Attentional orienting in virtual reality using endogenous and exogenous cues in auditory and visual modalities

Rébaï Soret, Christophe Hurter, Charra Pom, Peysakhovich Vsevolod

Attentional orienting in virtual reality using endogenous and exogenous cues in auditory and visual modalities

Rébaï Soret
ISAE-SUPAERO, Université de Toulouse, France
rebai.soret@isae-supaero.fr

Pom Charras
Université Paul-Valéry Montpellier 3, France
pom.charras@univ-montp3.fr

Christophe Hurter
ENAC, Toulouse, France
christophe.hurter@enac.fr

Vsevolod Peysakhovich
ISAE-SUPAERO, Université de Toulouse, France
vsevolod.peysakhovich@isae-supaero.fr

ABSTRACT

The virtual reality (VR) has nowadays numerous applications in training, education, and rehabilitation. To efficiently present the immersive 3D stimuli, we need to understand how spatial attention is oriented in VR. The efficiency of different cues can be compared using the Posner paradigm. In this study, we designed an ecological environment where participants were presented with a modified version of the Posner cueing paradigm. Twenty subjects equipped with an eye-tracking system and VR HMD performed a sandwich preparation task. They were asked to assemble the ingredients which could be either endogenously and exogenously cued in both auditory and visual modalities. The results showed that all valid cues made participants react faster. While directional arrow (visual endogenous) and 3D sound (auditory exogenous) oriented attention globally to the entire cued hemifield, the vocal instruction (auditory endogenous) and object highlighting (visual exogenous) allowed more local orientation, in a specific region of space. No differences in gaze shift initiation nor time to fixate the target were found suggesting the covert orienting.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; *HCI design and evaluation methods.*

KEYWORDS

attentional orienting, virtual reality, endogenous, exogenous

1 INTRODUCTION

The endless progress in virtual and augmented reality allows nowadays creating complex learning and training environments. These technologies are particularly useful in procedural training [Azimi et al. 2018; Webel et al. 2013] where participants have to acquire procedural knowledge consisting of several steps of objects manipulation (for instance, assembly operations). While directly manipulating objects fosters learning compared with passive viewing [Jang et al. 2017], the virtual reality systems can be inefficient due to excessive cognitive load induced by the simulation. Additionally, sometime, a head-mounted display can be as efficient as desktop simulation [Buttussi and Chittaro 2018]. Therefore, it is still unclear how to design optimal procedural training environments using virtual or mixed reality.

One key aspect of immersive training is selective attention. An efficient system should orient attention to the next object to attend or to manipulate with the least cognitive cost and the best performance. For example, Hoareau [2016] tested the effectiveness of visual guidance in virtual reality to learn a medical procedure for laboratory technicians. She observed that visual cueing of the subsequent objects to be manipulated improved learners' performance as indicated by reduced time to complete the procedure and a lower number of incorrect actions. In another study, Sheik et al. [2016] showed that cues of different modalities (visual/auditory) effectively directed a person's attention to a target character when viewing 360° video. Finally, Lin and colleagues [2017] have shown that the use of directional arrows can guide a person's attention and improve performance under certain conditions in 360° video. Nevertheless, if virtual reality can improve procedural learning through the attentional cueing, current knowledge on the subject is insufficient to determine how individuals direct their attention in an immersive 360° ecological environment (i.e., in conditions close to everyday life), nor what type of cues is the most effective for this purpose. The majority of studies are carried out in the laboratory using traditional desktop screens. In an immersive ecological environment, one could expect to obtain differential effects of exogenous and endogenous cueing of attention. For instance, Maringelli and colleagues [2001] suggested dissociation between attentional systems controlling the proximal (close to the subject) and the distal (far from the subject) visual space. Thus, virtual reality makes us one step closer to generalize the laboratory results on everyday activities. As Olk et al. [2018] put it: "Virtual reality allows building a bridge between traditional research in the laboratory and daily situations."

The most widespread experimental paradigm to assess the orienting of visuospatial attention is the one developed by Michael Posner [1980]. In the classical version of this paradigm, a fixation point, presented in the center of a computer screen, is surrounded by two empty rectangles placed on the right and on the left of the screen. The participants are instructed to detect as quickly as possible, a target appearing inside one of the two rectangles. The target presentation is preceded by a brief cue (illumination of rectangles). The cues can be valid, which indicates the position where the target will appear; invalid, which indicates an erroneous location of the target; or neutral, which does not provide information about the target's location. In the classical version of the paradigm, the use of a peripheral cue (the illumination of the rectangle to the left or right) automatically directs attention to the indicated location. This know as *exogenous* orienting of attention. Exogenous orienting imply bottom-up involuntary processing, which is used, for example, when you turn the head towards a sudden braking noise of a car. By using central cues in the Posner paradigm, which present, for example, an arrow symbol pointing to the left or the right : we can assess the individual's *endogenous* (voluntary) attentional orienting. Symbols require decoding and an intention to orient attention to the cued location. Endogenous orienting imply top-down processing, which is used, for instance, when you read a caution "wet floor" sign and then pay attention to the wet surface. Endogenous cueing requires cues to be predictive (i.e., more valid than invalid cues) of the target location to create expectations. When the percentage of valid cues is high (for instance, 80%), the participants elaborate endogenous guidance strategies. The typical results of this spatial orienting paradigm show that the cues improve the target's discrimination rate in both endogenous and exogenous versions (see, for instance, Berger et al. [2005]). The target identification is slower when preceded by invalid cues. These results have also been confirmed by other measurements than motor response times like physiological measurements [Handy et al. 1999; Hawkins et al. 1990].

Originally used to explore the mechanisms of covert attention, i.e., our ability to prepare the processing of information that may appear or be contained in a certain spatial location without eye movement, Posner Cueing Paradigm also makes it possible to determine the cue effectiveness. By subtracting the response times between the valid and invalid cues, we obtain a "cueing effect" which reflects the cue capacity to spatially direct attention and thus improve the accuracy and speed of information processing at that location. Similar response times between valid and invalid cues (no cueing effect) indicate that the cue is inefficient. Faster responses to cued targets (positive cueing effect) indicate that the cue is effective. Eventually, faster responses following an invalid cue (negative cueing effect) indicated the effect known as "inhibition of return" and additional processing cost — in this case, the cue is ineffective.

In this study, we adapted Posner's paradigm to an ecological and immersive environment using a virtual reality head-mounted display with an eye-tracking system. We aimed to determine the effectiveness of cues types (endogenous/exogenous) and sensory modality (visual/auditive) in a procedural sandwich preparation task. Does one type of cue and/or perceptual modality is more effective than the others? We also investigated the relevance of using Posner's paradigm in virtual reality when attention must

Figure 1: An example view of the virtual environment used in the study.

be directed in an immersive ecological environment and aimed to reproduce traditional effects obtained on a standard computer screen.

2 MATERIALS AND METHODS

2.1 Participants

Twenty-one participants (11 women, mean age±SD: 27±10), students and personnel of French Aerospace Engineering School (ISAE-SUPAERO, Toulouse), volunteered to participate in the study. In accordance with the Declaration of Helsinki, all participants gave their written consent before the experiment. Participants did not receive any contributions for their participation.

2.2 Apparatus

We used an HTC Vive virtual reality headset with an integrated Tobii eye-tracking system[1]. The headset is composed of a Dual AMOLED 3.6" diagonal screen, a resolution of 1080×1200 pixels per eye (2160×1200 combined) with a refresh rate of 90 Hz and a field of view of 110° (145° diagonally). Participants used the standard HTC Vive controllers to interact with the environment. Tobii eye-tracking system has a gaze data output frequency (binocular) of 120 Hz with an estimated accuracy of 0.5°. Trackable field of view of eye-tracker is 110° (full HTC Vive field of view). The experimental material consisted of a modified version of Posner's

---

[1]https://www.tobiipro.com/product-listing/vr-integration/

paradigm developed using the Unity3D game engine supporting C# programming and the plugins needed for virtual reality (OpenVR and SteamVR) as well as the plugin needed for Tobii eye-tracking, Tobii Pro SDK.

2.3 Stimuli

The virtual environment reproduces a fast-food restaurant principally composed of a countertop on which several food trays are placed. The figure 1 shows an example view of the virtual world. A central $14° × 25°$ tray contained a $6° × 23°$ piece of bread and $12° × 12°$ trays was placed at $24°$ eccentricity of the fixing point. Four different cues were designed:

	Auditory	**Visual**
Endogenous	Vocal instruction	Directional arrow
Exogenous	Spatialized sound	Object highlighting

The object highlighting cue corresponded to a 300 ms color change (from a metal material to unsaturated light gray and back). The central directional arrow cue was a light gray $3° × 12°$ arrow appeared in the center of the screen. The spatialized sound cue was a stereo pure tone with the duration of 300 ms. Each sound has been spatialized according to the location of the corresponding tray. Due to the relative closeness of the trays and to improve perceptual discrimination, the audio sources were shifted further away (1.20 m to the side for the bottom row and 1.06 m in depth for the top row). The vocal instruction cue was a 300 ms pronounced numbers between 1 and 4. The target was a small red semi-sphere with a diameter of 0.08 m appeared on top of the corresponding tray. We used different ingredients, similar in size, were randomly selected on each trial (bacon, steak, meat slice, salami, tomato, cucumber, mushroom, salad). A lid covered ingredients before target appearance to avoid the effects of their different colors and shapes. Figure 2 gives a schematic representation of valid and invalid (both ipsilateral and contralateral) cueing.

2.4 Procedure

Before the experiment, participants were asked to complete a consent form and a preliminary questionnaire. An instruction sheet on the conduct of the experiment was also provided. Participants were instructed to take an ingredient from the targeted tray and put it on the sandwich as quickly as possible. They were also informed that they would have been given an indication of where the target would most likely appear (cue) and that in 25% of cases the cue would be wrong (25% invalid trials, 75% valid trials). Several practice trials of free placement of ingredients in the bread were carried out in order to accustom the participant to the interaction. Then, all the cues used during the experimental phase were presented once to the participants in each validity condition (valid/invalid). In addition, a training corresponding to 8 trials of each block of the experimental phase was performed by the participant. This training introduces an additional placement instruction to control the participants' starting position when recording each trial.

Each block of the experimental phase began, once the participant was well positioned and focused on the fixing point, with the appearance of one of the 4 possible cues according to the blocks and the counterbalancing, during 300 ms. After an inter-stimulus interval (ISI = 300 ms) the target appeared on one of the 4 ingredient trays according to the validity conditions during 300 ms. The inter-stimulus interval has been defined to allow deployment of endogenous orienting while avoiding the effect inhibitions of return for exogenous cueing. As endogenous orientation is known to be deployed from about 300 ms and exogenous orientation is vulnerable to inhibition of return beyond about 300 ms [Chica et al. 2014]. Therefore, this inter-stimulus interval seemed to us to be the best compromise, one of the objectives being to compare the effect of an endogenous versus exogenous orienting on information processing and associated motor responses. Once the ingredient was placed in the bread, a minimum time interval of 3 seconds was respected before the next trial. Figure 3 illustrates the time course of a valid trial with endogenous visual cueing.

The experimental phase was divided into 4 blocks, each block corresponding to one of four cues ({auditory, visual}×{endogenous, exogenous}). The blocks were counterbalanced between subjects. Each block consisted of a total of 40 trials (30 valid, 10 invalid). Half of the invalid trials were ipsilateral, and the other half – contralateral. The experimental phase consisted of a total of 160 trials including 40 verbal indications, 40 directional arrows, 40 spatialized sounds and 40 object highlights. The entire experiment (with questionnaire and instruction time) lasted about 40 minutes.

2.5 Data analysis

We recorded two different motor responses: **Action Initiation**, i.e., the time interval between the appearance of the target and the time when the participant removes one of these hands from resting position, and **Lid Grip**, i.e., the time interval between the appearance of the target and the time when the participant takes the lid. We also measured the eye movements throughout the experiment and more particularly: **Eye Movement Initiation**, i.e., the time interval between the end of cue appearance and the moment when the participant stops fixating the central point; and **Gaze On Lid**, i.e., the time interval between the appearance of the target and the moment when the participant looks at one of the 4 covers. The accuracy of the Lid Grip, Eye Movement Initiation, and Gaze On Lid was also recorded. Behavioral measurements were recorded using collider (a virtual box that is triggered when an object touches or passes through this box) placed on different objects in the virtual scene. The eye movements were obtained as the gaze ray (determined by the eye-tracking system integrated into the helmet) collided with a virtual object.

Trials involving movement initiation before the target appeared were excluded (7%), because an earlier initiation of the movement according to the subjects could be at the origin of the observed results. Trials involving eye movement initiation before the end of cue appeared were excluded for Eye Movement Initiation (19%) and Gaze On Lid measures (36%). The latter percentage for Gaze On Lid is due to that often subject performed the action without gazing directly on the trays. Moreover, 3 participants for Eye Movement Initiation and 4 for Gaze On Lid were excluded due to a zero number of trials in one of the experimental conditions. Subject response times for Action Initiation and Lid Grip were computed from the onset of the target presentation. Subject response times for Eye Movement Initiation and Gaze On Lid were computed from the
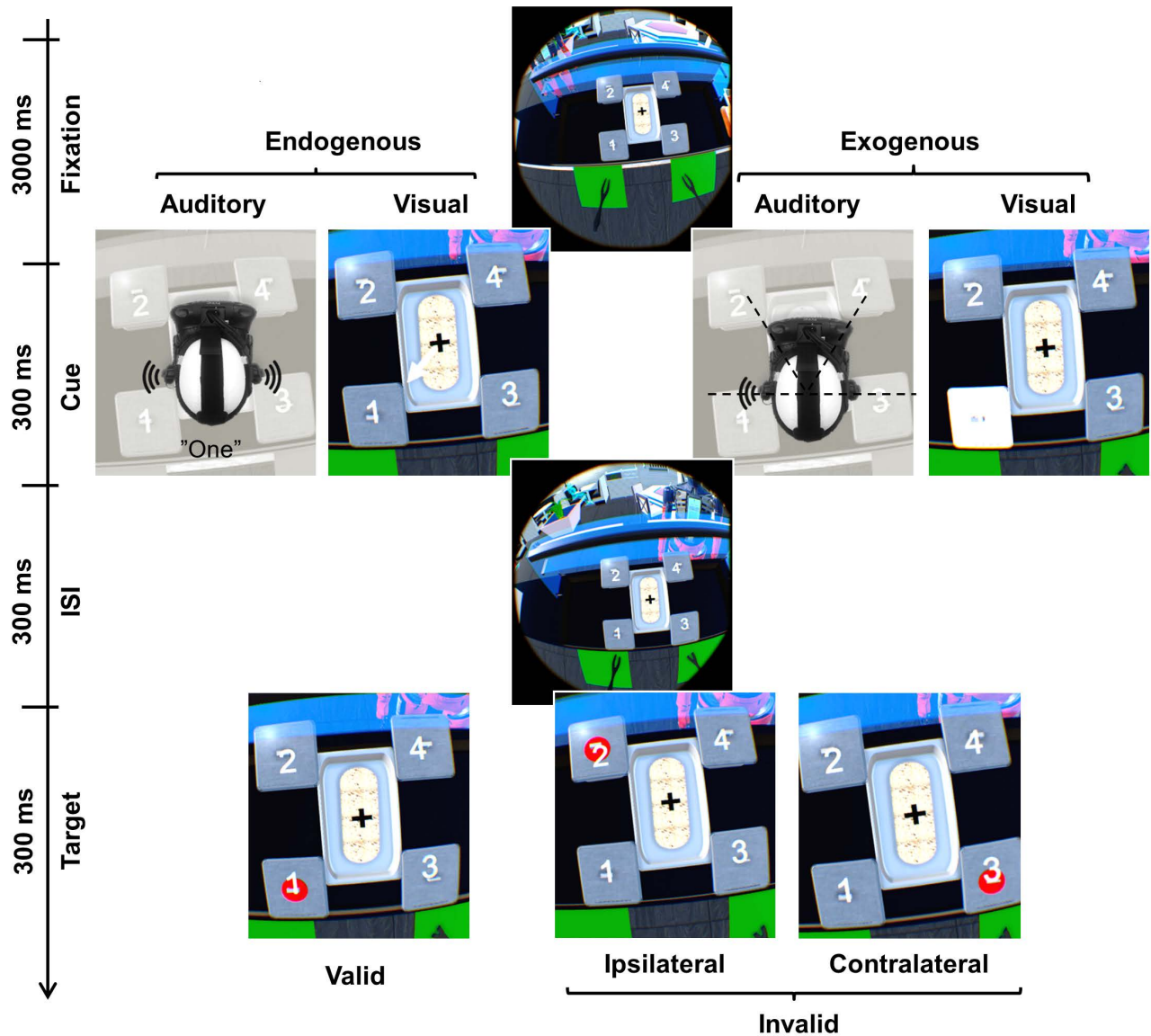
Figure 2: From top to bottom the time course of a trial: fixation of 3 seconds, cueing of 300 ms according to the block modality and type, 300 ms inter-stimulus interval, and 300 ms target presentation with a small red ball.

beginning of the trial, to take into account the trials when the first fixation on the lid was done during the target onset and because the gaze initiation often takes place before the target appearance. In total, for the statistical analysis we had 27.7 ± 2.3 out of 30 valid and 9.6 ± 0.7 out of 10 invalid trials for Action Initiation, 27.4 ± 2.4 valid and 9.4 ± 0.8 invalid trials for Lid Grip, 21.3 ± 5.7 valid and 7.7 ± 2.0 invalid trials for Eye Movement Initiation and 16.9 ± 4.8 valid and 6.2 ± 2.0 invalid trials for Gaze On Lid.

We performed a second analysis in order to compare which cues can direct attention within and across hemifield, implying a "cueing effect" for both the ipsilateral and contralateral targets; and which cues can direct attention only across hemifield, implying

a "cue validty effect" for the contralateral targets only. For this analysis, we excluded ipsilateral invalid cues, keeping valid cues and contralateral invalid cues. One participant for Action Initiation and Lid Grip, 4 participants for Eye Movement Initiation and 6 for Gaze On Lid were excluded due to a zero number of trials in one of the experimental conditions. For this second analysis the average number of recording was: 27.6 ± 2.3 valid out of 30 and 5.0 ± 1.4 out of 10 invalid trials for Action Initiation, 27.3 ± 2.4 valid and 4.9 ± 1.4 invalid trials for Lid Grip, 21.2 ± 5.7 valid and 4.1 ± 1.5 invalid trials for Eye Movement Initiation, and 16.0 ± 5.4 valid and 3.2 ± 1.4 invalid trials for Gaze On Lid. Statistical analysis indicated
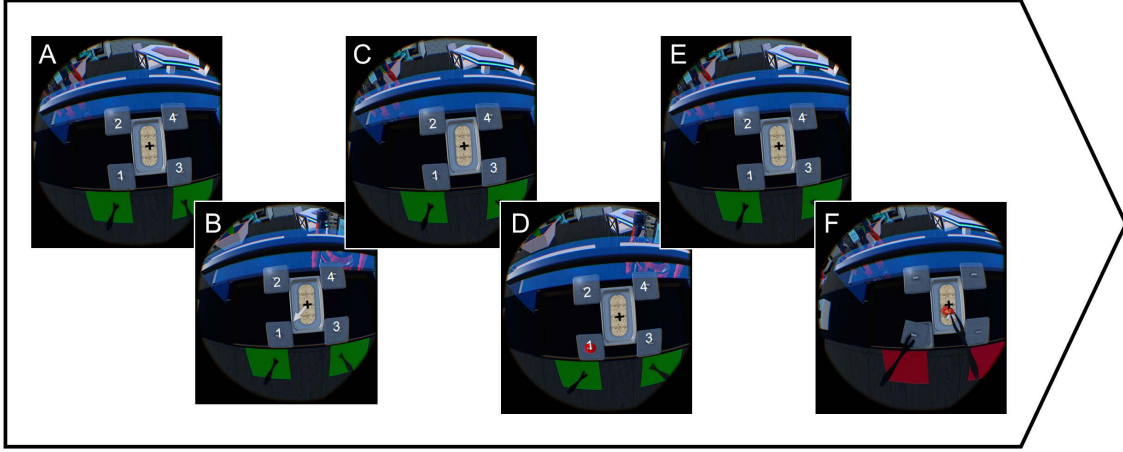
Figure 3: An example of one valid trial where the bottom left lid was cued using central directional arrow: A) fixation on the central cross to start the trial, the hands position controlled using interactive boxes that are green when the controller is inside; B) cueing of a tray (here an example of endogenous visual cueing of the tray 1); C) ISI; D) target presentation, the trial is valid; E) response time before the action initiation; F) the participant used the controller to take the indicated ingredient.

that the numbers of lost trials were independent of experimental condition.

We used the STATISTICA software for the statistical analysis. We performed three-way repeated-measures ANalyses Of VAriance (ANOVA) on the average RTs for each measure [Action Initiation, Eye Movement Initiation, Lid Grip and Gaze On Lid] to observe the effects of: Validity (Valid/Invalid), Modality (Auditory/Visual) and Cue Type (Exogenous/Endogenous). Fisher LSD test was used for post-hoc comparisons.

3 RESULTS

3.1 Eye Movement Initiation and Gaze On Lid

The ANOVA revealed no significant main effects nor interactions neither for both analyses including all invalid cues nor including contralateral invalid cues only (all $p > 0.1$) for both Eye Movement Initiation and Gaze On Lid variables.

3.2 Action Initiation

3.2.1 All invalid cues. The main effect of validity was significant, $F(1, 20) = 19.3, p < 0.001, \eta_p^2 = 0.49$. Participants initiated their actions significantly faster in the valid condition compared to the invalid condition. The main effect of modality was also significant, $F(1, 20) = 7.1, p < 0.05, \eta_p^2 = 0.26$. Participants were faster to react following auditory cueing independently of the validity of the trial.

The analysis also revealed a significant interaction between the three independent variables: validity, modality and cue type, $F(1, 20) = 13.7, p = 0.001, \eta_p^2 = 0.41$ (see Figure 4). The interaction indicated that the cueing effect on the action initiation was dependent on the type of cue and the modality used. Post-hoc comparison indicated that the cueing effect was significant for the endogenous auditory cue ($p < 0.001$) and for the exogenous visual cue ($p = 0.001$). No significant differences were observed for the exogenous auditory cue nor for the endogenous visual cue.

No other main effects nor interaction were observed (all $p > 0.1$).

3.2.2 Contralateral invalid cues only. As for the all invalid cues analysis, the ANOVA revealed significant main effects of validity, $F(1, 19) = 21.5, p < 0.001, \eta_p^2 = 0.53$, and modality, $F(1, 19) = 5.0, p < 0.05, \eta_p^2 = 0.21$.

The triple interaction was no longer significant, $F(1, 19) = 4.0, p = 0.064, \eta_p^2 = 0.17$. No other main effects nor interaction were observed (all $p > 0.1$).
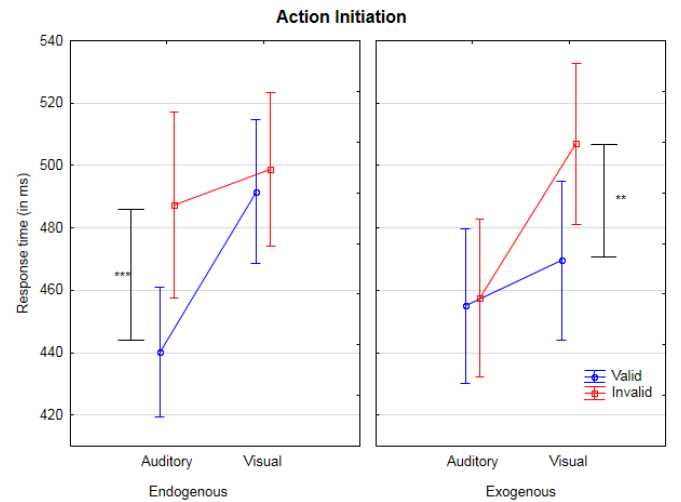

Figure 4: Mean response times for Action Initiation.
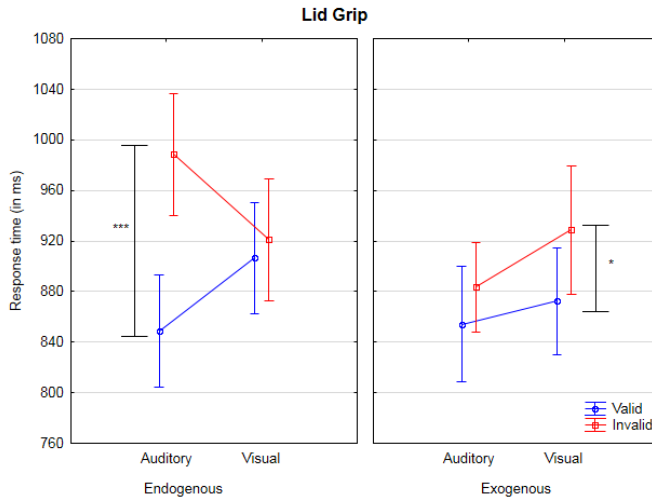$* = p < 0.05, ** = p < 0.01, *** = p < 0.001$

Figure 5: Mean response times for Lid Grip.
$* = p < 0.05, ** = p < 0.01, *** = p < 0.001$

3.3 Lid Grip

3.3.1 All invalid cues. ANOVA revealed significant the main effects of validity, $F(1, 20) = 37.9, p < 0.001, \eta_p^2 = 0.65$. Participants grab the lid significantly faster in the valid condition compared to the invalid condition. The main effect of cue type was also significant, $F(1, 20) = 6.8, p < 0.05, \eta_p^2 = 0.25$. Participants were faster to perform the action following exogenous cueing compared with endogenous cues.

The analyses also revealed a significant triple interaction, $F(1, 20) = 11.2, p < 0.01, \eta_p^2 = 0.36$ (see Figure 5). Post-hoc comparison indicated that the cueing effect was significant for the endogenous auditory cue ($p < 0.001$) and for the exogenous visual cue ($p < 0.05$). No significant differences were observed for the exogenous auditory cue nor for the endogenous visual cue.

No other main effects nor interaction were observed (all $p > 0.05$).

3.3.2 Contralateral invalid cues only. As for the all invalid cues analysis, the ANOVA revealed significant main effects of validity, $F(1, 19) = 39.6, p < 0.001, \eta_p^2 = 0.68$.

The triple interaction was no longer significant, $F(1, 19) = 1.7$, p=0.21, $\eta_p^2 = 0.08$. No other main effects nor interaction were observed (all $p > 0.1$).

4 DISCUSSION AND CONCLUSION

The objective of the current study was to determine the effectiveness of endogenous and exogenous cues in both auditory and visual modalities. A modified version of Posner's paradigm [1980] demonstrated that the use of perceptual cues to orient a person's attention could improve the performance and speed of visual information processing. Unfortunately, most of these studies are carried out in the laboratory on a standard computer screen, experimental conditions lacking ecological validity. We, therefore, designed an immersive ecological version of Posner's paradigm that allows us to determine through behavioral measures the effectiveness of different types of modality (auditory/visual) and different types of cues

(exogenous/endogenous) on information processing. The use of a virtual reality helmet equipped with an eye-tracking system also allowed us to control and evaluate the influence of overt orienting on the results obtained.

4.1 Cueing effect

As highly expected, we observed a main effect of the cue's validity. Participants process information more quickly when cue-target relation was valid, as compared to invalid. Participants processed information more quickly when the cue directed their attention to target location than when she direct their attention to the wrong place. The use of visual or sound cues can direct a person's attention and improve the information processing in virtual reality. It would, therefore, seem that the knowledge obtained in the laboratory through experimentation on a standard computer screen can be transposed into an immersive ecological environment. It may also be that the cognitive mechanisms of attention identified in laboratories can be transposed in conditions close to everyday life. However, the use of virtual reality may reveal differences not highlighted by these previous studies.

4.2 "Global" vs. "local" orienting

The main effect of validity without interaction in the second analysis indicates that the use of cues can direct a person's attention regardless of cues types (exogenous/endogenous) and modality (visual/auditory). This facilitation is observed on the subjects' premotor response times (action initiation) and motor (lid grip). However, this facilitation provided by all cues only exists if we exclude invalid ipsilateral trials, i.e., when the target appears in the same visual hemifield as the cue. If we consider all invalid trials, both ipsi- and contralateral, only the endogenous auditory cue (vocal instruction) and the exogenous visual cue (luminous flash) produce benefits (significant cueing effect), as indicated by the interaction effect. Therefore, we argue that the endogenous auditory cue (vocal instruction) and the exogenous visual cue (object highlighting) guide attention locally, which means that overt orienting of attention has been deployed to a specific region of the space. Emphasis is on the preservation of the cueing effect regardless of the target's hemifield of appearance for invalid trials (significant cueing effect within and across hemifield). On the other hand, the exogenous auditory cue (spatialized sound) and the endogenous visual cue (directional arrow) direct attention more globally, meaning that overt orienting of attention has been deployed on one full hemifield i.e., absence of cueing effect when cue and target occur on the same hemifield (significant cueing effect across hemifield only). Note that we interpret the lack of cueing effect within hemifield as a lack of processing cost for invalid trials in the ipsilateral condition (within hemifield) according to the literature [Mathôt et al. 2010; Umiltà et al. 1991]. However, without a proper neutral condition, we cannot infer whether the lack of cueing effect is due to the processing benefit or to an additional cost for valid trials in this condition (within hemifield). Further studies are needed to clarify this point.

A possible explanation of our results is the nature of the attentional selection. Indeed, two theories are proposed to describe the

processes of visual selection and attentional orienting. The first theory defends the idea that visual attentional selection takes place in a certain region of space [Eriksen and Hoffman 1973; Maringelli et al. 2001; Posner 1980] and is known as space-based attention. The benefits observed in Posner's paradigm would, therefore, be due to the processing preparation for a certain region of space independently of the objects contained in that region. A second theory defends the idea that selection is more based on objects rather than on a certain region of space [Egly et al. 1994a; Moore et al. 1998]. The selection would be spatial because an object necessarily occupies a certain region of space, but it is more the object than the region of space itself that would be selected. Numerous studies have shown that both modes of selection can influence the allocation of attention. However, it is still unclear which factors will lead to a focus on an object or a spatial location. Nevertheless, there is a consensus that both modes of selection coexist in the visual system [Mozer and Vecera 2005]. Possibly the cue type may determine the selection method. The hypothesis we then put forward is that the cues previously suggested as allowing a global attentional orienting (the directional arrow and spatialized sound) lack spatial accuracy. They provide unsufficiently precise information on the target's occurrence, leading to an object-based selection (the tray) rather than to a specific region of the space. We can assume that directional arrow and spatialized sound provide localization information with low spatial resolution (less spatially accurate) in comparison with tray highlighting or the verbal indication referring to a pre-defined spatial position. Therefore, if the directional arrow and the spatialized sound lead to an object-based orientation (due to their low spatial accuracy), considering hemifield processing differentiation, the processing benefits due to the object cueing can be extended to all identical objects in the same visual hemifield. This assumption is in line with the work done by Egly et al. [1994b] and Reuter-Lorenz et al. [1996], who suggested cerebral specialization as object-based attention. Besides, more recent work by Ozaki et al. [2009] showed that redirecting attention across hemifields, using a directional arrow, produces a significant cueing effect while redirecting within a hemifield does not. The fMRI analysis also revealed a dissociation of brain activation in the right posterior parietal region between an attentional reorientation within and across hemifield.

4.3 Covert vs. overt orienting

People can direct their attention to an object or a region of visual space covertly by allocating attention resources or overtly by performing an eye movement. One of the oldest questions on the attentional orienting is how much the attentional shift is independent of the gaze shift [Posner 1980]. Many studies suggested a close relationship between eye movement preparation and attention [Awh et al. 2006; Deubel and Schneider 2003; Rizzolatti et al. 1987]; or coupling of spatial attention and saccadic preparation, [Hoffman and Subramaniam 1995; Kowler et al. 1995] or dissociation according to exogenous or endogenous orienting [Smith and Schenk 2012]. Others supported a functional distinction between eye movement and attention [Juan et al. 2004; Posner 1980]. In our study, we found no significant effect on the validity of the cue on participants' eye movements. The cue validity does not impact the initiation of the gaze after target's occurrence or the first fixation on the lid. We

did not observe any difference in saccades latency (Eye Movement Initiation) as a function of cue validity, neither for exogenous nor for endogenous cueing. This result is consistent with the study of Juan et al. [2004] which showed that sensorimotor structures can direct attention covertly without preparing a saccade.

Therefore, we argue that, although eye movement and attention share a close relationship, there may be an attentional orienting without eye movement [Posner 1980]. Note that the reciprocal is false: we cannot have a gaze shift without a prior attentional shift [Hoffman and Subramaniam 1995; Peterson et al. 2004]. In the same way, our results also suggest that the eye movements of the subjects did not reveal the benefits provided by the cue, while the manual responses revealed a beneficial effect of the cue on the subject's information processing. The motor response times used to study the effects of attentional orienting cannot, therefore, be completely replaced by eye tracking (sometime qualified as a more direct measure, [Duc et al. 2008]). Eye tracking should preferably be combined with behavioral and/or physiological measurements as much as possible. Because covert attention benefit cannot always be measured by observing exclusively the eye movement of the subjects. Do not focus on visual information does not necessarily mean not paying attention to it. So, a visual pattern is not exactly the same as an attentional pattern. This is a well-known problem for eye-tracking researchers. Indeed, an eye-tracker can only follow the movements of overt attention and not those of the covered attention. This reminds us of the important implicit hypothesis of any attention research through the analysis of eye movements: "We assume that attention is linked to foveal gaze direction, but we acknowledge that it may not always be so" [Duchowski 2007]. While many studies accept this implicit assumption and emphasize it in the interpretation of their work [Bucher and Schumacher 2006], others confidently reject the impact that covered attention could have, in their research [Findlay and Gilchrist 1998]. Finally, this raises an interesting question for future studies on covert orienting, particularly for research in 360° immersive environments: how can we direct a person's attention outside the user's visual field and thus improve the performance of information processing ?

4.4 Implications for the procedural training

These results offer the possibility of designing virtual reality training courses, particularly about procedural learning. Indeed, procedural learning is guided by two essential factors: the transformation of declarative information into action and repetition through practice. As virtual reality allows unl imited repetition of a procedure, combined with attentional guidance, it could become a very effective procedure learning tool. Let's take the example of learning a checklist (a set of instructions to be performed by a pilot before a flight phase). The use of a voice instruction or a flash, for example, could enable the learner to better discriminate and identify the different tools and measuring instruments that the pilot must check before starting the flight phase and thus facilitate their indexation and entries in memory and consequently the transformation of the declarative information into action [Hoareau 2016]. In addition, a directional arrow or a spatialized sound could be presented first to direct attention to a part of the visual field (left or right) to improve the processing of the cue to be presented next. Finally, many studies

use a directional arrow to guide learners in virtual environment. If, as we suppose, directional arrows disallow local orientation under certain conditions, it would be interesting to reproduce some study using a flash or voice instruction to see if performance can be improved. As for example in the study by [Vembar et al. 2004], which used a directional arrow in addition to a 2D map to guide a person through a maze.

ACKNOWLEDGMENTS

REFERENCES

Edward Awh, Katherine M Armstrong, and Tirin Moore. 2006. Visual and oculomotor selection: links, causes and implications for spatial attention. *Trends in cognitive sciences* 10, 3 (2006), 124–130.

Ehsan Azimi, Alexander Winkler, Emerson Tucker, Long Qian, Jayfus Doswell, Nassir Navab, and Peter Kazanzides. 2018. Can mixed-reality improve the training of medical procedures?. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4065–4068.

Andrea Berger, Avishai Henik, and Robert Rafal. 2005. Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology: General* 134, 2 (2005), 207.

Hans-Jürgen Bucher and Peter Schumacher. 2006. The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media. *Communications* 31, 3 (2006), 347–368.

Fabio Buttussi and Luca Chittaro. 2018. Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE transactions on visualization and computer graphics* 24, 2 (2018), 1063–1076.

Ana B Chica, Elisa Martín-Arévalo, Fabiano Botta, and Juan Lupiáñez. 2014. The Spatial Orienting paradigm: how to design and interpret spatial attention experiments. *Neuroscience & Biobehavioral Reviews* 40 (2014), 35–51.

Heiner Deubel and Werner X Schneider. 2003. Delayed saccades, but not delayed manual aiming movements, require visual attention shifts. *Annals of the New York Academy of Sciences* 1004, 1 (2003), 289–296.

Albert Hoang Duc, Paul Bays, and Masud Husain. 2008. Eye movements as a probe of attention. In *Progress in brain research*. Vol. 171. Elsevier, 403–411.

Andrew T Duchowski. 2007. Eye tracking methodology. *Theory and practice* 328 (2007), 614.

Robert Egly, Jon Driver, and Robert D Rafal. 1994a. Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General* 123, 2 (1994), 161.

Robert Egly, Robert Rafal, Jon Driver, and Yves Starrveveld. 1994b. Covert orienting in the split brain reveals hemispheric specialization for object-based attention. *Psychological Science* 5, 6 (1994), 380–383.

Charles W Eriksen and James E Hoffman. 1973. The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics* 14, 1 (1973), 155–160.

John M Findlay and Iain D Gilchrist. 1998. Eye guidance and visual search. In *Eye guidance in reading and scene perception*. Elsevier, 295–312.

Todd C Handy, Amishi P Jha, and George R Mangun. 1999. Promoting novelty in vision: Inhibition of return modulates perceptual-level processing. *Psychological Science* 10, 2 (1999), 157–161.

Harold L Hawkins, Steven A Hillyard, Steven J Luck, Mustapha Mouloua, Cathryn J Downing, and Donald P Woodward. 1990. Visual attention modulates signal detectability. *Journal of Experimental Psychology: Human Perception and Performance* 16, 4 (1990), 802.

Charlotte Hoareau. 2016. *Elaboration et évaluation de recommandations ergonomiques pour le guidage de l'apprenant en EVAH: application à l'apprentissage de procédure dans le domaine biomédical*. Ph.D. Dissertation. Brest.

James E Hoffman and Baskaran Subramaniam. 1995. The role of visual attention in saccadic eye movements. *Perception & psychophysics* 57, 6 (1995), 787–795.

Susan Jang, Jonathan M Vitale, Robert W Jyung, and John B Black. 2017. Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education* 106 (2017), 150–165.

Chi-Hung Juan, Stephanie M Shorter-Jacobi, and Jeffrey D Schall. 2004. Dissociation of spatial attention and saccade preparation. *Proceedings of the National Academy of Sciences* 101, 43 (2004), 15541–15544.

Eileen Kowler, Eric Anderson, Barbara Dosher, and Erik Blaser. 1995. The role of attention in the programming of saccades. *Vision research* 35, 13 (1995), 1897–1916.

Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell me where to look: Investigating ways for assisting focus in 360

video. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2535–2545.

Francesco Maringelli, John McCarthy, Anthony Steed, Mel Slater, and Carlo Umilta. 2001. Shifting visuo-spatial attention in a virtual three-dimensional space. *Cognitive Brain Research* 10, 3 (2001), 317–322.

Sebastiaan Mathôt, Clayton Hickey, and Jan Theeuwes. 2010. From reorienting of attention to biased competition: Evidence from hemifield effects. *Attention, Perception, & Psychophysics* 72, 3 (2010), 651–657.

Cathleen M Moore, Steven Yantis, and Barry Vaughan. 1998. Object-based visual selection: Evidence from perceptual completion. *Psychological science* 9, 2 (1998), 104–110.

Michael C Mozer and Shaun P Vecera. 2005. Space-and object-based attention. In *Neurobiology of attention*. Elsevier, 130–134.

Bettina Olk, Alina Dinu, David J Zielinski, and Regis Kopper. 2018. Measuring visual search and distraction in immersive virtual reality. *Royal Society open science* 5, 5 (2018), 172331.

Takashi J Ozaki, Seiji Ogawa, and Tsunehiro Takeda. 2009. Dissociable neural correlates of reorienting within versus across visual hemifields. *Neuroreport* 20, 5 (2009), 497–501.

Matthew S Peterson, Arthur F Kramer, and David E Irwin. 2004. Covert shifts of attention precede involuntary eye movements. *Perception & psychophysics* 66, 3 (2004), 398–405.

Michael I Posner. 1980. Orienting of attention. *Quarterly journal of experimental psychology* 32, 1 (1980), 3–25.

Patricia A Reuter-Lorenz, Maxwell Drain, and Corinne Hardy-Morais. 1996. Object-centered attentional biases in the intact brain. *Journal of Cognitive Neuroscience* 8, 6 (1996), 540–550.

Giacomo Rizzolatti, Lucia Riggio, Isabella Dascola, and Carlo Umiltá. 1987. Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia* 25, 1 (1987), 31–40.

Alia Sheikh, Andy Brown, Zillah Watson, and Michael Evans. 2016. Directing attention in 360-degree video. (2016).

Daniel T Smith and Thomas Schenk. 2012. The premotor theory of attention: time to move on? *Neuropsychologia* 50, 6 (2012), 1104–1114.

Carlo Umiltà, Lucia Riggio, Isabella Dascola, and Giacomo Rizzolatti. 1991. Differential effects of central and peripheral cues on the reorienting of spatial attention. *European Journal of Cognitive Psychology* 3, 2 (1991), 247–267.

Deepak Vembar, Nikhil Iyengar, Andrew T Duchowski, Kevin Clark, Jason Hewitt, and Keith Pauls. 2004. Effect of visual cues on human performance in navigating through a virtual maze.. In *EGVE*. 53–60.

Sabine Webel, Uli Bockholt, Timo Engelke, Nirit Gavish, Manuel Olbrich, and Carsten Preusche. 2013. An augmented reality training platform for assembly and maintenance skills. *Robotics and Autonomous Systems* 61, 4 (2013), 398–403.