Comparing Effects of 2-D and 3-D Visual Cues During Aurally Aided Target Acquisition

Andrea C. Pierno, Andrea Caria, and **Umberto Castiello,** Royal Holloway, University of London, Egham, United Kingdom

The aim of the present study was to investigate interactions between vision and audition during a visual target acquisition task performed in a virtual environment. In two experiments, participants were required to perform an acquisition task guided by auditory and/or visual cues. In both experiments the auditory cues were constructed using virtual 3-D sound techniques based on nonindividualized head-related transfer functions. In Experiment 1 the visual cue was constructed in the form of a continuously updated 2-D arrow. In Experiment 2 the visual cue was a nonstereoscopic, perspective-based 3-D arrow. The results suggested that virtual spatial auditory cues reduced acquisition time but were not as effective as the virtual visual cues. Experiencing the 3-D perspective-based arrow rather than the 2-D arrow produced a faster acquisition time not only in the visually aided conditions but also when the auditory cues were presented in isolation. Suggested novel applications include providing 3-D nonstereoscopic, perspective-based visual information on radar displays, which may lead to a better integration with spatial virtual auditory information.

INTRODUCTION

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

A number of studies have investigated the use of 3-D auditory displays to reduce the workload associated with an overloaded visual modality and to enhance the acquisition of visual targets (Bolia, D'Angelo, & McKinley, 1999; Nelson et al., 1998; Perrott, Cisneros, McKinley, & D'Angelo, 1996). For example, Perrott et al. had participants sit at the center of a geodesic sphere and detect a visual target presented at 1 of 264 different locations. Results demonstrated that the addition of a 3-D virtual spatial auditory cue produced a significant reduction in target acquisition time. Similarly, Nelson et al. demonstrated the beneficial effects of 3-D virtually localized auditory cues on performance and perceived workload in a visual target acquisition task. In that study visual targets were presented on a head-mounted display (HMD) in three auditory cue conditions that differed in the amount of information provided: localized auditory cues, nonlocalized auditory cues, or no auditory cues. Results showed that the addition of localized auditory cues led to a significant improvement in target acquisition performance and to significant reductions in workload ratings, as compared with when auditory information was either nonlocalized or absent.

Other studies have investigated possible multisensory integration effects by comparing situations in which 3-D auditory cues were presented in isolation or accompanied by a visual cue (Flanagan, McAnally, Martin, Meehan, & Oldfield, 1998; Bronkhorst et al., 1996). Bronkhorst et al. tested the effectiveness of a 3-D virtual auditory cue presented in isolation or together with a

Address correspondence to Andrea C. Pierno, Department of General Psychology, University of Padua, Via Venezia, 8, 35131 Padova, Italy; andreacristiano.pierno@unipd.it. **HUMAN FACTORS,** Vol. 46, No. 4, Winter 2004, pp. 728–737.

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 3-D auditory cue could be as effective as a visual cue. A point worth mentioning is that in their study, the visual cue was a top-view radar display located away from the initial line of gaze. This implies that in order to utilize the visual cue, participants had to move their head and eyes. This may have extended the search time because participants had to look away from the search field to examine the visual cue (i.e., the radar screen).

To circumvent this potential problem, Flanagan et al. (1998) designed a 2-D exocentric visual cue in the form of an arrow always presented in the participants' line of sight. The base of the arrow was at the center of a circular visual display, whereas the head of the arrow indicated the direction in which the observer's head should be rotated in order to bring the target into view. This arrow provided information on both target azimuth and elevation. They administered a visual search paradigm in which participants were required to locate a target presented outside the initial field of view on an HMD. The target location was cued by 3-D auditory cues presented in isolation or together with the 2-D visual cue. Their results suggested that both visual and auditory spatial cues reduced search time dramatically, as compared with unaided search.

Although the visual cue designed by Flanagan et al. (1998) was in plain sight and thus easier to use than the off-center radar screen used by Bronkhorst et al. (1996), it nonetheless required a significant amount of cognitive processing. Specifically, in contrast to the auditory cues that are egocentric in nature, the information carried by the arrow needs to be extracted from an exocentric frame of reference 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 auditory cues were based on noncomparable reference frames, with the former being exocentric and the latter egocentric. The natural question is whether this discrepancy affects target acquisition performance. Consequently the present research was designed to investigate possible differences in target acquisition performance depending on whether visual and auditory cues rely on similar or different reference frames.

EXPERIMENT 1

Experiment 1 was aimed at replicating the effects found by Flanagan et al. (1998) using a 2-D exocentric visual cue. Although this experiment was very similar to that conducted by Flanagan et al., we felt that this partial replication was necessary in order to establish a basis for comparison with Experiment 2, in which a different type of visual cue was utilized. Therefore in Experiment 1 we investigated target acquisition performance in a situation in which target location was cued by virtual 3-D auditory cues presented in isolation or accompanied by a 2-D exocentric visual cue.

Method

Participants. Thirteen volunteers participated in the experiment. All were naive as to the purpose of the experiment and had no previous experience in using HMDs. Their ages ranged from 18 to 40 years, with a mean of 27 years. None of the participants was familiar with virtual 3-D sound or had skills in performing visual search tasks in virtual environments. All reported normal or corrected-to-normal vision and normal auditory functioning.

Apparatus and 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 ±80°, yaw ±180°, roll ±90°. 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 3-D sound. The RP2.1 audio module was equipped with a Scharc digital signal processor (DSP) running at 50 MHz and capable of synthesizing and processing wide-band signals in real time (24-bit, 100-kHz bandwidth) with a 110-dB signal-to-noise ratio.

Before presenting the sound to the participants' ears, we amplified each auditory cue 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 Hz 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 to 16 kHz, impedance 48 ohms, 1 kHz sensitivity, 108 dB SPL for a 4.0-volt input, maximum output 120 dB SPL, maximum continuous input 2.5 volts root mean square. When properly sealed in the ears, these earphones are able to provide 15 to 20 dB of external noise exclusion.

Visual displays and the visual cues were presented using a nonstereoscopic 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 Video Electronics Standards Association Super Video Graphics Array, vertical refresh frequency 75 Hz, horizontal refresh frequency 46.9 kHz). In order to avoid light or other visual distractors that might influence the perception of the presented visual displays, the HMD was inserted in a flexible rubber mask that sealed to the participant's skin.

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

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 circle was always presented in the center of the FOV of the HMD.

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 consequence, the target could be presented at only 11 of the 12 possible locations corresponding to the combination of the four azimuths and the three elevations.

Visual cue. The visual cue used to indicate target location was a 2-D continuously updated arrow presented on the sighting circle. The term *continuously updated* here means that the length and direction of the arrow were changed to indicate target location relative to the instantaneous head position. The arrow's tail was positioned at the center of the sighting circle, and the tip was at the mapped target azimuth. The orientation and length of the arrow were updated at a frequency of 75 Hz in order to pinpoint the target azimuth relative to the instantaneous head position. Thus the arrow acted as a visual cue providing continuously updated information about the position of the target with respect to the observers' head. The length of the arrow expressed the angular distance between the participant and the target (*minimum length* refers to targets at 0° azimuth and *maximum length* refers to targets at ±180° azimuth, regardless of elevation). The orientation of the arrow indicated the target's azimuth but not its elevation (see Figures 1a and 1b).

Type of auditory cues. Three types of auditory cues were delivered: noninformative cue, transient spatial cue, and updated spatial cue. All cues were played back at a conversion rate of 50 kHz and an intensity of 70 dB SPL measured at the eardrum. All cues consisted of a series of pulses of white noise 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 noninformative auditory cue was composed of three successive pulses of white noise, identically presented in both ears, that did not provide any spatial information regarding the target's position. The transient spatial auditory cue consisted of three successive pulses of white noise played at the target azimuth and elevation

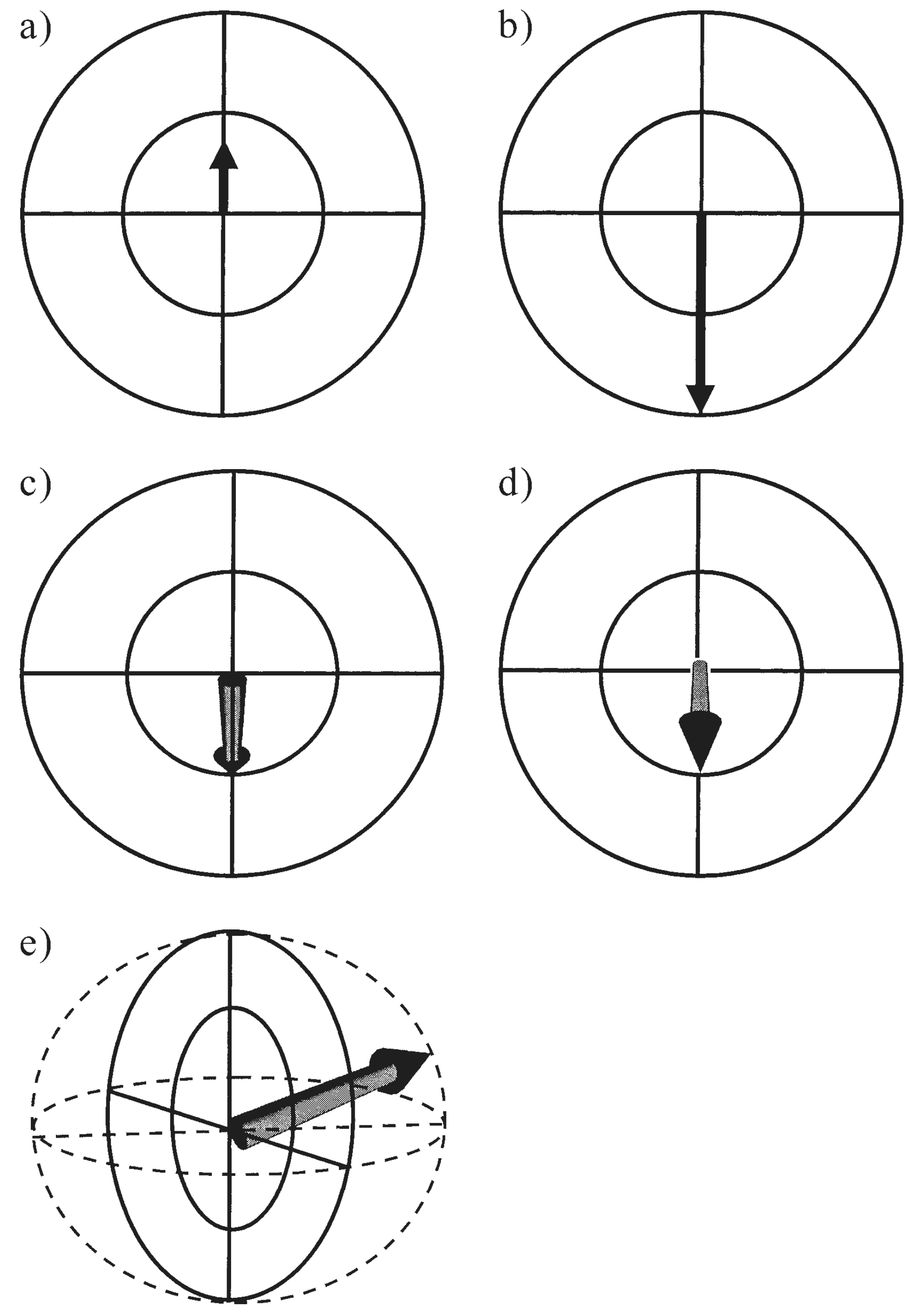

Figure 1. Panels a and b represent the visual display used in Experiment 1; panels c, d, and e represent the visual display used in Experiment 2. (a) The sight and the 2-D visual cue pointing to a target in front of the observer (azimuth 0°). (b) The sight and the 2-D visual cue pointing to a target located behind the participant (azimuth ±180°). (c) The sight plus the nonstereoscopic, perspective-based 3-D visual cue pointing to a target located in front of the observer (azimuth 0°) below the line of sight (–30° elevation). (d) The sight plus the nonstereoscopic, perspective-based 3-D visual cue pointing to a target located behind the observer (azimuth 180°) below the line of sight (–30° elevation). (e) A rotated view of both the sight and the nonstereoscopic, perspective-based 3-D visual cue pointing to a target located in front (azimuth 0°) of the observer above the line of sight (+30° elevation). The dashed lines represent the rotational frame on which the perspective-based visual cue indicated target's location.

and was able to provide information about the target's location. The updated auditory cue was a continuous train of successive pulses of white noise providing constantly updated information at a rate of 256 Hz about the target's location in relation to the head position. For each noise burst, the appropriate function simulating the target location relative to the instantaneous head position was selected to generate the proper spatial auditory cue from the library of the adopted head-related transfer functions (HRTFs; Knowles Electronics Manikin for Auditory Research, Tucker-Davis Technologies). To generate 3-D virtual spatial auditory cues, we adopted nonindividualized HRTFs.

Procedure. Participants performed the task in a room of approximately 3 × 3 m. They wore the HMD while sitting on a swivel chair. The task was to locate the target in the virtual environment and to rotate the head and body so as to overlap the sighting circle 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 and hold it in this location for 2.5 s. Subsequently, the fixation point disappeared and the color of the sighting circle changed from red to green. This was a warning signal to the participants that the trial was initiating. Following a delay that varied randomly from 1.5 to 3 s, during which participants were instructed to keep their head oriented toward the fixation point, the presentation of one of the auditory cues in isolation or together with the visual cue was the signal to start searching for the target.

To perform the task, participants were allowed to make head and body movements. Trials in which any head movement occurred before the presentation of the cue or cues were discarded and were presented again at random times during the block. The target was always presented outside the initial FOV of the HMD in 1 of the 11 possible locations. 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, any auditory and/or visual cues 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. The dependent measure was the acquisition time: the time from cue presentation to target acquisition.

Experimental conditions. Six experimental conditions were created by the crossing of the three auditory cues (noninformative, informative, and absent) and two arrow (present and absent) conditions: (a) the arrow was absent and the noninformative auditory cue was present; (b) the arrow was absent and the transient auditory cue was present; (c) the arrow was absent and the updating auditory cue was present; (d) both the arrow and the noninformative auditory cue were present; (e) both the arrow and the transient auditory cue were present; and (f) both the arrow and the updating auditory cue were present.

Each participant performed four randomly determined repetitions of the six conditions for all 11 target locations, resulting in a total number of 264 trials divided into four experimental blocks. Participants completed 33 trials of practice before the first block.

Data analysis. An analysis of variance with type of auditory cue (noninformative, transient, and updating sound) and visual cue (arrow present, arrow absent) as within-subjects factors was conducted. Post hoc comparisons were performed using Bonferroni-corrected t tests.

Results and Discussion

The main factor of auditory cue was significant, $F(2, 24) = 123.62$, $p < .0001$. For the updating sound conditions, acquisition time was faster than for the transient or the noninformative sound conditions (4.2, 4.5, and 6.0 s, respectively; ps < .05). The main factor of visual cue was also significant, $F(1, 12) = 206.22$, $p < .0001$. Acquisition time for the arrow-present conditions was faster than for the arrow-absent conditions (3.6 vs. 6.2 s). The interaction of Auditory Cue × Visual Cue was significant, $F(2, 24) = 132.55$, $p < .0001$. Bonferroni-corrected comparisons revealed that when the arrow was present, there were no differences in acquisition time with respect to the type of auditory cue (ps > .05; see Figure 2a).

When the arrow was absent, acquisition time was faster for the updating auditory cue condition than for the transient and the noninformative cue conditions (ps < .05; see Figure 2a). Interestingly, we also found that when the arrow

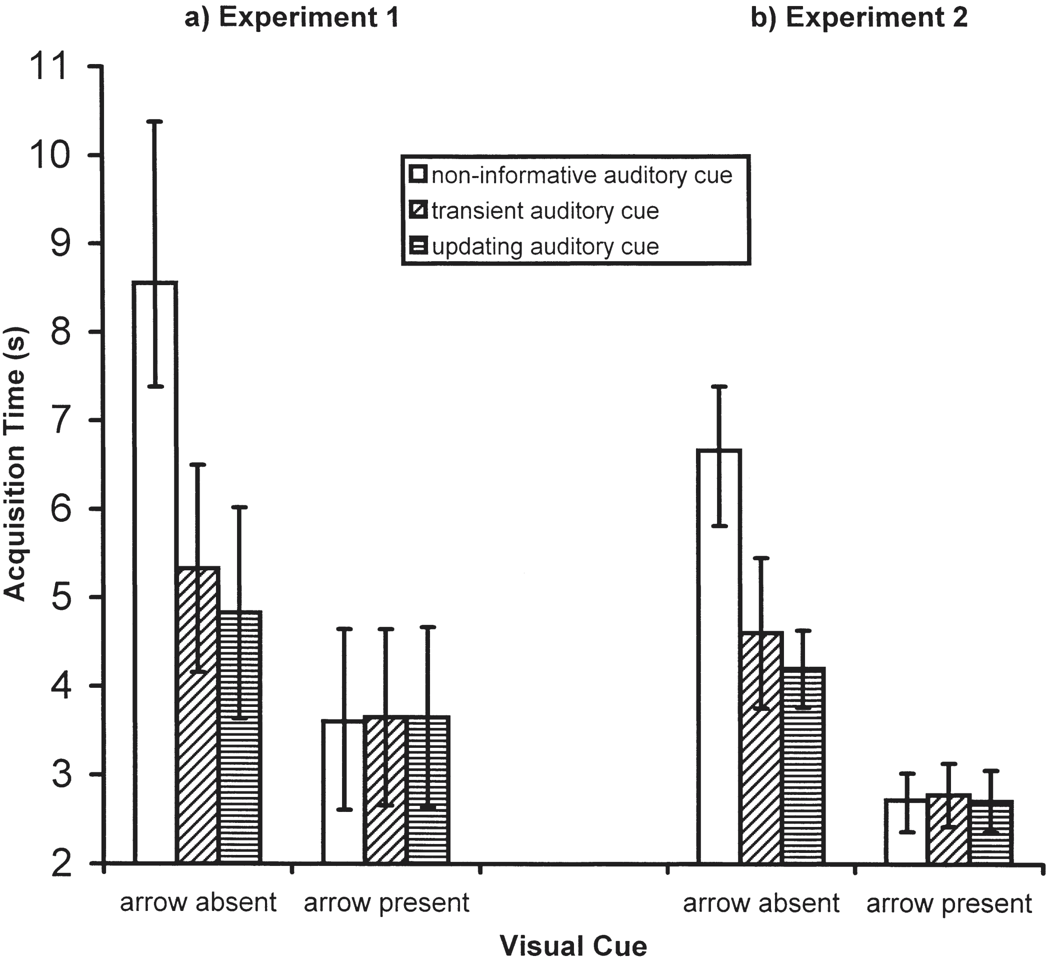

Figure 2. (a) Mean acquisition time as a function of the presence or absence of the 2-D visual cue and the type of auditory cue. (b) Mean acquisition time as a function of the presence or absence of the nonstereoscopic, perspective-based 3-D visual cue and the type of auditory cue.

was absent, acquisition time for the updating sound condition was significantly slower than that for all conditions in which the arrow was present (*p*s < .05). These results suggest that in the arrow-absent conditions, the 3-D updating auditory cue was more effective in reducing target acquisition time than were the transient and the noninformative auditory cues. The better performance for the 3-D updating auditory cue might be attributable to the fact that this type of cue provides the participant with continuously updated information regarding target location. This was not the case for the transient auditory cue because it provided only brief, nonupdated information about target location at the beginning of the trial. In the arrow-present conditions, target acquisition time was not affected by the presence of the different types of auditory cues. This may be because the presence of the visual cue attenuated the effects determined by the auditory cues that were found in the arrow-absent conditions.

In line with previous findings (Nelson et al., 1998; Perrott et al., 1996,) our results point out that even when generic rather than individualized HRTFs are used, 3-D virtual sounds are effective in cuing the location of visual targets presented out of the participant's FOV. Further, the data suggest that a constantly updated virtual sound seems to aid target acquisition performance better than a transient virtual sound does.

The present results are also in agreement with those obtained by Flanagan et al. (1998), who demonstrated that the presentation of a visual cue along the line of sight reduced acquisition time. In addition, the presence of both the visual and the auditory cues does did not enhance target acquisition performance in the present study (see Figure 2a). At first glance this may suggest that presenting similar information through multiple modalities does not always improve performance, in contrast to what Welch and Warren (1986) observed. However, two possible alternative explanations can also account for these results. The first is that the lack of additive effects of the auditory and visual cues could have been caused by redundancy in the information carried by the two types of cues. The second is that the lack of additive effects may be attributable to interference in processing information carried in conflicting frames of reference (i.e., egocentric vs. exocentric). All in all, these hypotheses point to the need to investigate situations in which the visual and auditory cues carry the same type of information (azimuth and elevation) and rely on similar frames of reference. The following experiment was aimed to clarify this issue.

EXPERIMENT 2

In Experiment 2 the 2-D visual cue was replaced by a nonstereoscopic, perspective-based 3-D visual cue that provided information about both target azimuth and elevation. The perspective-based visual cue also had the characteristic of relying on a reference frame that should, in principle, be similar to that on which the auditory cues are referenced. Thus auditory and visual cues should now provide participants with the same type of information in terms of azimuth and elevation and rely on similar frames of reference. These may be crucial factors in eliciting a better integration of the information carried by the two types of cues. Further, they may reveal cross talk between the visual and auditory modalities that would enhance target acquisition performance.

Method

Participants. Thirteen new participants with the same characteristics as those who took part in Experiment 1 were tested. Their ages ranged from 18 to 32 years, with a mean of 26 years.

Apparatus and material, stimuli, procedure, and data analysis. These were all similar to those used in Experiment 1, except that a perspective-based rather than a 2-D arrow was presented as a visual cue. The perspective-based 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 Figures 1c and 1d). Although not stereoscopically presented, they both were perspective-based 3-D Virtual Reality Modeling Language computer-generated objects.

The end of the arrow's body was fixed on the center of the sighting circle that was used as a rotation point. The arrow was free to rotate on the *x, y,* and *z* axes 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 (see Figure 1e). Therefore the arrow's head was always pointing to the exact target location with respect to the position of the observer's head. In contrast to the 2-D visual cue used in Experiment 1, the perspective-based visual cue was not referenced in a 2-D top-view display but operated within an egocentric 3-D frame. The length of the arrow remained constant for each trial regardless of target location.

Results and Discussion

The main factor of auditory cue was significant, $F(2, 24) = 220.29$, $p < .0001$. Acquisition time for the updating sound condition was faster than for the transient and the noninformative sound conditions (2.7, 3.5, and 5.6 s, respectively; $ps < .0001$). The main factor of visual cue was also significant, $F(1, 12) = 228.54$, $p < .0001$. Acquisition time for the arrow-present condition was faster than for the arrow-absent condition (3.3 vs. 4.5 s). The interaction between the type of auditory cue and visual cue was significant, $F(2, 24) = 46.23$, $p < .0001$. Bonferroni-corrected t tests revealed that when the arrow was present, there were no differences in acquisition time with respect to the type of auditory cue ($ps > .05$; see Figure 2b).

When the arrow was absent, acquisition time was faster in the updating 3-D auditory condition than in the transient and the noninformative auditory conditions (ps < .05; see Figure 2b). It was also found that acquisition time in the updating 3-D auditory condition was significantly slower in the arrow-absent condition than in all the conditions in which the arrow was present (ps < .05).

Comparison analysis. To test the possible differences in performance between the 2-D and the perspective-based 3-D arrow conditions, we conducted a mixed analysis of variance with type of arrow (2-D, perspective-based 3-D) as a between-subjects factor and type of auditory cue (noninformative, transient, and updating sound) and visual cue (arrow present, arrow absent) as within-subjects factors. The main effect of type of arrow was significant, $F(1, 24) = 8.908$, $p < .05$. Acquisition time for the perspective-based arrow was faster than that for the 2-D arrow (3.9 and 4.9 s, respectively).

The main factor of type of auditory cue was significant, $F(2, 48) = 223.49$, $p < .0001$. Acquisition time for the updating 3-D auditory conditions was faster than for the transient and the noninformative auditory conditions (3.8, 4.0, and 5.4 s, respectively). Significance of the main factor of visual cue, $F(1, 24) = 423.33$, $p < .0001$, indicated that acquisition time was faster for the arrow-present conditions than for the arrow-absent conditions (3.2 vs. 5.7 s). The interaction among the auditory cue, visual cue, and arrow (2-D, perspective-based 3-D) conditions was also significant, $F(2, 48) = 10.77$, $p < .001$. Post hoc Bonferroni-corrected comparisons revealed that for all the perspective-based arrow-present conditions, acquisition time was significantly faster than for the 2-D arrow-present conditions (ps < .05). Similarly, in the arrow-absent conditions, participants in Experiment 2 were faster than those in Experiment 1.

The perspective-based arrow used in Experiment 2 produced significantly faster acquisition times than did the 2-D arrow (see Figures 2a and 2b). This demonstrates that a perspective-based visual cue providing information on both target azimuth and elevation is more effective than a 2-D visual cue that provides information on only target azimuth. Further, in both experiments the presence of either type of arrow eliminated any possible effect of the different type of auditory cue (see Figures 2a and 2b).

Participants who experienced only the perspective-based arrow acquired the target faster in the conditions in which the arrow was absent than did participants who experienced only the 2-D arrow. As shown in Figures 2a and 2b, when the perspective-based visual cue was absent, target acquisition time was faster than when the 2-D visual cue was absent. This evidence suggests that the perspective-based visual cue may function as a 3-D trigger that also produces benefits for an exclusively 3-D aurally guided acquisition task. In other words, the experience of a perspective-based visual cue sets the participants within a frame of reference similar to that of the auditory cues. This may lead to the formation of a multisensory representation in which the information carried by the two cues is better integrated than in the condition in which the 2-D visual cue was presented. This better integration may allow for a more effective use of the information encoded in such a representation by the auditory modality.

CONCLUSIONS

The present study aimed to investigate the effectiveness of virtual auditory and visual cues for the acquisition of visual targets and to explore how multisensory integration between vision and audition may facilitate this process.

First and foremost, the present findings reveal that multiple sensory cues carrying similar types of information (i.e., azimuth and elevation) and relying on similar frames of reference can be integrated into a multisensory representation. This is line with previous literature suggesting that a common reference frame is required if stimuli from different senses are to elicit the same behavioral response (Zambarbieri, Beltrami, & Versino, 1995; Zambarbieri, Schmidt, Magenes, & Prablanc, 1982). The information contained within this representation is then accessible and available to both modalities independently. In other words, when presenting a nonstereoscopic, perspective-based 3-D visual cue along with 3-D virtual auditory cues, a cross-modal binding is elicited in terms of a long-lasting multisensory representation (at least for the duration of the experimental block). In this view the auditory modality can access this representation and use the multisensory trace even in conditions in which the visual modality is not directly involved.

This may explain why the target acquisition performance obtained in Experiment 2 for the arrow-absent conditions was superior to target acquisition performance obtained for the same conditions in Experiment 1.

A further point concerns the efficacy of the 3-D virtual sound. The present results are in line with previous evidence (Nelson et al., 1998; Perrott et al., 1996) that 3-D virtual spatial auditory cues generated by means of generic HRTFs can substantially reduce target acquisition time, as compared with conditions in which the target acquisition task is unaided. The updating spatial auditory cue led to shorter acquisition times than did the transient auditory cue, but both led to significantly faster acquisition times than did the noninformative auditory cue.

Two main conclusions can be drawn from the present study. The first concerns the applicability of these techniques in modern workstations. Our results suggest that a perspective-based visual cue may be easier to utilize than 2-D visual cues, such as those present on top-view radar displays. A possible reason for this is that our perspective-based visual cue is more easily remapped onto real-world coordinates.

In addition, the use of virtual sound rather than free-field listening techniques may allow the application of spatial auditory cues to prevent visual overloading in workstations such as the cockpit of an aircraft or a ground combat vehicle. Although it is thought that the use of individualized HRTFs might be an ideal means of providing 3-D virtual spatial information over headphones (Moller, Sorensen, Jensen, & Hammershoi, 1996), we have confirmed that adopting generic HRTFs allows a reasonably accurate spatialization of sounds that can be generalized across listeners (Wenzel, Arruda, Kistler, & Wightman, 1993; Wenzel, Wightman, Kistler, & Foster, 1988) and involves a noticeable reduction of time and technology resources, as compared with the adoption of individualized HRTFs.

The second conclusion is concerned with the nature of the multisensory information required to optimize visual search performance. In this respect, converging vision and audition onto similar frames of reference seems to be relevant for the exploitation of spatial information carried by virtual auditory cues.

Implications for Design

Data obtained in the present research suggest that the use of nonstereoscopic, perspective-based 3-D visual cues for the delivery of spatial information is more efficient than the top-view exocentric visual cues used by others (Flanagan et al., 1998; see also Bronkhorst et al., 1996). We believe that the level of egocentricity of such a visual display, along with the possibility for the visual cue to operate in three dimensions, could provide a more straightforward solution to the question of how to display 3-D information. Thus the results of the present study may be helpful in developing and optimizing the design of systems to aid visual target acquisition and in encoding and delivering 3-D perspective-based visual information so that it is more easily integrated into representations of the 3-D world.

ACKNOWLEDGMENTS

A preliminary report of the work described herein was presented at the 10th Annual Meeting of the Cognitive Neuroscience Society (March 30–April 1, 2003, New York). The authors wish to thank Andy Haswell and Philip Roberts for providing essential hardware support throughout all phases of this investigation. Scott Glover, Ken MacAnally, and Francesco Pavani 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 C. Pierno was supported by a Thomas Holloway scholarship to Prof. Umberto Castiello.

REFERENCES

Begault, D. R., & Pittman, M. T. (1996). Three-dimensional audio versus head-down traffic alert and collision avoidance system displays. *International Journal of Aviation Psychology, 6,* 79–93.

Bolia, R. S., D'Angelo, W. R., & McKinley, R. L. (1999). Aurally aided visual search in three-dimensional space. *Human 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. *Human 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. *Human Factors, 40,* 461–468.

Moller, H., Sorensen, M. F., Jensen, C. B., & Hammershoi, D. (1996). Binaural technique: Do we need individual recordings? *Journal of the Audio Engineering Society, 44,* 451–469.

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. *Human 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. *Human Factors, 38,* 702–715.

Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 1–36). New York: Wiley.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America, 94,* 111–123.

Wenzel, E. M., Wightman, F. L., Kistler, D. J., & Foster, S. H. (1988). Acoustic origins of individual differences in sound localization behavior [Abstract]. *Journal of the Acoustical Society of America, 84,* S79.

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 Research, 35,* 3305–3312.

Zambarbieri, D., Schmidt, R. A., Magenes, G., & Prablanc, C. (1982). Saccadic responses evoked by presentation of visual and auditory targets. *Experimental Brain Research, 47,* 417–427.

Andrea C. Pierno is a postdoctoral student in the Department of General Psychology, University of Padua, Italy. He received his Ph.D. in psychology at Royal Holloway, University of London, in 2004.

Andrea Caria is a staff member in the Department of Cognitive Sciences and Formation at the University of Trento, Italy. He received his master's degree in electronics-biomedical engineering in 2001 from the University of Genoa, Italy.

Umberto Castiello is a professor of psychology in the Department of General Psychology at the University of Padua, Italy and in the Department of Psychology at Royal Holloway, University of London. He received his D.Sc. in attention and motor control in 2002 from the University of Melbourne, Australia.

Date received: April 29, 2003
Date accepted: April 9, 2004