

Analyzing User's Task-Driven Interaction in Mixed Reality

Styliani Kleanthous
Cyprus Center for Algorithmic Transparency
Open University of Cyprus
RISE - Research Centre on Interactive media, Smart
systems and Emerging Technologies
Nicosia, Cyprus
styliani.kleanthous@ouc.ac.cy

Elena Matsi
Department of Computer Science
University of Cyprus
Nicosia, Cyprus
ematsi02@cs.ucy.ac.cy

ABSTRACT

Mixed reality (MR) provides exciting interaction approaches in several applications. The user experience of interacting in these visually rich environments depends highly on the way the user perceives, processes, and comprehends visual information. In this work we are investigating the differences between Field Dependent – Field Independent users towards their interaction behavior in a MR environment when they were asked to perform a specific task. A study was conducted using Microsoft HoloLens device in which participants interacted with a popular HoloLens application, modified by the authors to log user interaction data in real time. Analysis of the results demonstrates the differences in the visual processing of information, especially in visually complex environments and the impact on the user's interaction behavior.

CCS CONCEPTS

• CCS → Human-centered computing → Human computer interaction (HCI) Interaction techniques; HCI theory, concepts and models

KEYWORDS

Mixed Reality; visual behavior; task-driven interaction; human cognitive differences; field dependence-independence

ACM Reference format:

Styliani Kleanthous, Elena Matsi. 2019. Analyzing User's Task-Driven Interaction in Mixed Reality. In *Proceedings of ACM International Conference on Intelligent User Interfaces (IUI'2019)*. ACM, Marina del Rey, CA, USA, 5 pages.
<https://doi.org/10.1145/3301275.3302286>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

IUI '19, March 17–20, 2019, Marina del Rey, CA, USA
© 2019 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-6272-6/19/03..\$15.00
<https://doi.org/10.1145/3301275.3302286>

1 Introduction

Mixed reality (MR) provides exciting interaction approaches in several application scenarios e.g. automotive, education, manufacturing, design. This new paradigm in technological development brought experiences and challenges unknown before in the human computer interaction spectrum [1]. MR as a term is used in the literature to reflect the union of the real with the virtual world. MR can be used in a wide variety of areas. Applications already exist in marketing, entertainment, education, engineering, and design [2] among others. According to Azuma [3], MR must i) Combine real and virtual content: You can see both at the same time. ii) Interact in real time: Allow virtual content to interact with each other, iii) Represent in 3D, virtual objects can be firmly displayed in place within the space

The development of highly interactive applications, usually in 3D settings, opens new possibilities of natural interaction with the user's virtual and physical surroundings, through heavy flow of visual information [2]. The user experience of interacting in these visually rich environments depends highly on the way the user perceives, processes, and understands visual information [3]. Although, the concept of performing various types of tasks through MR seems exciting, we need to understand how the cognitive and visual differences that exist between people, may affect their interaction behavior hence the design of MR applications.

In this work we would like to explore the interaction of the user in these visually rich environments with respect to their cognitive characteristics. Previous research has already looked at the individual differences among users and particularly cognitive styles [4]. An established, widely researched and validated cognitive aspect is the construct of Field Dependence – Independence (FD-I) [5]. FD-I is mainly considered to capture the differences in visual perception or comprehension of information in a visually complex and demanding task [6, 7]. Consequently, FD-I construct classifies people as Field Dependent (FD), if they demonstrate a more holistic approach when processing information and appear to have difficulties in identifying simple objects in visually complex surroundings; and Field Independent (FI), when they follow a more analytical way of processing information, capture easily detailed information and are able to easily extract simple objects in visually complex environments.

Hence, the importance of this construct to MR interaction settings.

Cognitive styles [7], perceptual issues [6], processing of visual information [4] have long been researched in 2D settings. FD-I in 2D scenarios is associated to the visual perceptiveness of individuals with FD users taking longer to identify a simple figure that is embedded in a complex surrounding compared to a FI [4, 7]. Thus, FD users compared to FIs find it more difficult to abstract relevant information from visual or even textual material. This is true usually in more complex and difficult tasks also in 3D environments [8]. Furthermore, FIs demonstrate a more structured way for information searching based on recent eye-tracking studies, performed in simple task performance [9], cultural heritage [8] and user authentication scenarios [10], compared to FDs who have a more unstructured approach to visual search.

Existing research can benefit from an investigation of the individual differences in human cognitive abilities and how these affect the way individuals perceive and interact with visual holographic content in 3D environments in order to inform further HCI research [8, 10]. Understanding the cognitive and visual behavior of users in these upcoming and visually complex environments will allow for designing and developing applications that would take into consideration the diversity in visual information processing among users.



Figure 1 Executing the study (GEFT & HoloLens)

In this research we are aiming to investigate the differences between FD-I users towards their behavior and interaction in a mixed reality environment when they were asked to perform specific tasks with varied levels of difficulty. The contribution of this paper lays primarily on i) bringing new knowledge on FD-I differences in an emerging interface spectrum (e.g. MR environment), ii) pointing at important differences in the interaction of the user in these interfaces and how interaction is affected by the user's cognitive characteristic of FD/I and the difficulty of the task. The rest of the paper will describe a user study which is aiming to address the objectives of this paper in the form of hypotheses followed by a results section. Discussion on the findings will conclude the paper.

2 User Study

The study presented in this paper, aligned with our research goal, is an initial attempt to examine the following null hypotheses:

H01 There is no interaction effect between FD-I and difficulty level towards the time spent for achieving each target object.

H02 There is no interaction effect between FD-I and difficulty level towards the gaze behavior for achieving each target object.

Gaze behavior through Microsoft HoloLens has been selected to be examined in this work. Microsoft HoloLens uses head orientation as the primary method for interacting in MR applications, while a gazed control dot allows you to 'point' at objects. However, in a visually rich environment, the ability of the user to locate objects and perform actions depends highly on the user's perception of the visual cues. Hence, we are also interested in the time that a user needs to perform an action according to the cognitive aspect of FD-I. Interaction effects are being examined since we wanted to understand how the difficulty level of a task and the differences in the cognitive aspects of FD-I affects the interaction of the user towards the abovementioned variables.

Apparatus: Participants were asked to use Microsoft HoloLens Mixed Reality smart glasses (Figure 1) that allow the user to interact with holographic objects. HoloLens features holographic lenses, sensors that capture information on the user's interaction and the surrounding environment, built in speakers and speech recognition.

To measure the visual behavior of the users, we logged gaze interaction data (as provided by Microsoft HoloLens) and video recordings of users' interaction (Figure 2). HoloLens allows for interaction gaze data to be collected as the points the user has gazed at, since the primary method for deriving an object is through user's gaze. Gaze data collected by our logger as points of interest, thus the number of items the user gazed at (e.g. how many items the user gazed during his/her searching session prior to selecting Uranus). Furthermore, a video recording of user's interaction allowed us to view exactly what the user was seeing through HoloLens (Figure 2). Videos were recorded for all users and provided us with supplementary qualitative data on their interaction behavior that we used for evaluating the logged data.

Galaxy Explorer MS HoloLens Application: We selected the Galaxy Explorer HoloLens app¹, that allows the user to explore the Milky Way Galaxy in a full 3D holographic model via HoloLens, along with Earth's solar system and listen to some interesting information and facts about each object in our galaxy. This is an open source application allowing us to integrate the logger we developed for collecting interaction data through the app (gestures, gaze, speech, timestamps of events, hand location). Finally, the concept of Galaxy Explorer was voted on as the most popular application idea for the HoloLens during a contest held in 2016 at the Build 2016 conference, so we

¹ <https://www.microsoft.com/en-us/store/p/galaxy-explorer/9nblggh4q4jg>

considered this as a popular application on HoloLens that would be easily accessible by the general public around the world.

2.1 Procedure

Participants. Thirty-one subjects (17 males and 14 females), ranging in age from 20 to 29 (Mean: 22.35; SD: 1.743), took part in this study. Participants were recruited by announcements on specific groups on social media and email

requests at our University. The minimum requirements for participating were: i) not to have taken the GEFT before, ii) have no vision problems that would incur their interaction and iii) have no previous experience with HoloLens.

GEFT & Training Sessions. At arrival participants were given a short questionnaire collecting demographic information and the Group Embedded Figures Test (GEFT) (Figure 1). FD-I is determined using the GEFT tool which was designed by Oltan et al. [5]. It is a 'paper-and-pencil' test and is a time-administered tool (15 minutes long) in which users were asked to identify simple shapes within complex figures of diverse complexity. This is a validated tool for identifying FD-I differences in humans. The classification of participants into FD or FI is based on a threshold score and for the purposes of this study, we used a threshold score of 11 correct items, which is aligned with what is used extensively in practice [4,8,9]. The participants who scored 11 or lower were classified as FD, and those who scored 12 to 18 classified as FI.

Participants' scores in our sample ranged from 1 to 17 (Mean:12.26; SD: 4.058). In our sample 20 participants classified as FI and 11 participants as FD. It is important to mention that the frequencies of participant's GEFT scores in our sample are similar to general public GEFT scores [8-9].

To ensure that all participants had equal exposure to the basic operations of HoloLens (e.g. Gaze control, Gestures - Air Tap, Tap and Hold, Grab, Zoom, Tilt, Speech, Bloom) we asked participants to perform an approximately 10 minutes long training task. Each participant had to open the Holograms App through HoloLens and select any holograms of their preference, place the selected holograms anywhere in the space and manipulate those using the HoloLens operations.

The study design, environment and procedure maintained the same for all participants. Participants indicated that the training session was sufficient for them to feel comfortable in using the relevant operations and execute the task. Though, participants needed some time and adjustments before they physically feel comfortable with the HoloLens as a device.

Task execution Session. At this point participants had to follow certain steps to locate three target objects within the Galaxy Explorer application environment. Each target object had a different difficulty level (Easy, Moderate, Difficult) in terms of searching and locating (gaze behavior). However, the users were not aware of this classification. Screenshots from in-app videos are presented in Figure 2.

Uranus planet was the most difficult object to be located since it required the user to move his/her head or even body in the space and search for it in the holographic planetary system; Sagittarius A* was the easiest object since it was the first object the user could view when entered the Galactic Centre; and the Earth was considered moderate level since the users needed to search for it in the planetary system but did not need any movement of the head/body.

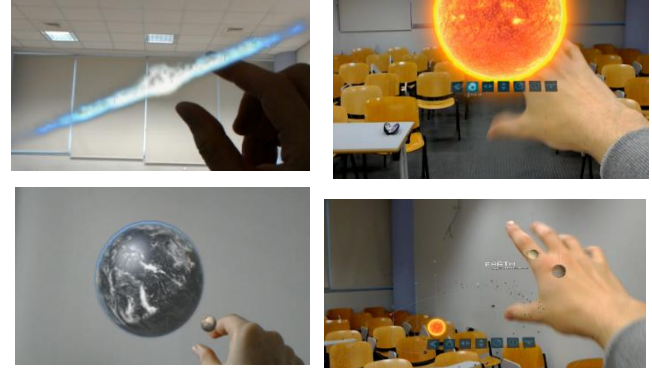


Figure 2 Participants interacting with the Galaxy Explorer application for executing the task

There was no time restriction when participants interacted with the Galaxy Explorer app for performing this task. We selected this task since it is a straight forward search, locate, select task without any other kind of distractions in the environment that could have affected the user's visual or cognitive load.

3 Results

The analysis of the data collected focuses on two aspects, the *time that each participant needed to select each target object*, the *gaze behavior they followed to locate the target object*. These were analyzed based on the participants cognitive characteristics (FD or FI) and the difficulty level of each target object (Easy, Moderate, Difficult).

Table 1 Means and Standard deviation of the completion time in seconds and number of objects viewed required by participants at each group and each level of difficulty to complete the task.

FD-I	Level	Mean and SD (Completion time in Seconds)		Mean and SD (Number of objects viewed)	
FD	1	414	306.52	15.10	12.80
FD	2	5526	4951.23	64.30	28.26
FD	3	3162	1120.01	112.30	45.87
FI	1	616.6	684.04	16.83	19.31
FI	2	4353.33	3009.74	49.38	35.65
FI	3	1710	1066.83	60.00	42.81

Descriptive statistics (Table 1) show that the mean time required by FD participants to complete the first two levels (Easy

and Moderate) was similar, although longer, to the time needed by FI participants. Similarly, the objects viewed by FD participants during the first two levels before locating the required object are close to the number of objects viewed by FI participants. However, at the difficult level, the mean time required by FD participants to locate the required object was almost double the mean time required by FI participants. This is also the case for the mean number of objects viewed by FDs during the difficult level compared to the FIs.

We then test our zero hypotheses and then we adopted a more exploratory approach to understand the behavior of our participants.

H01 There is no interaction effect between FD-I and difficulty level towards the time spent for achieving each target object.

A two-way ANOVA test, with the FD-I cognitive style and the level of difficulty as the independent variables and the time spent for achieving a target object as the dependent variable. The test met all assumptions of a two-way ANOVA thus, employed for evaluating the above hypothesis. The results indicate (Table 2) that at the easy and moderate levels FDs do not appear to have spent statistically more time than FIs to locate and select a target object. However, at the *difficult level*, FDs spent *substantially more time* to locate a target object (Uranus) than FIs ($F=5.899$; $p=0.017$; partial $\eta^2=0.063$). The mean difference in seconds between the two groups at the difficult level was 2085.27 (95% CI, 378.77 to 3791.77), $p = 0.017$.

The overall results indicated no statistically significant interaction effect between the cognitive styles of FD-I and difficulty level towards the *overall time* spent for completing the full task ($F=1.877$; $p=0.159$; partial $\eta^2=0.41$).

Table 2 Interaction effect between cognitive styles FD-I and difficulty level towards time spent

Time spent	F	P	Partial η^2
Easy Level	0.095	0.759	0.001
Moderate Level	1.405	0.239	0.016
Difficult Level	5.899	0.017	0.063
Overall	1.877	0.159	0.41

H02 There is no interaction effect between FD-I and difficulty level towards the gaze behavior for achieving each target object.

A similar statistical approach followed to examine H02, the two-way ANOVA test, with the FD-I cognitive style and the level of difficulty as the independent variables and the gaze behavior for achieving a target object as the dependent variable. The test met all assumptions and the results (Table 3) indicated a statistically significant interaction effect between FD-I cognitive style and difficulty level of the *overall task* towards the gaze behavior followed by participants ($F=7.937$; $p<0.001$; partial $\eta^2=0.154$). The mean difference of objects viewed overall between the two groups was 53.80 (95% CI, 46.5 to 61.03), $p = 0.001$. As the level of difficulty increased there appear *significant differences between FDs and FIs with FDs gazing at more objects* while

searching for the target object at the *difficult level* ($F=26.072$; $p<0.000$; partial $\eta^2=0.231$). The mean difference of objects viewed at the difficult level between the two groups was 64.35 (95% CI, 39.3 to 89.4), $p = 0.000$.

Table 3 Interaction effect between cognitive styles FD-I and difficulty level towards objects gazed

Objects Viewed	F	P	Partial η^2
Easy Level	0.097	0.756	0.001
Moderate Level	1.113	0.294	0.013
Difficult Level	26.072	0.000	0.231
Overall	7.937	0.001	0.154

Further to the results of the second hypothesis we were interested to examine if there was a difference between the gaze behavior of the two cognitive groups (FD-I) in the number of times they gazed at Uranus before they selected it, during the difficult level of the task.

Thus, a Mann-Whitney U test was run to determine if there were differences between FDs and FIs on the number of times they gazed the target object during their searching session before they selected it. Distributions of the number of times the Uranus gazed for FDs and FIs were not similar, as assessed by visual inspection. Number of times Uranus gazed for FDs (mean rank = 20.45) and for FIs (mean rank = 13.55) were statistically significantly different $U=61$; $z=-2.036$, $p = 0.044$. The results are indicating that FIs gazed at Uranus less times compared to FDs before they selected their target.

To assess the overall relationship between the difficulty level and the gaze behavior of the participants in both groups, we ran a Pearson correlation test. The analysis revealed a strong correlation between the difficulty level and the number of objects viewed in the gaze behavior (in terms of number of objects gazed) of participants ($r = 0.563$; $p = 0.000$). A correlation was also revealed between difficulty level and the time needed for the completion of each target object ($r=0.251$; $p=0.015$).

4 Implications

This paper is an initial attempt to investigate the differences between FD-I users towards their visual behavior and interaction in a mixed reality environment when they were asked to perform a specific task. Results revealed significant differences among field dependent and field independent individuals towards their visual interaction with the complex environment of mixed reality, especially as the task became more difficult where support would be needed for both cognitive groups.

FDs approached the identification of each target object following a more holistic strategy (gazed at more objects before locating and selecting the target object) which cost them more time. On the other hand, FIs using a more analytical approach (gazed at less objects during the execution of the task) they interacted more efficiently, locating the target object in less time. These differences were more obvious at the difficult level of the task compared to the easy and medium levels, indicating that as

the difficulty of the task increases, differences in the search behavior between FDs and FIs become more important in a 3D space. This is an interesting outcome especially since the target object at the difficult level of the task required the user to move his/her head and body through the space. Regarding the identification of the target object at the difficulty level we have seen that FDs gazed at their target object more times before they selected it compared to FIs. This is indicating that further studies are needed for us to understand perceptual and spatial memory [13] aspects between these two cognitive groups.

Related work emphasized the need for further studies that will help understand how users perceive information and are interacting in visually rich environments. The findings of this work can be of interest to designers of MR applications who prepare interactive material in MR settings. The analysis of data revealed that FIs are far more comfortable dealing with changing contexts (move from one area of the application to the other) for locating the target object, than FDs were. This can be considered when developing MR applications, not to favor users of one cognitive group over the other and in extend encumber their experience when performing a task. Specifically, participants in the FD group would benefit from timely guidance and clear directions during the task execution. Moving into more immersive and visually rich environments, the cognitive characteristics and specifically FD-I can be an additional parameter in the user model when designing personalized support.

The method followed in this work, logging user interaction data in real time directly from a single device, can be useful when considering temporal personalized and adaptive support. Support can take different forms (e.g. visual, auditory, sensory) according to the modelled characteristics of the user and can be provided at the time that the user appears to struggle with the material or task at hand.

5 Conclusion

In this work we demonstrated in an initial study the differences between FD-I users towards their visual behavior and interaction in a mixed reality environment. Validated psychometric test – GEFT- was employed for extracting participants' cognitive style towards field depended or field independent.

A holographic application – Microsoft Galaxy Explorer- was used through Microsoft HoloLens mixed reality wearable device. Participants asked to perform a task that required identifying three different objects while navigating through the application's environment. The objects that needed to be identified were of varied difficulty to be located by visual search. Participants' logged interaction data (time for task completion, gaze data, in-app interaction video) allowed us to investigate their visual and interaction behavior. Results revealed noticeable statistical differences among field dependent and field independent individuals towards their visual interaction with the complex environment of mixed reality. Differences were more obvious as the task became more difficult with field dependents requiring

more time to complete the task and having more difficulty locating the target object.

We acknowledge the small sample size that does not allow us to generalize the outcomes of this study. However, the above results revealed noticeable differences between FDs and FIs participants during their interaction with the holographic objects and in how they approached the task given, providing us with a good starting point for future work.

Future work is underway focusing on a larger scale study using a different application and tasks, that will allow us to investigate not only the differences between FD-I but also usability and gesture use and behavior of the user during his/her interaction in MR. In addition, we would like to study further the visual differences among individuals, considering demographic data as well (e.g. age differences). The aim of this direction is to allow for the definition of behavior patterns to be defined that will aid in personalized and adaptive support to be provided in real-time thought mixed reality environments.

ACKNOWLEDGMENTS

This project is partially funded by the European Union's Horizon 2020 research and innovation programme under grant agreements No. 739578 (RISE) and 810105 (CyCAT).

REFERENCES

- [1] Drascic, David, and Paul Milgram. "Perceptual issues in augmented reality." In *Stereoscopic displays and virtual reality systems III*, vol. 2653, pp. 123-135. International Society for Optics and Photonics, 1996.
- [2] Billinghurst, Mark, Adrian Clark, and Gun Lee. "A survey of augmented reality." *Foundations and Trends® in Human-Computer Interaction* 8, no. 2-3 (2015): 73-272.
- [3] Ronald, Azuma, T. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments* 6. 4 (1997): 355-385.
- [4] Charoula, Angeli, Nicos, Valanides, and Paul, Kirschner. Field dependence independence and instructional-design effects on learners' performance with a computer-modeling tool. *Computers in Human Behavior* 25,6 (2009) 1355-1366.
- [5] Philip, Oltman, K., Evelyn, Raskin, Herman, Witkin, A. *Group Embedded Figures Test*. Consulting Psychologists Press, Palo Alto CA, USA (1971).
- [6] Herman, Witkin, A., Donald, Goodenough, R., Stephen, Karp, A. Stability of cognitive style from childhood to young adulthood. *Journal of personality and social psychology* 7, 3, (1967) 291-300.
- [7] Herman, Witkin, A., Carol Ann Moore, Donald, Goodenough, R., Patricia Cox, W. *Field Dependent and Field-Independent Cognitive Styles and their Educational Implications*. ETS Research Bulletin Series 2, (1975), 1-64.
- [8] George, Raptis, E., Christos Fidas, and Nikolaos Avouris. Effects of Mixed-Reality on Players' Behaviour and Immersion in a Cultural Tourism Game: A Cognitive Processing Perspective. *International Journal of Human-Computer Studies* (2018). DOI: <https://doi.org/10.1016/j.ijhcs.2018.02.003>
- [9] Efi. Nisiforou, and Andrew Laghos. Field Dependence-Independence and Eye Movement Patterns: Investigating Users' Differences Through an Eye Tracking Study. *Interacting with Computers* 28, 4 (2016), 407-420.
- [10] Christina, Katsini, Christos Fidas, Marios Belk, Nikolaos Avouris, and George Samaras. Influences of users' cognitive strategies on graphical password composition. *Proceedings of the 2017 CHI Conference Extended Abstracts*, ACM, (2017), 2698-2705.
- [11] Chih-Ming, Chen and Yen-Nung Tsai. Interactive augmented reality system for enhancing library instruction in elementary schools. *Computers & Education* 59, 2 (2012), 638-652.
- [12] Iulian, Radu. Augmented reality in education: a meta-review and cross-media analysis. *Personal and Ubiquitous Computing* 18, 6 (2014), 1533-1543.
- [13] Markus Löchtefeld, Frederik Wiehr, and Sven Gehring. 2017. Analysing the effect of tangible user interfaces on spatial memory. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. ACM, New York, NY, USA, 78-81. DOI: <https://doi.org/10.1145/3131277.3132172>