University of Reading

Supporting people with dementia - understanding their interactions with Mixed Reality Technologies

Conference or Workshop Item

Published Version

Creative Commons: Attribution-Noncommercial 4.0

Open Access

Desai, Shital, Blackler, Alethea, Fels, Deborah and Astell, Arlene ORCID logoORCID: https://orcid.org/0000-0002-6822-9472 (2020) Supporting people with dementia - understanding their interactions with Mixed Reality Technologies. In: DRS2020, 11-14 Aug 2020, Brisbane. doi: https://doi.org/10.21606/drs.2020.266 Available at https://centaur.reading.ac.uk/96095/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

Identification Number/DOI: https://doi.org/10.21606/drs.2020.266 <https://doi.org/10.21606/drs.2020.266>

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

DRS2020
BRISBANE, 11–14 AUG
SYNERGY

Supporting people with dementia - Understanding their interactions with Mixed Reality Technologies

Shital Desai[a*], Alethea Blackler[b], Deborah Fels[c], Arlene Astell[d]

[a] Social and Technological Systems (SaTS) lab, School of Arts Media Performance and Design, York University, Canada
[b]Queensland University of Technology, Brisbane, Australia
[c]Ted Rogers School of Information Technology Management, Ryerson University, Canada
[d]Departments of Psychiatry and Occupational Sciences & Occupational Therapy, University of Toronto, Canada
[d] Dementia Ageing Technology Engagement Lab, University Health Network, Toronto, Canada
[d] School of Psychology & Clinical Language Sciences, University of Reading, UK

*Corresponding author e-mail: desais@yorku.ca
doi: 10.21606/drs.2020.266

Abstract:
Emerging technologies such as Mixed Reality Technologies (MRTs) could offer possibilities to support People with Dementia (PwD). This study examined interactions of PwD with two MRTs - HoloLens and Osmo. Ten participants (MoCA = 18 to 23, Age= 63 to 88 years) played a game of Tangram on Osmo. Six of these participants played Young Conker on HoloLens. The study found that PwD used gestural actions in the physical world more correctly than speech. Audio prompts in human voice were more correctly perceived than visual prompts. Physical affordances, embodied actions and familiarity to verbal instructions contribute to this success. Visual prompts such as text present promising opportunities to complement with audio prompts. Interaction with MRTs require prompts to direct PwD towards physical or virtual worlds. The research outcomes are significant as the focus on interactions of PwD could open up avenues for further research on actions and perceptions with emerging technologies.

Keywords: Mixed Reality, Assistive technologies, Dementia

1. Introduction

Assistive prompting technologies could support people with dementia (PwD) to continue instrumental activities of daily living (IADL) such as cooking, laundry and making a phone call. There is huge potential for technology-based prompting to support PwD through the sequences required to complete activities. Mihailidis, Boger, Craig, & Hoey (2008) and Pigot, Mayers, & Giroux (2003) developed intelligent systems using textual and auditory prompts, utilising sensors, computer vision and artificial intelligence. In a real-world deployment,

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Orpwood, Adlam, Evans, Chadd, & Self (2008) created a smart apartment equipped with passive sensors and light controls, bed occupancy monitor, tap and cooker monitors and voice prompting devices. Their trial with a single user with quite severe dementia, revealed that the technology helped the user retain a lot of independence, regain urinary continence and improved his hours of sleep and reduced night-time wanderings. To translate these promising results into scalable solutions for PwD to continue independent activities, Mixed Reality Technologies (MRTs) could offer affordable, adaptable solutions that can be easily adopted and deployed.

On a physical virtual continuum, MRTs are anything in between physical interfaces at one extreme end of the continuum and virtual interfaces at the other extreme end (Desai, Blackler, & Popovic, 2016). MRTs can either consist of augmentation of the physical world with virtual objects, as in Augmented Reality (AR) (Azuma et al., 2001) or the virtual world augmented with physical objects, as in Augmented Virtuality (AV) (Regenbrecht et al., 2004). AR and AV technologies could offer new technological solutions for PwD by generating prompts in response to their behaviour and actions. For example, AR technology, HoloLens is capable of tracking the gaze of people and can be programmed to generate prompts in response to their gaze and actions. This could be used to detect when people have become derailed during an activity and are searching for cues to get back on track. Similarly, AV technologies, Kinect and Osmo could track the behaviour of people through a network of sensors to determine when the person has derailed during the activity.

For MRTs to be able to deliver prompts that are effective and perceivable for PwD, firstly requires developing an understanding of their interactions with MRTs. This study thus investigates actions (e.g. gestures, speech, etc.) that are most suitable for PwD and forms of prompts that should be administered to PwD (e.g. such as auditory, visual, pictorial, etc.) in MR environments.

2. Background

MRTs could provide an adaptive and scalable solution that can be easily deployed in different applications. Chang, Chou, Wang, & Chen (2013) reported that a Kinect-based prompting system in a vocational setting improved task completion in two participants, one with early onset dementia and one with an acquired brain injury. In this study the prompts were delivered on a computer screen. Since HoloLens' launch in 2016, AR has been explored as a therapeutic tool for PwD. Ro, Park, & Han (2019) developed a projection based AR robot with 360 degrees of space for PwD, some of the applications targeted are therapy, entertainment, spatial art and mental care aids. Vovk, Chan, & Patel (2019) developed an AR HoloLens game as a memory assessment tool for early Alzheimers diagnosis. Aruanno, Garzotto, & Rodriguez (2017), Aruanno, Garzotto, Torelli, & Vona (2018) and Garzotto, Torelli, Vona, & Aruanno (2019) developed three holographic activities using HoloLens in cooperation with neurologists to stimulate memory functions affected by Alzheimers disease, with the goal of delaying the cognitive decline. PwD did not have problems using

the technology, however the success with which they completed the holographic activities depended on the design features that they interacted with. Although, AR and VR have been explored in research to support manual tasks such as maintenance, assembly and surgery for over five decades, there is a need to research interaction requirements of PwD.

2.1 Actions and prompts

Perception and action are inseparable processes which form the basis of any interaction. Perception is about recovering information about the interface such as shape, colour, motion and materiality through the use of sensory systems of vision, touch, auditory, taste and smell and the prompts presented in the interface (Gibson, 1979a). This information is processed by reasoning and decision-making to determine actions to be carried out on the interfaces. These actions could change the features and behaviour of the interface, communicating feedback, status or/and attention to the user, ready to be perceived again for subsequent actions (Maier, 2015). An interaction comprises of several such perception action loops. If for some reason there is a failure to understand the prompts or to act on the interface, the perception action loop is broken, people could find it difficult to get back on track. Desai, Blackler, & Popovic (2019) refer to interfaces that facilitate continuous perception action loops as intuitive in nature.

Tangible interactions have been explored for reminiscence in PwD. Huber, Berner, Uhlig, Klein, & Hurtienne (2019) using contextual design process identified needs of PwD and developed three tangible prototypes – a pyramid, set of drawers and a juke box. All three prototypes used object manipulation for gestural interaction. Their study found that the tangibles were intuitive to use where the form was familiar to the participants. Even though when the memory of PwD slipped away, they were still exploring the contents of the tangibles in their hands.

Natural User Interfaces (NUIs) have been successfully used in reminiscence therapy for PwD (Loureiro & Rodrigues, 2011). NUIs use natural interactions of humans such as touch, speech, gestures, gaze, facial expressions to interact with interfaces. Older adults find interactions with mouse and keyboard difficult (Czaja, 1997). On the other hand touchscreen gestures have found to be easy to use for PwD (Barban et al., 2016; Buiza et al., 2009; Gamberini et al., 2006). Gündogdu, Bejan, Kunze, & Wölfel (2017) evaluated use of NUIs in interactive multimedia systems as means of reminiscence therapy in PwD. The study found that while touch gestures and tangible interactions were successful in generating positive feelings and activating mental states, the interactions with NUIs depends on the person's stage of cognitive impairment and the individual form on the day.

Carlomagno, Pandolfi, Marini, Di Iasi, & Cristilli (2005) studied gestures of PwD in their daily communications with other people and they found that irrespective of the level of impairment, people were using gestures without any trouble. Bewernitz, Mann, Dasler, & Belchior (2009) compared machine administered verbal and visual prompts in PwD carrying out three tasks, drinking water, brushing and eating. Visual cues were not effective during

the brushing activity. This was due to the placement of the screen at the sink level height which made it difficult to look at while looking into the mirror. But they were able to complete the task of brushing which is an indication that the prompts were probably not needed for that activity. The audio cues were equally effective for all the three activities. These studies are limited in terms of exploration of types of prompts and actions for PwD because the focus is on the design of systems followed with user evaluations to test the designed systems rather than understanding the interactions with different types of modalities. Also, MR environments pose an additional challenge of dealing with the transition between the physical and virtual environments. A focussed study on understanding actions and prompts that work for PwD interacting with MR environments is thus required.

3. Method

3.1 Research design

This research was an observational study conducted at Memory and Company, a memory health club for seniors with dementia. Game play was used as a probe to elicit natural behaviour in the participants. Games are part of daily day programs at Memory and Company to engage PwD in hands on activities that simulate their mind and challenge their mental, functional and physical abilities. Thus, play-based research design allowed us to integrate the study as part of the day program at Memory and Company. It also acted as an ice breaker and made the participants feel more comfortable around technologies. The study was approved by the Research Ethics Board. Informed written consent was obtained from all participants and their care givers.

Ten people in their early stages of dementia with low cognitive impairment (MoCA = 18 to 24, Age= 63 to 88 years) participated in the study. The Montreal Cognitive Assessment (MoCA) is a rapid screening tool to detect cognitive dysfunction for early diagnosis of Alzheimers disease. It consists of 30 questions targeting attention, concentration, executive functions, memory, language, visuo-constructional skills, conceptual thinking, calculations, and orientation. Each question is scored as per the MoCA scoring guidelines. The total possible score is 30 points. A score of 26 or above means that there is no cognitive impairment. As the score gets lower, the level of impairment increases.

We used two off the shelf MRTs – HoloLens from Microsoft and Osmo from Tangible Play as case studies to study interactions of PwD with MRTs. Using off the shelf existing technologies is an effective way to understand technology needs of people and their perception action behaviour (Blackler, Popovic, & Mahar, 2010; Desai et al., 2019; Lawry, Popovic, Blackler, & Thompson, 2019). The two technologies represented AR (HoloLens) and AV (Osmo) types of MRT and were easily available for the study off the shelf. Participants played Tangram on Osmo and Young Conker on HoloLens. These games were chosen after exploring several games available on these technologies. The criteria for the selection was

that they should present different forms of prompts (visual, audio, animations, text) and opportunities to carry out various actions (gestural and speech interactions). Pilot studies with 2 older adults in their early stages of dementia (MoCA = 18 and 22) were carried out before the actual study to determine the suitability of the technologies and games for the study.

3.1.1 Osmo and Tangram

Osmo, AV technology, with distinct physical and virtual environments to interact with separately, allows physical play with a virtual environment on a tablet. It comes with a reflector, a stand to place the tablet on, and games (Figure 1).

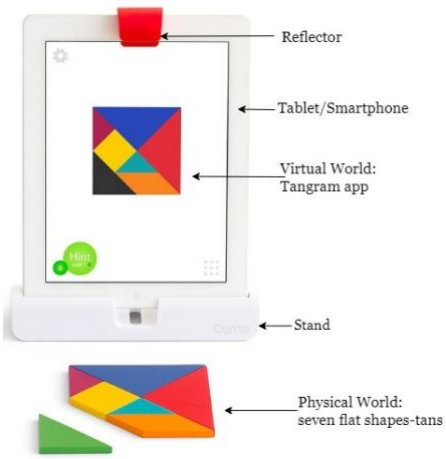


Figure 1 Osmo Setup and Tangram

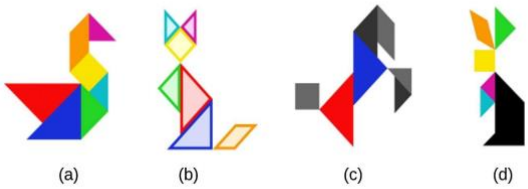


Figure 2 Tangram pieces arranged in shapes. (a) Seven tans arranged in the shape of a swan. (b) Easy level shape presented in the app (virtual world) consists of shape and colour information. (c) Medium level shape presented in the app (virtual world) consists of shape and colour tone information. (d) Hard level shape presented in the app (virtual world) consists of partial shape information

The camera on the tablet captures any physical activity performed in front of the tablet. The captured information is fed back into the virtual world on the tablet, and integrated with added virtual elements in an app. Tangram is played with seven flat shapes, called tans, which are put together to form shapes (Figure 2). The objective of the game is to form a specific shape using all seven pieces, without overlapping each other. A shape is presented to the player on the tablet screen through the Tangram app. The player is expected to arrange the seven flat shapes in physical space to match the shape on the tablet. When the correct flat shape is placed in the right place in the physical space, the corresponding flat shape in the app on the screen is filled with the corresponding colour. This is an indication that the placement of the shape in the physical space is correct. The app provides visual and

audio prompts to the player, as shown in Table 4. PwD are familiar with solving puzzles and thus tangram was an ideal choice amongst the games available with Osmo.

3.1.2 Hololens and Young Conker

HoloLens, AR technology with overlapped physical and virtual spaces is a holographic compu ter in the form of a headset (Figure 3). HoloLens uses the physical world to overlay holograms for the user (who wears the headset) to interact with them, see and hear them within their environment.

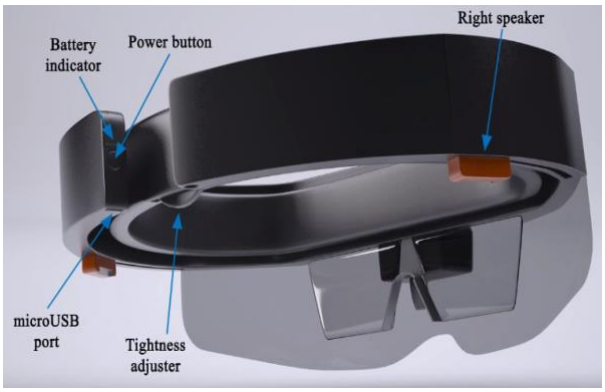


Figure 3 HoloLens - AR technology

Young Conker transforms existing real-world setting into a platform to go on a mystery adventure. The game starts with a scan of the physical space to detect walls and other objects. The game consists of different levels referred in the game as missions. Each mission involves looking for generated holograms in the space. A holographic squirrel character, Conker interacts with the user through speech. The player is required to guide Conker through their gaze movements to the holograms in each mission. The game prompts the player through visual texts, graphics, animations and Conker's speech to accomplish a set mission in the game.

3.2 Data Collection

Ten participants consented for the study who played Tangram on Osmo. Of these, six also played Young Conker on HoloLens. Four participants did not show up for the play session with HoloLens. The study was conducted in three sessions on three different days of the day program. Cognitive impairment of the participant was recorded using MoCA assessment tool in the first session followed by game play with Osmo (Figure 4) and HoloLens (Figure 5) in each of the next two sessions. On the day of the game play with MRT, participants were explained how each technology works and how to play the games on the MRT. Participants were prompted by the researcher when they were unable to proceed with the game play or/and when they asked for help. Each play session lasted for maximum 60 minutes. Although this was a small sample observation study, a 60-minute game play allowed us to

collect enough interactions (actions and prompts) to compare types of prompts and gestures.

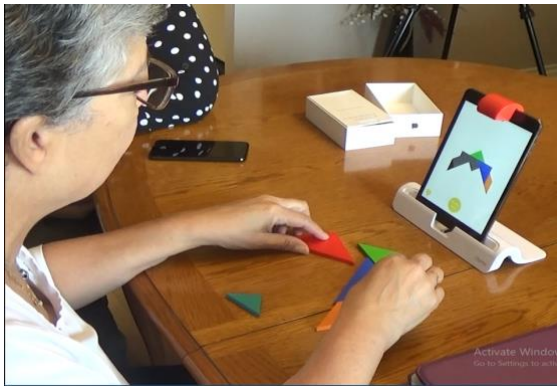


Figure 4 Participant P1_2000 playing a game of Tangram on Osmo

(a) *(b)*

Figure 5 participant P9_2009 playing a game of Young Conker on HoloLens (a) using gaze to move young Conker, the squirrel from one place to another in the game (b) Screen capture of what participant sees through the headset

3.3 Analysis

The video recordings were analysed in Noldus Observer XT 14.0, a software for analysis of observational data that facilitates coding and description of participant behaviour. The videos were coded for two events: prompts presented by technology, opportunities to perform actions on the technology. These coded events were assigned a sub-code. The coding heuristics for the codes and sub-codes is shown in (

Table 1). The coded data was exported to Microsoft Excel and analysed in SPSS. Nonparametric statistical methods were used as small-n data does not comply with the normality tests. Nonparametric methods also do not rely on the estimation of statistical parameters such as mean and standard deviation which makes them suitable for small-n data analysis (Gibbons & Chakraborti, 2011).

Table 1 Coding Heuristics

Codes for an event	Sub-codes for a coded event	Description
Prompts presented		Instant when a prompt is presented to the participant.
	Type of prompts	The coded event, Prompts presented was coded for the type of prompts – *Visual, Audio*.
	Correct use	The coded event, Prompts presented was coded as *Correct use* when the prompt was correctly perceived. This was determined through a combination of verbal protocols and participant behaviour in the videos.
	Incorrect use	The coded event, Prompts presented was coded as *Incorrect use* when the prompt was incorrectly perceived. This was determined through a combination of verbal protocols and participant behaviour in the videos.
	Undetected prompt	The coded event, Prompts presented was coded as *Undetected prompt* when the prompt was presented by the technology but not detected by the participant. This was determined through a combination of verbal protocols and participant behaviour in the videos.
Opportunity for Actions		Instant when the technology presents an opportunity for the user to perform an action is coded as *Opportunity for Action*.
	Correct action	When the user performs a correct action in response to *Opportunity for Action* by the technology, it is coded as *Correct action*.
	Incorrect action	When the user performs a incorrect action in response to *Opportunity for Action* by the technology, it is coded as *Correct action*.
	Type of actions	When the user performed a *Correct action, t*he type of action was coded – *Gestural, Speech*.

The Independent Variables (IVs) for the analysis were *Type of prompts* and *Type of actions*. The total prompts presented by both the technologies, total opportunities to perform actions with each of the technologies, total correct actions performed and total prompts correctly perceived were computed from the coded events. We also coded for incorrect use of prompts and actions and undetected prompts, but only included correct use of prompts

and actions in the analysis to compare the gestures and prompts. The dependent variables for the analysis were: Percentage of prompts correctly perceived and Percentage of actions correctly performed.

4. Results

HoloLens

HoloLens allows two *Types of actions* – Gestural and Speech. Two types of prompts were presented to the participants by the AR technology, HoloLens – Audio and Visual (Table 2).

Table 2 Types of action and prompts in HoloLens

Interaction/Type	Interaction	Description
Prompt/Visual - Graphic		Visual Graphic prompt, a tap icon, prompting to select an object in the virtual world
Prompt/Visual - Text		Visual Text prompt, a tap icon, prompting to select an object in the virtual world
Prompt/Visual Graphic **Prompt/Audio – Animated voice**		Visual Graphic prompt (microphone) to use speech as an action. An Audio prompt in the form of the squirrel's animated voice.
Prompt/Visual Graphic		Visual Graphic prompt (arrows) indicate the direction in which to look for the objects in each of the game missions.
Action/Gestural – Air tap		Gestural action, air tap to select an object in the virtual world. This is derived from the familiar gesture of tapping on a mouse to select in a desktop computer.

Interaction/Type	Interaction	Description
Action/Gestural - Bloom	Step 1 Step 2	A bloom gesture is used to reset HoloLens to Home screen.
Action/Gestural - Gaze		Gaze is used to guide (move) the squirrel, Young Conker in the direction of the object which is being searched in the individual missions.

A comparison between the *Types of actions* - Gestural and Speech and the *Types of prompts* – Audio and Visual in HoloLens was first carried out using nonparametric statistical methods. Of the 6 participants who played Young Conker with HoloLens, all showed high percentage of correct use of gestural actions as compared to speech to interact with the technology. A Wilcoxon signed-rank test determined that there was a statistically significant increase in percentage of correct use of actions when subjects were using Gestural actions (*Mdn* = 80.31%) compared to Speech actions (*Mdn* = 23.61%), $z = -2.201$, $p = 0.028$, $p < 0.050$.

Similarly, analysis for *Type of prompts* perceived correctly using HoloLens found that out of 6 participants, there was higher percentage of correct perception of audio prompts as compared to visual prompts in 5 participants whereas 1 participant used visual prompts more effectively than audio prompts. A Wilcoxon signed-rank test determined that there was a statistically significant increase in percentage of correct perceptions of prompts when subjects were using Audio prompts (*Mdn* = 60%) compared to Visual prompts (*Mdn* = 19.52%), $z = -1.992$, $p = 0.046$, $p < 0.050$.

Further comparative analysis of types of gestural actions and types of visual prompts in HoloLens was carried out. Although audio prompts were more effective than visual prompts, we wanted to know the effectiveness of various visual prompts. Descriptive statistics for these gestures and visual prompts are shown in

Table 3.

Table 3 Descriptive statistics for Types of gestural actions and visual prompts

	Percentage of correct actions and perceptions		
Type of Gestures and Visual Prompts in HoloLens	**Mean (%)**	**Median (%)**	**Standard Deviation** (%)
Gestural - Gaze	76.52	74.50	6.43
Gestural - Bloom	18.33	0.0	28.57
Gestural - Air tap	76.47	92.96	37.85
Visual Prompts - text	36.48	44.43	18.75
Visual Prompts - Graphics	14.17	16.25	8.60

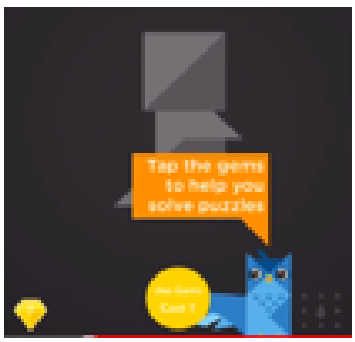
Gestural – Air tap was most effective while Gestural – Bloom was most ineffective with PwD. Visual text prompts were more effective in terms of correct perception than visual graphic prompts. A Friedman test was run to determine if there were differences in percentage of correct actions with the following types of gestures: Gestural-Gaze, Gestural-Bloom and Gestural-Air tap. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons (SPSS Statistics, 2016). Percentage of correct actions were statistically significantly different for the types of gestures, $\chi^2(2) = 8.96$, $p = 0.011$ ($p < .050$). Post hoc analysis revealed statistically significant differences in percentage of correct gestures from Gestural-Bloom (*Mdn* = 0.0%) to Gestural-Air tap (*Mdn* = 92.96%) ($p = 0.012$, $p < 0.05$), but not Gestural-Bloom & Gestural-Gaze and Gestural-Gaze & Gestural-Air tap.

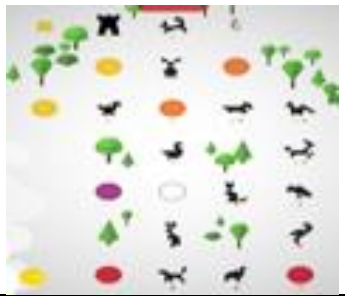
A comparison between the two types of visual prompts – Text and Graphic revealed that out of the 6 participants who played Young Conker with HoloLens, 5 participants showed high percentage of correct perception of visual text prompts as compared to visual graphic prompts. One participant used visual graphic prompts more effectively than the text prompts. A Wilcoxon signed-rank test determined that there was a statistically significant increase in percentage of correct perception when participants were using visual text prompts (*Mdn* = 44.43%) compared to graphic prompts (*Mdn* = 16.25%), $z = -1.992$, $p = 0.046$, $p < 0.050$.

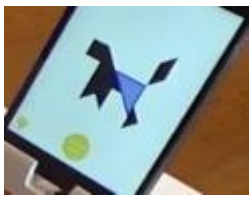
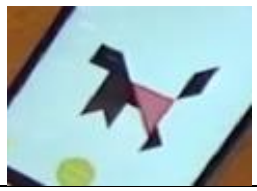
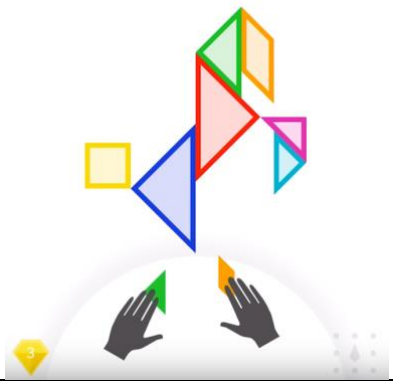
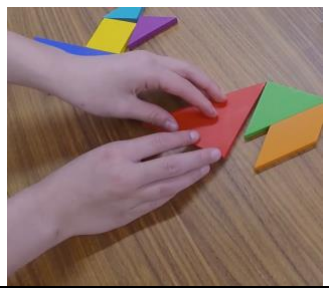
Osmo

Osmo offered only Gestural actions as a modality to interact with Osmo (Percentage of correct Gestural actions: Mean = 70.06%, Median = 69.13%, Std Deviation = 5.53). Two types of prompts were presented to the participants – Audio and Visual (Table 4).

Table 4 Types of action and prompts in Osmo

Interaction/Type	**Interaction**	**Description**
Prompt/Visual		Visual properties of the blocks – shape and size were used to put the pieces together in a puzzle.
Prompt/Visual – Animation Prompt/ Visual - Graphic		A Visual - animation of coins emerging from a gem on completing a puzzle indicate to the players that they have earned coins. These coins could be used by the players to release more prompts in the game. A Visual – graphic (symbolic) prompt of a tick metaphorically relates to a tick on a list of things to do meaning 'it is done'. In the game, the tick means that the player has successfully completed the level.
Prompt/ Audio Prompt/ Visual – Text Prompt/ Visual - Graphical		Audio prompt in the form of an owl in a female voice, prompts the player to use the gems collected during the game play to release a prompt. This audio prompt is complemented with a Visual – Text prompt in a speech bubble.

Interaction/Type	**Interaction**	**Description**
		Visual – Graphical prompt with a text prompt in the form of a number inside a gem, indicates number of gems available to the player. Visual – Graphical prompt complemented with a Visual – Text prompt in the form of a yellow circle with a text "Use Gems Cost #", where # is the number of gems required to release a prompt for the current game, prompts the player to use gems to release more prompts to solve the Tangram puzzle.
Prompt/ Visual - Graphical		Visual – Graphical prompt in the form of yellow, orange and red ellipses representing easy, medium and hard levels respectively. Purple represents the highest level, which the participants did not play in the study.
Prompt/ Visual – Text Prompt/Visual - shapes		Visual – Text 'flip' prompts players to turn over the physical orange tangram piece when it is not aligned correctly with the rest of the pieces. Visual – shapes represent prompts that include shapes filled with colour, shapes with coloured outlines.
Prompt/ Audio	Music	Audio prompt in the form of a tone prompts players about the progress of the puzzle solving task as they put the pieces together.

Interaction/Type	**Interaction**	**Description**
Prompt/ Visual – Flickering	 	Visual – Flickering prompt in the form of flickering red and blue colour for a block prompts the players to use either of the two colours for the same shape in that position.
Prompt/Visual - Animations		Visual – Animation prompt shows how to complete the puzzle. Two hands emerge from the bottom of the screen showing how to place each piece in the puzzle.
Action/Gestural		Gestural actions performed through manipulations of the Tangram pieces.

A comparison between the two Types of prompts – Audio and Visual in Osmo was first carried out using nonparametric statistical methods. Of the 10 participants who played Tangram, there was higher percentage of correct perception of audio prompts as compared to visual prompts in 8 participants whereas 2 participants used visual prompts more effectively than audio prompts. A Wilcoxon signed-rank test determined that there was a statistically significant increase in percentage of correct perceptions of prompts when subjects were using Audio prompts (*Mdn* = 20.68%) compared to Visual prompts (*Mdn* = 10.76%), $z = -2.090$, $p = 0.037$, $p < 0.050$.

Further comparative analysis of types of Gestural actions, Audio prompts and Visual prompts in Osmo was carried out. Descriptive statistics for these gestures and prompts are shown in Table 5.

Table 5 Descriptive statistics for Types of gestural actions and prompts

	Percentage of correct actions and perceptions		
Type of Gestures and Prompts in Osmo	Mean (%)	Median (%)	Standard Deviation (%)
Gestural – Object Manipulation	90.0	90.51	3.00
Gestural – Touchscreen swipe	54.3	63.33	29.47
Gestural – Touchscreen tap	39.53	40.0	3.90
Audio Prompts – Human voice	82.40	91.15	29.50
Audio Prompts – Music tone	11.64	10.14	7.23
Visual Prompts - Animations	15.92	16.10	1.70
Visual Prompts - Shapes	38.14	38.63	2.19
Visual Prompts – Shapes flickering	3.27	3.27	0.53
Visual Prompts - Text	25.80	25.17	2.48
Visual Prompts - Graphic	5.14	5.36	1.02

A Friedman test was run to determine if there were differences in percentage of correct actions with the following types of gestures: Gestural-Object manipulation, Gestural-Touchscreen swipe and Gestural-Touchscreen tap. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons (SPSS Statistics, 2016). Percentage of correct actions were statistically significantly different for the types of gestures, $\chi^2(2) = 16.80$, *p = 0.0 (p* < .050). Post hoc analysis revealed statistically significant differences in percentage of correct gestures between Gestural-Object manipulation (*Mdn* = 90.51%) and Gestural-Touchscreen swipe (*Mdn* = 63.33%) ($p = 0.022$, $p < 0.05$) and Gestural-Touchscreen tap (*Mdn* = 40.0%) ($p = 0.0$, $p < 0.05$), but not Gestural-Touchscreen swipe & Gestural-Touchscreen tap.

A comparison between the two types of audio prompts – human voice and music tone, revealed that out of the 10 participants who played Tangram with Osmo, 9 participants showed high percentage of correct perception of human voice as compared to music tone prompts. One participant used music tones more effectively than the human voice prompts. A Wilcoxon signed-rank test determined that there was a statistically significant increase in percentage of correct perception when participants were using human voice prompts (Mdn = 91.15%) as compared to music tones (Mdn = 10.14%), $z = -2.7$, $p = 0.007$, $p < 0.050$.

A Friedman test was run to determine if there were differences in percentage of correct perception with the visual prompts- animations, shapes, flickering shapes, graphic and text. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons (SPSS Statistics, 2016). Percentage of correct perceptions were statistically significantly different for the types of visual prompts, $\chi^2(4) = 39.28$, *p = 0.0 (p* < .050). Post hoc analysis revealed statistically significant differences in percentage of correct perceptions between Visual-animations (*Mdn* = 16.10%) and Visual-shapes (*Mdn* = 38.63%) ($p = 0.047$,

$p < 0.05$), Visual-flickering shapes (*Mdn* = 3.27%) and Visual-shapes (*Mdn* = 38.63%) ($p = 0.0$, $p < 0.05$), Visual-flickering shapes (*Mdn* = 3.27%) and Visual-text (*Mdn* = 25.17%) ($p = 0.0$, $p < 0.05$), Visual- shapes (*Mdn* = 38.63%) and Visual-graphic (*Mdn* = 5.36%) ($p = 0.0$, $p < 0.05$), Visual- text (*Mdn* = 25.17%) and Visual-graphic (*Mdn* = 5.36%) ($p = 0.03$, $p < 0.05$), but not between Visual-animations and Visual-flickering shapes, Visual-graphic and Visual-flickering shapes, Visual-graphic and Visual-animations, Visual-animations and Visual-text, Visual-shapes and Visual-text.

5. Discussion

The study found that PwD can use gestural actions correctly and successfully. These gestures could range from the ones that are familiar (e.g. touchscreen tap) to others that are completely unfamiliar (e.g. air tap). Participants found the game of Tangram on Osmo easy to use, one participant expressed that the game was for children and wanted to play a game for grownups.

> "this game looks like for children, I want to play a game for grownups". P10

Participant P10 went on to play longer than the maximum allocated play time for the study. This is an indication that the participant found the game easy to use and was engaged in the game.

The air tap gesture was most effective in HoloLens, while PwD used object manipulation effectively with Osmo. Dementia can affect parts of the brain that control language (Bayles, 1982) which supports effectiveness of non-verbal forms of responses such as gestures. Communicative functions are impaired, specifically in semantic systems. The semantic capacities are dissociated from the systems of syntax and phonology. This means that PwD have difficulties in forming complete sentences which acts as a barrier in using speech as a modality to perform actions on the technology. HoloLens employs single word commands for speech interaction such as 'select', 'stop', etc. It can be programmed for specific commands to be used in an application. However, this means that PwD are expected to learn new commands and remember them short term as well as long term for ongoing use. Dementia is associated with decline and/or loss of memory and reasoning and this is due to both an impairment in the short term memory and additional difficulty in creating new memories in long term memory (Miller, 1973). This means that PwD would find it difficult to remember the commands resulting in them abandoning the technology altogether.

The participants used air tap gesture in HoloLens to select virtual holograms in the application. The technology requires the index finger to be within the line of view of the headset cameras for the gesture to be tracked by the technology. Participants had to be verbally prompted by the researcher to keep the index finger in front of their face. The gaze gesture in HoloLens is used to move virtual holographic young conker from one place to another. This is embodied in the act of looking around for something in the physical environment. In the virtual environment, it in turn results in the act of looking for virtual objects and in turn move the cursor for object selection. Embodied interactions are intuitive

in nature and result in successful interactions (Desai et al., 2019). Beaudouin-Lafon (2000) suggested that if the extent of similarity between the user actions on a domain object and the resulting response (degree of conformance) are kept higher, it could result in successful and correct interactions. With the gaze gesture in HoloLens, the movement of the head and the eye gaze results in the movement of the virtual object in the same direction in the virtual world.

PwD used three gestures with Osmo, object manipulation and touchscreen tap and swipe gestures. In object manipulation, people were using the natural and material properties of the physical objects, shape, size and colour, as prompts to put the puzzle pieces together. In other words, they were using physical affordances of the objects to determine how they may interact with it and what manipulations may be possible with the objects. People use their sensorimotor knowledge to decide on these actions and manipulations, which is the most accessible form of knowledge in humans and is established early in life (Blackler, Desai, McEwan, Dieffenbach, & Popovic, 2019). Gibson (1979b) emphasized that affordances depend on an individual's ability to perform actions in the environment. This means that although people will unconsciously look for physical affordances to interact with the objects, the interactions with the same objects could lead to different actions and manipulations in children, older adults and PwD depending on their abilities and sensorimotor knowledge.

The touch interactions were effective, contributing to 40 - 65% of the total presented opportunities. Touch screens allow more than one person to interact with the interface, both visually as well as by touch. Instead of limiting the interaction to one person through single control device such as a mouse, PwD become equal partners in the interaction with the researcher in the study but with their care givers in the larger context (Joddrell & Astell, 2019). The touch screen tap was the least effective, but this was due to lack of affordances in coloured circles in the app to tap on them to select the level of game. This prompt either went unnoticed or PwD used incorrect gesture with it. PwD could not interpret the expectation to shift their attention from the physical world of object manipulation to the virtual world and tap on the circles. MRTs should offer affordances to shift attention between the virtual and physical worlds. Since people are expected to interact with the physical and virtual elements, both at the same time, participants often were unable to determine when they should interact with the virtual elements and when with the physical objects.

The study found that audio prompts were more effective in correct perception of the prompts in comparison to the visual prompts for both technologies, despite of the fact that the total number of visual prompts presented by the technologies were much higher than the audio prompts. PwD were unable to understand the visual prompts. Many visual prompts remained unnoticed. Visual cognition which involves ability to determine on the basis of the retinal input, the presence of particular shapes, configurations of shapes,

objects, scenes, and their properties are before us, is impaired in PwD (Pinker, 1984). This supports our observation that some visual prompts were not noticed by PwD.

Animated audio prompts were most effective in HoloLens, followed by visual text and then visual graphics. Audio prompts in human voice were most effective in Osmo followed by music tones and then with animated voice. Participants expressed their dislike towards the animated voice of Young Conker in HoloLens,

> "I don't like to talk to this guy [young conker]?" P3
>
> "I don't like this voice…it is too much for my ears" P6.

Participants preferred the human voice in Osmo but had no problems with the visual character of an owl in Osmo. Effectiveness of audio prompts and specific preference to human voice could be due to familiarity of receiving instructions from their care givers at home and in day programs. All participants in the study attended day programs for varying lengths.

Shapes with coloured outline, filled with colour and filled with black and grey colours, were most successfully perceived followed by text prompts. This is despite impairments in visual cognition that make it difficult for PwD to recognize shapes and colours and reason and remember their properties to inform decision making process. We believe that this success is credited to the action of object manipulation that accompanies visual perception process in Osmo Tangram game. It was also observed that participants would often not look up at the tablet for the visual cues but rely on the object manipulation to figure out the right arrangement. Animations and flickering shapes were most ineffective in HoloLens. Dementia affects attention, working memory and visual-perceptual ability to varying extents depending on the type of dementia (Calderon et al., 2001). People with certain types of dementia are also prone to visual hallucinations. This could explain participants' lack of attention to visual animations and flickering shapes.

Participants read all text prompts presented to them by both technologies. One participant said while playing Young Conker in HoloLens,

"They should write what we have to do." P5

However, words and language when using text prompts or audio prompts should be simple, that does not require PwD to reason the meaning. Use of population stereotypes, idioms and metaphors requiring access to past experience and knowledge are difficult to perceive for PwD due to decline or loss of memory and reasoning. The use of word 'Flip' to turn over the orange piece in the tangram puzzle was confusing for all the participants. The researcher had to intervene with the word 'Turn Over' which the participants immediately understood

the meaning and translate it into a correct object manipulation. Similarly, colour coded circles representing level of the puzzle could not be deciphered by PwD.

Design Implications

The findings from this study have defined use of interaction modalities in the design and development of MRTs for PwDs, resulting in a design framework shown in **Error! Reference source not found.**.

Actions:

The Actions layer (at the bottom of Interaction Modalities) represents actions that should be used in MRTs for PwDs. Gestures, including those that PwD are unfamiliar with, are effective in generating correct actions. The actions should demonstrate high degree of conformance, that is the action in the physical world should result in the same functionality in the virtual world as in the physical world. Embodied actions such as object manipulations are effective with PwD. Other forms of embodied actions such as full body interactions (e.g. bowling) should be explored in future studies. In the context of IADL, effective correct actions mean that PwDs have control and independence in carrying out the activities instead of them following instructions from the technology.

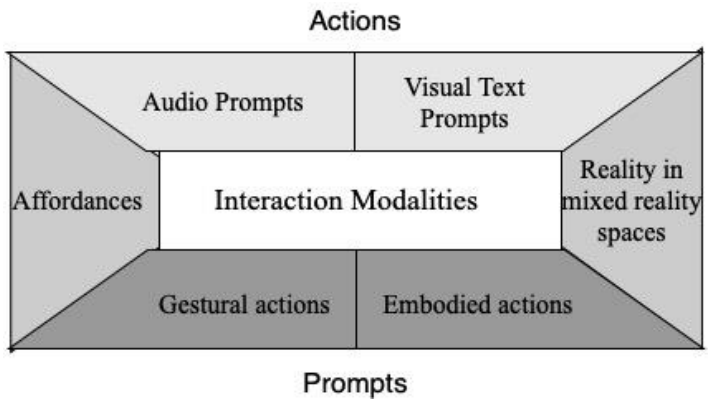


Figure 6 Framework for interaction modalities in MRTs for PwDs

Prompts:

The prompts layer (at the top of Interaction Modalities) represents modalities of prompts that should be generated in MRTs to provide clues to PwDs to generate appropriate actions. Audio prompts in human voice are more effective over visual prompts of all forms. Amongst visual prompts, text prompts were desirable and preferred by PwD. Visual prompts and possibility of complementing visual with audio prompts needs to be investigated in future research. In the context of IADL, correct perception and interpretation of prompts means that PwD remain on track in executing tasks in an activity. This in turn would ensure that PwD adopt and accept these technologies in their everyday use. The effectiveness of the

prompts and the actions depend on the sense of reality in physical and virtual worlds and affordances.

Reality:

In AV technologies, specific prompts or affordances may be required to shift attention between physical and virtual worlds. In general, PwDs should be aware of the world in which they are acting and perceiving, irrespective of the coupling mechanism, whether AR or AV (Benyon, 2012). This ensures that PwD respond to the correct prompts that are presented to them and in turn respond with the correct actions. Segregating reality from virtuality in IADL activities is important for the safety of PwD in some activities, but also to determine in which world they should be executing actions and perceiving.

Affordances:

Affordances play an important role in execution of actions as well as perception of prompts. Metaphors or idioms or population stereotypes are difficult for PwD to process and reason the meaning out of them. Physical affordances on the other hand offer natural and intuitive form of interaction for PwD (Blackler, Li-Hao, Desai, & Astell, 2020). The speed at which prompts are perceived and actions executed, will determine ongoing engagement of PwDs in IADL.

6. Conclusions

The study has found that PwD interact with MRT effectively through gestural interactions and audio prompts in human voice. Affordances and embodied actions contribute to the success of these actions and prompts. Text prompts that used simple and direct language were the most effective prompts amongst visual prompts. MRTs should incorporate affordances that allow PwD to differentiate interactions with virtual and physical elements. These findings are significant as they will drive further human centred studies into understanding prompting mechanisms for technology development for PwD. The outcomes could be applied to other interactive systems for PwD. Future research will investigate audio and visual prompts, specifically how to complement these prompts with gestural actions for effective interactions for PwD with MRTs.

7. Acknowledgements

This research was supported by AGE-WELL Catalyst grant. We are very thankful to all participants and the staff at Alzheimers Society of Durham and Memory and Company for facilitating recruitment of participants and data collection. We thank Vishakha for providing logistic support in data collection.

8. References

Aruanno, B., Garzotto, F., & Rodriguez, M. C. (2017). HoloLens-based mixed reality experiences for subjects with Alzheimer's disease. *ACM International Conference Proceeding Series*, *Part F1313*.

https://doi.org/10.1145/3125571.3125589

Aruanno, B., Garzotto, F., Torelli, E., & Vona, F. (2018). Holo learn: Wearable mixed reality for people with neurodevelopmental disorders (NDD). *ASSETS 2018 - Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, (1), 40–51. https://doi.org/10.1145/3234695.3236351

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, *21*(6), 34–47. https://doi.org/10.1109/38.963459

Barban, F., Annicchiarico, R., Pantelopoulos, S., Federici, A., Perri, R., Fadda, L., … others. (2016). Protecting cognition from aging and Alzheimer's disease: a computerized cognitive training combined with reminiscence therapy. *International Journal of Geriatric Psychiatry*, *31*(4), 340–348.

Bayles, K. A. (1982). Language function in senile dementia. *Brain and Language*, *16*(2), 265–280.

Beaudouin-Lafon, M. (2000). Instrumental interaction: an interaction model for designing post-WIMP user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 446–453.

Benyon, D. (2012). Presence in blended spaces. *Interacting with Computers*, *24*(4), 219–226. https://doi.org/10.1016/j.intcom.2012.04.005

Bewernitz, M. W., Mann, W. C., Dasler, P., & Belchior, P. (2009). Feasibility of machine-based prompting to assist persons with dementia. *Assistive Technology*, *21*(4), 196–207.

Blackler, A., Desai, S., McEwan, M., Dieffenbach, S., & Popovic, V. (2019). Perspectives on the nature of intuitive interaction. In Blackler (Ed.), *Intuitive Interaction: Research and Application*. CRC Press.

Blackler, A., Li-Hao, C., Desai, S., & Astell, A. (2020). Intuitive Interaction Framework in User-product Interaction for People Living with Dementia. In G. Brankaert, Rens, Kenning (Ed.), *HCI and Design in the Context of Dementia*.

Blackler, A., Popovic, V., & Mahar, D. (2010). Investigating users' intuitive interaction with complex artefacts. *Applied Ergonomics*, *41*(1), 72–92. https://doi.org/10.1016/j.apergo.2009.04.010

Buiza, C., Soldatos, J., Petsatodis, T., Geven, A., Etxaniz, A., & Tscheligi, M. (2009). HERMES: Pervasive computing and cognitive training for ageing well. *International Work-Conference on Artificial Neural Networks*, 756–763.

Calderon, J., Perry, R. J., Erzinclioglu, S. W., Berrios, G. E., Dening, Tr., & Hodges, J. R. (2001). Perception, attention, and working memory are disproportionately impaired in dementia with Lewy bodies compared with Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *70*(2), 157–164.

Carlomagno, S., Pandolfi, M., Marini, A., Di Iasi, G., & Cristilli, C. (2005). Coverbal gestures in Alzheimer's type dementia. *Cortex*, *41*(4), 535–546.

Chang, Y.-J., Chou, L.-D., Wang, F. T.-Y., & Chen, S.-F. (2013). A kinect-based vocational task prompting system for individuals with cognitive impairments. *Personal and Ubiquitous Computing*, *17*(2), 351–358.

Czaja, S. J. (1997). Computer technology and the older adult. In *Handbook of human-computer interaction* (pp. 797–812). Elsevier.

Desai, S., Blackler, A., & Popovic, V. (2016). Intuitive interaction in a mixed reality system. *Design Research Society*, 16.

Desai, S., Blackler, A., & Popovic, V. (2019). Children's embodied intuitive interaction--Design aspects of embodiment. *International Journal of Child-Computer Interaction*.

Gamberini, L., Alcaniz, M., Barresi, G., Fabregat, M., Ibanez, F., & Prontu, L. (2006). Cognition, technology and games for the elderly: An introduction to ELDERGAMES project. *PsychNology Journal*, *4*(3), 285–308.

Garzotto, F., Torelli, E., Vona, F., & Aruanno, B. (2019). HoloLearn: Learning through mixed reality for people with cognitive disability. *Proceedings - 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality, AIVR 2018*, 189–190. https://doi.org/10.1109/AIVR.2018.00042

Gibbons, J. D., & Chakraborti, S. (2011). Nonparametric Statistical Inference. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 977–979). https://doi.org/10.1007/978-3-642-04898-2_420

Gibson. (1979a). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum.

Gibson. (1979b). *The Theory of Affordances in the Ecological Approach to Visual Perceptual*. Houghton Mifflin.

Gündogdu, R., Bejan, A., Kunze, C., & Wölfel, M. (2017). Activating people with dementia using natural user interface interaction on a surface computer. *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 386–394.

Huber, S., Berner, R., Uhlig, M., Klein, P., & Hurtienne, J. (2019). Tangible Objects for Reminiscing in Dementia Care. *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 15–24.

Joddrell, P., & Astell, A. J. (2019). Implementing Accessibility Settings in Touchscreen Apps for People Living with Dementia. *Gerontology*, 1–11.

Lawry, S., Popovic, V., Blackler, A., & Thompson, H. (2019). Age, familiarity, and intuitive use: An empirical investigation. *Applied Ergonomics*, *74*, 74–84. https://doi.org/https://doi.org/10.1016/j.apergo.2018.08.016

Loureiro, B., & Rodrigues, R. (2011). Multi-touch as a natural user interface for elders: A survey. *6th Iberian Conference on Information Systems and Technologies (CISTI 2011)*, 1–6.

Maier, J. (2015). On the computability of affordances as relations. *AI EDAM*, *29*(Special Issue 03), 249–256. https://doi.org/10.1017/S0890060415000207

Mihailidis, A., Boger, J. N., Craig, T., & Hoey, J. (2008). The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study. *BMC Geriatrics*, *8*(1), 28.

Miller, E. (1973). Short- and long-term memory in patients with presenile dementia (Alzheimer's disease). *Psychological Medicine*, *3*(2), 221–224. https://doi.org/10.1017/S003329170004856X

Orpwood, R., Adlam, T., Evans, N., Chadd, J., & Self, D. (2008). Evaluation of an assisted-living smart home for someone with dementia. *Journal of Assistive Technologies*, *2*(2), 13–21.

Pigot, H., Mayers, A., & Giroux, S. (2003). The intelligent habitat and everyday life activity support. *Proc. of the 5th International Conference on Simulations in Biomedicine, April*, 2–4.

Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, *18*(1–3), 1–63. https://doi.org/10.1016/0010-0277(84)90021-0

Regenbrecht, H., Lum, T., Kohler, P., Ott, C., Wagner, M., Wilke, W., & Mueller, E. (2004). Using Augmented Virtuality for Remote Collaboration. *Presence: Teleoperators and Virtual Environments*, *13*(3), 338–354. https://doi.org/10.1162/1054746041422334

Ro, H., Park, Y. J., & Han, T.-D. (2019). *A Projection-based Augmented Reality for Elderly People with Dementia*. Retrieved from http://arxiv.org/abs/1908.06046

SPSS Statistics. (2016). *SPSS Statistics*. Armonk, NY: IBM Corp.

Vovk, A., Chan, D., & Patel, A. (2019). Augmented reality for early Alzheimer's disease diagnosis. *Conference on Human Factors in Computing Systems - Proceedings*, 1–6.

https://doi.org/10.1145/3290607.3313007

About the Authors:

Shital Desai is an Assistant Professor in Interaction Design at York University where she heads SaTS lab. Her research focusses on role of interactive and emerging technologies in supporting people with accessibility needs, using human centred and systems design methods.

Alethea Blackler is Discipline Leader for Industrial Design, Interaction Design, Visual Communication and Fashion Design within the School of Design at QUT. Her principle area of research interest is intuitive interaction, in which she is one of the world leaders, older people and technology and design for dementia (including an international ERC Horizon 2020 grant project). She has worked with partners in government, community and industry on various projects and has had three Australian Research Council grants. She has published more than 80 peer reviewed papers, been invited to give presentations internationally and is the recipient of several awards, including a 2018 ig-nobel.

Deborah Fels has a PhD (1994) in Human Factors from Industrial Engineering at the University of Toronto. She is currently employed as a professor in the Ted Rogers School of Information Technology Management, and the Director of the Inclusive Media and Design Centre at Ryerson University. Her research interests involve inclusive design, access to media and technology for people with disabilities and older adults, and inclusive business. She has published over 150 articles on inclusive technologies and applications and received three patents.

Arlene Astell is a psychologist and currently Professor of Neurocognitive Disorders at the University of Reading, UK and Associate Professor at the University of Toronto, Canada, where she directs the Dementia, Aging and Technology Engagement (DATE) lab.