Bridging the Gap between HCI and DHM: The Modeling of Spatial Awareness within a Cognitive Architecture

Bryan Robbins[1,2], Daniel Carruth[1], and Alexander Morais[1,2]

[1] Human Factors and Ergonomics Group, Center for Advanced Vehicular Systems,
Box 5405, Mississippi State, MS 39762-5405
[2] Department of Computer Science and Engineering, Mississippi State University,
Box 9637, Mississippi State, MS 39762-9637
`{bryanr,dwc2,amorais}@cavs.msstate.edu`

Abstract. In multiple investigations of human performance on natural tasks in three-dimensional (3D) environments, we have found that a sense of space is necessary for accurate modeling of human perception and motor planning. In previous work, we developed ACT-R/DHM, a modification of the ACT-R cognitive architecture with specific extensions for integration with 3D environments. ACT-R/DHM could leverage existing extensions from the ACT-R community that implement the spatial sense, but current research seems to indicate that an "egocentric-first" approach is most appropriate. We describe the implementation of a custom spatial module in ACT-R/DHM, which allows for the consideration of spatial locations by adding a single ACT-R module that performs a very small set of operations on existing location information. We demonstrate the use of the 3D, egocentric-first spatial module to simulate a machine interaction task.

Keywords: Digital Human Modeling, Human Performance Modeling, Spatial Cognition, Cognitive Modeling, Cognio simulatetive Architecture, ACT-R/DHM, ACT-R.

1 Introduction

The interdisciplinary field of Digital Human Modeling (DHM) has much to gain from integration efforts. As DHM research continues to realize the need for the simulation of human cognition, cognitive architectures, as first defined by Newell [1] and now implemented by many [2, 3, 4] seem to be a logical next step in integration efforts. However, many cognitive architectures, because of their heritage in Human-Computer Interaction (HCI) research, provide only marginal support for the consideration of the three-dimensional (3D) virtual environments common in DHM applications. The consideration of the human sense of space (or "spatial sense") is critical in DHM applications, but does not play a vital role in HCI, and thus is not a strong component of existing cognitive modeling architectures.

In previous work [5, 6], the ACT-R cognitive architecture [2] has been extended for use with DHM research as ACT-R/DHM. The goal of ACT-R/DHM is to leverage

V.G. Duffy (Ed.): Digital Human Modeling, HCII 2009, LNCS 5620, pp. 295–304, 2009.

the ACT-R modeling architecture's theory of cognition and its decades of development, improvement, and validation for the purposes of DHM research by adding theory-based and architecturally consistent extensions. To-date, ACT-R/DHM has extended the existing visual and motor modules of ACT-R and added a kinesthetic and proprioceptive (KP) module in addition to the spatial module described herein.

1.1 The ACT-R Theory and Implementation

Before elaborating on the implementation of the spatial sense and other extensions of ACT-R/DHM, we describe the original ACT-R theory [2] and its implementation in the ACT-R 6 software. ACT-R's model of knowledge is based on a separation between declarative memory (facts) and procedural memory (actions). Models of human sensory and perception systems convert features of the external environment to an internal representation suitable for processing. Strictly typed "chunks" of information serve as the basic building block for ACT-R's internal representation. Chunks of declarative memory are manipulated by action elements of rules in procedural memory.

The central core of the ACT-R software implements a small number of critical cognitive functions. Additional modules supplement the core with memory, perceptual, and motor capabilities. ACT-R is implemented as a production system, with procedural memory elements constructed by the modeler as If-Then rules called productions. Execution of productions is accomplished by matching the "If", or Left-Hand Side (LHS), of the production against the current state of the modules and, if a match is found, executing the "then", or Right-Hand Side (RHS) of the production.

The modular construction of the ACT-R architecture allows for the extension of existing capabilities and the creation of new modules. The theoretical concepts underpinning the ACT-R architecture are enforced by the interface between the core and the modules. Important architectural constructs include the modules, the module buffers, the module requests and the chunk. The module itself, implemented as a LISP object, encapsulates the model of the system being represented. For example, the vision module simulates human vision, sensory memory, feature integration, attention and any other aspects associated with human vision, within a single module. A module's buffer(s) makes available the current state of its module providing a window to the module environment. Module requests provide mechanisms for updating the module's state via productions. Finally, chunks, as the basic building block of the architecture, hold declarative information in a strictly defined type, known as the chunk type.

As mentioned, the constructs of ACT-R are more than implementation considerations – they enforce the underlying ACT-R theory. Any new capability added to ACT-R, including the extensions in ACT-R/DHM, must follow the required structure. If an extension deviates from the architectural standards, it gains little from ACT-R's well established psychological validity. For this reason, we describe the extensions of ACT-R/DHM in terms of the modules, buffers, requests, and chunks affected.

1.2 ACT-R/DHM - Current Implementation

ACT-R/DHM, prior to the development of a spatial module, extended ACT-R in a number of ways. First and most importantly, the vision module of ACT-R, which has primarily been used for stimuli presented on a two-dimensional (2D) computer screen, was expanded to consider 3D space. No additional module requests or buffers were necessary, but instead of storing flat (X, Y) coordinates, new visual chunks were derived from the original structures that encode spherical coordinates: pitch, azimuth and distance. The pitch, azimuth, and distance, or "PAD", encoding is also reflected in the spatial module and elsewhere in ACT-R/DHM as a consistent representation of space.

In addition, ACT-R/DHM includes a kinesthetic and proprioceptive (KP) module that allows for the consideration of avatar movements and body part locations in a cognitive model. The KP module adds a single buffer, "kp", that holds the current state, (position and movement), for a single, attended body part. The kp position representation is consistent with the PAD from the vision module, spatial module, and other modules, as mentioned above. The movement state simply indicates whether or not the body part is currently in motion. Elaboration on the details of the KP module's implementation is outside the scope of this paper.

1.3 The Need for Spatial Functions

Relatively simple DHM task models clearly illustrate the need for spatial functionality in the majority of DHM scenarios. The following example is based on previous efforts [6, 7] using ACT-R/DHM with a wrapper interfacing to the Virtools™ 3D Virtual Environment and the Santos™ avatar. Santos™ is a digital human model developed at the Center for Computer Aided Design at the University of Iowa for the Virtual Soldier Research Project [8]. Santos includes a full-body avatar, with a skeleton and a posture prediction algorithm that serves as the basis for KP body part and movement information. The virtual environment provides the remaining environmental feature information (i.e. visual feature descriptions) to ACT-R/DHM.

Figure 1 shows the virtual environment setup for a vending machine interaction task [6]. Participants were given 10 coins to purchase a beverage of their choice. This task involved a series of human-machine interactions in a large visual field. Participants must learn the layout of the interface, deposit their coins, choose from opions visually presented via labels, select their drink, and retrieve their drink from the machine. As the model performs the physical motions necessary to accomplish the task, the head moves and key features of the interface drop in and out of the visual field. In Figure 1a, many of the key features are not visible when the avatar is standing before the machine (initial position). In this position, the model must have some mechanism to access the current egocentric spatial position of the target feature in order to shift the view towards the target. Figure 1b shows one view encountered during interaction with the machine, and is filled with a number of critical machine-interaction features, such as buttons, the coin deposit, and other features.

To accomplish tasks without repeatedly scanning for visual features any time they drop out of the current visual field, spatial memory must be available during task performance and is critical in the most basic of real-world maneuvers. To

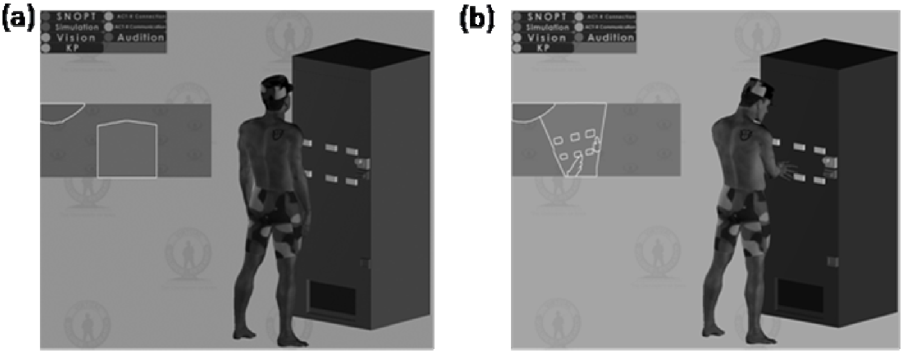

Fig. 1. The availability of visual features changes as body movements are made in a simple machine interaction task. The initial view (a) contains no objects, but reliable motor planning is possible via spatial cognition. Another view (b) may have a completely different set of features available.

appropriately model human-machine interaction, DHM cognitive architectures must include models of spatial cognition. A significant obstacle to successful modeling of spatial cognition is accounting for storage, processing and recall of spatial knowledge derived from egocentric sensory data. Below, we consider the psychological evidence for spatial awareness and existing ACT-R modeling approaches before arriving at the egocentric-first spatial module now implemented in ACT-R/DHM.

2 Modeling Spatial Cognition

We propose two limitations on the implementation of the spatial module of ACT-R/DHM. First, the new implementation should be based on current theory of human spatial cognition. Second, the implementation should conform to current ACT-R theory and implementation framework.

2.1 Spatial Theory

There exists significant debate in the spatial cognition literature regarding the nature of spatial representations and processes. Our current implementation draws primarily from the theory of McNamara [8], as supported by Rock [9] and Sholl and Nolin [10]. The following theorems integrate concepts from all three authors' theories.

Theorem 1, expressed in the work of McNamara [8] and Sholl and Nolin [10], is that human spatial competence depends on both egocentric (first-person) and allocentric (third-person) representations. Sholl and Nolin define the egocentric representations as "self-to-object" relationships and the allocentric representations as "object-to-object" relationships. Theorem 2, elaborated by Sholl and Nolin, states that human spatial competence requires the interaction of the egocentric and allocentric systems. Fundamentally, the egocentric system is based on instantaneous experiences. Theorem 3 states that all spatial relationships must begin as egocentric relationships because they are derived from egocentric percepts. We also find, however, that egocentric representations are generally inadequate for reuse. Once the egocentric origin moves in the environment (e.g. head or other body movement in the world), previously encoded egocentric relationships are no longer valid for recall. Rather than postulate a continuous update of spatial memory, we opt to implement Theorem 4: spatial relationships are stored as allocentric object-to-object relationships.

McNamara considers the storage of spatial relationships in memory [8]. He outlines a theory that emphasizes the role of orientation and frames of reference in spatial encoding. He appeals to the work of Rock [9] regarding orientation detection of new objects, which states that a variety of cues lead to the ultimate determination of object orientation, and that an object's encoded orientation depends on many factors, including its environment, intended use, shape, and others. McNamara argues that some persistent cues, which he calls environmental cues, are present across multiple egocentric views, and provide a critical link between egocentric views. We extend this idea as Theorem 5, which states that all spatial relationships require a frame of reference and an orientation, and the orientation must be expressed relative to the frame of reference. Note that Theorem 5 applies to egocentric relationships which use the origin of the spatial system as a frame of reference and a native orientation for each object in the egocentric view. Any object may provide a frame of reference for any other object in an object-to-object relationship. However, certain environmental objects are more likely to be used as referents due to location, size, saliency, permanence, etc. No claim is made as to how actual detection and/or recognition of an object's native or intrinsic orientation should be modeled.

Theorems 1-5 summarize fundamental building blocks of human spatial cognition. Thus, any implementation of the spatial sense in ACT-R/DHM should hold to the theoretical claims of these theorems.

2.2 Other ACT-R Implementations of Spatial Cognition

Previous efforts have modeled the spatial sense within ACT-R implementations. Before developing a custom module for the spatial sense in ACT-R/DHM, we considered existing extensions from the ACT-R community.

Gunzelmann and Lyon offer a spatial implementation that covers many of the elements of spatial theory identified in section 2.1 [11]. Specifically, they propose adding three buffers to the visual system: an egocentric buffer which holds 3D, egocentric location information, an environmental frame of reference buffer that tracks a frame of reference, as suggested by Theorem 5, and an episodic buffer integrates the egocentric and frame of reference information with existing visual information as needed. Gunzelmann and Lyon also propose a spatial module that makes spatial information accessible across multiple ACT-R modules and provides spatial processing for mental transformations, magnitude estimations, and magnitude calculations [11]. Additional ACT-R implementations have been proposed by Best and Lebiere [12] and Harrison and Schunn [13]. Best and Lebiere's implementation, designed for the development of intelligent agents, preprocesses the environment to directly provide an allocentric representation to the agent. Harrison and Schunn implement a "configural" buffer that associates a visual object with its orientation, then a system to update up to three "behaviorally significant" egocentric relationships based on the direction of body motion (e.g. walking). No persistent allocentric representation is used.

To summarize, our implementation of spatial cognition in ACT-R differs from other current efforts either in underlying theoretical claims (i.e. Harrison and Schunn and Best and Lebiere) or architectural implementation (i.e. Gunzelmann and Lyon).

3 The ACT-R/DHM Spatial Module

The following section describes ACT-R/DHM's spatial module, and how this module is used to support modeling of spatial cognition.

3.1 Module Implementation

Previous ACT-R/DHM work extended the vision module of ACT-R to support PAD encoding. The spatial implementation requires a single new module, simply named the spatial module. The module provides only one buffer to the environment, also named the spatial buffer. The spatial buffer should only hold chunks of the type *spatial-relationship*, and only three module requests are provided as operations on spatial-relationship chunks: *ego-relate*, *chain-relate*, and *mid-relate*. The spatial module has no member data of its own and derives from ACT-R's generic module class. The spatial-relationship chunk type and the module requests of the ACT-R/DHM spatial module capture many of the

The spatial-relationship chunk type is detailed in Table 1. ACT-R chunks hold smaller pieces of information in slots. The slots of the spatial-relationship chunk hold a frame of reference in the reference slot, an object in the object slot, the position of the object in the pitch, azimuth, and distance slots, and finally the orientation of the object as three axis vectors, <xdir, ydir, zdir>, <xup, yup, zup>, and <xright, yright, zright>. The position and orientation are relative to the frame of reference.

The three spatial module requests operate on spatial-relationship chunks to allow for the modeling of spatial competence. Specifications for each module request are included in Table 2. For the ACT-R/DHM spatial module, all operations occur in the spatial buffer. The *ego-relate* request takes an object as encoded by a sensory/perception module, produces an egocentric spatial relationship to the object,

Table 1. The spatial relationship chunk includes information about both the object and its frame of reference

SPATIAL RELATIONSHIP CHUNK	
Slot	***Description***
referent	The referent object
object	The encoded object from a sensory/perception module
pitch, azimuth, distance	Object position in egocentric spherical coordinates
xdir, dir, zdir xup, yup, zup xright, yright, zright	Object orientation axes

Table 2. The spatial module provides three requests that support spatial reasoning. Requests are used by the cognitive modeler to simulate human performance. The outcome of spatial requests is always placed in the spatial buffer.

SPATIAL MODULE REQUESTS		
Request	***Input Arguments***	***Output to Spatial Buffer***
ego-relate	visual percept *obj*	**spatial relationship *self-obj*:** referent = SELF object = *obj*
mid-relate	spatial relationship *ref-obj* spatial-relationship *ref-tar*	**spatial relationship *obj-tar*:** referent = object of *ref-obj* object = object of *ref-tar*
chain-relate	spatial-relationship *ref-obj* spatial-relationship *obj-tar*	**spatial relationship *ref-tar*:** referent = referent of *ref-obj* object = object of *obj-tar*

and places the new spatial-relationship chunk in the spatial buffer. The object chunks themselves must provide some native orientation information. In the case of the vision system, we have extended visual object chunks to include orientation information. When a visual-object chunk is passed to *ego-relate*, orientation information is passed to the new spatial-relationship.

The *mid-relate* and *chain-relate* functions build on egocentric spatial-relationship chunks to produce object-to-object relationships. Mathematically, *mid-relate* and *chain-relate* are simply vector addition functions, while psychologically they correspond to a "single-step" mental rotation (see Gunzelmann and Lyon [11] and also Kosslyn [14]). The *mid-relate* request takes two spatial-relationships as arguments, e.g. Self->A and Self->B, and creates an object-to-object spatial-relationship chunk A->B, where A serves as the frame of reference. Similarly, *chain-relate* takes as input any two spatial-relationship chunks of the form A->B and B->C and creates a spatial-relationship chunk A->C, where A serves as the frame of reference.

3.2 Enforcing Constraints

We now review the implementation relative to the previously described theorems and the ACT-R architecture.

Theorem 1 states that both egocentric and allocentric encoding is necessary for human spatial competence. The ACT-R/DHM implementation provides a generic representation, the spatial-relationship chunk type, that allows both types of encoding.

Theorem 2 states that the interaction of egocentric and allocentric representations is essential to human spatial cognition. The *mid-relate* function provides a direct transformation from egocentric to allocentric representations. The *chain-relate* function can also be used for the inverse operation, if the first spatial-relationship in the "chain" is an egocentric relationship, e.g. Self->A + A->B = Self->B.

Theorem 3 states that all spatial relationships must begin as egocentric. Sensory percepts are encoded into objects that include egocentric pitch, azimuth, and distance.

This egocentric information can be used to generate egocentric spatial-relationships using *ego-relate*. Theorem 4 states that all relationships are stored in an allocentric form. The *mid-relate* mechanism converts egocentric spatial-relationships with two objects into a single object-to-object relationship. In our implementation object-to-object relationships are only useful when tied to an egocentric relation. This requirement is based on Theorems 2 and 3.

The spatial-relationship chunk also allows for the encoding of a frame of reference and of object orientation, as required by Theorem 5. While the current implementation relies on Rock [9] and McNamara [8] to support the requirement for orientation encoding, it makes no claim to model the determination of orientation by visual or declarative memory methods. This is an interesting question for our implementation that deserves future work as the assumption that orientation can be encoded underpins the spatial module's encoding of allocentric, object-to-object relationships.

3.3 Spatial Modeling with ACT-R/DHM

With the ACT-R/DHM implementation of the spatial sense now specified, the application of the new spatial modeling capability is perhaps most interesting to the DHM community. We now describe the use of the new spatial module to improve the performance of the previously introduced model –the vending machine interaction task. As mentioned previously, ACT-R/DHM to-date has been used to drive the Santos™ digital human model, and Santos™ exists in a high-fidelity 3D virtual environment.

The vending machine interaction task now uses the module requests of the spatial module. The model assumes that the human subject is familiar with the parts of the vending machine (e.g. buttons, labels, coin slot, etc.) but has not seen or used this specific machine before. Thus, as the avatar approaches the machine, he encodes the layout of the machine relative to the machine's background (a large environmental cue) using a series of *ego-relate* and *mid-relate* requests. The machine's background, known in the model as the "CokeBox", is visible at all times during the machine interaction task, and is therefore an ideal object for the construction of object-to-object relationships. For example, the model encodes egocentric relationships Self->Button1 and Self->CokeBox then uses Mid-Relate with these two egocentric relationships as arguments to create and store the object-to-object relationship CokeBox->Button1. To utilize the stored spatial relationship, the model must relocate CokeBox in the visual field and use *chain-relate* to program egocentric movements for machine interaction.

After encoding the machine layout, the model programs numerous motor movements from the avatar's hand to the coin slot of the machine, simulating the deposit of coins. Note that as the avatar looks at his hand, the coin slot drops from the visual field. The position of the coin slot relative to the CokeBox must then be recalled from declarative memory and an egocentric spatial-relationship chunk constructed via *chain-relate* in order to relocate the coin slot and continue depositing.

4 Conclusions and Future Enhancements

The implementation of a spatial module in ACT-R/DHM resolves significant issues related to knowledge of object locations in 3D environments and provides the capability to model human performance for many dynamic tasks.

To make the spatial modeling capability more accessible and accurate, a number of additional enhancements are necessary. For example, the link between the visual and spatial systems should be explored with regard to attention and autonomous behavior. It seems feasible, as Harrison and Schunn have suggested [13], that at least one or more currently attended spatial relationships may be updated automatically as the body moves. In fact, ACT-R/DHM's KP system provides some functionality that could be used to update spatial knowledge based on kp movement information. The enforcing of limitations on spatial cognition is also an area that needs additional research and implementation. If, as Kosslyn suggests [14], spatial reasoning must occur egocentrically, then only egocentric spatial relationships should be available to the spatial buffer. This could be accomplished by implementing mid-relate and chain-relate functionality within the spatial module and exposing only egocentric spatial relationships via the spatial buffer.

While much future work remains to extend this implementation, compare our implementation with alternative implementations, and validate against human spatial cognition data, the ACT-R/DHM spatial module provides significant functionality based on spatial cognition theory and within the existing ACT-R framework.

References

1. Newell, A.: Unified Theories of Cognition. Harvard University Press, Cambridge (1999)
2. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An Integrated Theory of the Mind. Psychological Review 111(4), 1036–1060 (2004)
3. Laird, J.E., Newell, A., Rosenbloom, P.S.: SOAR: An Architecture for General Intelligence. Artificial Intelligence 33(1), 1–64 (1987)
4. Kieras, D.E., Meyer, D.E.: An Overview of the EPIC Architecture for Cognition and Performance with Application to Human-Computer Interaction. Human-Computer Interaction 12(4), 391–438 (1997)
5. Carruth, D., Robbins, B., Thomas, M., Letherwood, M., Nebel, K.: Symbolic Model of Perception in Dynamic 3D Environments. In: Proceedings of the 25th Army Science Conference, Orlando, Florida (September 2006)
6. Carruth, D., Thomas, M., Robbins, B.: Integrating Perception, Cognition and Action for Digital Human Modeling. In: Duffy, V.G. (ed.) HCII 2007 and DHM 2007. LNCS, vol. 4561, pp. 333–342. Springer, Heidelberg (2007)
7. Abdel-Malek, K., Yang, J., Marler, R.T., Beck, S., Mathai, A., Zhou, X., Patrick, A., Arora, J.: Towards a New Generation of Virtual Humans: Santos. International Journal of Human Factors Modeling and Simulation 1(1), 2–39 (2006)
8. McNamara, T.: How Are the Locations of Objects in the Environment Represented in Memory? In: Freksa, C., Brauer, W., Habel, C., Wender, K.F. (eds.) Spatial Cognition III. LNCS, vol. 2685, pp. 174–191. Springer, Heidelberg (2003)
9. Rock, I.: Orientation and Form. Academic Press, New York (1973)

10. Sholl, M.J., Nolin, T.L.: Orientation Specificity in Representations of Place. Journal of Experimental Psychology: Learning, Memory, and Cognition 23(6), 1494–1507 (1997)
11. Gunzelmann, G., Lyon, D.R.: Mechanisms for Human Spatial Competence. In: Barkowsky, T., Knauff, M., Ligozat, G., Montello, D.R. (eds.) Spatial Cognition 2007. LNCS (LNAI), vol. 4387, pp. 288–307. Springer, Heidelberg (2007)
12. Best, B.J., Lebiere, C.: Spatial Plans, Communication, and Teamwork in Synthetic MOUT Agents. In: Proceedings of the 12th Conference on Behavior Representation In Modeling and Simulation (2003)
13. Harrison, A.M., Schunn, C.D.: ACT-R/S: Look Ma, no “cognitive map"! In: Detje, F., Doerner, D., Schaub, H. (eds.) Proceedings of the Fifth International Conference on Cognitive Modeling, pp. 129–134. Universitats-Verlag, Bamberg (2003)
14. Kosslyn, S.M.: Image and Brain: The Resolution of the Imagery Debate. MIT Press, Cambridge (1994)