See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/332928276

Perceptual Limits of Optical See-Through Visors for Augmented Reality Guidance of Manual Tasks

5 authors, including:

Sara Condino
Università di Pisa
72 PUBLICATIONS **484** CITATIONS

SEE PROFILE

Marina Carbone
Università di Pisa
41 PUBLICATIONS **483** CITATIONS

SEE PROFILE

Roberta Piazza
Università di Pisa
19 PUBLICATIONS **72** CITATIONS

SEE PROFILE

Vincenzo Ferrari
Università di Pisa
154 PUBLICATIONS **1,786** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   VOSTARS View project

Project   HTA of Da Vinci View project

Perceptual Limits of Optical See-Through Visors for Augmented Reality Guidance of Manual Tasks

S. Condino[*], M. Carbone, R. Piazza, M. Ferrari, V. Ferrari

Abstract—*Objective:* **The focal length of available optical see-through (OST) head-mounted displays (HMDs) is at least 2 m therefore, during manual tasks, the user eye cannot keep in focus both the virtual and real content at the same time. Another perceptual limitation is related to the vergence-accommodation conflict (VAC), this latter being present in binocular vision only. This paper investigates the effect of incorrect focus cues on the user performance, visual comfort and workload during the execution of augmented reality (AR) guided manual task with one of the most advanced OST HMD, the Microsoft HoloLens.** *Methods:* **An experimental study was designed to investigate the performance of 20 subjects in a connect-the-dots task, with and without the use of AR. The following tests were planned: AR guided monocular and binocular; Naked-eye monocular and binocular. Each trial was analyzed to evaluate the accuracy in connecting dots; NASA Task Load Index and Likert questionnaires were used to assess the workload and the visual comfort.** *Results:* **No statistically significant differences were found in the workload, and in the perceived comfort between the AR guided binocular and monocular test. User performances were significantly better during the Naked eye tests. No statistically significant differences in performances were found in the monocular and binocular tests. The maximum error in AR tests was 5.9 mm.** *Conclusion:* **Even if there is a growing interest in using commercial OST-HMD, for guiding high-precision manual tasks, attention should be paid to the limitations of the available technology not designed for the peripersonal space.**

Index Terms — **Augmented Reality, Focus Cues, Head-mounted Displays, Optical See-Through, Vergence-Accommodation Conflict.**

I. INTRODUCTION

Wearable augmented reality (AR) devices, commonly referred to as head-mounted displays (HMDs), offer the most ergonomic solution for manual tasks performed by the user under direct vision since they provide an egocentric viewpoint and allow the operator to work hands-free.

AR HMDs can be categorized according to the see-through paradigm they implement: video see-through (VST) HMDs and optical see-through (OST) HMDs [1]. In standard OST HMDs, the user's direct view of the real world is augmented through the projection of the computer-generated content on semi-transparent displays in front of the user's eyes. Differently, in VST HMDs the real view of the world is acquired by one or two external cameras anchored to the visor and presented on the HMD after being coherently merged with the virtual content. Thus, in these systems, the augmented view is presented sharing the camera's viewpoint.

OST systems have the main advantage of providing the user with a natural view of the real world with its own full resolution. The fundamental OST paradigm of HMDs (Google Glass, Microsoft HoloLens, Epson Moverio, Lumus optical, Optinvent Ora-2, Meta 2) has not changed from the one firstly described by Benton in 2001 [2]. A straightforward implementation of the OST paradigm comprises the employment of a half-silvered mirror or a beam combiner to merge the real view and virtual content. The user's own view is herein augmented by rendering the virtual content on a two-dimensional (2D) micro display mounted outside the user field of view, and by sending the display images to the user eye through a beam combiner. Lenses can be placed between the beam combiner and the display to focus the virtual image so that it appears at a comfortable viewing distance on a semitransparent surface of projection (SSP) [3, 4].

OST HMDs have been at the leading edge of the AR research for a long time [5], and a lot of commercial products were developed. Literature is now concerned on facing the technological and perceptual limitations they present [6]. To mention a few of them: presence of a small augmentable field of view, obtrusiveness of the device, need for frequent recalibrations of the HMD for an accurate spatial registration, low luminance of the micro-displays, and perceptual conflicts between three-dimensional (3D) real world and 2D virtual image on the SSP. Some of these technological limitations, like the small field of view, the overall bulkiness of the device, and the technology behind the displays, are being solved through technological progress. The remaining two limitations are harder to cope with.

This paper is concerned with perceptual issues with the

*S. Condino (correspondence email: sara.condino@endocas.org), M. Carbone, R. Piazza, V. Ferrari, are with the Information Engineering Department, University of Pisa, Italy, and with the EndoCAS Center, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Italy.
M. Ferrari is with the EndoCAS Center, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Italy, and with the Vascular Surgery Unit, Cisanello University Hospital AOUP, Pisa, Italy.

current generation of consumer-level OST HMDs which are only capable of presenting the virtual content at a single fixed focal distance: the possible mismatch in the accommodation stimulus for the virtual content and the real-world scene, and the vergence-accommodation conflict (VAC).

The ocular accommodation is the process by which the refractive power of the eye is adjusted to ensure a clear retinal image of a target object. In natural vision the accommodation stimulus is strictly coupled with the vergence action. This latter, necessary to obtain/maintain a single binocular vision, is the simultaneous movement of the eyes, in opposite directions, to bring the visual axes of the two eyes to the same point. The fixation point is sharply focused on the retina, but points away from fixation are increasingly blurred [7] (see Materials and Methods for details).

When viewing the AR content with an OST device, the user eye is cued to accommodate at the SSP to bring the virtual information in focus, and, at the same time, it is forced to accommodate and converge to the depth of the real 3D object onto which the digital information is overplayed. As explained in the following section, the distance gap between the 2D virtual image on the SSP and real-object can be beyond what the human eyes can accommodate simultaneously, inducing a "focal rivalry" (FR) [8] when the user attempts to view the real world and virtual reality (VR) data together [9, 10].

The focal length of commercially available AR HMDs is generally more than 2 m, therefore, during manual tasks, the user eye cannot keep in focus both the virtual and real content at the same time. For this reason, the use of OST HMDs with a focal distance greater than the arm's length, seems practical for the guidance of manual tasks when the VR content is not strictly coupled with the real world, as in the case of textual information or simple icons. In such cases indeed the user can change fixation point and alternately focus on real world and VR information, as a driver who alternately looks the street and the tachymeter in the cockpit.

However there is a growing interest also in using commercial OST HMDs to guide manual tasks using VR data which should be accurately aligned to the real target, as in AR guided surgery [11, 12]. By way of example: a virtual cutting line/plane registered with the patient anatomy to guide the surgical incision/cutting [13], and a virtual axis displayed over a bone surface for the alignment of a surgical reamer according to a preoperative plan [14]. In such cases, the VR content is superimposed to the physical target, and the visual experience can be impaired by the blurring caused by simultaneously viewing virtual content and physical objects at differing focal distances.

Recent feasibility studies demonstrated that OST HMDs based on light field displays can potentially solve the above mentioned accommodative difficulties intrinsic to standard optical see-through HMDs [9, 15]. However, this technology is not ready for deployment and only traditional OST HMDs are today available on the market.

Many psychophysical studies [16-20] have examined the adverse consequences of incorrect focus cues in stereoscopic displays and HMDs which may contribute to: distorted depth perception, diplopic vision, visual discomfort and fatigue, and degradation in oculomotor response [21]. However, to the best of our knowledge, no previous work has investigated the effect of OST FR and VAC, on the user performance, visual comfort and workload during the execution of AR guided manual task.

The present study directly addresses this gap in the literature, presenting a quantitative evaluation of user performance and a qualitative study of perceived visual comfort and subjective workload, with one of the most advanced OST HMD available today, the Microsoft HoloLens.

II. MATERIAL AND METHODS

The following section is structured as follows: first of all details to clarify perceptual issues of standard OST-HMDs are reported, then the experimental protocol, and metrics to evaluate user performance, perceived visual comfort and subjective workload are explained.

A. The Human Eye Optical System and Focus Cues with OST-HMDs

For a given accommodation state of the eye, there is only one target position which produces an image perfectly focused on the retina. If the target point is brought closer to the eye its image will be focused behind the retina; on the contrary, if it is brought farther away, it will be focused in front of the retina. In both cases the light received from the target object will be distributed on the retinal plane in a circular shape, producing a blur circle.

The distance between the image plane and the retinal plane which can be tolerated without incurring "in an objectionable lack of sharpness in focus" [22] defines the depth-of-focus, which is conventionally expressed in diopters. Related to depth-of-focus is the depth-of-field (DOF), which is the "projection of the dioptric interval of the depth-of-focus into free space" [22].

DOF extends in front and behind the fixation point. As long as the displacement of the target point is within the proximal (*NearDOF*) and distal (*FarDOF*) boundaries of the depth-of-field, the target point will be perceived as being in focus. DOF varies according to the fixation distance: the eye can tolerate much larger intervals of depth when the fixation point is far then when it is near. For each fixation distance (*P*), the depth-of-field range, *NearDOF-FarDOF*, can be derived from the depth-of-focus (ΔD) as in (1) [23]:

$$\text{NearDOF} = \frac{1}{\frac{1}{P} + \Delta D} \qquad \text{FarDOF} = \frac{1}{\frac{1}{P} - \Delta D} \qquad (1)$$

Literature studies [23] on human factors of stereoscopic 3D displays refer to Wang and Ciuffreda work [22] to estimate the depth-of-field. According to their work, which reports a summary of experimental findings on the human eye, the average depth of focus is of the order of ±0.50 D. Based on this value, one can easily verify that for fixation points within the peripersonal space (intended as the space that contains reachable objects), the focal distance of commercially available AR HMDs (which, as already said, is generally more

than 2 m) is completely outside the depth-of-field: e.g assuming for manual tasks a fixation distance (P) of 0.50 m, (1) predicts a depth-of-field ranging from 0.40 m to 0.67 m.

From this analysis we can conclude that during manual tasks the user eye cannot keep in focus both the VR and real content at the same time.

Fig. 1 schematically illustrates possible visual cue conflicts in OST AR, highlighting discrepancies between the accommodation for real objects (equal to the vergence distance) and the accommodation for VR data, being the HMD focal plane out of the user DOF.
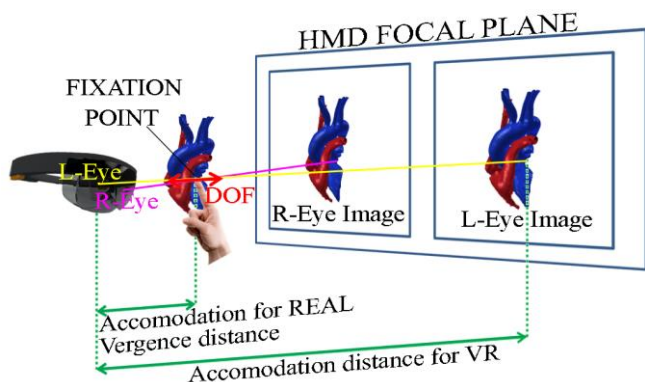

Fig. 1. Exemplification to illustrate possible visual cue conflicts with OST technology. The two eyes converge to the forefinger (real object). The figure shows the discrepancy between accommodation distances for the forefinger and the VR heart model, and the mismatch between the vergence distance and accommodation distance for the VR model. The red arrow, extending in front and behind the fixation point, indicates the user DOF.

Additional information on the depth-of-focus is reported in the appendix, which attempts at proposing a mathematical derivation of the depth-of-focus according to its dependence on the pupil diameter, and the resolution of the HMD display.

B. Design of the Experimental Study

The experimental study was designed to investigate the effects of the "focal rivalry" and VAC together and separately, and to limit the influence of registration errors which can negatively affect the user performance.

The following tests were planned:
1) *AR guided binocular test (AR_{Bin})* to evaluate the effect of the FR and VAC together.
2) *AR guided monocular test (AR_{Mon})* to evaluate the effect of the FR only.
3) *Binocular Naked eye test (NK_{Bin})*.
4) *Monocular Naked eye test (NK_{Mon})*.

Each of the above mentioned tests was planned to be repeated three times by each subject enrolled in the study (for a total of twelve trials for each subject).

To limit the influence of registration errors, a simple task, which does not require a superimposition of the virtual scenario with a physical counterpart, was designed for this study.

The selected task consists of connecting a sequence of numbered dots using a ruler to draw lines on a physical support.

During AR tests (AR_{Bin}, AR_{Mon}) numbered dots were showed in AR and the subject was asked to draw lines on a paper (A4). This task forces the user's eye to contemporaneously focus on the virtual content (numbered dots) and real objects (pen, ruler, and paper).

For Naked eye tests (NK_{Bin} and NK_{Mon}) the sequences of numbered dots were printed on a paper (A4) and the subject was asked to draw lines on an overlaying tracing paper.

A total of twelve "connect-the-dots" sequence (CDS) were designed. Each CDS consists of fifteen dots scattered within a square area of 15 cm x 15 cm (Fig. 2).
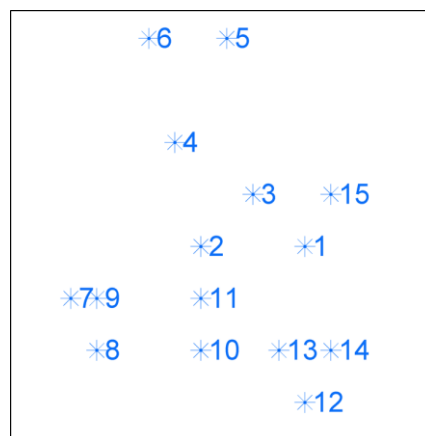

Fig. 2. An example of a "connect-the-dots" sequence

The spatial distribution of dots was defined in order to standardize the difficulty of the task according to the following criteria:
1) lines connecting pairs of consecutive dots (connecting lines) cannot intersect other connecting lines;
2) the distance of two consecutive dots is not fixed, thus connective lines are of different lengths {L1, L2,…, L14};
3) {L1, L2,…, L14} is the same for all the twelve CDS.

C. Hardware and Software Materials

The HoloLens technology is based on an undisclosed Intel 32-bit processor, with a custom-built Microsoft Processing Unit (HPU 1.0) which supports Universal Windows Platform (UWP) apps. It is equipped with 2 GB of RAM (1GB for CPU, and 1GB for HPU) and 64 GB of flash memory; it features network connectivity via Wi-Fi 802.11ac and Bluetooth 4.1 LE wireless technology. The HPU processes data from multiple sensors: 4 *"environment understanding cameras"*; 1 depth camera; and 1 world-facing photo/video camera (2 MP); 1 inertial measurement unit (IMU) to track head movements; 4 microphones; 1 ambient light sensor. As already said HoloLens features a fixed focal length of 2.0 m [24], and the field of view (FOV) is about 35°. The weight is 579 g.

Unity 5.6 engine was used to build the HoloLens application, and Vuforia 6.1 SDK was employed for registration and tracking functionalities. A custom Image

Target (an image with features that Vuforia SDK can detect and track) was used to anchor and display the virtual content at a fixed and stable position in the space (see Fig. 3), avoiding any perceptible spatial jitter which could affect the user performance.

The Image Target was created starting from the native Image Target Samples "Lion Flakes" provided by Vuforia.

D. Subjects

Twenty subjects, seven male and thirteen female, were recruited from technical employees (engineers) and students of the University of Pisa.

The accommodative ability decreases steadily from later childhood greatly due to the age-related changes of the eye; however it remains more than adequate until the onset of presbyopia at the age of about 40 years [25]. Participants included in this study were aged between 18 and 36.

Table I reports the demographics of the participants. Besides demographic data we also asked the participants to rate their experience with AR methods and their experience with HMDs to get the opportunity to correlate these with their performance and workload.

TABLE I
DEMOGRAPHICS OF PARTICIPANTS

Gender (male; female; non-binary)	(7; 13; 0)
Age (min; max; mean; STD)	(18; 36; 26; 5)
Visual Acuity (normal, corrected to normal)	(11,9)
Stereo Acuity (normal ≤ 63 arcsec, low > 63 arcsec [26])	(20,0)
VR experience (none; limited; familiar; experienced) *	(10, 3, 4, 3)
AR experience (none; limited; familiar; experienced) *	(9, 4, 5, 2)
Hololens experience (none; limited; familiar; experienced) *	(10, 6, 2, 2)
Handedness (left; right; ambidextrous)	

*none=technology never used; limited=technology used less than once a month; familiar=technology used about once a month; experienced=technology used several times a month.

All participants claimed to have normal vision acuity or corrected-to-normal visual acuity with the use of contact lenses.

Visual acuity was verified for left eye, right eye and binocular vision, using the Digital Acuity LogMAR Charts from Chart2020. A testing distance of 6 m, and the standard termination rule of stopping after four or more mistakes on a line, were adopted. The stereo acuity of participants was measured using the Random Dot Stereo Acuity Test from Chart2020. Tests were conducted using Polarized/Color Tinted Anaglyphs with a testing distance of 6 m. Each subject provided written consent before proceeding with the study.

E. Protocol of the Study

The experimental setting is shown in Fig. 3.

An Image Target has been used to standardize the positioning of the "connect-the-dots" images during the different trials: it has been attached in a fixed position on a desk worktop to register the "connect-the-dots" virtual images on a A4 paper, at the same position as the physical images during the *Naked eye tests*. Moreover, a "chin rest" was used to fix the target/user eye distance to 0.5 m; more in particular, each subject was asked to sit in front of the desk, to place his/her chin on the "chin rest", and to tilt his/her head in such a way as to visualize the entire "connect-the-dots" image avoiding the need of any head movement during the task completion. This also reduced the risk of instability in virtual content placement due to inaccurate headset tracking during the user's head movements.

Fig. 3. Experimental setting during a video recorded AR_{Bin} trial. The picture shows the Vuforia Image Target used for registering the virtual content in correspondence with the A4 paper.

A slow-motion camera (GoPro Hero4, 240Hz) was used to record the experimental sessions.

Before starting the study the subject was asked to calibrate the HoloLens using the Calibration app (by Microsoft), based on the user interpupillary distance.

As described in Section B, each subject performed the "connect-the-dots" task three times for each presentation modalities (PM): AR_{Bin}, AR_{Mon}, NK_{Bin}, NK_{Mon}. A blind fold was used for monocular tests. The twelve "connect-the-dots" sequences (CDS) were randomly associated with each of the four PM, this randomization was repeated for each subject. The presentation modalities were administered in random order.

Subjects were instructed that the priority of the experiment was to accurately connect all the dots, within 3 minutes maximum. Moreover subjects were asked to report any perceptible spatial jitter or drift of the virtual content and to stop the task if any.

Video-recorded trials were analyzed for a quantitative evaluation of user performance, moreover participants were asked to complete anonymous questionnaires according to the protocol described in the following paragraph and summarized in Fig. 4.
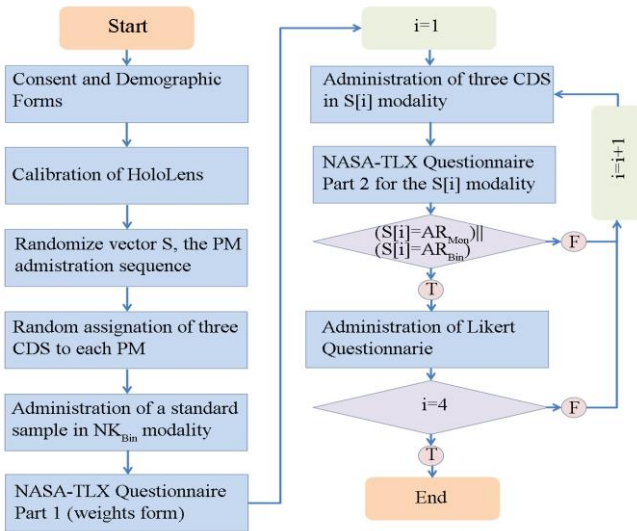
Fig. 4. Flowchart of the experiment for each subject included in the study. S is the array: S={ARBin, ARMon, NKBin, NKMon}.

1) Qualitative Evaluation of AR Presentation Modalities

The study includes subjective workload assessments with the NASA Task Load Index (NASA-TLX) questionnaire, and subjective evaluation of focus cues with a Likert questionnaire.

NASA-TLX is a multi-dimensional rating procedure that provides an overall workload score, between 0 and 100, based on a weighted average of ratings on six subscales [27]:
1) mental demands ("How mentally demanding was the task?");
2) physical demands ("How physically demanding was the task?") ;
3) temporal demands ("How hurried or rushed was the pace of the task?");
4) own performance ("How successful were you in performing the task?") ;
5) effort ("How hard did you have to work to achieve your level of performance?") ;
6) frustration ("How insecure, discouraged, irritated, stressed, and annoyed were you?").

NASA-TLX questionnaire was administrated to investigate any perceived workload difference among different AR presentation modalities, and to examine workload levels of users with differing characteristics (Experience with AR, with Hololens and with VR).

The Likert questionnaire, which is reported in Table III, comprises 8 items, each evaluated using a five-point monotone Likert scale (from 1 = strongly disagree, to 5 = strongly agree).

F. Quantitative Evaluation

Each trial was analyzed by a blinded observer to measure the accuracy in connecting dots and time to completion ($T_{COMPLETION}$).

The evaluation of the accuracy of each trial required the following step:

1) digitization of the A4 paper/tracing paper with connecting lines drawn by the subject;
2) digital image filtering for contrast adjustment;
3) automatic corner detection (Harris Corner Detector [28]);
4) manual numbering of the endpoints ($Start_i$, End_i) of each line i among features automatically detected during step 3 (by using the Matlab function ginput).

Digital Image Processing was performed with MATLAB® Version R2017b.

The task accuracy was evaluated in terms of:

1) difference in length (E_i) of each line l_i (14 lines for each trial) drawn by the subject and the ideal segment (whose endpoints coordinates have been defined according to the procedure described in Section B);

2) gap (G_{ij}) between the endpoints End_i and $Start_j$ of each pair (i,j) of consecutive lines.

More in particular, for each test session (each 3 trials performed by the subject with the same modality), the following parameters were calculated: maximum and mean difference in length (E_{MAX} and E_{MEAN}), maximum and mean gap (G_{MAX} and G_{MEAN}).
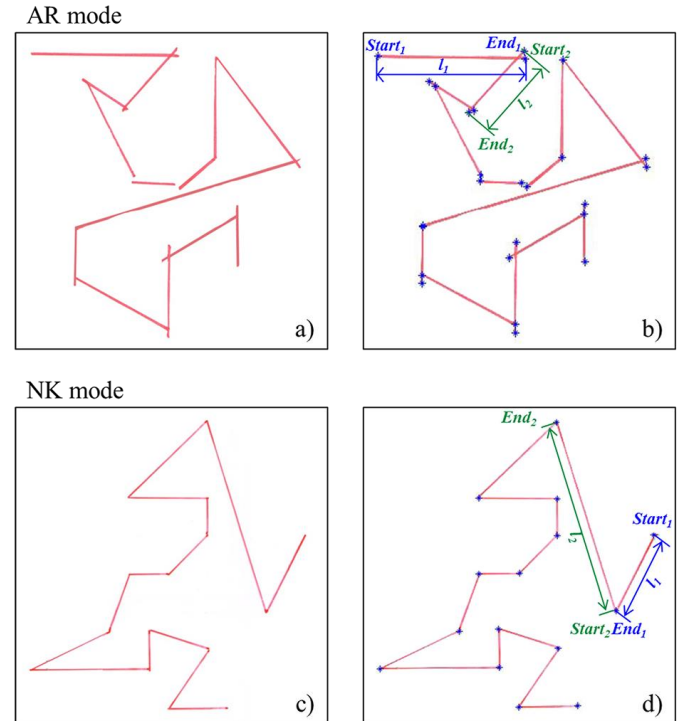

Fig. 5. On the left: examples of two tasks respectively performed with AR (a) and without AR (c). On the right (b and d): corresponding images processed for accuracy evaluation: the endpoints ($Start_i$, End_i) of each line l_i are represented with blue stars; the length of two segments (l_1 and l_2) is showed as an example.

G. Statistical Analysis

Statistical analysis of data was performed using the SPSS® Statistics Base 19 software.

Results of the NASA-TLX questionnaire were summarized in terms of means, and standard deviation. The central tendencies of responses to Likert items were summarized by using median with dispersion measured by interquartile range.

Wilcoxon Signed-Rank test was then carried out to see if there were differences in perceived workload and visual comfort based on the AR PM. The Kruskal-Wallis test was used to understand whether the answering tendencies differ based on "Experience with AR" / "Experience with Hololens" / "Experience with VR". A p-value<0.05 was considered statistically significant.

Quantitative results are presented as the average value and standard deviation of the accuracy in connecting dots and time to completion. Data were analyzed with the Friedman test to examine whether the PM has an effect on the subject performance; post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at p<0.012. Moreover, for each presentation modality, the Mann-Whitney U test was used to evaluate whether the subject performance differ based on gender. A p-value<0.05 was considered statistically significant.

III. RESULTS

A. Qualitative Evaluation Results

Table II shows the results of the subjective workload scores from the NASA-TLX questionnaire. No statistically significant differences were found in the total perceived workload and in the individual workload domains between the AR_{Bin} and AR_{Mon} modalities: Mental Demands (p=0.758), Physical Demands (p=0.320), Temporal Demands (p=0.867), Own Performance (p=0.679), Effort (p>0.959), Frustration (p=0.465), Total Workload (p=0.687).

Moreover, no statistically significant differences were found between groups with different experience with AR, VR and HoloLens.

TABLE II
NASA QUESTIONNAIRE RESULTS

	AR_{Bin}		AR_{Mon}	
	Mean (STD)	p-value Exp. with AR, VR and Holo	Mean (STD)	p-value Exp. with AR, VR and Holo
Mental Demands	46.21 (25.88)	0.386 0.526 0.292	47.14 (31.81)	0.391 0.611 0.168
Physical Demands	43.36 (30.23)	0.871 0.711 0.842	46.07 (32.22)	0.627 0.910 0.407
Temporal Demands	41.29 (24.86)	0.092 0.862 0.187	42.57 (25.17)	0.304 0.732 0.111
Own Performance	35.86 (31.87)	0.885 0.884 0.807	39.64 (31.75)	0.885 0.884 0.807
Effort	55.21 (29.43)	0.395 0.989 0.869	60.79 (28.41)	0.142 0.711 0.958
Frustration	55.21 (23.94)	0.280 0.398 0.680	54.00 (19.28)	0.298 0.526 0.657
Total	50.58 (16.77)	0.908 0.410 0.710	51.19 (19.28)	0.542 0.610 0.743

Table III reports results of Likert questionnaire. Users stated to have perceived VR content (dots and numbers) as clear and sharp in both AR_{Bin} and AR_{Mon}. However, for both modalities they expressed a neutral opinion regarding: the ability to contemporaneously focus at the VR content and real objects, the perception of visual discomfort due to blur, the comfort of using the AR guidance for the selected task, the trustability of the AR modality to successfully guide manual task, their confidence in the precision they can reach in manual tasks guided by the AR modality. Finally, users stated to have experienced visual fatigue (intended as one or more of the following symptoms: eyestrain, dried mucus or tears around the eyelids, discomfort when the eyes are open, hot eyes, and headaches) during the AR_{Bin} tests, while they expressed a neutral opinion regarding the visual fatigue during the AR_{Mon} tests. However this difference in answering is not statistically significant (*p*=0.1).

For all items there was no statistically significant difference (*p*>0.05) in answering tendencies among participants with different level of experience with VR, AR, and HoloLens (see Table III for *p* values).

TABLE III
LIKERT QUESTIONNAIRE RESULTS

	AR_{Bin}		AR_{Mon}	
Item	Median (25°~75°)	p-value Exp. with AR, VR and Holo.	Median (25°~75°)	p-value Exp. with AR, VR and Holo.
I perceived VR dots as clear and sharp	4 (3~5)	0.090 0.421 0.234	4 (3~5)	0.902 0.305 0.193
I perceived VR numbers as clear and sharp	4 (4~5)	0.055 0.421 0.234	4 (3~5)	0.627 0.217 0.183
I was able to contemporaneously focus at the VR Content and Real Objects	3 (3~4)	0.236 0.421 0.234	3 (2~4)	0.854 0.456 0.600
I did not perceived any visual discomfort due to blur	3 (2~4)	0.155 0.421 0.234	3 (2~3)	0.945 0.691 0.406
I felt comfortable using this AR guidance modality for the selected task	3 (3~4)	0.537 0.421 0.234	3 (2~4)	0.497 0.500 0.598
I did not experience visual fatigue	2 (2~4)	0.206 0.421 0.234	3 (2~4)	0.436 0.908 0.368
I can trust this AR modality to successfully guide manual task	3 (3~4)	0.172 0.421 0.234	3 (3~4)	0.353 0.461 0.090
I am confident of the precision of manual tasks guided by this AR modality	3 (2~4)	0.077 0.421 0.234	3 (3~4)	0.312 0.844 0.461

B. Quantitative Evaluation Results

All the user successfully completed the twelve tasks and did not reported any perceivable spatial displacement (including jitter and drift) of the virtual content. Mean and standard deviation values of E_{MAX}, E_{MEAN}, G_{MAX}, G_{MEAN} and $T_{COMPLETION}$ are reported in Table IV.

There was a statistically significant difference in mean E_{MAX}, E_{MEAN}, G_{MAX}, G_{MEAN} depending on PM, ($p<0.001$ for E_{MAX}, $p<0.001$ for E_{MEAN}, $p=0.001$ for G_{MAX}, $p<0.001$ for G_{MEAN}).

Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p<0.012$. As reported in Table V, there were no significant differences in E_{MAX}, E_{MEAN}, G_{MAX}, and G_{MEAN} between the two NK guided modalities (NK_{Mon} vs NK_{Bin}) and the two AR guided modalities (AR_{Mon} vs AR_{Bin}); however, there was a statistically significant increase in E_{MAX}, E_{MEAN}, G_{MAX}, and G_{MEAN} in the two binocular tests (NK_{Bin} vs AR_{Bin}), and the two monocular tests (NK_{Mon} vs AR_{Mon}).

TABLE IV
QUANTITATIVE EVALUATION RESULTS

		NK		AR	
		NK_{Bin}	NK_{Mon}	AR_{Bin}	AR_{Mon}
E_{MAX} [mm]	MEAN (STD)	2.7 (1.0)	2.8 (1.2)	5.8 (2.0)	5.9 (1.7)
	p-value gender	0.817	0.536	0.393	0.097
E_{MEAN} [mm]	MEAN (STD)	0.9 (0.2)	0.9 (0.2)	2.2 (0.8)	2.3 (0.6)
	p-value gender	0.157	0.311	0.699	0.938
G_{MAX} [mm]	MEAN (STD)	1.8 (1.0)	2.3 (1.2)	3.9 (2.1)	5.8 (3.2)
	p-value gender	0.135	0.351	0.699	0.938
G_{MEAN} [mm]	MEAN (STD)	11.9 (2.8)	12.1 (3.1)	31.4 (11.1)	31.6 (8.8)
	p-value gender	0.135	0.351	0.699	0.938
$T_{COMPLETION}$ [s]	MEAN (STD)	71 (15)	73 (22)	71 (25)	60 (18)
	p-value gender	1.000	1.000	0.517	0.833

Finally, the Mann-Whitney U test revealed that there were no significant differences in accuracy performances between male and female participants, in terms of E_{MAX}, E_{MEAN}, G_{MAX}, and G_{MEAN} (see p-value reported in Table IV).

Moreover, the Friedman test revealed a statistically significant difference in $T_{COMPLETION}$ depending on PM ($p=0.045$), but post hoc analysis with Wilcoxon signed-rank tests on the different combinations of PM, showed that the only statistical significant difference occurred between the AR_{Mon} and NK_{Bin} ($p=0.009$). No statistically differences indeed occurred between: the two AR guided modalities (AR_{Mon} vs AR_{Bin}), the two NK guided modalities (NK_{Mon} vs NK_{Bin}), the two binocular tests (NK_{Bin} vs AR_{Bin}), and the two monocular tests (NK_{Mon} vs AR_{Mon}).

No significant differences were also found between male and female participants in terms of mean $T_{COMPLETION}$ (see p-value reported in Table IV).

TABLE V
POST HOC ANALYSIS WITH WILCOXON SIGNED-RANK TESTS: P-VALUES

	BIN *VS* MON		NK *VS* AR	
	NK_{Bin} *VS* NK_{Mon}	AR_{Bin} *VS* AR_{Mon}	NK_{Bin} *VS* AR_{Bin}	NK_{Mon} *VS* AR_{Mon}
E_{MAX}	0.556	0.654	**<0.001**	**<0.001**
E_{MEAN}	0.723	0.386	**<0.001**	**<0.001**
G_{MAX}	0.089	0.016	**0.001**	**0.001**
G_{MEAN}	0.085	0.747	**<0.001**	**0.001**
$T_{COMPLETION}$	0.575	0.037	0.799	0.028

IV. CONCLUSION

With the current generation of OST HMDs, the AR experience is impaired: in binocular conditions by the VAC conflict and the FR between the real scene and virtual content, and in monocular condition only by the FR.

Results of our experimental study with Microsoft HoloLens show no statistically significant differences between the AR guided binocular and monocular tests neither in terms of perceived workload and comfort, nor in user performances.

Thus we can assume that, for the proposed AR guided manual task, the workload/comfort and the user performance do not significantly change if only the FR is present or also the VAC conflict occurs.

To further investigate the contribution of VAC and FR separately, future tests will aim at comparing the OST with the VST technology, the latter being influenced only by VAC. Given that the AR text readability is affected by the screen resolution, and refresh rate, and the workload for OST-HMD visualization is affected by the ergonomics of the HMD, it is best to perform these future tests by using a single HMD devices able to switch between the OST and VST modality, so not to introduce bias which can affect the obtained results. Such kind of device, a hybrid video-optical see through HMD, is being developed to address human factors for the achievement of optimal ergonomics and perfect usability in surgical applications [29].

The qualitative analysis of this study shows that the user is not fully aware of the accuracy that can be achieved with the OST-HMD AR guide and is not able to express whether or not to trust the proposed AR mode.

The quantitative results suggest that HoloLens should not be used for high-precision manual tasks: on average, participants made errors of up to 5.9 mm in length (versus 2.8 mm during the naked eye tasks) with gaps up to 5.8 mm. The findings of this work are strengthened by results of a previous

study with HoloLens [30], showing targeting errors up to 5.2 mm in the peripersonal space.

We can hypothesize that such inaccuracies may depend not only on eye-to-display calibration errors, but also on the afore mentioned accommodation issues which impair the contemporaneous view of the real world and the virtual target. In this work the official Microsoft app was used to calibrate the HoloLens. The app does not offer a complete user-based OST calibration procedure, but solely determines the interpupillary distance [31], and no official documentation is available assessing the obtainable accuracy. Eye-to-display calibration errors may led to a distortion of the perceived virtual content, and therefore they can play a major role in the misperception of line lengths. However we can argue that the measured gaps between the endpoints of the lines cannot be related to calibration inaccuracies, moreover since no spatial jitter/drift of the virtual content was noticed by participants, we can hypothesize that they are directly related with the perceptual issues which impair the contemporaneous view of the real world and the virtual target.

At present only a qualitative assessment of spatial jitter/drift was performed, and a future quantitative study is necessary for a deeper understanding of the quantitative results of this work: e.g. a feature detection algorithm could be used to quantify any potential displacement of the virtual content by accessing the stream of the HoloLens cameras. Another study limitation is the relative small study sample, thus future studies will include a larger number of subjects to see if similar trends appear. Moreover brain activity recordings could be used complementary to qualitative questionnaires to investigate visual fatigue induced by the VAC conflict and the FR between the real scene and virtual content.

Despite the above mentioned limitations of the study, obtained results suggest that, although there is increasing interest in using commercial OST-HMDs to guide high-precision manual tasks that require accurate alignment of VR data to the actual target (such as surgical tasks), attention must be paid to the current limitations of available technology that is not designed for peripersonal space.

APPENDIX

The depth-of-focus is affected by several factors, thus there are discrepancies in its empirical value: this can be due to differences in the stimuli/methodology used by the experimenter, or in the criterion on blurriness. Below a mathematical derivation of the depth-of-focus is proposed, enriched with specific considerations for the AR vision trough HMD.

Although the eye is a complex optical system composed of several elements, it can be considered as a single refracting surface, situated behind the cornea, separating two media of refractive indices $n_0=1$ and $n_1=1.336$ [32] as depicted in Fig. 6. As already said, the depth-of-focus is the distance between the image plane and the retinal plane (δ) (see Fig. 7) which can be tolerated without incurring "in an objectionable lack of sharpness in focus" [22].

The depth-of-focus (ΔD) is conventionally expressed in diopters, and it is calculated as the change in power required by the optical system to change its focal length (f_p) by an amount equivalent to δ [33, 34], as in (2) [34]:

$$\Delta D = \pm \left(\frac{n_1}{f_p+\delta} - \frac{n_1}{f_p} \right) = \pm \frac{n_1 \delta}{f_p(f_p+\delta)} \tag{2}$$
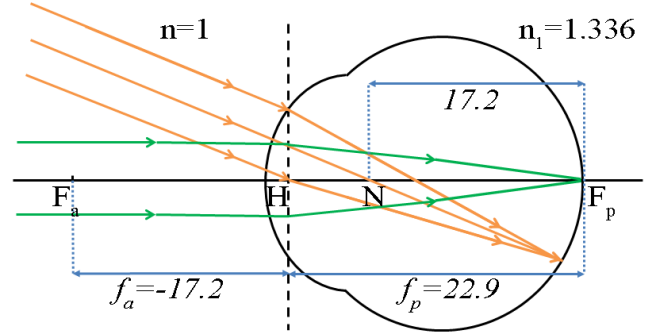

Fig. 6. The reduced model of a relaxed eye proposed by Davson [32] (not drawn to scale). The principal plane is situated 1.5 mm behind the cornea. Fa and F_p represent the two focal points (anterior and posterior), the distances from the principal point H to their respective focal points being the anterior (f_a) and the posterior (f_p) focal length. N represents the nodal point. A ray of light from the height of any given object passes, undeviated, through N.

Fig. 7 shows the optical diagram of a relaxed eye with the distal depth-of-field, *FarDOF,* at an infinite distance away from the principal plane. As long as the target point is at a distance greater than proximal boundary of the depth-of-field, *NearDOF*, it will be perceived as being in focus: it will produce on the retina a blur spot whose diameter cannot be resolved by the human eye. This diameter corresponds to the diameter (*c*) of the so-called "circle of confusion".
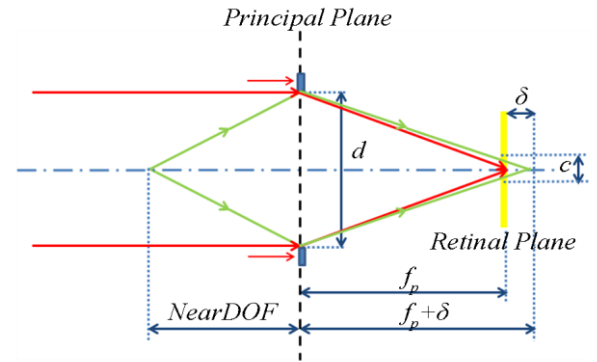

Fig. 7. Optical diagram of a relaxed eye. The depth-of-focus depends on the size of the pupil (*d*), thus on the action of the iris. A point at an infinite distance away is perfectly focused on the retina. A point at the proximal boundary of the depth-of-field (*NearDOF*), produces on the retina a blur spot with a diameter (*c*) which cannot be resolved by the human eye.

Observing Fig. 7, a simple relation (3) [34] can be found (by method of similar triangles) between the circle of confusion diameter (*c*), the diameter of the pupil (*d*), the posterior focal length of the eye (f_p) and ($f_p + \delta$) . This latter is the distance at which a point at the proximal boundary of the DOF is focused.

$$\frac{c}{d} = \frac{\delta}{f_p+\delta} \tag{3}$$

Substitution of (3) in (2) results in (4):

$$\Delta D = \frac{n_1 c}{f_p d} \qquad (4)$$

The circle of confusion can be estimated considering that people with normal visual acuity are able to resolve features which subtend a single arc minute of visual angle (α=1 arcmin). According to (5), an arc minute of visual angle corresponds to a circle of about 5 μm in diameter on the retina.

$$c = \overline{NF}_p \ tg\,\alpha \ \cong 5 \ \mu m \qquad (5)$$

where $\overline{NF}_p$ (see Fig.6) is equal to the distance of the nodal point to the focal plane, which is 17.2 mm according to reduced eye model proposed by Davson [22].

Thus, considering (as in [33, 34]) a pupil diameter of 3 mm, a focal length of 22.9 mm, and a refractive index n_1=1.336, (4) predicts a depth-of-focus of about ±0.10 D, which is consistent to several reports in the literature [34-37].

Given a depth-of-focus of ±0.1 D, and assuming for manual tasks a fixation distance (P) of 0.50 m, (1) predicts a depth-of-field ranging from about 0.48 m to about 0.53 m.

During an HMD AR application, the minimum angle of resolution (α in (5)) is bigger than the smallest visual angle that a person with normal visual acuity can see clearly, thus the circle of confusion diameter increases. For example, considering the angular resolution of HoloLens (2500 light points per radian), the circle of confusion is about 7 μm, (4) gives a depth-of-focus of about ±0.13 D and (1) predicts a depth-of-field ranging from about 0.47 m to about 0.54 m.

This further analysis reinforces the conclusions drawn in section A of "Materials and Methods", analyzing the contribution of HMD resolution to the DOF.

REFERENCES

[1] J. P. Rolland, and H. Fuchs, "Optical versus video see-through mead-mounted displays in medical visualization," *Presence-Teleoperators and Virtual Environments,* vol. 9, no. 3, pp. 287-309, Jun, 2000.

[2] S. A. Benton, *Selected Papers on Three-dimensional Displays*: SPIE Optical Engineering Press, 2001.

[3] J. P. Rolland, and O. Cakmakci, "The past, present, and future of head-mounted display designs." *Proc SPIE*, vol. 5638, pp. 368-377, 2005.

[4] N. S. Holliman *et al.*, "Three-Dimensional Displays: A Review and Applications Analysis," *IEEE Transactions on Broadcasting,* vol. 57, no. 2, pp. 362-371, 2011.

[5] F. Zhou *et al.*, "Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR," *7th Ieee International Symposium on Mixed and Augmented Reality 2008, Proceedings*, vol. 2, pp. 193-202, 2008.

[6] F. Cutolo *et al.*, "Perspective Preserving Solution for Quasi-Orthoscopic Video See-Through HMDs," *Technologies,* vol. 6, no. 1, Mar, 2018.

[7] D. Vishwanath, and E. Blaser, "Retinal blur and the perception of egocentric distance," *J Vis,* vol. 10, pp. 1-16, 2010.

[8] K. Oshima *et al.*, "SharpView: Improved clarity of defocussed content on optical see-through head-mounted displays." *2016 IEEE Virtual Reality (VR)*, pp. 253-254, Mar, 2016.

[9] H. Hua, and B. Javidi, "A 3D integral imaging optical see-through head-mounted display," *Opt Express,* vol. 22, no. 11, pp. 13484-13491, Jun, 2014.

[10] H. Hua, "Enabling Focus Cues in Head-Mounted Displays," *Proceedings of the Ieee,* vol. 105, no. 5, pp. 805-824, May, 2017.

[11] P. Shao *et al.*, "Designing a Wearable Navigation System for Image-Guided Cancer Resection Surgery," *Ann Biomed Eng*, vol. 42, no. 11, pp. 2228–2237, Jul, 2014.

[12] O. J. Muensterer *et al.*, "Google Glass in pediatric surgery: an exploratory study," *Int J Surg,* vol. 12, no. 4, pp. 281-289, 2014.

[13] F. Ricciardi *et al.*, "An Augmented Reality System for Maxillo-Facial Surgery." *Augmented Reality, Virtual Reality, and Computer Graphics.* Part of the Lecture Notes in Computer Science book series, vol. 10325, pp. 53-62, Jun, 2017.

[14] G. Turini *et al.*, "A Microsoft HoloLens Mixed Reality Surgical Simulator for Patient-Specific Hip Arthroplasty Training," *Augmented Reality, Virtual Reality, and Computer Graphics.* Part of the Lecture Notes in Computer Science book series, vol. 10851, pp. 201-210, Jun, 2018.

[15] D. Lanman, and D. Luebke, "Near-Eye Light Field Displays," *Acm Transactions on Graphics,* vol. 32, no. 6, Nov, 2013.

[16] E. Peli, "The visual effects of head-mounted display (HMD) are not distinguishable from those of desk-top computer display," *Vision Res,* vol. 38, no. 13, pp. 2053-2066, Jun, 1998.

[17] J. P. Rolland, and K. Arthur, "Studies of depth judgments in a see-through head-mounted display," *Head-Mounted Displays Ii,* vol. 3058, pp. 66-75, 1997.

[18] S. J. Watt *et al.*, "Focus cues affect perceived depth," *J Vis,* vol. 5, no. 10, pp. 834-862, Dec, 2005.

[19] D. M. Hoffman *et al.*, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *J Vis,* vol. 8, no. 3, pp. 33.31-3330, 2008.

[20] J. P. Wann *et al.*, "Natural problems for stereoscopic depth perception in virtual environments," *Vision Res,* vol. 35, no. 19, pp. 2731-2736, 1995/10/01/, 1995.

[21] X. Hu, and H. Hua, "Design and Assessment of a Depth-Fused Multi-Focal-Plane Display Prototype," *Journal of Display Technology,* vol. 10, no. 4, pp. 308-316, 2014.

[22] B. Wang, and K. J. Ciuffreda, "Depth-of-focus of the human eye: theory and clinical implications," *Surv Ophthalmol,* vol. 51, no. 1, pp. 75-85, Jan-Feb, 2006.

[23] R. E. Patterson, *Human Factors of Stereoscopic 3D Displays*: Springer London, 2015.

[24] B. C. Kress, and W. J. Cummings, "Optical architecture of HoloLens Mixed Reality Headset," *Digital Optical Technologies 2017,* vol. 10335, 2017.

[25] W. N. Charman, "The eye in focus: accommodation and presbyopia," *Clinical and Experimental Optometry,* vol. 91, no. 3, pp. 207-225, May, 2008.

[26] D. Gadia *et al.*, "Assessing stereo blindness and stereo acuity on digital displays," *Displays,* vol. 35, no. 4, pp. 206-212, Oct, 2014.

[27] S. G. Hart, and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psychology*, P. A. Hancock and N. Meshkati, eds., pp. 139-183: North-Holland, 1988.

[28] C. Harris, and M. Stephens, "A combined corner and edge detector," *Proc. of Fourth Alvey Vision Conference*, pp. 147-151, 1988.

[29] M. Carbone *et al.*, "Proof of Concept: Wearable Augmented Reality Video See-Through Display for Neuro-Endoscopy," *Augmented Reality, Virtual Reality, and Computer Graphics.* Part of the Lecture Notes in Computer Science book series, vol. 10851, pp. 95-104, Jun, 2018.

[30] S. Condino *et al.*, "How to Build a Patient-Specific Hybrid Simulator for Orthopaedic Open Surgery: Benefits and Limits of Mixed-Reality Using the Microsoft HoloLens," *Journal of Healthcare Engineering*, vol. 2018, pp. 1-12, Nov, 2018.

[31] J. Grubert *et al.*, "A Survey of Calibration Methods for Optical See-Through Head-Mounted Displays," *IEEE Trans Vis Comput Graph,* vol. 24, no. 9, pp. 2649-2662, Sep, 2018.

[32]   H. Davson, *The Physiology of The Eye*: Elsevier Science, 2012.

[33]   M. Yanoff *et al.*, *Ophthalmology*: Mosby Elsevier, 2009.

[34]   D. G. Green *et al.*, "Depth of focus, eye size and visual acuity," *Vision Res,* vol. 20, no. 10, pp. 827-835, 1980.

[35]   J. Tucker, and W. N. Charman, "The depth-of-focus of the human eye for Snellen letters," *Am J Optom Physiol Opt,* vol. 52, no. 1, pp. 3-21, Jan, 1975.

[36]   R. J. Jacobs *et al.*, "Effect of defocus on blur thresholds and on thresholds of perceived change in blur: comparison of source and observer methods," *Optom Vis Sci,* vol. 66, no. 8, pp. 545-553, Aug, 1989.

[37]   M. Rosenfield, and J. A. Abraham-Cohen, "Blur sensitivity in myopes," *Optom Vis Sci,* vol. 76, no. 5, pp. 303-307, May, 1999.