The Ominidirectional Attention Funnel: A Dynamic 3D Cursor for Mobile Augmented Reality Systems

Frank Biocca†, Arthur Tang†*, Charles Owen*, Xiao Fan*

†Media Interface and Network Design (M.I.N.D.) Laboratories
**Media and Entertainment Technologies Laboratory*
Michigan State University, East Lansing, MI 48824
Web-page: http://mindlab.org
E-mail: {biocca, tangkwo1, cbowen, xiaofan}@msu.edu

Abstract

Knowledge of objects, situations, or locations in the environment may be productivity enhancing, generally useful, or even life-critical for mobile augmented reality (AR) users. Users may need assistance with: (1) dangers, obstacles, or situations requiring attention, (2) visual search, (3) task sequencing, and (4) spatial navigation. The omnidirectional attention funnel *is a general purpose AR interface technique that rapidly guides attention to any object, person, or place in the space. The attention funnel dynamically draws user attention with a set of strong bottom-up spatial attention cues. In a study comparing the attention funnel to other attention directing techniques such as highlighting and audio cueing, we found that the attention funnel increased search speed by over 50%, and decreased perceived cognitive load by 18%. The technique has applicability as a general 3D cursor or cue in a wide array of AR applications that require visual search, emergency warning and alerts to specific objects or obstacles, or as a method for 3D navigation to objects in a room or guidance through space.*

1. Challenge of directing attention in mobile, augmented reality environments

Augmented reality (AR) interfaces are being incorporated into a mobile, wearable computers and integrated into pervasive computing systems [1]. Mobile AR systems offer the possibility of overlaying large volumes of context aware, virtual information onto the user's environment. Information objects such as labels, overlays, 3D objects, and other data are integrated into the physical environments. Objects, tasks, and locations can be cued when appropriate to support navigation and tasks of a mobile active user. In such systems we can foresee the continuous usage while mobile of 3D virtual objects, labels, and other information populating an already cluttered physical environment.

Mobile information-rich applications of AR begin to push up against a fundamental human factors limitation, the limited attention capacities of humans. The attention demands of more limited interfaces such as cell phones appear to contribute to accidents [2, 3]. One form of this challenge can be characterized as follows: How can an AR system successfully manage and guide visual attention to places in the environment where critical information or objects are present, even when they are not within the visual field? This question is part of a larger set of issues that we call ***attention management and augmentation*** in mobile AR and Virtual Reality (VR) interfaces.

1.1. Example scenarios where visual spatial cueing must support the user

To illustrate the benefits of AR system management of visual attention, consider the following common scenarios:

1.1.1. Telecollaborative spatial cueing. An emergency technician wears a camera and an AR Head-mounted Display (HMD) while collaborating with a remote doctor during a medical emergency. The remote doctor needs to indicate a piece of equipment that the technician must use next. What is the quickest way to direct her attention to the correct tool among a large and cluttered set of alternatives, especially if she is not currently looking at the tool tray and doesn't know the technical term for the tool?

1.1.2. Object search. A warehouse worker uses a mobile AR system to manage inventory, and is searching for a specific box in an aisle where dozens of virtually identical boxes are stacked. Systems integrated into the warehouse detect that the box is stored on a shelf behind the user using inventory records, an RFID tag, or other marker. What is the most efficient way to signal the location to the user?

1.1.3. Procedural cueing during training. A trainee repair technician uses an AR system to learn a sequence of steps where parts and tools are used to repair complex manufacturing equipment. How can the computer best indicate which tool and part to grab next in the procedural sequence, especially when the parts and tools may be distributed throughout a large workspace?

1.1.4. Spatial navigation. A tourist with a personal digital assistant (PDA) equipped with Global Positioning System (GPS) is looking for an historic building in a street with many similar buildings. The building is around the corner down the street. What is the fastest way to signal a path to the front door?

2. Attention management

Attention is one of the most limited mental resources [4]. Human cognitive capacity is limited. Attention is used to focus the limited cognitive capacity on a certain sensory input so that the brain can concentrate on processing the information of interest. Attention is primarily directed internally, from the "top down" according to the current goals, tasks, and larger dispositions of the user. Attention, especially visual attention, can also be cued by the environment. For example, attention can be user driven, i.e., "find the screwdriver," collaborator driven "use this scalpel now," or system driven "please use this tool for the next step."

Attention management is a central human-computer interaction issue in the design of interfaces and devices [5, 6]. For example, the attention demands of current interfaces such as cell phones and PDAs play a significant role in car accidents. The scenarios from the previous section illustrate various cases where attention must be guided, augmented or managed by the AR system or by another user communicating remotely.

2.1. Attention Cueing in Existing Interfaces

Users and interface designers have evolved various ways to direct visual attention in interpersonal interaction, architectural settings, and standard interfaces.

Attention cueing during interpersonal interaction. In interpersonal interaction, there are various sets of cues that are labeled ***indexical cues***. The phrase comes from the most obvious cue to visual attention, the pointing of an index finger directing the eyes to "look there." Similarly, we learn very early in life to monitor movement of other people's eyes, "drawing" a mental vector to the spatial location of the person's visual attention. These virtual vectors create an implicit cue of "look there." Gestures, eye movement and various other linguistic cues help disambiguate otherwise confusing spatial terms in languages such as "this," "that," "over there" and vague descriptive references to objects or locations in space.

Spatial linguistic cues can be the most ambiguous spatial cues. The meaning of spatial language (e.g., "left," "here," "in front of") varies with respect to the spatial reference frame of the speaker, listener and the environment. For areas that need accuracy, for example boating, theater, etc., conventions are used (e.g. stage left, stage right, dolly in) to partially solve this ambiguity problem.

The ambiguity of spatial language creates major communication problems when a computer system needs to communicate spatial content to a user, or when another person communicates to the user remotely through an AR system. Neither natural language nor non-verbal interactions in current interfaces are sufficient for complex and remote interactions.

Spatial cueing in windows interfaces. WIMP (window, icon, menu, and pointer) interfaces benefit from the assumption that user's ***visual attention is directed to the screen.*** Therefore, visual cues such as flashing cursors, pointers, radiating circles, jumping centered windows, color contrast, or content cues are used to direct user's visual attention to spatial locations on the screen surface. The integration of audio with visual cues helps draw attention even when vision is not directed to the screen. Of course, these systems work within the confines of a very limited amount of screen real estate, an area most users can scan very quickly. These techniques cannot easily cue

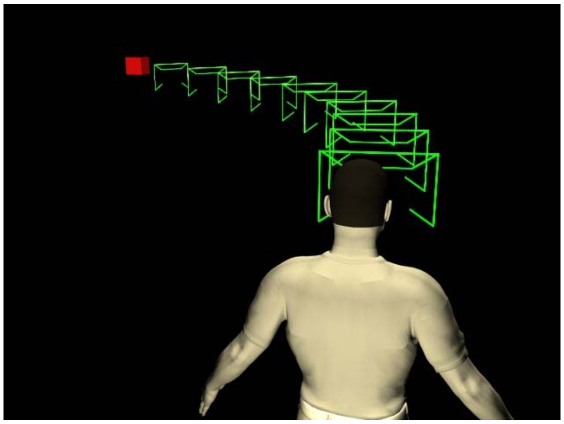

Figure 1. The attention funnel links the head of the viewer directly to an object anywhere around the body.

objects in the 3D environment around the user, for example pointing at an object behind the user.

Spatial cueing techniques used in interpersonal communication, WIMP interfaces, and architectural environments are not easily transferred to AR systems. In mobile AR environments, attention is shared across many tasks. Tasks in the virtual space are often not the primary user task. This is very different from typical computer tasks such as word processing in standard WIMP interfaces. Individuals may be walking freely in space, working with physical tools and objects, and interacting with others. The user may not be at the correct location in the scene, or looking at the correct spatial location or object needed to accomplish a task. When communicating with remote users, the indexical cues of interpersonal communication are not available or presented in a decreased modality, so finger pointing and eye gaze are useless and linguistic references "this," "that" and "over there" are even more ambiguous than in direct communication.

2.3. Spatial Cursors and Cueing Techniques in Augmented Reality Systems

In mobile AR environments, the volume of information is large and omnidirectional. AR environments have the capacity to display large amount of informational cues to physical objects in the environment.

Responsiveness is important for mobile multitasking computing environment. In a mobile multitasking setting, user's ability to detect a specific virtual or physical object at the appropriate time is limited. Visual attention is even more limited, since the system may have information about objects anywhere in an omnidirectional working environment around the user. Visual attention is limited to the field of view of human eyes ($<200^0$), and this limitation is further narrowed by the field of view of an HMD ($< 80^o$).

Most current AR systems adopt WIMP cursor techniques or visual highlighting to direct user's attention to an object (e.g., [7], [8]). These techniques may not be effective for mobile AR systems. System designer for AR interfaces cannot assume that the user is looking to the direction of the cued object. Audio can be used to cue the user to perform a search. But the cue provides limited spatial information. Spatialized audio and the human auditory system do not have the spatial resolution to inform spatial location precisely.

3. Omnidirectional Attention Funnel: A cursor paradigm for mobile 3D interaction.

The review of the need and limited implementation of a general technique for directing visual attention in 3D space suggests that interface design in a mobile AR system presents two basic challenges in managing and augmenting attention of the user:

(1) Omnidirectional cueing challenge. How to quickly and successfully cue visual attention to any location of physical or virtual objects when there is an immediate need, and

(2) Minimal attention demands. How to keep the demands of the virtual information from consuming or interfering with attention to tasks, objects, or navigation in the physical environment.

The **attention funnel** interface techniques have been designed as a general purpose interface paradigm to address the attention management challenge of mobile AR systems. The attention funnel is a spatial interface concept in the mobile infospaces project, a project (http://mindlab.org/web2/research/mobile.htm) examines human factors issues in the design of high volume, mobile AR systems.

The Omnidirectional Attention Funnel is an AR display technique for rapidly guiding visual attention to any location in physical or virtual space. The basic components of the attention funnel are illustrated in Figure 1. The most visible component is the set of dynamic 3D virtual objects linking the view of the user directly to the virtual or physical object. In spatial cognitive terms, the attention funnel visually links a head-centered coordinate space directly to an object centered coordinate space, funneling focal spatial attention of the user to the cued object. The

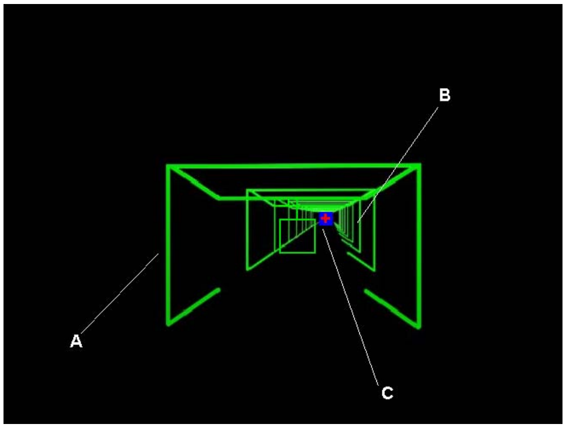

Figure 2. Three basic patterns are used to construct a funnel: (A) the head centered plane includes a bore sight to mark the center of the pattern from the user's viewpoint, (B) funnel planes, added in a fixed pattern (approximately every 0.2 meters) between the user and the object, and (C) the object marker pattern that includes a red cross hairs marking the approximate center of the object.

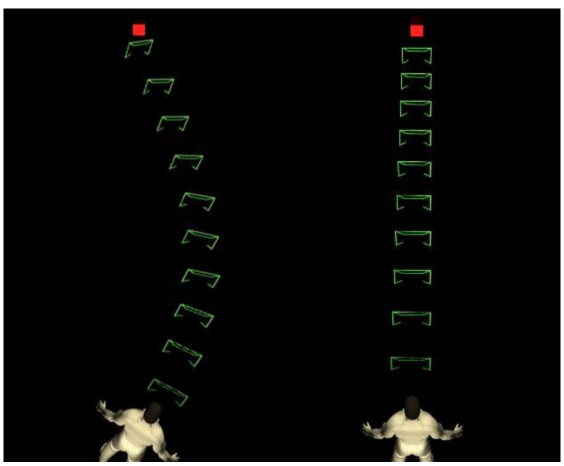

Figure 3. As the head and body move, the attention funnel dynamically provides continuous feedback. Affordances from the perspective cues automatically guide the user towards the cued location or object. Dynamic head movement cues are provided by the skew (e.g., left, right, up, down) of the attention funnel. The level of alignment (skew) of the funnel provides an immediate intuitive sense of how much the body or head must turn to see the object.

attention funnel takes advantage of spatial cueing techniques impossible in the real world, and AR's ability to dynamically overlay 3D virtual information onto the physical environment.

Like many AR components the AR funnel paradigm consists of: (1) a display technique, the attention funnel, combined with (2) methods for tracking and detecting the location of objects to be cued.

3.1. Components of the attention funnel

The attention funnel is realized as an interface widget in an augmented reality program named ImageTclAR, developed by the authors The attention funnel interface component (arwattention) and is one component in a set of user interface widgets being designed for mobile AR applications in the ImageTclAR development environment [9]. This widget provides a mechanism for drawing visual attention to locations, objects, or paths in an AR environment.

The basic components of the attention funnel, as illustrated in Figure 2, are: (a) a view plane pattern with a virtual bore sight in the center, (b) a dynamic set of attention funnel planes, (c) an object plane with a target graphic, and (d) a curved path linking the head or viewpoint of the user to the object. Along this path are placed patterns that are repeated in space and normal to the line. We refer to this line and the repeated patterns as an attention funnel. The path is defined by a Hermite curve [10]. A Hermite curve is a cubic curve segment defined by a start location, end location, and derivative vectors at each end. The curve follows a path from the starting point in the direction of the starting end derivative vector. It ends at the end point with the curve approaching the end point in the direction of the derivative vector. As a cubic curve segment, the curve presents a smoothly changing path from the start point to the end point with curvature controlled by the magnitude of the derivative vectors. Hermite curves are a standard cubic curve method discussed in any computer graphics textbook. Figure 3 clearly illustrates the curvature of the funnel from a bird's eye perspective.

The starting point of the Hermite curve is located at some specified distance in front of the origin in a frame defined to be the viewpoint of the user (the center of projection for a single viewpoint or average of two viewpoints for stereo viewers). The curve terminates at the target. The derivative vector of the Hermite curve at the starting point is in the –z direction and the derivative vector of the Hermite curve at the ending point is a vector from the starting point to the ending point. The curvatures of the starting and ending points are specified in the application.

The orientation of each pattern along the visual path is obtained by spherical linear interpolation of

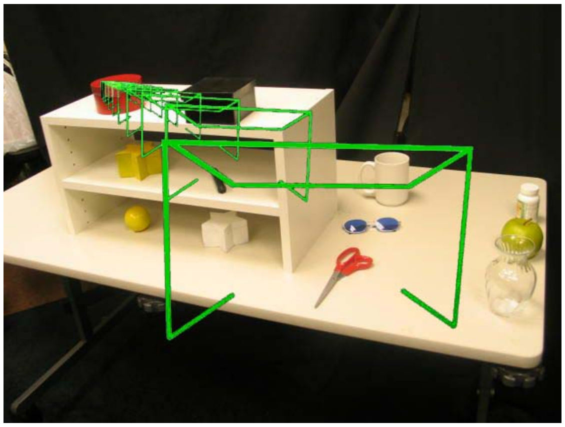

Figure 4. Example of the attentional funnel drawing attention of the user to an object on the shelf, the red box.

the up direction of the source frame and the up direction of the target frame. Spherical linear interpolation was introduced to the computer graphics society by Shoemake [11], and it is different from linear interpolation in the way that the angle between each interval is constant, i.e. the changes of orientations of the patterns are smooth. The formula used is:

$$\vec{v}(t) = \vec{v}_1 \frac{\sin((1-t)\theta)}{\sin(\theta)} + \vec{v}_2 \frac{\sin(t\theta)}{\sin(\theta)}$$

In this equation, $t \in [0,1]$, and θ is the angle between $\vec{v}_1$ and $\vec{v}_2$ computed as:

$$\theta = \cos^{-1}(\vec{v}_1 \cdot \vec{v}_2).$$

The computational cost of this method is very small, involving the solution of the cubic curve equation (three cubic polynomials), the spherical interpolation equation, and a rotation matrix for each pattern display location.

The purpose of an attention funnel is to draw attention when it is not properly directed. When the user is looking in the desired direction, the attention funnel becomes superfluous and can cause visual clutter and distraction. The solution to this case is to fade the funnel as the dot product of the source and target derivative vector approaches 1, indicating the direction to the target is close to the view direction.

3.2. Affordances in the attention funnel that guide navigation and body rotation

The attention funnel uses various overlapping visual cues that guide body rotation, head rotation, and gaze direction of the user.

Building on an attention sink pattern introduced by Hochberg [12], the attention funnel uses strong **perspective cues** as shown in Figure 4. Each attention funnel plane has diagonal vertical lines that provide depth cueing towards the center of the pattern. Each succeeding funnel plane is placed so that it fits within the preceding plane when the planes are aligned in a straight line. Increasing degrees of alignment cause the interlocking patterns to draw visual attention towards the center. Three basic patterns are used to construct a funnel: (1) the head centered plane includes a bore sight to mark the center of the pattern from the user's viewpoint, (2) funnel planes, added in a fixed pattern (currently every 12cm) between the user the object, and (3) the object marker pattern that includes a red bounding box marking the approximate center of the object. Patterns 1 and 3 are used for dynamically cueing the user that they have "locked onto" the object (see below).

As the head and body moves, the attention funnel provides continuous feedback that indicates to the user how to turn their body and/or head towards the cued location or object. Continuous dynamic head movement cues are provided by the skew (e.g., left or right) of the attention funnel. The pattern of the funnel provides an immediate intuitive sense of the location of object relative to the head. For example, if the funnel skews to the right then the user knows to move their head to the right (e.g., more skewing suggests that more body rotation is needed to see it). The funnel continuously changes, providing a dynamic cue that one is getting closer to being "in sync" and locked onto the cued object. When looking directly at the object, the funnel fades so as to minimize visual clutter. A target behind the user is indicated by a funnel that moves forward for visibility, then turns and heads behind the user - a clear visual cue.

3.3. Methods for Sensing or Marking Targets Objects or Locations

Attention funnels are applicable to any augmented vision display technology capable of presenting 3D graphics including head-mounted displays and video see-through devices such as tablet PC's or handheld computers. The location of target objects or locations in the environment may be known

Figure 5. Test Environment: The user sat in the middle of test environment for the visual search task. It consisted of an omnidirectional workspace assembled from four tables each with 12 objects (6 primitive shapes and 6 general office objects) for a total of 48 target search objects.

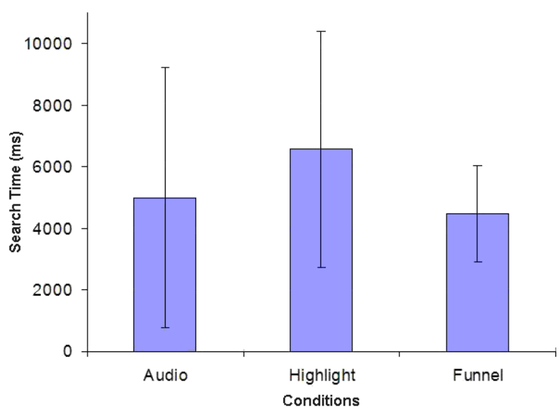

Figure 6. Search time and consistency by experimental condition. Attentional funnel decreased search time by 22% on average (28% when reach time is subtracted) and increased search consistency (decreased variability) by 65%.

to the system because they are: (1) virtual objects in tracked 3D space, (2) tagged with sensors such as visible markers or RFID tags, or (3) predefined spatial locations as in GPS coordinates. Virtual objects in tracked 3D space are the most straightforward case, as the attention funnel can link the user to the location of the target virtual object dynamically. Objects tagged with RFID tags are not necessarily detectable at a distance, but local sensing in a facility may be sufficient to indicate a position sufficient for attention direction.

In some cases, the location of the object is detected by sensors and is not known ahead of time. An implementation we are currently exploring involves the detection of visible markers with omnidirectional cameras, which can be implemented in a video see-through or optical see-through system. (Note that this implementation is different from the traditional video see-through system, where the only camera used represents the viewpoint of the user). The head-mounted omnidirectional camera detects markers in a 360^o environment around the user. The relation of the camera to the user's viewpoint is known. Detected objects can be cued for the user based on task needs or search requests by the user (i.e., "find the tool box").

4. Evaluation of User's in a Visual Search and Retrieval Tasks

Does the attention funnel truly direct user attention more efficiently than the most common techniques used in current AR interfaces? We briefly report some initial findings from a study evaluating the effectiveness of the attentional funnel in guiding attention around the immediate space of the user [13].

A very common task for an AR cursor system in a mobile setting will be to guide a user to an object that user needs to retrieve in the immediate environment. The attention funnel paradigm was tested against two other alternative techniques: (1) a commonly used AR highlighting technique, where the target object is cued by a surrounding green bounding box, and (2) a control condition mimicking interpersonal interaction, where the object to be found is indicated only by its name (e.g., "pick up the screwdriver"). A 360 degree omni-directional workspace was created with 48 objects distributed over four tables (12 objects each). See Figure 5. Half of these objects were primitive geometric objects of different colors and the other half recognizable tools (e.g., screwdriver, stapler, and notebook).

When compared to standard cueing techniques such as visual highlighting and audio cueing, we found that the attention funnel **decreased the visual search by 22%** overall, or approximately 28% for visual search time, and 14% over its next fastest. See Figure 6. While increased speed is valuable in some applications of augmented reality, such as medical emergency and other high risk applications, it may be critical that the system support the user's consistent performance. The attention funnel had a very robust effect on making the user search consistently (decreased standard error). The interface **increased user's consistency by 65%** on average, and 56% over the next best interface.

Earlier we indicated than a key criterion was the need for **minimal attention demand**. In cases where augmented reality environments are used for emergency services, repair work, other time critical

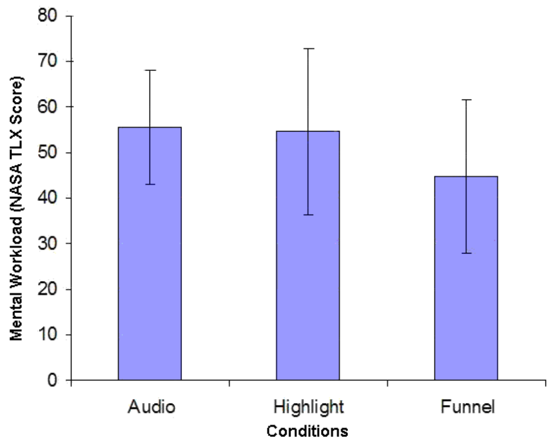

Figure 7. Mental workload measured by NASA TLX for each experimental condition.

and attention demanding applications, search time may cost mental effort. The effects of interface type of mental workload are illustrative. See Figure 7. Supporting users with only audio, which involved holding the object in memory, cost additional mental workload. But visual highlighting techniques, which demand less memory, demanded additional mental workload, possibly because of the uncertainty of where to search. The attention funnel, which placed limited demand on memory and which directed search immediately and continuously, provided an 18% decrease in mental workload.

In summary the attention funnel led to faster search and retrieval times, greater consistency of performance, and decreased mental workload when compared to verbal cueing and visual highlighting techniques.

5. Application of the Attention Funnel of Various Mobile and 3D Interfaces

The attention funnel paradigm involves basic techniques that have potentially broad applicability in AR and VR interfaces: A user's attention has to be directed to objects or locations in order to accomplish tasks.

Broadly, the attention funnel techniques can support user performance in the following generic classes of fundamental AR tasks:

Physical object selection. Situations where a user may be looking for a physical object in a space, for example a tool in a workbench, a box in a warehouse, a door in space, the next part to assemble during object assembly, etc. The system can direct the user to the correct object.

Virtual object selection. An AR system may insert labels or 3D objects inside the environment. These may be within or outside the current view of the user. Attention funnels can cue them to look at the spatially registered label, tool, or cue.

Visual search in a cluttered space. The user may be searching in a highly cluttered natural or artificial environment. An attention funnel can be used to cue them to the correct location to view, even if they are not looking in the right place.

Navigation in near space. The system might also need to direct the walking path of the individual through near space (e.g., through aisles, etc.). A directional funnel path (slightly different implementation than the attention funnel above) can be used to indicate and cue the user's direction, and provide dynamic cues as to path accuracy.

Navigation in far space. An attention funnel can direct users to distant landmarks. As an example, someone walking towards an office several blocks away must maintain a link to the landmark as they navigate through an urban environment, even when landmarks are obscured.

With the success of AR systems, designers will seek to add potentially rich, even unlimited layers of virtual information onto physical space. As AR systems are used in various real, demanding, mobile applications such as an manufacturing assembly, warehousing, tourism, navigation, training, and distant collaboration, interface techniques appropriate to the AR medium will be needed to manage the mobile user's limited attention, improve user performance, and limit cognitive demands for more optimal spatial performance. The AR attention funnel paradigm represents an example of cognitive engineering interface techniques for which there is no real world equivalent, and which is specifically adapted for users of AR systems navigating and working in information and object rich environments.

6. Authors Note

This project is part of the Mobile Infospaces project and supported in part by a grant from the National Science Foundation IIS-0222831. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science

Foundation. The authors would like to acknowledge the assistance of Bradly Alicea, Zena Biocca, and Corey Bohil in the preparation of this manuscript.

7. References

1. Moran, T.P. and Dourish, P. (2001). Introduction to This Special Issue on Context-Aware Computing. Human-Computer Interaction, 16: p. 87-95.
2. Redelmeier, D.A. and Tibshirani, R.J. (1997). Association between Cellular Telephone Calls and Motor Vehicle Collisions. New England Journal of Medicine, 336(7): p. 453-458.
3. Strayer, D.L. and A., J.W. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone. Psychological Science, 12(6): p. 462-466.
4. Shiffrin, R. (1979). Visual processing capacity and attentional control. Journal of Experimental Psychology: Human Perception and Performance, 5: p. 522-526.
5. Horvitz, E., Kadie, C., Paek, T. and Hovel, D. (2003). Models of attention in computing and communication: from principles to applications. Communication of the ACM, 46(3): p. 52-59.
6. McCrickard, D. and Chewar, C. (2003). Attentive user interface: attuning notification design to user goals and attention costs. Communication of the ACM, 46(3): p. 67-72.
7. Feiner, S., MacIntyre, B. and Seligmann, D. (1993). Knowledge-baed augmented reality. Communication of the ACM, 36(7): p. 52 - 62.
8. Mann, S. (2000). Telepointer: Hands-Free Completely Self Contained Wearable Visual Augmented Reality without Headwear and without any Infrastructural Reliance. In Proceedings of Fourth International Symposium on Wearable Computers.
9. Owen, C., Tang, A. and Xiao, F. (2003). ImageTclAR: a blended script and compiled code development system for augmented reality. In Proceedings of STARS2003, The International Workshop on Software Technology for Augmented Reality Systems. Tokyo, Japan.
10. Hearn, D. and Baker, M.P. (1996). Computer Graphics, C Version, ed. n. Edition: Prentice Hall.
11. Shoemake, K. (1985). Animating rotation with quaternion curves. In Proceedings of 12th Annual Conference on Computer Graphics and Interactive Techniques.
12. Hochberg, J. (1986). Representation of motion and space in video and cinematic displays, in Handbook of Perception and Human Performance, Vol. 1, K. Boff, L.T. Kaufman, and J. Thomas, Editors. Wiley: New York, NY.
13. Biocca, F., Tang, A., Owen, C. and Xiao, F. (2005). Augmented Attention Funnel for Visual Attention, Technical Report #MI2005a: Media Interface & Network Design Laboratories, East Lansing, Michigan.