications of AR in this context. HCI literature has extensively studied age-related differences [47], but little work has focused on older adults' responses to AR prompts. Indeed, previous research has noted that this user group is often not considered the target for AR development [23]. As the world's population ages, more work must be carried out to understand how older users interact with AR technology, allowing designers of AR technologies to design more effective tools for older users.

In this paper, we argue that a better understanding is required regarding the types of visual AR prompts that can be effective for enabling older users to complete tasks in everyday, non-industrial contexts. We see this as a necessary first step towards developing effective AR task prompting systems to support older adults with daily living. We aim to answer the following research questions:

- **RQ1** What types of visual AR prompts are suitable for task prompting to support activities of daily living?
- **RQ2** Does the suitability of AR prompts depend on the types of actions that need to be performed?

We conducted an experiment to assess the efficacy of different visual prompts in terms of how they affected: (a) the *successful* *performance* of actions; (b) *user confidence* in performing actions correctly; (c) the *time it takes* users to perform actions; and (d) *user preferences* for augmentation types. Whilst our study could only cover a small part of the wide design space for AR prompting, we chose augmentations that have been identified as promising by related works [15, 42, 44, 55], but have not been evaluated with older adults before.

By answering these questions, we work towards creating a vocabulary of augmentations that can be adapted for different tasks and contexts. An AR task prompting tool that employs these AR actions could support someone to complete a task by overlaying important information directly onto the objects that are required for that task, providing both a visual reminder of the task and guidance on how to complete it. In this paper we:

(1) provide a comparison of AR prompts to support older adults (an under-represented population in other studies) with ADLs;
(2) provide insights into what types of visual AR prompts impact the successful completion and completion time of actions, and user confidence when performing actions, and discuss user preferences for different types of visual AR prompts;
(3) describe an experimental setup for testing AR prompts that could be built upon to explore more complex interactions and newly emerging augmentation designs.

2 RELATED WORK

2.1 Task Prompting for Older Adults

There is an association between aging and declining physical and cognitive abilities, such as learning, memory and executive function [17]. Older adults may have more trouble completing tasks, develop memory problems, or have difficulties recognising objects [2, 41]. These deficits can affect a person's ability to successfully complete ADLs, like preparing meals or managing finances, many of which include multiple steps that must be completed in the correct order [4]. Assistive technologies (ATs) have been developed to help people with memory difficulties and there is evidence that ATs can be effective at improving the independence of those that use them [19]. Technological support for task completion for people living with dementia has been identified [32], and has been addressed through non-AR prompting systems such as prompting aids [29] and assistive robots [48].

One particular type of AT includes prompting devices or task prompters [50], which aim to guide someone through a task by breaking it down into its individual components. For example, work using audio or audio and video prompts to support hand washing in people with moderate-to-severe dementia found an improvement in independence as well as less need for caregiver interaction to complete the task when using the prompting system [30]. Other research exploring verbal prompts to support older adults with mild cognitive impairment shows that prompting technology is considered helpful by this population [45]. Boyd et al. [8] compared four different prompt types and showed that text prompts and audio prompts separately can be effective at guiding someone through a task, and that the combination of these two formats is particularly effective. They also explored picture prompts and video prompts, which were found to be less effective than the text prompts and audio prompts, although they note how picture prompts in conjunction with text prompts can be helpful for tasks that need the person being prompted to distinguish between similar parts of an object, like buttons on a CD player. Other comparisons of prompt types for different populations also suggest that audio prompts are more effective and most preferred when used for directing adults with acquired brain injury [12] and to help students with moderate intellectual disability with shopping tasks [7]. In contrast, video prompts were found to be more effective than picture prompts in a study comparing these prompt modes to teach daily living skills to autistic teenagers [51], highlighting how prompt efficacy can depend on the user group being supported.

2.2 AR for Task Prompting

There has been much work exploring the use of AR to support assembly task completion in industry and maintenance contexts. A comprehensive survey [53] categorised existing literature into three areas, namely AR assembly guidance, AR assembly training and AR assembly simulation, design and planning. Research has shown that AR can lead to fewer unsolved errors when completing novel assembly tasks when compared with other types of prompt [13], and can also improve the time it takes to locate parts of a task [15]. Other work suggests that the complexity of visual AR prompts should match the complexity of the step of the task [38]. The survey also identified five areas on which to focus in future work, including real-time 3D reconstruction of the user's workspace, more intelligent prompting systems that can react to a user's mental state, and collaborative AR assembly systems [53]. An approach not discussed in the survey is demonstrated by *GhostHands UX* [42], a study that focused on user experience of both a trainee and a remote instructor who was providing telementoring using an AR copy of the instructor's hands and forearms, with verbal communication too. The instructor could interact with what the trainee could see, and the trainee could see the ghosted version of the instructor's hands overlaid on a view of the task objects on a computer screen in front of the task. Both instructor and trainee liked this arrangement, with trainees feeling self-confident and safe when using the *GhostHands*. A notable commercial example of an AR prompting system in an industrial setting is ScopeAR [44], which uses a combination of text, beacons, and arrows superimposed on real objects, as well as remote assistance. This incorporates many of the above approaches, except the use of ghost hands, and is used in current industry situations.

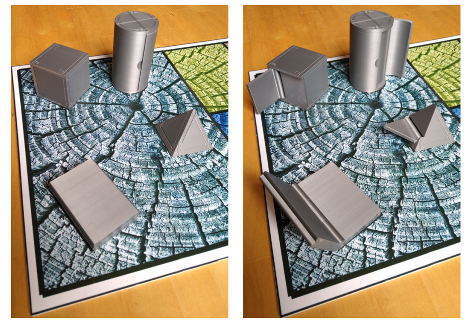
However, we are interested in task prompting in an everyday or domestic context, which differs from the technical setting of assembly tasks in that the types of activities being carried out could be less localised (moving around a room, or between rooms) and involve less precise placement of objects. Recent work exploring the use of AR in domestic settings identified that participants could see the benefit of AR for task assistance [21]. Existing examples of AR in non-industrial settings include: AR as a tool for domestic maintenance tasks, such as mountain bike assembly using QR codes [27]; a comparison of video and paper-based user manuals with AR user guides to operate household appliances [31]; HoloHome, a concept that envisions the use of AR headsets in smart homes, for example to turn lights on or off, or to give visual information to help someone to find things [26]; and projected AR to augment different areas of a kitchen [6]. Although these works cover a range of examples, they only evaluate with a small population size, or contain no evaluation at all. AR has also been shown to be helpful to support activity completion for other user groups. For example, an AR app was successfully used to help improve the focus of children with autism [11] and to improve vocational job skills for people with cognitive difficulties [9]. There are also examples of AR as a tool to support task prompting in the home for people with memory difficulties, for example, [18] and [40].

2.3 AR for Older Adults

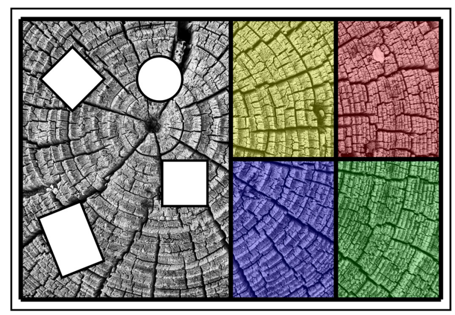
Research has begun exploring the use of virtual reality (VR) and AR for older adults, including to: envision fall prevention modifications to older adults' homes [25]; help others understand what it might be like living with cognitive impairments such as dementia [1]; diagnose early stages of memory loss [52]; reduce navigation- and distraction-related errors with in-car navigation systems [20]; and provide training for improving spatial direction [22]. The benefit of AR as a prompting tool for ADLs has also been identified [49], and this concept is starting to be explored in more depth. Earlier work proposed the concept of visually prompting in a kitchen context, projecting lights to highlight objects of importance for making a cup of coffee in an unfamiliar environment [18]. Correspondingly, Rohrbach et al. [40] implemented *Therapy Lens*, a HoloLens application that uses text displayed in AR and animations to instruct users through a tea-making task [40]. Although participants successfully used the hardware, the use of the *Therpay Lens* application did not affect the ability of participants to complete the task. However, *Therapy Lens* did not augment the real objects involved, rather virtual 'copies' of the objects were presented to participants. Similar work uses the HoloLens to support cooking tasks, with two versions of the application: one for the person affected by dementia and the other for a carer [56]. The carer is able to add AR elements to a room that has been scanned using the HoloLens, and the person affected by dementia can use the application to be aided when cooking. Arrows point to instructions that are not currently in view. These studies show the initial feasibility of using AR in an ADL context, but do not investigate which of the features implemented were effective and why. Furthermore, a review of prompting technologies to support older adults with cognitive impairment did not contain any work relating to AR [46]. The authors note how adoption of assistive technology depends on a positive experience with new technology, so it is important to learn what healthy older adults' views are of AR as a baseline for creating AR prompting technologies for older adults with cognitive impairments.

3 DESIGN DECISIONS

In this study, we wanted to explore the ability of purely visual AR prompts to prompt performance of actions. We believe that visual prompts could be a less intrusive method of prompting, and would not rely on someone's ability to comprehend spoken or written language, thus making them accessible to a wider audience. In the following sections, we describe the resources that were produced for the study and how these were improved by informal pilot testing.



(a) The 3D printed objects produced for our study in their closed (left) and open (right) states, in position on the foam board base.



(b) The image target used for the foam board base. The larger left rectangle is grayscale, and the four small rectangles are coloured clockwise from the top left yellow, red, green and blue.

Figure 1: The objects and image target.

3.1 Objects

Previous work recommends testing augmentations for a specific object and purpose within a specific context, since this affects the interpretation of the augmentation [24]. However, we wanted to account for some of the cognitive difficulties that older adults may face, such as difficulties identifying and recognising objects [37]. We designed a set of 3D-printed objects with unfamiliar affordances that would be unfamiliar to participants, to mitigate effects of participants having prior experience about how to manipulate the objects. It also simplified AR development, since we had the 3D models of the objects. In this way, we could test how well the visual augmentations were able to prompt someone to perform the action it was representing on the unfamiliar object.

The objects were: a cube (side length: 6 cm); a cylinder (diameter: 6 cm, height: 10 cm); a square-based pyramid (base side length: 6 cm, side angle: 60 degrees); and a book shape (6 cm × 8.4 cm × 2 cm). Each of the objects could be opened in some way: most objects had an 'open-able' flap, and the book shape was hinged to be open or closed. The objects are shown in Fig. 1(a) in their open and closed states. The objects were placed on a 3 mm thick foam board base of size 55 cm × 38 cm, which had the image in Fig. 1(b) attached to it. This image was used as an image target for producing the AR visual prompts (see Section 3.3). The significance of the coloured areas is described later.

3.2 Choosing Actions

Prior work [55] exploring prompting solutions for problems that can arise when completing kitchen tasks used Schwartz's Action Coding System (ACS) ([43]) to code video data with actions performed during ADLs. The ACS uses four high-level commands to describe any type of action: GIVE, TAKE, MOVE and ALTER. We used these commands as a basis for deciding which actions to test in our study, combined with our own semi-structured interviews with professionals with experience of working with older adults, including those with cognitive difficulties such as dementia (two occupational therapists, a psychologist, and two community workers). The professionals described typical ADLs and techniques that they employ to support someone to carry out these ADLs, including how a common approach is to deconstruct ADLs into simpler steps. Using this data, we checked that the ACS was capable of fully describing common ADLs that require support, such as 'making a cup of tea'.

In the following descriptions, '[OBJECT]' is used to mean one of the objects described above, and '[LOCATION]' is used for the coloured areas in Fig. 1(b), for example, the top-left coloured region is 'the yellow area'. In the ACS, TAKE implies becoming in possession of an object that a person does not have, so we decided to use the action 'pick up [OBJECT]'. GIVE implies the relinquishing of an object that a person already has, so we decided to use the action 'put down [OBJECT] in [LOCATION]'. MOVE differs from a combination of TAKE and GIVE since the action occurs in one movement, so the person never 'keeps' the object. We chose to use the action 'move [OBJECT] to [LOCATION]'. ALTER implies a change of state, for example, switching on a button, or opening a door. There were multiple possibilities, but we decided to use two commands: 'open [OBJECT]' and 'close [OBJECT]'. We chose these because they commonly arose in our interviews with professionals as steps in ADLs, such as opening and closing cupboard doors or tea boxes. To summarise, the five actions tested were:

(1) Pick up [OBJECT] (**PU**);
(2) Put down [OBJECT] in [LOCATION] (**PD**);
(3) Move [OBJECT] to [LOCATION] (**M**);
(4) Open [OBJECT] (**O**); and
(5) Close [OBJECT] (**C**).

We decided to only consider single steps of a task rather than composite actions, to explore whether certain augmentations would perform differently for different actions. For example, instead of 'pick up the cube and put it down in the blue area', we considered 'pick up the cube' and 'put down the cube in the blue area' separately.

3.3 Developing the Augmentations

In order to augment the four objects with visual AR prompts we designed a large image target (Fig. 1(b)) to be tracked using Vuforia in Unity. The objects had a 'home' position on this image target, and the augmentations were implemented relative to the target in such a way that they coincided with where the objects were placed. A high-resolution, highly textured image was used to improve the image tracking, ensuring that augmentations were accurately aligned with the positions of the real-world objects. The four coloured areas in the image served as distinct, bounded areas where objects could be placed when required. The following four augmentations types were developed using Unity:

a) **arrow**s pointing at target objects or locations (ARROW);
b) a moving transparent ghost of the **object** (OBJECT);
c) a pulsating transparent ghost of the object, which remained in its position but **highlight**ed the object or location (HIGHLIGHT);
d) a **ghost hand** model[1] together with a transparent ghost of the object (GHOSTHAND).

Arrows have been used in prior work for AR task prompting [15]. Both arrows and the use of a moving transparent ghost of the object were inspired by existing commercial applications, such as *ScopeAR* [44], and were also suggested as potential improvements to the *GhostHands UX* [42, p.242]. The pulsating transparent ghost of the object was based on suggestions to use a flashing light as an implicit way of identifying objects that relate to each other when completing kitchen tasks [55]. We took the concept of a ghost hand directly from the *GhostHands UX* study [42], using an opaque, white hand in combination with the moving transparent object. An illustration of each type of augmentation for the 'move' action is given in Fig. 2. A video summary and text description of these and all other combinations of action and augmentation can be found in the supplementary material.

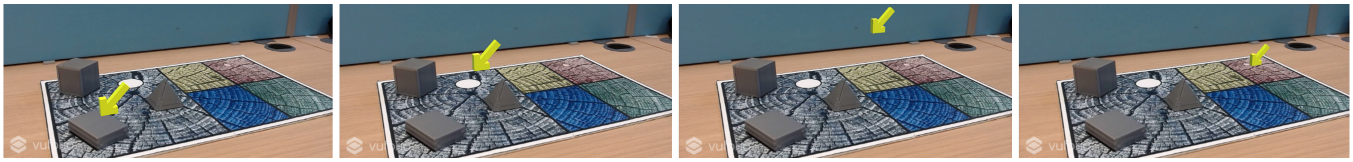
One design recommendation from the *Therapy Lens* study was that animations are useful when using AR prompts because of their eye-catching qualities [40]. Therefore, all of the augmentations developed for our study included animated elements. We did not include any words or audio prompts used for the augmentations because we focussed on visual prompts. However, as a control condition, we used an explicit prompt (AUDIO+TEXT) based on an existing non-AR ADL prompting tool that has been tested with people living with dementia and their carers [8]. The prompt was displayed as text on the screen, and an audio recording of a computer-generated voice saying the exact same words was played once, for example, "Pick up the cylinder".

3.4 Pilot Testing

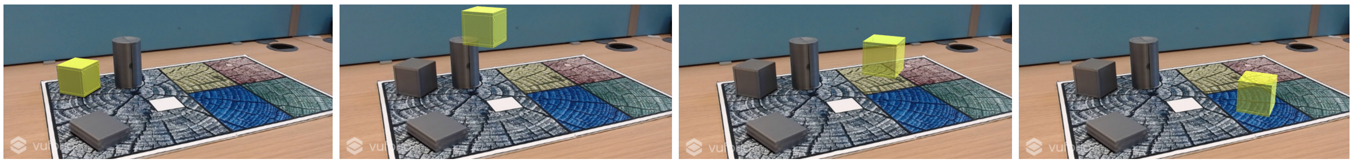
Informal pilot testing was carried out during the design phase to validate that the augmentations were as clear as possible. Testing was carried out with eight people, including the authors not involved with any technical development, colleagues of the authors, and doctoral students at the University of Bath.

The testing resulted in some minor improvements to the arrow augmentations. Originally, the arrows used the same transparent material as the moving objects, but this was changed to use an opaque yellow to make arrows stand out. The orientation of the arrows was also changed: originally, for the 'move' command, the arrow's orientation followed the direction of the arc as it moved. During testing, this was interpreted as changing the object's orientation as well as moving it, that is, flipping it over, rather than just moving it. We changed the arrow to be like a cursor, as shown in Fig. 2. The other augmentations did not need any changes.

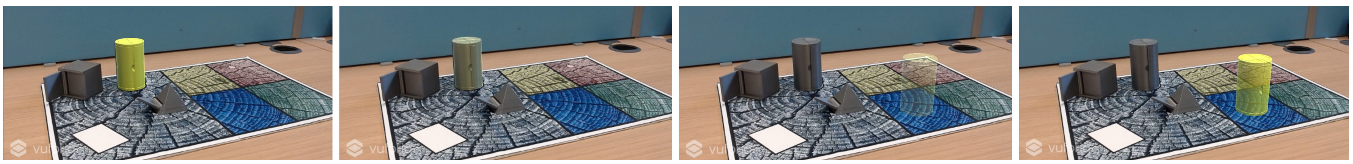
[1]The Oculus hand models from the Unity Sample Framework were used because they were freely available on the Unity asset store (https://assetstore.unity.com/packages/tools/integration/oculus-integration-82022).



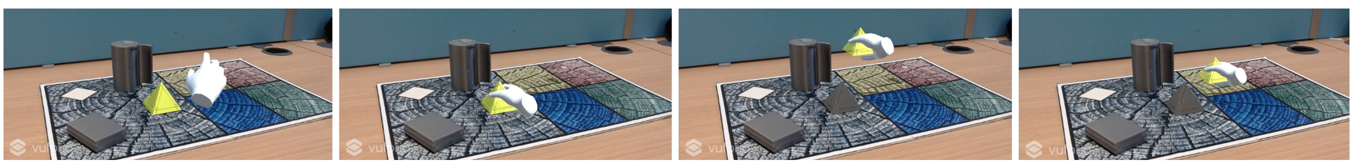
(a) The arrow begins by pointing at the object in its original position, then moves in a smooth arc to the intended final location and rests there for 1 second.



(b) A transparent yellow model of the object moves in a smooth arc from the object's original position to the intended final location, resting there for 1 second.



(c) A transparent, yellow model of the object pulses twice in the starting location (the maximum and minimum opacity are shown in left-most images), waits for half a second, then pulses twice in the intended location, resting for one second at its maximum opacity in the intended location (right-most image).



(d) The opaque, white hand begins above the object, moving smoothly to a 'gripping' position. Then, the hand and object move in the same way as for (b).

Figure 2: The move action is used to illustrate the four augmentation types: (a) ARROW, (b) OBJECT, (c) HIGHLIGHT, and (d) GHOSTHAND. The leftmost image shows the first keyframe, and the rightmost image shows the final keyframe of the animation. The images between these show intermediate keyframes. The descriptions below each set of four keyframes describe the full animation.

The timings of animations were changed as a result of user feedback. We tried to be consistent with both speed of moving objects and lengths of animations, but due to the different types of augmentation, it was not possible to make all animations the same length. We updated all animations to have a *cue length* (the time for one complete animation cycle) of between 3 and 4.2 seconds, except for the arrow and highlight animations for 'pick up' and 'put down', which were 1.3 seconds. This rate of movement was perceived by pilot participants as comfortable, that is, not too urgent.

The pilot testing also helped us to refine the procedure for the study. During informal testing, it became clear that there needed to be a physical distinction between a 'researcher space' and a 'participant space', so that participants could interact with the board and objects without the researcher obstructing them, and so that the researcher could arrange the objects correctly and easily.

After the changes from informal pilot testing had been made, the procedure as described in the next section was carried out with two people not involved with the previous testing. These were not participants of the study, but followed the same procedure participants would go through. These trial runs verified that the procedure could be carried out practically by the researcher and that the changes from informal pilot testing had been successfully incorporated into the study design.

4 METHODS

Our experiment had three independent variables, namely: Action (the five actions described in Section 3.2); Augmentation (the four augmentation types described in Section 3.3 plus the AUDIO+TEXT control); and Trial Number (the first time or second time participants encountered a certain combination of Action and Augmentation). These were controlled in a within-participant design based on a balanced Latin square to mitigate order effects. Each trial comprised a participant being prompted to complete a single Action for one object, using a particular Augmentation. Participants were shown 50 trials in total, in three blocks. They were shown 20 trials (5 Actions $\times$ 4 Augmentations) in block one for the first Trial Number, and then the same 20 trials in a different order for the second Trial Number. The order in which participants saw the trials was randomised and counter-balanced using a balanced Latin-square of size 20 in both blocks. After blocks one and two, participants were

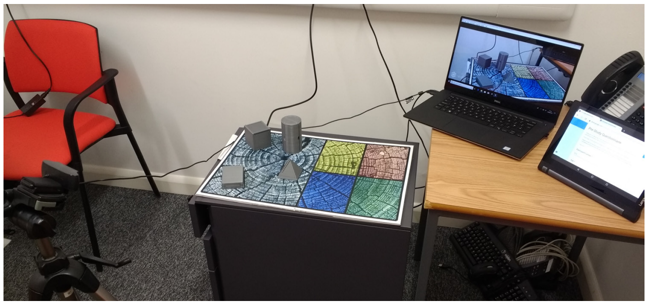


Figure 3: The experiment setup from the participant's point of view. Participants viewed the Augmentations on the laptop screen (top right), which showed a view of the board and objects from the point of view of the webcam (bottom left). The tablet in the right of the image was used to record participant responses during the study.

shown 10 control trials in block three. These consisted of 5 audio and text prompts (one prompt for each of the 5 Actions) repeated twice, to reflect the method for the first two blocks. The control trials were also counter-balanced using balanced Latin squares. The control trials were performed after the Augmentations to avoid additional learning effects, since the audio and text prompts made the Actions explicit.

Sessions were video recorded using a camcorder. During the experiment, a screen capture of the laptop screen was taken, which also recorded audio. Ethical approval was given by the Research Ethics Approval Committee for Health (REACH) at the University of Bath (Reference: EP 18/19 003).

4.1 Apparatus

An image of the typical setup is shown in Fig. 3. A 64-bit Dell Windows laptop with 32.0 GB of RAM and a 2.80 GHz Intel Core i7-7700HQ processor with 3840 x 2160 resolution running Unity version 2017.4.30f1 (64-bit) and a custom scene was placed on a table to the right of the image target and objects. A Logitech HD Pro Webcam C920 was connected to the laptop, and was used to track the image target. This was placed on a tripod to the left of the objects. Participants sat so they could easily reach the objects on the board and see the laptop screen, as well as touch the tablet computer, which was used by participants during the study to record their responses (see **Procedure**). Weiss et al. [54] showed that analysing augmentations through different empirical methods or modalities (such as headset and screen-based displays) leads to comparable results, so our approach of using screen-based AR is sufficient to answer our research questions. Furthermore, screen-based AR is an important modality to consider as it is currently more widely used in home settings than AR glasses and has been widely studied in prior work.

4.2 Participants

The experiment was conducted with adults who were 50 years of age or older, as adults in this age range are considered 'older users' [17, p. 32]. In total, 20 participants were recruited by using mailing lists of academic and teaching staff at the University of Bath and advertising the opportunity on social media. Participants were given a £10 shopping voucher for their participation.

The average age was 60.65 ± 7.84 years (Min: 50, Max: 74). There were 13 female and 7 male participants. 17 participants were right-handed, 3 were left-handed. All participants wore glasses or contact lenses, but none reported having a type of colour-blindness. When asked to choose the highest qualification obtained, most participants responded with a university degree (6 Doctorates, 5 Master, 2 Bachelor), 2 had secondary or further education qualifications, and 5 responded with 'other' (degree level courses).

Participants were asked to rate their experience of the following four concepts on a 7-point Likert scale (1="Not at all experienced", 7="Very experienced"). The ratings were as follows: Information Technology (IT) [Average: 4.9±1.33, Min: 2, Max: 7]; Virtual Reality (VR) [Average: 1.8±1.32, Min: 1, Max: 6]; Augmented Reality (AR) [Average: 1.6±1.23, Min: 1, Max: 6]; and Task prompting people with a memory or cognitive impairment [Average: 1.5±1.24, Min: 1, Max: 6]. Hence, our participants generally identified as being IT literate, but with limited experience of AR/VR technologies and task prompting.

4.3 Procedure

Participants signed a consent form before filling out the demographic questionnaire. Participants were then briefed on the process of the experiment. The setup of the board, computer and camera as seen in Fig. 3 was explained to the participants.

It was made clear that the researcher would be controlling what was shown on the screen by using a wireless keyboard and that, for each trial, the screen would be black to start with. It was explained that participants would be shown visual information on the laptop screen about an action that they should perform on one of the objects. Participants were told to perform the action as soon as it was clear to them what the action was, and to indicate that they were finished by saying "I'm done" or "I'm finished".

The researcher then explained that they would ask two questions after each completed action:

(1) What do you think the action was that you had to perform?
(2) On a scale of 1 to 5, where 1 means "Not confident at all" and 5 means "Very confident", how confident are you that you have performed the correct action?

The first question was answered verbally and the second question was answered by the user on a tablet computer.

For the start state of each trial, one of the four objects was given to the participant to ensure that the Action "Put Down" was available to them as an option for the action to perform and the remaining objects on the board were set to be either open or closed. The object to be given to the participant, the states of the objects on the table, the object that would be acted upon, and the coloured region to be used (when applicable) were all generated randomly by our software during the experiment. We did not counterbalance these factors in addition to the counterbalancing of Actions and Augmentations, since the random selection meant that there was rarely any repetition in the directions and distances of the motions to be carried out. The researcher counted down from three out loud before revealing the Augmentation to the participant (on 'zero'). The Augmentations repeated in a loop. The researcher then waited for the participant to finish performing the action (at which point the researcher pressed a button to record the time for completion) and for the participant to indicate they had finished before making the screen blank again.

4.3.1 Measures (Dependent Variables). After participants completed each trial, the researcher recorded whether the participant had performed that Action correctly as True or False (**DV: Successful Completion**). The time (in seconds) it took the participant to perform the Action (from the Augmentation being revealed until the person completed performance) was also recorded by the researcher, if it was performed correctly. However, as noted earlier, the *cue length* varied slightly depending on the Action and Augmentation, (within a range from 1.333 to 4.167 seconds). For analysis the **DV: Completion Time** is defined as the time taken to perform the correct Action *minus the cue length* for that Action and Augmentation combination. The *cue length* for the text and audio prompts for the control trials is equal to the average length of the audio recordings for each Action (between 1.25 and 2.7 seconds); an average recording length was used because there were multiple recordings for each Action in order to make every combination of object and location available and therefore there was necessarily some small variation in the lengths of those recordings due to the variation in length of object names and colours of the locations. The **DV: Number of Cues** is defined as the number of repetitions of the visual cue that was shown, and calculated as the time taken for the participant to perform the correct Action *divided by the cue length*. For the text and audio prompts, the **Number of Cues** was set to 1, since the audio prompt was only given once. Participants also rated their **DV: Confidence** that they had performed the correct action on a 5-point Likert scale after each trial. Participants were not told if they had performed the correct action after each trial.

After all trials had been completed, participants were asked to rank the Augmentations (including the AUDIO+TEXT prompt) for each Action in order of **DV: Preference**, using a 'most preferred' to 'least preferred' ranking. Participants were shown the four AR Augmentations for each Action before doing this to remind them what each Augmentation looked like for that Action. Participants were allowed to rank more than one Augmentation at the same level. In these cases, participants were asked to place them at the appropriate point of the ranking scale.

5 RESULTS

In the following sections, we will use the following shorthand to refer to the three independent variables: Action (pick up = **PU**, put down = **PD**, move = **M**, open = **O**, and close = **C**), Augmentation (ARROW, OBJECT, HIGHLIGHT, GHOSTHAND and AUDIO+TEXT), and Trial Number (first trial or second trial). Descriptions for each Action and Augmentation are given in Section 3.2 and Section 3.3, respectively.

5.1 Effect on Successful Completion

Figure 4 shows a comparison of the success rates (that is, when a participant performed the correct action on the correct object) of each Action-Augmentation combination between the first trial and second trial. A Durbin test shows that the effects of Action ($\chi^2(4) = 82.416, p < .001^{**}$) and Augmentation ($\chi^2(4) = 93.736, p < .001^{**}$)

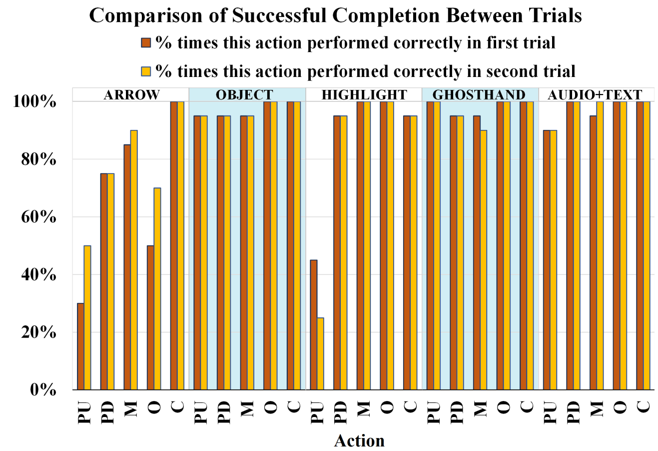


Figure 4: Comparison of Successful Completion between first trial and second trial.

on **Successful Completion** were significant, and that the effect of Trial Number was not significant ($\chi^2(1) = 0.149, p = .700$). Pairwise Conover post-hoc comparisons with Holm correction between Actions ($T(976) \leq 2.223, p \geq .264$) and Augmentations ($T(976) \leq 2.058, p \geq .398$) were not significant. Although not statistically significant, the results seem to indicate that ARROW augmentations were successfully completed less often than all other augmentation types (73% of the time across all actions for ARROW). For OBJECT, HIGHLIGHT, GHOSTHAND and AUDIO+TEXT, successful completion rates were consistently above 90% for all Actions, with the exception of the HIGHLIGHT Augmentation for the **PU** action. Given the high success rates in the first trial, there was no significant improvement on **Successful Completion** in the second trial.

5.2 Effect on Completion Time, Number of Cues and Confidence

For the dependent variables **Completion Time** and **Number of Cues**, the effects of Action and Augmentation were compared using two-way repeated measures ANOVAs. These was carried out using data from the second trial only, as there were too many missing values from the first trial due to incorrect actions being completed. We report the main and interaction effects for these dependent variables. For all post-hoc comparisons paired t-tests with Holm correction were used.

To further explore the effect of all independent variables, we also carried out regression analyses. For the dependent variables **Completion Time**, **Number of Cues** and **Confidence** the effect of Action, Augmentation and Trial Number was compared using multilevel linear regression models in R through the nlme package [5], because of their power and flexibility in handling missing values when analysing repeated measures data [10, 36]. Action, Augmentation and Trial Number were set up as fixed effects. Participant number was set up as grouping factor, so that Action, Augmentation and Trial Number were treated as within-participant effects. Regression coefficients B for a condition indicate the estimated value by which an outcome variable is greater (for $B > 0$) or smaller (for

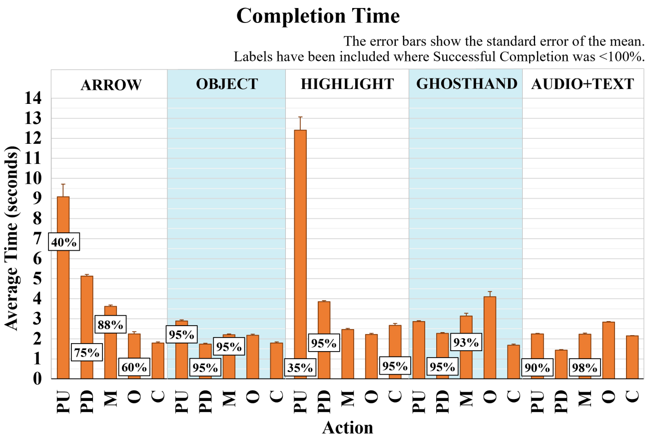


Figure 5: The average Completion Time for each Action-Augmentation combination.

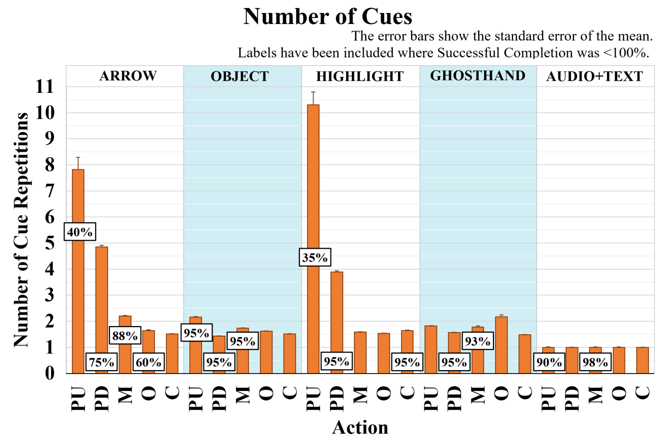


Figure 6: The average Number of Cues for each Action-Augmentation combination.

Table 1: The results of multilevel linear regression for dependent variable Completion Time.

Independent Variable	B	95% *CI*	$t(889)$	p-value
Intercept	2.06	[1.21,2.92]	4.76	$< .001^{**}$
Second Trial	-1.13	[-1.92,-0.35]	-2.84	$.014^{*}$
PU	1.95	[1.58,2.31]	10.44	$< .001^{**}$
PD	0.94	[0.57,1.31]	4.99	$< .001^{**}$
M	0.66	[0.29,1.04]	3.52	$.002^{**}$
O	0.66	[0.33,0.99]	3.96	$< .001^{**}$
ARROW	1.56	[0.99,2.13]	5.37	$< .001^{**}$
OBJECT	0.08	[-0.47,0.63]	0.28	.779
HIGHLIGHT	1.21	[0.64,1.77]	4.16	$< .001^{**}$
GHOSTHAND	0.29	[-0.22,0.80]	1.12	.527

Table 2: The results of multilevel linear regression for dependent variable Number of Cues.

Independent Variable	B	95% *CI*	$t(889)$	p-value
Intercept	0.64	[0.20,1.08]	2.86	$.017^{*}$
Second Trial	-0.41	[-0.81,-0.01]	-2.04	.126
PU	1.55	[1.34,1.76]	14.27	$< .001^{**}$
PD	1.01	[0.79,1.22]	9.19	$< .001^{**}$
M	0.19	[-0.02,0.41]	1.76	.157
O	0.15	[-0.04,0.34]	1.51	.157
ARROW	2.14	[1.82,2.46]	13.09	$< .001^{**}$
OBJECT	0.75	[0.44,1.06]	4.72	$< .001^{**}$
HIGHLIGHT	1.82	[1.50,2.15]	11.14	$< .001^{**}$
GHOSTHAND	0.84	[0.56,1.13]	5.74	$< .001^{**}$

$B < 0$) for the respective condition compared to the overall average. The baseline condition (the 'Intercept' in all tables) was the control AUDIO+TEXT with the **C** Action for the first trial. The **C** action was chosen as it was the least ambiguous command; Figure 4 shows how all participants performed this Action correctly in all trials. We report the 95% confidence intervals *CI* of coefficients and test them for significance at $\alpha = .05$ using Holm post-hoc correction.

5.2.1 Effect on Completion Time. A graph showing the **Completion Time** is shown in Fig. 5. The labels on the bottom of the bars show the percentage of times that Action was performed correctly in both trials if less than 100%. A two-way ANOVA did not show significant main effects for Action ($F(4, 4) = 1.931, p = .270, \omega^2 = 0.042$) or Augmentation ($F(4, 4) = 1.870, p = .280, \omega^2 = 0.035$). There was a significant interaction effect between Action and Augmentation ($F(16, 16) = 6.137, p < .001^{**}, \omega^2 = 0.149$). Post-hoc comparisons show that participants took longer for the combination of HIGHLIGHT and **PU** when compared with GHOSTHAND for PD ($t(19) = 5.348, p = .044^{*}$) and GHOSTHAND for PU ($t(19) = 5.686, p = .039^{*}$).

Table 1 shows a summary of the regression analysis results for the **Completion Time**. Participants decreased their **Completion Time** (that is, took less time) on average by 1.13 seconds from the first trial to the second trial ($p = .014^{*}$). Participants increased their **Completion Time** (that is, took longer) on average by 1.95 seconds for the **PU** Action ($p < .001^{**}$), by 0.94 for the **PD** Action ($p < .001^{**}$), by 0.66 for the **M** Action ($p = .002^{**}$) and by 0.66 for the **O** Action ($p < .001^{**}$). Participants took longer on average by 1.56 seconds for the ARROW Augmentation ($p < .001^{**}$) and by 1.21 seconds for the HIGHLIGHT Augmentation ($p < .001^{**}$). No significant effects were found for the OBJECT ($p = .779$) or GHOSTHAND ($p = .527$) Augmentations.

5.2.2 Effect on Number of Cues. Figure 6 shows the **Number of Cues** seen by the participant before completion of the Action. A two-way ANOVA did not show significant main effects for Action ($F(4, 4) = 4.909, p = .076, \omega^2 = 0.324$) or Augmentation ($F(4, 4) = 5.398, p = .066, \omega^2 = 0.388$). There was a significant interaction effect between Action and Augmentation ($F(16, 16) = 6.051, p < .001^{**}, \omega^2 = 0.426$). Post-hoc comparisons show that the **Number of Cues** was significantly larger for the combination of HIGHLIGHT and **PU** when compared with all other Actions for the OBJECT ($t(19) \geq 5.234, p \leq .030^{*}$) and GHOSTHAND ($t(19) \geq 5.337, p \leq .025^{*}$) Augmentations and the AUDIO+TEXT control ($t(19) \geq 6.445, p \leq .004^{**}$). **Number of Cues** was also significantly larger for the combination of HIGHLIGHT

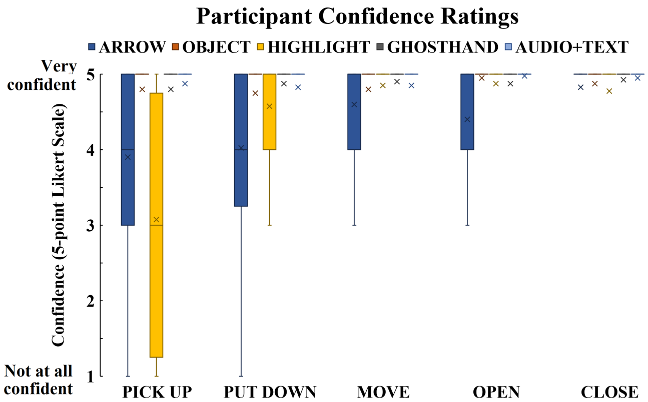


Figure 7: Summary of participants' Confidence ratings after performing each Action. The ×s show the mean.

Table 3: The results of multilevel linear regression for dependent variable participant Confidence

Independent Variable	B	95% CI	$t(990)$	p-value
Intercept	5.03	[4.88,5.18]	64.97	$< .001^{**}$
Second Trial	0.13	[0.02,0.24]	2.33	.101
PU	-0.57	[-0.69,-0.45]	-9.52	$< .001^{**}$
PD	-0.25	[-0.38,-0.12]	-3.79	$< .001^{**}$
M	-0.07	[-0.19,0.06]	-1.01	1.000
O	-0.05	[-0.17,0.06]	-0.86	1.000
ARROW	-0.54	[-0.70,-0.38]	-6.75	$< .001^{**}$
OBJECT	-0.01	[-0.17,0.15]	-0.09	1.000
HIGHLIGHT	-0.54	[-0.70,-0.38]	-6.67	$< .001^{**}$
GHOSTHAND	-0.07	[-0.22,0.08]	-0.89	1.000

and **PU** than the ARROW Augmentation for the **C**, **O** and **M** Actions ($t(19) \geq 5.203, p \leq .031^{*}$) and the HIGHLIGHT Augmentation for the **C**, **O** and **M** Actions ($t(19) \geq 6.827, p \leq .002^{**}$).

Table 2 shows a summary of the regression analysis results for the **Number of Cues**. No significant change was found for **Number of Cues** between the first trial and second trail ($p = .126$). The **Number of Cues** was 1.55 higher on average for the **PU** Action ($p < .001^{**}$) and 1.01 higher on average for the **PD** Action ($p < .001^{**}$). No significant changes were observed for the **M** ($p = .157$) or **O** ($p = .157$) Actions. The **Number of Cues** was higher on average by 2.14 for the ARROW Augmentation ($p < .001^{**}$), by 0.75 for the OBJECT Augmentation ($p < .001^{**}$), by 1.82 for the HIGHLIGHT Augmentation ($p < .001^{**}$), and by 0.84 for the GHOSTHAND Augmentation ($p < .001^{**}$).

5.2.3 Effect on Confidence. Figure 7 shows a boxplot of the participants' **Confidence** responses, grouped by Action. Table 3 shows a summary of the regression analysis results for **Confidence**. No significant change was found for **Participant Confidence** between the first trial and second trail ($p = .101$). Participants were less confident on average by 0.57 for the **PU** Action ($p < .001^{**}$) and by 0.25 for the **PD** Action ($p < .001^{**}$). There was no significant change for the **M** ($p = 1.000$) or **O** ($p = 1.000$) Actions. Participants

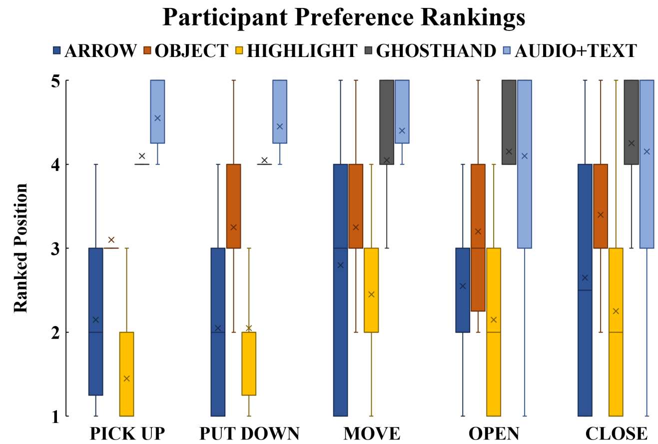


Figure 8: Summary of participant rankings of the Augmentations (including the AUDIO+TEXT prompt) for each Action. The ×s show the mean. 1 = Ranked Lowest, 5 = Ranked Highest.

were less confident on average by 0.54 for the ARROW Augmentation ($p < .001^{**}$) and by 0.54 for the HIGHLIGHT Augmentation ($p < .001^{**}$). No significant change was observed for the OBJECT ($p = 1.000$) or GHOSTHAND ($p = 1.000$) Augmentations.

5.3 Effect on Preference

Figure 8 shows a boxplot of participants' **Preference** rankings of the Augmentations grouped by Action. A Durbin test showed that the effect of Augmentation on **Preference** across all Actions was significant ($\chi^2(4) = 184.353, p < .001^{**}$). Pairwise Conover post-hoc comparisons with Holm correction showed that participants significantly preferred the AUDIO+TEXT prompt and the GHOSTHAND Augmentation over the HIGHLIGHT Augmentation ($T(476) \geq 4.489, p \leq .001^{**}$) and the ARROW Augmentation ($T(476) \geq 3.693, p \leq .002^{**}$). All other comparisons were not significant ($T(476) \leq 2.565, p \geq .064$).

6 DISCUSSION AND DESIGN CONSIDERATIONS

In our study, we wanted to learn what types of visual AR prompts were suitable for task prompting to support activities of daily living (RQ1) and if the suitability of AR prompts depended on the types of actions that needed to be performed (RQ2). We selected AUDIO+TEXT as the baseline (non-AR prompt) prior to analysing the results of our experiment, since this is a commonly used modality in existing task prompting systems [8, 50] and enabled comparison with the visual AR prompts we developed for our experiment. We did not have clear expectations about which prompt would perform best. When considering our results on the whole, it appears that participants were able to interpret many of the visual AR prompts that they were shown, to varying degrees.

Control trials were frequently performed well by our participants and resulted in high **Successful Completion**. It is important to note that the control trials were seen after the visual AR prompts for all participants, and so the performance for the control condition could be in part due to practice. For example, participants were likely to have seen all of the moving parts prior to the AUDIO+TEXT instructions; if they had been told to 'open the cylinder' at the start of the experiment, they might have struggled to find the flap due to the symmetry of the shapes. However, because of this we can have confidence that the visual prompts were suitable at prompting actions, since the **Successful Completion** was high for most combinations of Action and Augmentation. We might expect to see lower levels of successful completion for some Actions in the control trials if this block of trials was carried out first, or in isolation without the context given by the AR blocks.

There were some differences in significant results for **Completion Time** and **Number of Cues** due to how these variables were defined. **Completion Time** was measured as the time taken to complete the trial, beyond the time taken to watch one cycle of the cue. **Number of Cues** reflected the number of times that the cue was repeated until the trial was completed. For a set amount of time, Augmentations with a longer cue length had a relatively lower **Number of Cues** than Augmentations with a shorter cue length. In that sense, **Number of Cues** is a better indicator of how easy it is for someone to understand an Augmentation; it is a better standardisation of the times we recorded. Based on this, we think the **Number of Cues** is most telling about how easy an Augmentation was to interpret and the **Completion Time** is less useful in this respect.

With that in mind, we did not find significant main effects of Action or Augmentation in the ANOVA analyses for either **Completion Time** or **Number of Cues**. However, we did find a significant interaction between the effects of Action and Augmentation for both of these dependent variables. Hence, how long it takes someone to complete a task after being prompted once (**Completion Time**) and how easy an augmentation is to understand (represented by **Number of Cues**) both depend on the combination of task type (Action) and the visual AR prompt used (Augmentation). Our regression analysis provides some further insights. **Completion Time** was affected significantly by each of the Actions, but only the ARROW and HIGHLIGHT Augmentations. In contrast, **Number of Cues** was significantly affected by each of the Augmentations, but only the **PU** and **PD** Actions. Therefore, **Completion Time** (across all Augmentations) was mainly affected by Action, as some actions physically took longer to perform than others, but the ARROW and HIGHLIGHT Augmentations took more time to interpret; whereas **Number of Cues** was mainly affected by Augmentation, as some Augmentations (notably ARROW and HIGHLIGHT) were not immediately understood by participants and had to be re-watched, although we can also say that the PU and PD Actions were more difficult to interpret than the other Actions.

The GHOSTHAND and OBJECT Augmentations both performed comparably with the baseline AUDIO+TEXT prompt in terms of **Successful Completion**, and users only needed to see the cue between one and just over two times before completing the Action (Fig. 6). This implies that purely visual prompts can be effective for prompting action. In some cases, participants responded quicker to the GHOSTHAND and OBJECT Augmentations than to the baseline prompt (for example, the close (**C**) Action in Fig. 5), although we did not find any significant effects on **Completion Time** for these Augmentations. Our results did not show that participants were significantly more confident for the GHOSTHAND or OBJECT Augmentations than the baseline, but the GHOSTHAND Augmentation was the most preferred Augmentation amongst our participants. Further to this, participants commented on how they could 'relate' to the hand and it was 'obvious' what had to be done. This is in line with how participants of the *GhostHands UX* user study responded [42], which supports the use of a GHOSTHAND Augmentation outside of an industrial or maintenance context. Although the OBJECT and GHOSTHAND Augmentations were most preferred, these are also the most difficult Augmentations to develop; 3D models of the objects are required for high-fidelity augmentations, and alignment of AR prompts with real objects is an on-going challenge in task prompting [53], though some existing commercial applications are capable of correct placement of AR prompts, as seen in *ScopeAR* [44].

The arrows were the simplest Augmentation to develop, and only require recognition of objects, rather than 3D models of objects. Previous work [15] has used arrows successfully to indicate objects and draw attention to them, but, based on our results, we do not recommend using arrows to prompt actions because of the lower success rate and confidence ratings observed in our study. In all ARROW cases, users recognised the object with which they had to interact, but participants did not respond by performing the correct Action for most of the actions we tested in our study. Therefore, our results suggest that arrows for the use of action assistance is inadvisable, but could be used for attention guiding [39].

This recommendation is also relevant for the HIGHLIGHT Augmentation. Despite reasonable success rates for this Augmentation, this was the least preferred type of Augmentation amongst our participants, which can be explained by considering the pick up (**PU**) Action, one notable case that was performed incorrectly (on average less than 30% **Success Completion** for HIGHLIGHT). Participants knew what object to interact with, but did not know what to do with it. This was reflected in their responses when answering the question about what they thought the Augmentation was prompting them to do: Some participants thought they should *"just register its presence"* (P03), *"leave it there"* (P12), or *"just to take notice of it but not to do anything with it"* (P13). Most participants explained that they would rank the Augmentations by using clarity as their guide, and said that this was the most difficult Augmentation to understand because it was ambiguous. Some participants also mentioned how it took longer for the HIGHLIGHT Augmentations to convey the Action (participants had to see approximately 10 cues on average before completing the Action). This has implications for further development of experiential augmentations [24] since, at least in the context of performing a task, our study suggests that users prefer prompts that have a clear and direct meaning.

Audio and text prompts were considered to be the most clear for our participants, and were ranked as the most preferred prompt overall. Work exploring prompting technology for people with dementia also concluded that a combination of audio and text prompts is most effective at prompting someone through a task [8], though picture prompts were considered useful for situations that required choosing between a number of similar objects, for example, choosing the correct button on a remote control. Audio was also the most preferred prompt in a study exploring different prompting methods for navigation guidance [12]. However, given the cases in our study where the visual AR prompts performed on a par with the baseline and the cases that were considered ambiguous, it could be beneficial to combine visual prompts with audio and text prompt, though care would need to be taken not to overload the user with too much information [40].

The displaced camera and screen locations caused no difficulties for the majority of participants. This could have implications for future research since if this setup is sufficient, it could be more familiar than using AR headsets for older users, who may be unfamiliar with this novel technology. However, one participant did not interpret the depth of Augmentations correctly, resulting in that person holding the objects above the board for move (**M**) and put down (**PD**) Actions. Other participants sometimes found it difficult to determine the depth of the arrows on the screen in relation to the real physical environment. This could be resolved if the camera could move or by using head-mounted AR, though the latter introduces added complexity when aligning the 3D augmentations.

We also observed examples of people performing unintended actions, even in cases where the same Augmentation type was observed by the same person on the same object, such as: rotating objects when observing the ARROW Augmentation for the open Action (this was not observed for the close Action); leaving objects alone when observing the HIGHLIGHT Augmentation for the pick up Action (as discussed above); lifting objects into the air, instead of picking them up and bringing the object to themselves for the ARROW Augmentation and pick up Action. These examples highlight more evidence that the affordances of objects matter, and alter the interpretation of the Augmentation, as found in [24]. Sometimes, the orientation of AR objects for the OBJECT Augmentations caused confusion or seemed important to some participants, with participants wondering if the object should be in that orientation or not.

Although the AR techniques we tested are well-known, they have not been evaluated comparatively in prior work. For example, *GhostHands* has only been evaluated in a small qualitative, non-comparative study [42]. Similarly, arrows and animated sequences have been tested against non-AR baselines [15] but not against each other. Our novel contribution is that we perform this comparative, controlled study, and that we situate this comparison in an older adult context. We believe the results may have been different for younger participants. Ball and Hourcade [3] found that younger adults often out-perform older adults when given unfamiliar tasks, while older adults perform better when using previously learned behaviours. This suggests that younger adults would have been better able to deal with unfamiliar prompts, emphasising the importance of identifying prompts for older adults that are consistent with their previous experiences. Furthermore, the design space of possible Augmentations is vast and contains many small variations. Despite these Augmentations appearing to be similar, our results suggest clear differences between them. Our study contributes helpful directions for HCI researchers and designers interested in applying AR in an ADL task prompting context. That said, our findings would likely be affected by cognitive impairments (on the basis of related works on task prompting for people with cognitive impairments [8, 9]). For example, users with a cognitive impairment would likely have struggled even more with abstract Augmentations such as HIGHLIGHT and would have preferred more familiar elements such as GHOSTHAND. Our study could inform future explorations of AR for people with cognitive impairments.

6.1 Limitations

Our study was carried out in a single, small environment, that is, there was no movement between rooms or parts of a room. The Augmentations tested in our study could be more useful in larger environments, for example, using arrows to indicate movement between rooms, or to direct someone to a different part of the same room, a technique that is used in existing prototypes [56]. Additionally, we did not explore hybrid combinations of AR prompts. It is possible that a combination of ARROW and HIGHLIGHT prompts would be more effective at prompting action than either of these alone, which would be easier to develop from a technical point of view than the GHOSTHAND Augmentation. However, we cannot assume the effects of individual prompts would combine in an additive manner and so this is an opportunity for future work.

For simplicity in augmenting the objects, we used a single large image target to position our Augmentations, rather than tracking individual objects. This may not be practical in everyday situations, but indicates that this relatively simple and adaptable method is suitable for testing visual prompts. While the experimental setup worked for our study, the static camera introduced a constraint that would not be present with a head-mounted AR device. The position of the AR visualisation may be a factor that affects task performance, however this was not explored within our study.

Our measures did not capture all of the qualities that are likely to be desirable for task completion. For example, ADLs will likely involve the need for safety, so slow and considered performance of an action may sometimes be preferable to being quick and efficient. Although Augmentations could enable successful completion, the impact of AR task prompting on feelings of independence and self-efficacy would need to be explored in order to assess the impact of such a technology. Our insights are relevant for people without cognitive difficulties, but there are doubtless other aspects to take into account for people living with more specific or complex needs. For example, *GhostHands* could be disconcerting for someone who is experiencing visual hallucinations, or a disembodied voice could cause distress to someone with auditory hallucinations. More subtle clues, such as 'suggestive sounds' (e.g. of cutlery to prompt eating [48]), could be incorporated into an unobtrusive prompting system that takes advantage of attention grabbing Augmentations, rather than explicit prompts.

We only considered relatively simple Actions. There are other, more complicated ALTER-type commands that may form part of many ADLs, such as turning dials to the correct setting on a washing machine [8]. Some actions might require more precise execution (putting a CD into a CD player) or for objects to follow a certain path. Similarly, we did not consider inter-object interactions, which are common in many ADLs, for example, using a spoon to pick up sugar, or stirring tea with the spoon. We did not explore these nuanced actions, and prior work hypothesises that these actions are likely to require more complex augmentations in order to be completed correctly [38].

Our experiment was carried out in an abstract context, not a real ADL context. Specific contexts and objects are likely to introduce assumptions and affordances that will have an effect on the interpretation of prompts. There was also no context provided in terms of the "goal" of the task (i.e. situating each single action that participants were prompted to perform within a larger task made up of multiple steps). We did this deliberately to simplify the experimental design and to account for some of the challenges associated with declining cognitive abilities, like difficulties with task sequencing (linking together individual steps of a task) or memory difficulties (forgetting what is being done). However, this context may impact the interpretation of augmentations. Our work provides a starting point from which these effects can be isolated and explored.

7 CONCLUSION AND FUTURE WORK

In this paper we presented a study exploring the suitability of visual AR prompts (augmentations) for task prompting to support activities of daily living. We have provided a novel, comparative study of four AR visual prompts. The results of our study reveal several implications for design, for example suggesting that arrow and highlight augmentations are less effective than moving object or ghost hand augmentations, for a range of actions that apply to many tasks. Our study also indicates that the ghost hand prompt was both easy to understand and generally the most preferred augmentation for older adults.

These results support findings from existing AR assembly task prompting research [42, 53], showing that purely visual augmentations are suitable as a prompting method, and we demonstrated this in a non-industrial context. If these visual augmentations were to be combined with existing prompting tools like those for people with cognitive difficulties [8, 30], there is the potential to further improve the independence of older adults in later life. Our work provides a foundation for future research, which could extend the scope to explore more complex, multi-step tasks in a specific ADL context, such as making a cup of tea.

ACKNOWLEDGMENTS

Thank you to all of the participants of this study. The first author would like to thank Rob Hanson and Jess Ridgers for their help with producing the physical resources for the study. This research was funded by the EPSRC Centre for Doctoral Training in Digital Entertainment (Grant Code: EP/L016540/1).

REFERENCES

[1] Alzheimer's Research UK. 2019. A Walk Through Dementia. https://www.awalkthroughdementia.org/
[2] Alzheimer's Society. 2017. What is dementia? Factsheet 400LP. https://www.alzheimers.org.uk/sites/default/files/migrate/downloads/what{_}is{_}dementia.pdf
[3] Robert Ball and Juan Pablo Hourcade. 2011. Rethinking Reading for Age From Paper and Computers. *International Journal of Human-Computer Interaction* 27, 11 (Nov 2011), 1066–1082. https://doi.org/10.1080/10447318.2011.555319
[4] Cornelia Beck, Patricia Heacock, Carla Gene Rapp, and Susan O. Mercer. 1993. Assisting cognitively impaired elders with activities of daily living. *American Journal of Alzheimer's Care and Related Disorders & Research* 8, 6 (Nov 1993), 11–20. https://doi.org/10.1177/153331759300800602
[5] Paul Bliese. 2006. Multilevel Modeling in R (2.2)–A Brief Introduction to R, the multilevel package and the nlme package.
[6] Leonardo Bonanni, Chia-Hsun Lee, and Ted Selker. 2005. CounterIntelligence: Augmented Reality Kitchen. In *CHI 2005*. ACM Press, Portland, Oregon, USA, 1–7.
[7] Emily C Bouck, Rajiv Satsangi, Whitney Bartlett, and Pei-Lin Weng. 2012. Promoting Independence through Assistive Technology: Evaluating Audio Recorders to Support Grocery Shopping. *Education and Training in Autism and Developmental Disabilities* 47, 4 (Dec 2012), 462–473. http://www.jstor.org/stable/23879639
[8] Hazel C Boyd, Nina M Evans, Roger D Orpwood, and Nigel D Harris. 2017. Using simple technology to prompt multistep tasks in the home for people with dementia: An exploratory study comparing prompting formats. *Dementia* 16, 4 (May 2017), 424–442. https://doi.org/10.1177/1471301215602417
[9] Yao-Jen Chang, Ya-Shu Kang, and Po-Chiao Huang. 2013. An augmented reality (AR)-based vocational task prompting system for people with cognitive impairments. *Research in Developmental Disabilities* 34, 10 (Oct 2013), 3049–3056. https://doi.org/10.1016/j.ridd.2013.06.026
[10] Avital Cnaan, Nan M Laird, and Peter Slasor. 1997. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in medicine* 16, 20 (1997), 2349–2380.
[11] Lizbeth Escobedo, Mónica Tentori, Eduardo Quintana, Jesus Favela, and Daniel Garcia-Rosas. 2014. Using augmented reality to help children with autism stay focused. *IEEE Pervasive Computing* 13, 1 (2014), 38–46. https://doi.org/10.1109/MPRV.2014.19
[12] Stephen Fickas, McKay Sohlberg, and Pei Fang Hung. 2008. Route-following assistance for travelers with cognitive impairments: A comparison of four prompt modes. *International Journal of Human Computer Studies* 66, 12 (Dec 2008), 876–888. https://doi.org/10.1016/j.ijhcs.2008.07.006
[13] Nirit Gavish, Teresa Gutiérrez, Sabine Webel, Jorge Rodríguez, Matteo Peveri, Uli Bockholt, and Franco Tecchia. 2015. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments* 23, 6 (Nov 2015), 778–798. https://doi.org/10.1080/10494820.2013.815221
[14] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 6 (Jun 2017), 1706–1724. https://doi.org/10.1109/TVCG.2016.2543720
[15] Steven J. Henderson and Steven Feiner. 2009. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *Science and Technology Proceedings - IEEE 2009 International Symposium on Mixed and Augmented Reality, ISMAR 2009*. IEEE, Orlando, FL, USA, 135–144. https://doi.org/10.1109/ISMAR.2009.5336486
[16] Ann Horgas and Gregory Abowd. 2004. The Impact of Technology on Living Environments for Older Adults. In *National Research Council (US) Steering Committee for the Workshop on Technology for Adaptive Aging*, RW Pew and SB Van Hemel (Eds.). National Academies Press (US), Washington (DC), Chapter 9, 230–251. https://www.ncbi.nlm.nih.gov/books/NBK97336/
[17] Felicia Huppert. 2003. Designing for older users. In *Inclusive Design: Design for the Whole Population*, John Clarkson, Simeon Keates, Roger Coleman, and Cherie Lebbon (Eds.). Springer London, London, Chapter 1, 30–49. https://doi.org/10.1007/978-1-4471-0001-0_2
[18] Sei Ikeda, Zeeshan Asghar, Jaakko Hyry, Petri Pulli, Antti Pitkanen, and Hirokazu Kato. 2011. Remote Assistance Using Visual Prompts for Demented Elderly in Cooking. In *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL '11)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/2093698.2093744
[19] Matthew Jamieson, Breda Cullen, Marilyn McGee-Lennon, Stephen Brewster, and Jonathan J. Evans. 2014. The efficacy of cognitive prosthetic technology for people with memory impairments: A systematic review and meta-analysis. *Neuropsychological Rehabilitation* 24, 3-4 (Jul 2014), 419–444. https://doi.org/10.1080/09602011.2013.825632
[20] SeungJun Kim and Anind K Dey. 2009. Simulated Augmented Reality Windshield Display as a Cognitive Mapping Aid for Elder Driver Navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 133–142. https://doi.org/10.1145/1518701.1518724
[21] Pascal Knierim, Paweł W. Woźniak, Yomna Abdelrahman, and Albrecht Schmidt. 2019. Exploring the potential of augmented reality in domestic environments. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2019*, Vol. 19. Association for Computing Machinery, Inc, Taipei, Taiwan, 1–12. https://doi.org/10.1145/3338286.3340142
[22] I. Jui Lee, Chien Hsu Chen, and Kuo Ping Chang. 2016. Augmented reality technology combined with three-dimensional holography to train the mental rotation ability of older adults. *Computers in Human Behavior* 65 (Dec 2016), 488–500. https://doi.org/10.1016/j.chb.2016.09.014
[23] Li Na Lee, Mi Jeong Kim, and Won Ju Hwang. 2019. Potential of Augmented Reality and Virtual Reality Technologies to Promote Wellbeing in Older Adults. *Applied Sciences* 9, 17 (Aug 2019), 3556. https://doi.org/10.3390/app9173556
[24] Dixon Lo, Dan Lockton, and Stacie Rohrbach. 2018. Experiential Augmentation: Uncovering the Meaning of Qualitative Visualizations when Applied to Augmented Objects. In *Proceedings of the 2018 CHI Conference on Human Factors*

in Computing Systems - CHI '18. ACM Press, New York, New York, USA, 1–8. https://doi.org/10.1145/3173574.3174064

[25] Michael Lo Bianco, Sonja Pedell, and Gianni Renda. 2016. Augmented reality and home modifications: A tool to empower older adults in fall prevention. In *Proceedings of the 28th Australian Computer-Human Interaction Conference, OzCHI 2016*. Association for Computing Machinery, Inc, Launceston Tasmania Australia, 499–507. https://doi.org/10.1145/3010915.3010929

[26] Atieh Mahroo, Luca Greci, and Marco Sacco. 2019. HoloHome: An Augmented Reality Framework to Manage the Smart Home. In *6th International Conference on Augmented Reality, Virtual Reality and Computer Graphics (AVR 2019)*, Vol. 11614 LNCS. Springer Verlag, Santa Maria al Bagno, Italy, 137–145. https://doi.org/10.1007/978-3-030-25999-0_12

[27] Jorge Martín-Gutiérrez and Irene Inés Santos Pérez. 2011. Augmented reality approach to domestic maintenance tasks. In *HCI 2011: HCI International 2011 – Posters' Extended Abstracts*, Vol. 174 CCIS PART 2. Springer, Berlin, Heidelberg, Orlando, FL, USA, 125–129. https://doi.org/10.1007/978-3-642-22095-1_26

[28] Martina Merten. 2007. Erweiterte Realität: Verschmelzung zweier Welten. *Deutsches Ärzteblatt* 104, 13 (2007), A840. https://www.aerzteblatt.de/archiv/55014/Erweiterte-Realitaet-Verschmelzung-zweier-Welten

[29] Alex Mihailidis, Jennifer Boger, Marcelle Canido, and Jesse Hoey. 2007. The use of an intelligent prompting system for people with dementia. *interactions* 14, 4 (Jul 2007), 34. https://doi.org/10.1145/1273961.1273982

[30] Alex Mihailidis, Jennifer N Boger, Tammy Craig, and Jesse Hoey. 2008. The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study. *BMC Geriatrics* 8, 1 (Dec 2008), 28. https://doi.org/10.1186/1471-2318-8-28

[31] Lars Müller, Ilhan Aslan, and Lucas Krüßen. 2013. GuideMe: A mobile augmented reality system to display user manuals for home appliances. In *10th International Conference on Advances in Computer Entertainment Technology (ACE 2013)*, Vol. 8253 LNCS. Springer, Cham, Boekelo, The Netherlands, 152–167. https://doi.org/10.1007/978-3-319-03161-3_11

[32] R. Orpwood, J. Chadd, D. Howcroft, A. Sixsmith, J. Torrington, G. Gibson, and G. Chalfont. 2010. Designing technology to improve quality of life for people with dementia: user-led approaches. *Universal Access in the Information Society* 9, 3 (Aug 2010), 249–259. https://doi.org/10.1007/s10209-009-0172-1

[33] George Papagiannakis, Sebastien Schertenleib, Brian O'Kennedy, Marlene Arevalo-Poizat, Nadia Magnenat-Thalmann, Andrew Stoddart, and Daniel Thalmann. 2005. Mixing virtual and real scenes in the site of ancient Pompeii. *Computer Animation and Virtual Worlds* 16, 1 (Feb 2005), 11–24. https://doi.org/10.1002/cav.53

[34] Jon Peddie. 2017. Introduction. In *Augmented Reality*. Springer International Publishing, Cham, 1–28. https://doi.org/10.1007/978-3-319-54502-8_1

[35] Jon Peddie. 2017. Software Tools and Technologies. In *Augmented Reality*. Springer International Publishing, Cham, 165–182. https://doi.org/10.1007/978-3-319-54502-8_7

[36] James L Peugh. 2010. A practical guide to multilevel modeling. *Journal of school psychology* 48, 1 (2010), 85–112.

[37] K S Purdy, P A McMullen, and M Freedman. 2002. Changes to the object recognition system in patients with dementia of the Alzheimer's type. *Brain and cognition* 49, 2 (2002), 213–216.

[38] Rafael Radkowski, Jordan Herrema, and James Oliver. 2015. Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty. *International Journal of Human-Computer Interaction* 31, 5 (May 2015), 337–349. https://doi.org/10.1080/10447318.2014.994194

[39] Patrick Renner. 2018. Prompting Techniques for Guidance and Action Assistance Using Augmented-Reality Smart-Glasses. In *25th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018 - Proceedings*. Institute of Electrical and Electronics Engineers Inc., Reutlingen, Germany, 820–822. https://doi.org/10.1109/VR.2018.8446292

[40] Nina Rohrbach, Philipp Gulde, Alan Robert Armstrong, Linda Hartig, Anas Abdelrazeq, Stefan Schröder, Johanne Neuse, Timo Grimmer, Janine Diehl-Schmid, and Joachim Hermsdörfer. 2019. An augmented reality approach for ADL support in Alzheimer's disease: a crossover trial. *Journal of NeuroEngineering and Rehabilitation* 16, 1 (Dec 2019), 66. https://doi.org/10.1186/s12984-019-0530-z

[41] Jennifer Rusted and Linda Sheppard. 2002. Action-based memory in Alzheimer's disease: A longitudinal look at tea making. *Neurocase* 8, 1-2 (2002), 111–126. https://doi.org/10.1093/neucas/8.1.111

[42] Giuseppe Scavo, Fridolin Wild, and Peter Scott. 2015. The GhostHands UX: telementoring with hands-on augmented reality instruction. In *Workshop Proceedings of the 11th International Conference on Intelligent Environments*, D. Preuveneers (Ed.) (Ed.). IOS Press, Prague, Czech Republic, 236–243. https://doi.org/10.3233/978-1-61499-530-2-236

[43] Myrna F. Schwartz, Edward S. Reed, Michael Montgomery, Carolyn Palmer, and Nathaniel H. Mayer. 1991. The quantitative description of action disorganisation after brain damage: A case study. *Cognitive Neuropsychology* 8, 5 (Sep 1991), 381–414. https://doi.org/10.1080/02643299108253379

[44] Scope AR. 2018. Scope AR Brings the Next-Generation of Augmented Reality to the Enterprise with New AR Platform that Delivers Both Real-time Remote Assistance and AR Smart Instructions. https://www.prnewswire.com/news-releases/scope-ar-brings-the-next-generation-of-augmented-reality-to-the-enterprise-with-new-ar-platform-that-delivers-both-real-time-remote-assistance-and-ar-smart-instructions-300657204.html

[45] Adriana M. Seelye, Maureen Schmitter-Edgecombe, Diane J. Cook, and Aaron Crandall. 2013. Naturalistic assessment of everyday activities and prompting technologies in mild cognitive impairment. *Journal of the International Neuropsychological Society* 19, 4 (2013), 442–452. https://doi.org/10.1017/S135561771200149X

[46] Adriana M. Seelye, Maureen Schmitter-Edgecombe, Barnan Das, and Diane J. Cook. 2012. Application of cognitive rehabilitation theory to the development of smart prompting technologies. , 29–44 pages. https://doi.org/10.1109/RBME.2012.2196691

[47] Samuel Silva, Daniela Braga, and António Teixeira. 2014. AgeCI: HCI and Age Diversity. In *Universal Access in Human-Computer Interaction. Aging and Assistive Environments. UAHCI 2014.*, Constantine Stephanidis and Margherita Antona (Eds.), Vol. 8515 LNCS. Springer International Publishing, Cham, 179–190. https://doi.org/10.1007/978-3-319-07446-7_18

[48] Hugo Simão and Tiago Guerreiro. 2019. MATY: Designing An Assistive Robot for People with Alzheimer's. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3313016

[49] Daniel Sonntag. 2015. Kognit: Intelligent Cognitive Enhancement Technology by Cognitive Models and Mixed Reality for Dementia Patients. https://www.aaai.org/ocs/index.php/FSS/FSS15/paper/view/11702/11476

[50] Katherine M Tsui and Holly A Yanco. 2010. Prompting Devices: A Survey of Memory Aids for Task Sequencing. In *Proc. of Quality of Life Technologies (QoLT) Intl. Symp.: Intelligent Systems for Better Living, held in conjunction with RESNA*. Las Vegas, 1–7. http://robotics.cs.uml.edu/fileadmin/content/publications/2010/2010-06-qolt-ktsui-hyanco.pdfhttps://api.semanticscholar.org/CorpusID:15663680

[51] Toni Van Laarhoven, Erika Kraus, Keri Karpman, Rosemary Nizzi, and Joe Valentino. 2010. A Comparison of Picture and Video Prompts to Teach Daily Living Skills to Individuals With Autism. *Focus on Autism and Other Developmental Disabilities* 25, 4 (Dec 2010), 195–208. https://doi.org/10.1177/1088357610380412

[52] Alla Vovk, Ameera Patel, and Dennis Chan. 2019. Augmented Reality for Early Alzheimer's Disease Diagnosis. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland, UK, 1–6. https://doi.org/10.1145/3290607.3313007

[53] X. Wang, S. K. Ong, and A. Y.C. C Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4, 1 (Mar 2016), 1–22. https://doi.org/10.1007/s40436-015-0131-4

[54] Maximilian Weiss, Katrin Angerbauer, Alexandra Voit, Magdalena Schwarzl, Michael Sedlmair, and Sven Mayer. 2020. Revisited: Comparison of Empirical Methods to Evaluate Visualizations Supporting Crafting and Assembly Purposes. *IEEE Transactions on Visualization and Computer Graphics* Early Access (Oct 2020), 1–1. https://doi.org/10.1109/tvcg.2020.3030400

[55] Joseph P. Wherton and Andrew F. Monk. 2010. Problems people with dementia have with kitchen tasks: The challenge for pervasive computing. *Interacting with Computers* 22, 4 (Jul 2010), 253–266. https://doi.org/10.1016/j.intcom.2010.03.004

[56] Dennis Wolf, Daniel Besserer, Karolina Sejunaite, Anja Schuler, Matthias Riepe, and Enrico Rukzio. 2019. cARe: An Augmented Reality Support System for Geriatric Inpatients with Mild Cognitive Impairment. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia (MUM '19)*. Association for Computing Machinery, Pisa Italy, 11. https://doi.org/10.1145/3365610.3365612