ELSEVIER

Applied Ergonomics 36 (2005) 335–343

APPLIED ERGONOMICS

www.elsevier.com/locate/apergo

Effects of increasing visual load on aurally and visually guided target acquisition in a virtual environment

Andrea C. Pierno[a], Andrea Caria[a], Scott Glover[a], Umberto Castiello[a,b,*]

[a]*Department of Psychology, Royal Holloway– University of London, Egham, Surrey, UK*
[b]*Dipartimento di Psicologia Generale, Università di Padova, Via Venizia 8, 35131 Padova, Italy*

Received 17 December 2003; accepted 30 November 2004

Abstract

The aim of the present study is to investigate interactions between vision and audition during a target acquisition task performed in a virtual environment. We measured the time taken to locate a visual target (acquisition time) signalled by auditory and/or visual cues in conditions of variable visual load. Visual load was increased by introducing a secondary visual task. The auditory cue was constructed using virtual three-dimensional (3D) sound techniques. The visual cue was constructed in the form of a 3D updating arrow. The results suggested that both auditory and visual cues reduced acquisition time as compared to an uncued condition. Whereas the visual cue elicited faster acquisition time than the auditory cue, the combination of the two cues produced the fastest acquisition time. The introduction of secondary visual task differentially affected acquisition time depending on cue modality. In conditions of high visual load, acquiring a target signalled by the auditory cue led to slower and more error-prone performance than acquiring a target signalled by either the visual cue alone or by both the visual and auditory cues.

Keywords: Target acquisition; Visual load; Dual task; Virtual reality; Head related transfer functions; Three-dimensional sound; Multisensory integration

1. Introduction

Recent advances in digital signal-processing technology and the development of electromagnetic position trackers have enabled the construction of virtual three-dimensional (3D) audiospatial displays. These displays have been used to aid visual target acquisition in many modern workstations including aircraft cockpits and training simulators (Bronkhorst et al., 1996; Begault and Pittman, 1996).

A number of studies have investigated the use of 3D auditory displays to reduce the workload and enhance the acquisition of visual targets (Nelson et al., 1998; Perrott et al., 1996). For example, in Perrott et al. (1996), participants sat at the centre of a geodesic sphere and detected a visual target presented at one of 264 different locations. Results demonstrated that the addition of a 3D virtual auditory cue produced a significant reduction in target acquisition time. Similarly, Nelson et al. (1998) demonstrated the beneficial effects of 3D virtually localised auditory cues on performance and perceived workload in a visual target acquisition task. In this study, visual targets were presented on a head-mounted-display (HMD) in three auditory cue conditions that differed in the amount of information provided: (1) localised auditory cues; (2) non-localised auditory cues; or (3) no auditory cues. The addition of localised auditory cues led to a significant improvement in target acquisition performance and to significant reductions in workload ratings as compared to conditions in which auditory information was either non-localised or absent.

---

[*] Corresponding author. Dipartimento di Psicologia Generale, Università di Padova, Via Venizia 8, 35131 Padova, Italy.
E-mail address: umberto.castiello@unipd.it (U. Castiello).

0003-6870/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apergo.2004.11.002

Other studies have focused on possible multisensory integration by including situations in which 3D auditory cues were presented together with a visual cue (Flanagan et al., 1998; Bronkhorst et al., 1996). For example, Bronkhorst et al. (1996) tested the effectiveness of a 3D virtual auditory cue presented in isolation or together with a visual cue in a flight simulation experiment in which participants had to locate and track a target aircraft as quickly as possible. The results of their experiment indicated that a 3D auditory cue could be as effective as a visual cue.

More recently, Flanagan et al. (1998) administered a target acquisition paradigm in which participants were required to locate a target presented outside the initial field of view on a HMD. The target location was cued by 3D auditory cues presented in isolation or together with a visual cue. Their results suggested that both visual and 3D auditory spatial cues reduced acquisition time dramatically as compared to unaided search.

Although all of these studies provided clear evidence that 3D auditory cues have beneficial effects on target detection performance, they did not test the effectiveness of 3D auditory cues depending on visual workload. However, a study by Bolia et al. (1999) did investigate this issue. In their study, participants searched, with or without the aid of a 3D auditory cue, for a visual target presented in isolation or in the presence of a variable number of visual distractors. Results indicated that the addition of the 3D auditory cue significantly decreased search time without a corresponding increase in error rate. Further, although search time linearly increased depending on the number of distractors, it was always faster than search time obtained in the control condition in which no cue was present.

Thus, various studies have examined the use of 3D auditory cues during visual target acquisition. Included among these are studies that have gone beyond the use of a single cue modality to address multisensory integration. Other studies have investigated the effects of using only 3D auditory cues to alleviate visual workload when visual search in a cluttered environment was performed. However, as of yet the effects of multisensory cues on target acquisition performance in situations of variable visual load have not been investigated. We believe that the use of multisensory cues and visual workload are interrelated issues that should be addressed together.

Consequently, the aim of the present study was to determine whether cues presented in different modalities might differentially affect target acquisition performance depending on the level of visual workload. To this end, we examined the acquisition of a visual target in a task in which the target was signalled by auditory and visual cues that could be presented either together or in isolation. The visual workload was varied by introducing a secondary visual task. Thus, the experimental question was whether in conditions of high visual load there would be a differential effect of cue modality on target acquisition performance. It could be hypothesised that overloading vision by means of a visual secondary task should decrease the efficacy of a visual cue. This is because visual resources might be mainly engaged in performing the secondary task. Therefore, it is reasonable to expect target acquisition performance using only a visual cue to be more affected by the introduction of a secondary task than performance using only an auditory cue or both a visual and an auditory cue.

2. Methods

2.1. Participants

All 8 participants were volunteers and naïve as to the purpose of the experiment. Their ages ranged from 19 to 41 years with a mean of 26 years. All reported normal or corrected to normal vision and normal auditory functioning.

2.2. Materials

A computer equipped with a Pentium III processor was used to present the visual cues and to record the time taken by participants to locate the target. Head orientation was monitored by a three-degrees-of-freedom head tracker (Intertrax^2, Intersense) that sampled head orientation at 256 Hz with the following angular range: pitch $\pm 80^\circ$, yaw $\pm 180^\circ$, roll $\pm 90^\circ$. The auditory cues were presented by means of an audio-module (TDT RP2.1 Real Time Processor, Tucker-Davis Technologies) designed for the delivery of 3D sound. The RP2.1 audio module was equipped with a Scharc digital signal processor (DSP) running at 50 MHz able to synthesise and process wide-band signals in real time (24-bit, 100 kHz bandwidth) with a 110 dB signal-to-noise ratio.

Each auditory cue was amplified through a precision power amplifier (TDT HB7 Headphone Buffer, Tucker-Davis Technologies) capable of delivering up to 1 W of power to headphones or other transducers. The HB7 is a stereo device with excellent channel separation, low signal distortion and a flat frequency response ranging from 10 to 100 kHz. The output gain can be set between 0 and -27 dB in 3 dB steps, which allows for matching of dynamic range to the desired output level. The signal to noise ratio is 117 dB (20 Hz–80 kHz).

The sound was delivered through a pair of earphones (ER-6 Isolator, Etymotic Research) with the following specifications: frequency response 20 Hz–16 kHz, impedance 48 Ω, 1 kHz sensitivity 108 dB SPL for a 4.0 V input, maximum output 120 dB SPL, maximum continuous input 2.5 V RMS. When properly sealed in the ears, these earphones are able to provide 15–20 dB of external noise exclusion.

Visual displays and the visual cues were presented using a non-stereoscopic binocular HMD (Glasstron Sony PLM-S700E). The HMD provided a 30° horizontal and 22.5° vertical field of view (FOV) with an 800×600 pixel resolution (display mode VESA SVGA, vertical refresh frequency 60 Hz, horizontal refresh frequency 49.6 Hz). In order to avoid light or other visual distractors that might have influenced the perception of the presented visual displays, the HMD was inserted in a flexible rubber mask that sealed hermetically to the participant's skin.

The total system latency, that is the time elapsed from the transduction of a real event or action until the consequences of that event or action are perceivable within the virtual environment, was 26 ms (SD 6.3 ms).

2.2.1. *Visual virtual environment*

The virtual environment was a sphere centred on the observer's head. The portion of the sphere where the targets could be presented extended over 360° in azimuth (from 0° to 360°) and 60° in elevation (from −30° to 30°). The starting position was defined as 0° azimuth and 0° elevation. Negative azimuths were to the left of the starting position, positive to the right. Negative elevations were below the starting position, positive were above.

2.2.2. *Visual display*

Two nested circles, the outer with a radius of 3.8° of visual angle and the inner with a radius of 1.9° of visual angle served as a sight attached to the participant's head. The sighting circles were always presented in the centre of the FOV of the HMD. A square-shaped window with a side length of 2° of visual angle was always presented together with the sight at 0.4° of visual angle above the vertical diameter of the outer circle (see Fig. 1, panel 'a').

In some conditions, a series of sequentially presented numbers was displayed inside the window. Each number subtended 1.3° height and 0.6° width of visual angle. The sequence was composed of numbers ranging from 4 to 9. The numbers 0–3 were excluded for methodological reasons as explained in the Section 2.3. Each number was presented for 150 ms followed by a 200 ms blank interval before the onset of the subsequent number.

2.2.3. *Visual target*

The visual target was a white plus sign that subtended 1.6° of visual angle both vertically and horizontally. The possible target location in the virtual environment was defined by the combination of four azimuths (0°, 90°, 180° and −90°) and three elevations (0°, 30° and −30°). Given that the 0° azimuth and 0° elevation position served as the starting position, this position was excluded as a target location. Thus, the target could be presented at 11 of the 12 possible locations corresponding to the combination of the four azimuths and the three elevations reported above.

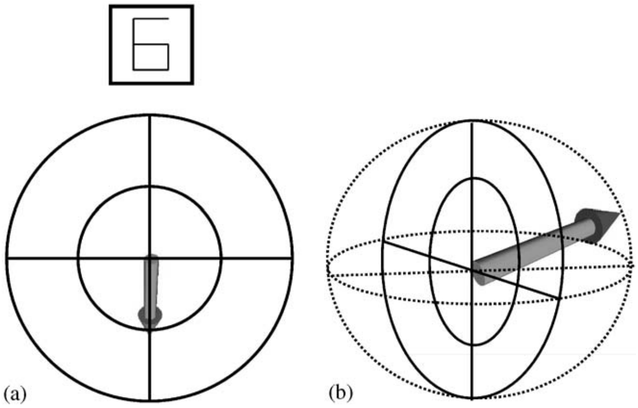

Fig. 1. Panel 'a' shows the sight, the square-shaped window in which the sequence of numbers was presented, and the 3D visual cue pointing to a target located in front of the observer (azimuth 0°) below the line of sight (−30° elevation). Panel 'b' shows a rotated view of both the sight and the 3D visual cue pointing to a target located in front (azimuth 0°) of the observer above the line of sight (30° elevation). The dashed line represents the rotational frame on which the 3D visual cue indicated target's location.

2.2.4. *Visual cue*

The visual cue used to indicate target location was a 3D continuously updated arrow presented on the sighting circle. The term "continuously updated" here means that the direction of the arrow was changed to indicate target location relative to the instantaneous head position. The 3D arrow was composed of a cylindrical green body with a radius of 0.2° of visual angle and a conical red head with a radius of 0.4° of visual angle (see Fig. 1, panels 'a' and 'b'). Both were 3D VRML computer-generated objects. The end of the arrow's body was fixed on the centre of the sighting circle. The arrow was free to rotate on the x, y, and z axis and was designed to update its position in the virtual environment as a function of the target's position (azimuth and elevation) and the position of the participant's head. Therefore, the arrow's head was always pointing to the exact target location with respect to the position of the observer's head. The length of the arrow remained constant for each trial regardless of the target location.

2.2.5. *Auditory cue*

Before describing the auditory cue utilised in the present study a brief description of the mechanisms by which humans localise sounds will be provided. Sound localisation in humans is enabled by the auditory system's ability to detect small changes in interaural arrival time and sound level, as well as spectral modification to the incident sound wave by the head, torso, and pinnae. By means of 3D audio virtual techniques all the differences in the interaural arrival time as well as the spectral modifications can be simulated by filtering a sound with head related transfer functions (HRTFs). The HRTFs allow for the rendering of binaural sounds delivered through headphones in any desired location.

The 3D updating auditory cue adopted in this experiment consisted of a continuous train of pulses of white noise each with a rise time of 5 ms and a white noise duration of 60 ms followed by a fall time of 5 ms separated by a silence lasting 70 ms. The 3D auditory cue was played back at a conversion rate of 50 kHz and an intensity of 70 dB SPL measured at the eardrum. This cue provided information about the target's azimuth and elevation, and was updated at a rate of 7 Hz. For each noise burst, the appropriate function simulating the target location relative to the instantaneous head position was selected from the library of the adopted HRTFs (Kemar HRTFs, TDT). To generate 3D virtual spatial auditory cues non-individualised HRTFs were adopted.

2.3. *Procedure*

The participants performed the task in a room of approximately 3 × 3 m while sitting on a swivel chair wearing the HMD. The task was to locate the target in the virtual environment and to rotate the head and body so as to overlap the sight over the target for 500 ms. At the beginning of each trial participants were required to overlap the inner circle of the sight over the fixation point (0° azimuth and 0° elevation), and hold it in this location for 2.5 s. Subsequently, the fixation point disappeared and the colour of the sighting circle changed from red to white. This was a warning signal to the participants that the trial was about to begin. Following a delay that varied randomly from 0.8 to 1.5 s, during which participants were instructed to maintain the head orientated toward the fixation point, a colour change of the sighting circle from white to green was the signal to start searching for the target. Simultaneously with the start signal, the 3D auditory and the 3D visual cue were presented together or in isolation. A control condition was included in which no cues were given. To perform the task participants were allowed to make head and body movements on the swivel chair. Trials in which any head movement was recorded before the start signal occurred were discarded and re-presented at random times during the block. The target was always presented outside the initial FOV of the HMD in one of the eleven possible locations.

2.3.1. *Visual workload conditions*

In three different blocks three types of visual workload task were administered to the participants. In the control visual workload task, no numbers were presented within the window displayed above the sighting circles. In the passive visual workload task, a sequence of numbers was presented within the window above the sighting circles, but no response to these numbers was required. In the active visual workload task, a sequence of numbers was also presented, but a secondary task had to be performed. This secondary visual task was to detect and verbally report at the end of each trial how many times a certain number (target number) had appeared in the sequence. We only used numbers ranging from 4 to 9 to minimise possible interference between the numbers and the counting process. Given that the target number typically appeared 0–3 times, using the numbers 0–3 as target numbers could potentially have caused interference with the counting process.

Regardless of the type of task, participants were required to overlap the inner circle of the sight with the target and maintain that position for 500 ms. Following this the target disappeared, the presentation of any auditory/visual cues present was ceased, and the HMD background turned yellow. To start a new trial, participants were required to return to the starting position and relocate the sighting circle on the fixation point. In both the passive and active visual workload blocks, the sequence of numbers was initiated simultaneously with the start signal and the onset of the cue/s if any. When the sight centre was less than 10° away from the target the number presentation was ceased to facilitate the overlapping manoeuvre. The dependent measure was the acquisition time: the time from cue presentation to target acquisition. In the active visual workload block of trials the number and type of errors made in the secondary task were also recorded.

2.3.2. *Cue conditions*

There were four cue conditions: (i) absent, in which no cue was presented; (ii) auditory, in which only the auditory cue was presented; (iii) visual, in which only the visual cue was presented and (iv) combined, in which both the auditory and the visual cues were presented.

2.4. *Experimental design*

There were twelve experimental conditions created by the crossing of the three types of task (control, passive, and active) and four types of cue (absent, auditory, visual, and combined). Each participant completed 6 randomly ordered blocks: two for the control visual workload task, two for the passive load task and two for the active load task. Each block consisted of 44 randomly determined trials, 11 for each type of cue.

2.5. Data analysis

An analysis of variance (ANOVA) with types of cue (absent, auditory, visual, and combined) and types of task (control, passive, and active) as within-subjects factors was conducted. Planned post-hoc comparisons were performed using *t*-tests.

3. Results

The main factor type of cue was significant, $F(3,21) = 185.12$, $p < 0.0001$. Acquisition time was 2.850 s for the combined cue condition, 3.153 s for the visual cue condition, 4.396 s for the auditory cue condition, and 6.618 s for the absent cue condition (see Fig. 2). The main factor type of visual workload task was also significant $F(2,14) = 26.713$, $p < 0.0001$. Acquisition time was 3.886 s for the control visual workload task, 3.957 s for the passive visual workload task, and 4.920 s for the active visual workload task. The interaction between type of cue and type of task was also significant, $F(6,42) = 5.511$, $p < 0.0001$.

t-Tests revealed that within each type of task acquisition time was fastest for the combined cue conditions ($p_s < 0.05$). The auditory cue in isolation produced a slower acquisition time than the visual cue in isolation but a faster acquisition time than the absent cue condition ($p_s < 0.05$, see Fig. 2). *t*-Tests also revealed that acquisition time for each cue condition in the control visual load task was not significantly different from acquisition time for each cue condition in the passive visual load task ($p_s > 0.05$). These results indicate that no differences were present between the control and the passive visual workload task conditions. However, *t*-tests revealed that acquisition time for each cue condition was significantly faster for the passive than for the active visual load task (see Fig. 2, $p_s < 0.05$).

To further explore the interaction, acquisition times obtained for each cue condition in the passive visual load task were subtracted from acquisition times obtained for each cue condition in the active visual load task. These four differences were calculated: (D1) the difference between acquisition times obtained for the absent cue conditions in the active and in the passive visual load tasks (1.571 s); (D2) the difference between acquisition times obtained for the auditory cue conditions in the active and in the passive visual load tasks (1.111 s); (D3) the difference between acquisition times obtained for the visual cue conditions in the active and in the passive visual load tasks (0.609 s); and (D4) the difference between acquisition times obtained for the combined cue conditions in the active and in the passive visual load tasks (0.560 s).

These differences were then compared using planned comparisons to test whether during target acquisition the weight carried by the visual and auditory cues varied with respect to the type of task. A significant difference was found between D2 and D3 ($p < 0.05$). This result supports the idea that the secondary task has a different effect on acquisition time depending on the type of cue and may indicate, in contrast to our hypothesis, that it is more difficult to perform a secondary visual task when the auditory cue is presented than when the visual cue is presented.

To support this statement an ANOVA with type of cue as a within subjects factor was conducted on the errors made by participants in performing the active visual workload task. The ANOVA revealed a significant effect of the main factor type of cue [$F(3,21) = 5.153$, $p < 0.05$] with the highest percentage of error for the auditory cue condition followed by the absent, the combined and the visual cue conditions

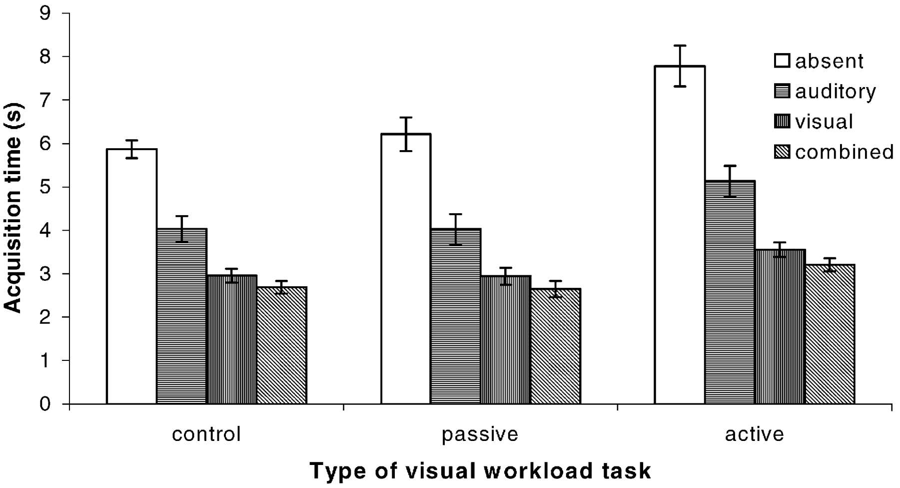

Fig. 2. Mean acquisition time as a function of the four types of cue and the three types of visual workload task.

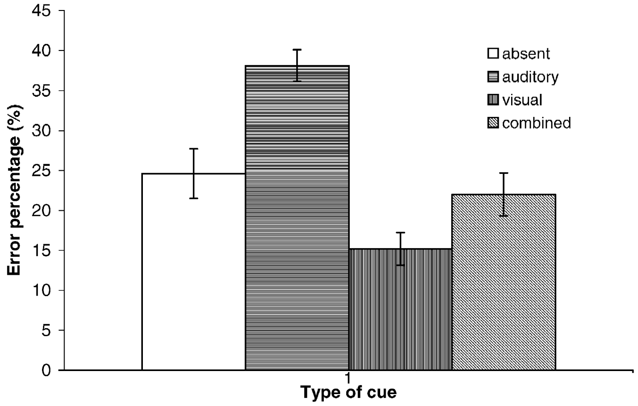

Fig. 3. Error rates obtained in the active visual workload task as a function of the different types of cue.

(38.2%, 24.6%, 22%, and 15.2%, respectively, for auditory, absent, combined, and visual cue conditions). Planned comparisons revealed that the percentage of errors obtained in the auditory cue condition was significantly higher than that obtained in the visual cue condition and in the combined cue condition (see Fig. 3, $p_s < 0.05$). This would confirm that an active visual load differentially affects target acquisition performance depending on cue modality. In particular, performing the secondary visual task in the presence of an auditory cue is more difficult than performing the same task in the presence of a visual cue or in the presence of both the auditory and visual cues.

4. General discussion

The aim of the present study was to investigate the effects of cues presented in different modalities (i.e. vision and audition) during a target acquisition task involving different level of visual load.

There were three main findings of the present study. First, the combination of the auditory and the visual cues led to better performance in target acquisition than when the two cues were presented in isolation. Second, regardless of the level of visual workload both cues produced faster acquisition time than that obtained in the condition in which no cue was provided. However, the presence of the visual cue elicited better performance than the auditory cue. This may be related to the use of non-individualised HRTFs. It is important to point out that the spectral modulations produced by the listener's head, torso, and pinnae on the incident sound wave are highly individualised. The use of generic rather than individualised HRTFs might have reduced the participants' ability to correctly localise the auditory cues. Further, it has been demonstrated that when non-individualised HRTFs are employed front-back discrimination errors are magnified with participants more likely to perceive sounds located in the front hemifield as though they were in the rear hemifield (Wenzel et al., 1993). Third, increasing visual workload produced a generalised slowing of acquisition time for all cue conditions. Nevertheless, target acquisition performance was most affected by an increase in visual load when the auditory cue was presented in isolation.

The present results are in line with the Bolia et al. (1999) study in which it was reported that, although aurally guided search performance was better than unaided performance, it became progressively worse with a corresponding increase in the number of visual distractors. However, it is worth noting that there are fundamental differences between Bolia et al. (1999) and the present study. For instance, the Bolia et al. study lacked a condition in which visual search is guided by a visual cue and participants are not required to perform a secondary task. We believe that the findings of the present experiment extend the work of Bolia et al. (1999) as our paradigm allowed us to demonstrate not only the utility of 3D auditory cues in aiding target acquisition but also to compare effects elicited by cues delivered in different modalities under different visual load conditions.

Our original hypothesis predicted that the visual cue would be less effective in aiding target acquisition in high rather than low visual load conditions. Further, we also expected acquisition time with the auditory cue to be less affected by the secondary visual task than acquisition time obtained with the visual cue. This is because overloading vision by means of a visual secondary task should result in a decrease in the ability to process the visual cue but not the auditory cue. In other words, given that the auditory channel was otherwise unencumbered, the processing of the auditory cues should have been less affected by the introduction of the secondary visual task.

In contrast to our prediction the increase in visual workload had its greatest effect on target acquisition time in conditions in which the auditory cue was presented in isolation. We propose two possible explanations to account for these findings. The first posits a common pool of attentional resources that shifts across space and sensory modalities. When participants are required to perform the secondary task it is reasonable to hypothesise that they allocate the great majority of attentional resources to the small square window where the stimuli for the secondary task is presented. The onset of the exogenous auditory cue originating from the target would capture part of these resources to its location leaving a small amount available for the secondary task. Because in this experiment the cue/s and the secondary task are present for the entire duration of the trial, we suggest that this shifting process may be responsible for both the high number of errors and the slowing of acquisition times.

These results are in line with previous studies in which voluntary orienting of attention towards a central spatial location was interrupted by reflexive orienting to random peripheral flashes (Müller and Rabbitt, 1989). Our results extend these findings to the crossmodal level given that peripheral auditory cues seem to be able to produce similar effects as random peripheral flashes.

The second explanation also refers to the notion of a common pool of attentional resources. If the secondary task requires a considerable amount of attentional resources, then less resources might be available to interpret or process the auditory cue when the secondary task is performed. This explanation may appear in contrast to the fact that even in the high visual workload condition the beneficial effects triggered by the simultaneous presentation of both the auditory and the visual cues were still obtained. If it is true that fewer resources are available to interpret the auditory cue in high visual load conditions, then this should be the case even when the auditory cue is presented together with the visual cue. To explain why benefits are still found when the two cues are combined we propose that the presence of the visual cue may play a role in disambiguating the spatial information carried by the auditory cue. Thus when the visual cue is also present the amount of attentional resources allocated to the auditory cue would be smaller than when the visual cue is not present. This is in line with previous evidence suggesting limitations in the localisation of virtual as compared to real auditory cues (Wenzel, 2001). If this is the case, it may be argued that participants needed more time and attentional resources to process the virtual auditory cue especially if presented in isolation.

It should also be pointed out that the low level of performance obtained in the active visual load task when the auditory cue was presented in isolation may be interpreted in terms of multiple resource theory (Wickens, 1984). For example, it is possible that the auditory cue tapped the verbal resources required for both the digit reading and the counting task, in addition to the spatial resources required for acquiring the visual target. On the other hand, verbal resources would not likely be tapped by the arrow cue. This would explain why performing the visual secondary task was more difficult when the auditory cue was presented in isolation.

In terms of multisensory integration, our results suggest that the simultaneous presentation of visual and auditory cues may be the best means of signalling target location. This evidence is in contrast with a similar study conducted by Flanagan et al. (1998) in which the simultaneous presentation of the auditory and the visual cues did not result in any noticeable benefit. It is worth noting that in Flanagan et al. (1998), the visual cue was a 2D arrow that indicated target location by rotating on a top view radar display. The auditory cues, however, were 3D sounds similar to those used in the present investigation. In contrast to the auditory cues that are egocentric in nature, the information carried by the arrow utilised by Flanagan et al. (1998) needs to be extracted from an exocentric reference frame and subsequently transformed to an egocentric reference frame that can be used to guide movements. Thus in the Flanagan et al. (1998) study the visual and the auditory cues were based on non-comparable reference frames, with the former being exocentric and the latter egocentric. Instead, we utilised a visual cue that was able to indicate target location in three dimensions as it was free to rotate on the x, y, and z axes. Further, our 3D visual cue relied on a reference frame that in principle should be similar to that of the auditory cues. This is because the rotational frame on which the arrow acted was very similar to the frame on which the auditory cues were encoded. Thus, auditory and visual cues in the present study relied on more similar frames of reference that may have been be easier to integrate than those in Flanagan et al. (1998). It could be speculated that this is a crucial factor in eliciting the benefits obtained in our study when both the visual and the auditory cues were simultaneously presented.

Although we are aware that the visual and the auditory cues used in our study are preliminary encoded in eye centred and head centred reference frames respectively, we suggest that it is the similarity in terms of egocentricity that leads to the formation of a common multisensory representation. The information contained within this representation is accessible and can be used by the two modalities. In this respect, it is generally agreed that a common reference frame is required if stimuli from different senses are to elicit the same behavioural response (Zambarbieri et al., 1982, 1995; Welch and Warren, 1986).

Four main conclusions can be drawn from the present study. The first conclusion is concerned with the manipulation of visual load. Surprisingly, the efficacy of 3D auditory cues in aiding target acquisition was reduced by the introduction of a secondary visual task. The second conclusion is methodological in nature. The use of a target acquisition task performed in a virtual environment allowed participants to get constant feedback of their actions. This could lead to new conditions for the investigation of multisensory integration. This new technique could potentially reveal mechanisms concerned with the search for objects out of the field of view, and could prove valuable in the study of complex actions involving head and body movements. The third conclusion is that 3D virtual auditory cues generated by means of non-individualised HRTFs are effective in cueing the location of visual targets presented out of the participant's FOV, but a 3D visual cue seems to be more effective in guiding visual search than a 3D auditory cue generated by means of non-individualised HRTFs. Wheras Martin et al., (2001) demonstrated that localisation with individualised virtual auditory cues can be as efficient as free-field localisation, our results suggest that this statement can not be extended to conditions in which auditory cues are synthesised using non-individualised HRTFs. Finally, the fourth conclusion is that the audio-visual information provided on similar frames of reference by the two cues can be bound together to elicit a cooperation between the two modalities.

5. Implication for design

Data obtained in the present research suggest that, although 3D auditory cues can be effectively used to aid visual target acquisition, they probably require a certain amount of attentional resources to extract and resolve the spatial information they carry. Thus the use of 3D auditory cues in situations in which attentional resources are required for the execution of other concurrent tasks is not the best means of aiding target acquisition. However, presenting auditory and visual cues that rely on similar reference frames not only elicits audio-visual integration, but also may reduce the amount of attentional resources needed to extract the spatial information carried by 3D auditory cues. Therefore, it seems that having a combination of audio-visual information is the most efficient way to aid target acquisition in conditions of high visual load.

Results of the present study may be helpful in the optimisation of design of systems to aid visual target acquisition in fighter jets or in ground combat vehicle cockpits in which the operator's visual channel is often overloaded and the need of providing efficient relevant spatial information can be of vital importance. For example, in a combat situation, numerous displays are often monitored simultaneously, and there may be a need to orient quickly to another display. In this case, it would appear that using a combination of both visual and auditory cues would lead to the most efficient acquisition of the critical display.

Finally additional research could be conducted to provide a more comprehensive picture. Given that the nature of the adopted auditory cue might have interfered with the secondary visual task it may be interesting to run a similar experiment in which the secondary task would not require any verbal resources. One possible solution is to replace the string of number with a string of meaningless geometric shapes and ask participants to detect the presence or absence of a target shape. This would eliminate both the counting and the reading components involved in the adopted secondary task. Another possible solution is to radically change the way of increasing the visual load. This could be done by cluttering the visual scene by means of a variable number of visual distractors sharing one or more feature with the visual target. Whereas in such a situation the participants would not be required to actively perform a secondary visual task, it would be interesting to examine target acquisition performance as a function of the different types of cue when the complexity of the visual scene is increased by the presence of visual distractors.

Acknowledgements

A preliminary report of the work described herein was presented at the 10th Annual Meeting of the Cognitive Neuroscience Society (March 30th–April 1st, 2003, New York City). The authors would like to thank Andy Haswell and Philip Roberts for providing essential hardware support throughout all the phases of this investigation. Francesco Pavani and Massimo Turatto are also thanked for their useful comments on previous versions of this manuscript. Andrea Caria was supported by a Leverhulme Trust grant awarded to Prof. Umberto Castiello, and Andrea Pierno was supported by a Thomas Holloway scholarship to Prof. Umberto Castiello. Scott Glover was supported by a grant from the Economic and Social Research Council.

References

Begault, D.R., Pittman, M.T., 1996. Three-dimensional audio versus head-down traffic alert and collision avoidance system displays. Int. J. Aviat. Psychol. 6, 79–93.

Bolia, R.S., D'Angelo, W.R., McKinley, R.L., 1999. Aurally aided visual search in three-dimensional space. Hum. Factors 41, 664–669.

Bronkhorst, A.W., Veltman, J.A., van Breda, L., 1996. Application of a three-dimensional auditory display in a flight task. Hum. Factors 38, 23–33.

Flanagan, P., McAnally, K.I., Martin, R.L., Meehan, J.W., Oldfield, S.R., 1998. Aurally and visually guided visual search in a virtual environment. Hum. Factors 40, 461–468.

Martin, R.L., McAnally, K.I., Senova, M.A., 2001. Free-field equivalent localization of virtual audio. J. Audio Eng. Soc. 49, 14–22.

Müller, H.J., Rabbitt, P.M.A., 1989. Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. J. Exp. Psychol.: Hum. Percept. Performance 15, 315–330.

Nelson, W.T., Hettinger, L.J., Cunningham, J.A., Brickman, B.J., Haas, M.W., McKinley, R.L., 1998. Effects of localized auditory information on visual target detection performance using a helmet-mounted display. Hum. Factors 40, 452–460.

Perrott, D.R., Cisneros, J., McKinley, R.L., D'Angelo, W.R., 1996. Aurally aided visual search under virtual and free-field listening conditions. Hum. Factors 38, 702–715.

Welch, R.B., Warren, D.H., 1986. Intersensory interactions. In: Boff, K.R., Kaufman, L., Thomas, J.P. (Eds.), Handbook of Perception and Human Performance, vol. 1. Wiley, New York pp. 25-1–25-36.

Wenzel, E.M., 2001. Effect of increasing system latency on localization of virtual sounds with short and long duration. Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland.

Wenzel, E.M., Arruda, M., Kistler, D.J., Wightman, F.L., 1993. Localization using nonindividualized head-related transfer functions. J. Acoust. Soc. Am. 94, 111–123.

Wickens, C.D., 1984. Processing resources in attention. In: Parasuraman, R., Davies, D.R. (Eds.), Varieties of Attention. Wiley, New York, NY.

Zambarbieri, D., Schmidt, R.A., Magenes, G., Prablanc, C., 1982. Saccadic responses evoked by presentation of visual and auditory targets. Exp. Brain Res. 47, 417–427.

Zambarbieri, D., Beltrami, G., Versino, M., 1995. Saccade latency toward auditory targets depends on the relative position of the sound source with respect to the eyes. Vision Res. 35, 3305–3312.