

# Spatial Representations From Perception and Cognitive Mediation

## The Case of Ultrasound

Roberta L. Klatzky,<sup>1</sup> Bing Wu,<sup>1,2</sup> and George Stetten<sup>2,3</sup>

<sup>1</sup>Department of Psychology, Carnegie Mellon University; <sup>2</sup>Robotics Institute, Carnegie Mellon University; and

<sup>3</sup>Department of Bioengineering, University of Pittsburgh

---

**ABSTRACT**—*Spatial representations can be derived not only by direct perception, but also through cognitive mediation. Conventional or ex-situ ultrasound displays, which displace imaged data to a remote screen, require both types of process. To determine the depth of a target hidden beneath a surface, ultrasound users must both perceive how deeply the ultrasound transducer indents the surface and interpret the on-screen image to visualize how deeply the target lies below the transducer. Combining these perceptual and cognitive depth components requires a spatial representation that has been called amodal. We report experiments measuring errors in perceptual and cognitively mediated depth estimates and show that these estimates can be concatenated (linked) without further error, providing evidence for an amodal representation. We further contrast conventional ultrasound with an in-situ display whereby an ultrasound image appears to float at the precise location being imaged, enabling the depth of a target to be directly perceived. The research has the potential to enhance ultrasound-guided surgical intervention.*

**KEYWORDS**—*perception; visualization; spatial cognition; ultrasound; medical imaging*

---

Any textbook in sensation and perception describes cues by which the visual system computes the locations of objects in space. They include oculomotor cues from converging the eyes toward the target and focusing the lens, pictorial cues like linear perspective and occlusion, movement-induced cues like paral-

lax, and binocular cues like stereopsis. Collectively, these mechanisms richly convey the spatial layout of objects, a process that we will call *direct perception*. In contrast, when sight is not available, another process relies on cognitive mediation to construct a representation of space; we call this spatial *visualization*.

Visualization has been shown to be effective for a variety of spatially directed behaviors. Studying a map enables people to accurately report the distance and direction from one point to another in the corresponding physical environment (Richardson, Montello, & Hegarty, 1999). Visualization from language, perhaps surprisingly, can sometimes guide action equivalently to direct perceptual control. For example, people can view a target and then walk to it directly or indirectly without vision, and they do likewise when the target location is described in spatial language like “2 o’clock, 8 feet” (Avraamides, Loomis, Klatzky, & Golledge, 2004; Klatzky, Lippa, Loomis, & Golledge, 2003; Loomis, Lippa, Klatzky, & Golledge, 2002).

### PERCEPTION VERSUS VISUALIZATION IN ULTRASOUND

Our interest in spatial visualization stems from applied research on ultrasound guidance of common surgical interventions, like placing peripheral catheters or performing breast biopsies. Ultrasound imaging works on the same principle as sonar: Narrow beams of high-frequency sound waves are sent out sequentially from a transducer in a fan-shaped pattern to interrogate a region of space. As the sound waves encounter a sound-reflecting object, echoes are sent back to the transducer with a delay indicating the object’s range. These signals are converted into a 2-D image, which shows the spatial locations of sound-reflecting objects in the scanned slice of environment.

---

Address correspondence to Roberta L. Klatzky, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213; e-mail: klatzky@cmu.edu.

Typically, the slice is displayed on a local monitor, with important consequences: Several aspects of the display lead to cognitive mediation being invoked to map the ultrasound data to locations in external space. First, because the visual display cannot physically occupy the space in front of the transducer, it is displaced from the region being scanned. Second, the axes in the displayed image are aligned only with the transducer's frame of reference; they do not correspond to any fixed frame in external space. Specifically, the vertical axis of the image is generally in the direction of the central ultrasound beam, while the horizontal axis is transverse to that beam. Hence, as the transducer is moved or tilted, the image on the screen corresponds to different locations and/or orientations. Third, the scale of the display is variable. The user can zoom to enlarge or reduce the screen image, with the objective size indicated by a scale on the monitor. While rescaling is convenient for viewing objects, it means that there is no consistent correspondence between the sizes of display space and external space. Collectively, these factors mean that in order to build a representation of target location from the ultrasound image, the user must invoke spatial cognitive processes such as mental rotation, translation, and rescaling. This constitutes visualization, not direct perception.

We have contrasted cognitively mediated and perceptually directed action by comparing conventional ultrasound to a novel ultrasound display, which employs "augmented reality" to portray hidden targets at their physical 3-D locations (Stetten & Chib, 2001). This device uses a tool familiar to psychologists, the half-silvered mirror, to project an ultrasound image directly into the area being scanned (Fig. 1). A small video display is embedded into the handle of an ultrasound transducer, and a half-

silvered mirror is mounted above its shaft. Light rays from the display that contact the mirror are reflected back toward the eyes of the user, providing normal visual distance cues including convergence, accommodation, and stereopsis. This line of sight creates the same percept as if the rays came from the space imaged by the transducer itself. Because light rays also pass directly through the mirror (that's the beauty of half-silvering), the user can see not only the reflected ultrasound image but also the real world in front of the transducer.

The result of the augmented-reality display is an illusion in which the ultrasound image appears to float within the space that is being scanned, at the precise location of the imaged data. We call this form of viewing *in-situ*, in contrast to the *ex-situ* visualization of the image on a conventional ultrasound display. In-situ viewing allows a representation of the location of ultrasonically detected targets to be encoded with basic perceptual processes; *ex-situ* visualization requires a cognitive overlay.

### EX-SITU ULTRASOUND AND THE AMODAL SPATIAL HYPOTHESIS

Ex-situ medical-imaging modalities like ultrasound require a representation of space that is sufficiently abstract that it relies jointly on perception and cognitive mediation, yet is capable of directing action. One candidate spatial representation is amodal; that is, not linked to any one sensory system. Previous evidence for amodal representations comes from research on the effectiveness of visualization in guiding action (Bryant, 1992; Loomis et al., 2002). Our work provides a further test: If there is a common representation that can be encoded both perceptually

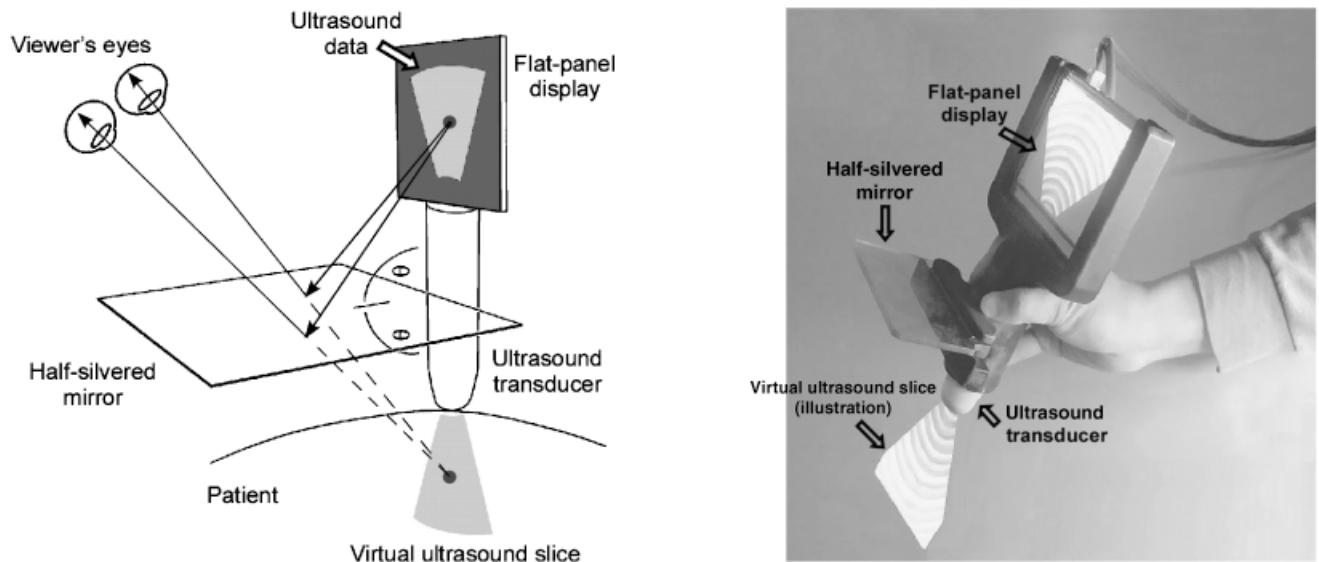


Fig. 1. The augmented-reality ultrasound visualization device. The ultrasound image on the flat-panel display is reflected in a half-silvered mirror and projected to the user's eyes as if it comes from the target area on the patient's body. The image thus looks as if the transducer is actually illuminating the inner structures being scanned. Reprinted from "Psychophysical Evaluation of In-Situ Ultrasound Visualization," by B. Wu, R.L. Klatzky, D. Shelton, & G. Stetten, 2005, *IEEE Transactions on Visualization and Computer Graphics*, 11, pp. 685 and 686. Copyright 2005, Institute of Electrical and Electronics Engineers. Reprinted with permission.

and through cognitive mediation, it should be possible to combine inputs from the two sources into a unified spatial continuum.

Unifying spatial inputs is a fundamental ability in many species. Early in perception, the brain combines information across successive saccades (e.g., Irwin & Andrews, 1996). Higher-level integration processes operate across successive views of space. For example, if a viewer sees only a patch of far ground surface through a small aperture, misperception of surface slant distorts the perceived distance of an object placed there. This error is eliminated, however, when the aperture is scanned from near the viewer to the far location, allowing integration of successive patches into an accurate global surface representation (Wu, Ooi, & He, 2004). Still more global processes combine independently learned “cognitive maps” on the basis of overlap and intersection (Blaisdell & Cook, 2005; Golledge, Smith, Pellegrino, Doherty, & Marshall, 1985; Holding & Holding, 1988; Moar & Carleton, 1982; Sturz, Bodily, & Katz, 2006).

Our studies have used ultrasound-based imaging to seek evidence of amodal representations. We tested for a specific form of integrating perception and cognitive mediation—namely, whether people could link together contiguous spatial components.

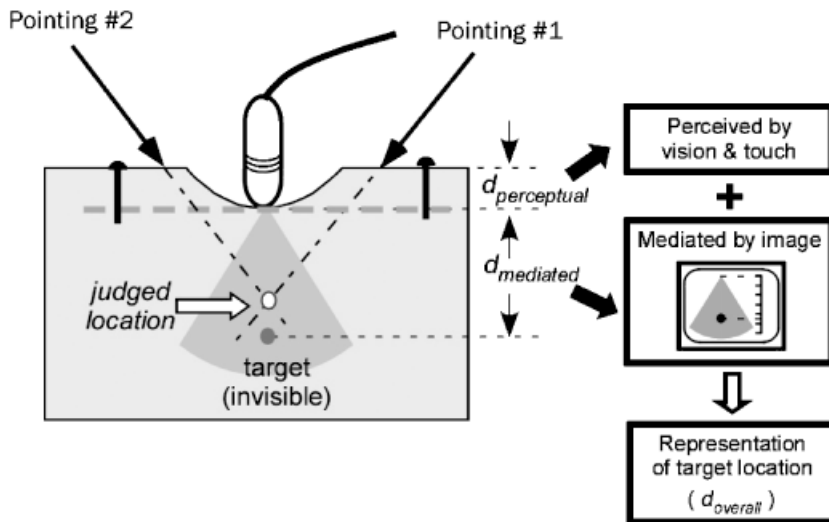
### TESTING THE AMODAL HYPOTHESIS

Ultrasound users must take into account two different spatial components in order to direct action toward a hidden target. One is the distance from the tip of the ultrasound transducer to the

target, which is displayed pictorially and numerically in the ultrasound image and interpreted with cognitive mediation, as just described. It is necessary, however, to consider a second component as well—namely, the indentation of the transducer into the surface it contacts. As the transducer meets the skin of the patient, it depresses the surface and moves toward the target. The target image accordingly shifts upward on the viewing monitor so that it appears to lie more shallowly. If a surgical tool enters the surface outside the indented area, the target is deeper relative to the tool than it is relative to the transducer, requiring that the transducer indentation be taken into account when aiming.

As a result, the depth of an ultrasound-viewed target from a surgical tool is the sum of (a) the distance from unindented surface to transducer tip and (b) the distance from transducer tip to target. The first quantity is conveyed perceptually, jointly by vision of the indentation and by haptic (touch) perception of resisting forces. The second quantity is portrayed by the ultrasound image and requires cognitive mediation. We ask whether the two depth representations can be mentally concatenated, as needed for guidance of a surgical tool.

We call the method to test this hypothesis “error-pass-through” (Wu, Klatzky, Shelton, & Stetten, 2008): Errors in each depth component are induced and measured independently (Fig. 2). Participants then engage in a composite task, which requires estimation of both components. We test whether errors in the composite task are entirely predicted by the sum of errors in the independent components. This would indicate that a mental process could combine perceived indentation with



**Fig. 2.** Decomposition of represented target depth ( $d_{\text{overall}}$ ) into two components: (a) transducer indentation, perceived from vision and touch ( $d_{\text{perceptual}}$ ); and (b) transducer-relative depth, visualized from ultrasound ( $d_{\text{mediated}}$ ). The  $d_{\text{perceptual}}$  component is measured by having subjects draw the indentation, and the  $d_{\text{mediated}}$  component is measured by the intersection of multiple pointing responses. From “Mental Concatenation of Perceptually and Cognitively Specified Depth to Represent Locations in Near Space,” by B. Wu, R.L. Klatzky, D. Shelton, & G. Stetten, 2008, *Experimental Brain Research*, 184, p. 296. Copyright 2007, Springer-Verlag. Reprinted with permission.

cognitively mediated depth from the ultrasound display, without adding further error. Such a result would lend support to the hypothesized amodal representation of space at which perception and cognitive mediation converge.

### ERRORS IN COGNITIVELY MEDIATED DEPTH

Our experimental paradigm for measuring errors in the cognitively mediated component is shown in Figure 2: Subjects use the ultrasound transducer to find a target (a bead) hidden in a tank filled with opaque fluid. They then point to it with a stylus from the rim of the tank. We use the intersection of the pointing responses from multiple (three or more) locations to determine where in 3-D space the target is localized. An initial study (Wu, Klatzky, Shelton, & Stetten, 2005) contrasted localization based on direct sight of the target bead (i.e., the tank was emptied of fluid) with localization by ultrasound. When pointing, subjects using ultrasound consistently underestimated the depth of the target. Moreover, when asked to guide a needle to the target, they initially aimed toward the same too-shallow location that had been computed from pointing. We reasoned that the underestimation tendency resulted from errors in cognitive mediation and, possibly, from subjects' failure to take into account deformation of the tank lid under the pressure of the ultrasound transducer.

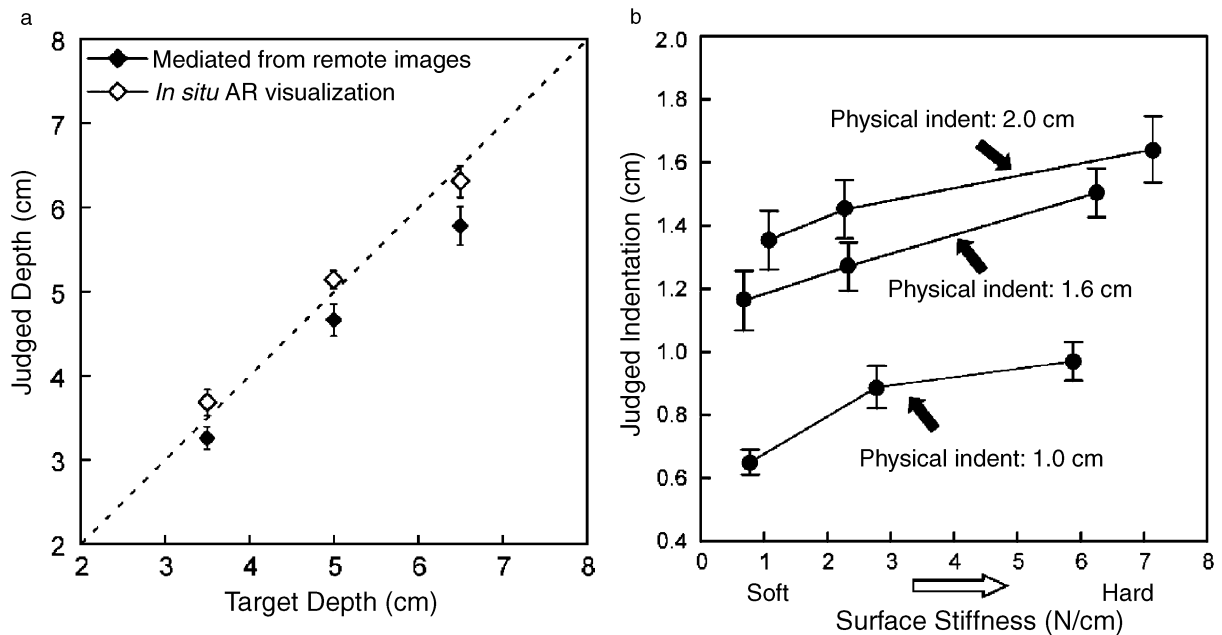
For the error-pass-through test, it was first necessary to use the pointing task to isolate and measure cognitively mediated errors, by employing a tank with a rigid lid that resisted indentation. As

was just described, subjects used the transducer to expose a target bead in the center of a fluid-filled tank, then pointed to it from multiple locations around the rim. For each of several objective target depths, a localization depth was computed from the intersection of the pointing responses. As shown in Figure 3a, subjects again underestimated the depth of the target (though less than with a deformable lid). These data, then, provide one component of the total depth that can be input to the error-pass-through test.

### ERRORS IN PERCEIVED INDENTATION

Error-pass-through also requires that we measure the perceived indentation of the transducer. For this purpose we created a set of tank lids; the center portion of each lid was indented by a specified amount and covered with a rubber cross-hatched surface. Participants pushed the ultrasound transducer into the center of the lid until it "bottomed out," then drew a line to show how deeply the transducer tip was depressed relative to the outer portion of the lid. The indentation was signaled by cues of visual depth perception, including visible deformation in the cross-hatched pattern.

To induce error in the indentation judgments, we added haptic cues to the visual ones. Elastic bands were mounted beneath the surface covering the tank center, creating resisting force. We hypothesized that resistance against the transducer would increase the judged indentation, constituting a haptic illusion. Indeed, as shown in Figure 3b, the more the surface pushed back



**Fig. 3.** Ultrasound-localized target depth as a function of true depth (a), and perceived indentation as a function of physical indentation and resisting force (b). The filled diamonds in (a) show that subjects tend to underestimate target depth when using remote images (the diagonal indicates correct performance); comparison data (open diamonds) are shown from the same task, but with an augmented-reality device (AR) that allows the target to be viewed in situ. Graph (b) shows that a stiffer surface leads to larger estimates of indentation. From "Mental Concatenation of Perceptually and Cognitively Specified Depth to Represent Locations in Near Space," by B. Wu, R.L. Klatzky, D. Shelton, & G. Stetten, 2008, *Experimental Brain Research*, 184, pp. 300 (3a) and 298 (3b). Copyright 2007, Springer-Verlag. Reprinted with permission.

against the transducer tip, the greater the subjects judged the indentation to be. The resulting errors in perceived indentation provide the second component for the test of the amodal spatial hypothesis.

### CONCATENATING PERCEIVED AND MEDIATED DEPTH COMPONENTS

The third experiment in this series combined the previous manipulations in order to perform the error-pass-through test: A target was placed in a tank covered with an indented, resisting center surrounded by an unindented rim. The subject pressed the transducer against the indentation, visualized the target, and then pointed to it from multiple locations around the rim. The target depth relative to the pointing location combined the two depth components: (a) indentation of the transducer, perceived from visual and haptic cues, and (b) depth of the target relative to the transducer tip, visualized from ultrasound. As we have seen, each component has its own error tendencies, and performance in the combined task turned out to be predictable from both together: The pointing responses indicated that target depth was underestimated overall, but less so as the indentation was made more salient by larger resisting forces.

Most importantly, the errors in the composite task of the third study were predicted by the sum of the errors observed for the same conditions in the first two studies, which measured cognitively mediated and perceived depth in isolation. This result supports the hypothesis of amodal spatial representations by indicating that a mental process combined perceived indentation with visualized depth from the ultrasound display. Moreover, that mental process produced no further error. A control experiment showed that this process was not simply mental addition of numerical estimates: Numerically specifying total target depth (e.g., “3 centimeters” along with a centimeter scale) was not sufficient to enable people to point accurately; instead they exhibited errors similar to those observed when visualizing the target with ultrasound. The inadequacy of numerically specified target depth indicates that perceived indentation and ultrasound-visualized depth are combined to support action by a process that doesn’t just sum numbers; it is intrinsically spatial.

### APPLIED CONSIDERATIONS

Our program of study exemplifies a symbiosis between basic and applied research. Consideration of medical contexts is an important motivation behind this work. To the extent that localization of targets by ultrasound can be improved, the success of clinical outcomes such as catheter placement could be enhanced.

We have conducted a number of experiments showing that in-situ ultrasound avoids errors found with the conventional ex-situ

display. Subjects using the in-situ device localize targets with accuracy equivalent to direct sight (Fig. 3a, from Wu et al., 2005). Moreover, because users see the target at its true location in externalized space, deformation has no measurable impact (Wu et al., 2008). Pushing the transducer against a deformable surface will change the position of the ultrasound slice in the external world, but the image of the target will remain co-located with the target itself.

A future goal is to understand the source of the systematic error in target localization that we have demonstrated untrained subjects exhibit with ex-situ ultrasound, even controlling for deformation. One possible approach is to evaluate radiologists at different stages in training. Another research direction is to extend the use of ultrasound to visualization of 3-D structures under the skin. We believe in-situ imaging can help users to integrate a series of slices obtained by sweeping the transducer through space, which could expand the applicability of ultrasound in clinical practice.

---

### Recommended Reading

- Hegarty, M., Keehner, M., Cohen, C., Montello, D.R., & Lippa, Y. (2007). The role of spatial cognition in medicine: Applications for selecting and training professionals. In G. Allen (Ed.), *Applied spatial cognition* (pp. 285–315). Mahwah, NJ: Erlbaum. An overview of spatial processing in medicine with emphasis on individual differences.
- Klatzky, R.L., & Wu, B. (2008). The embodied actor in multiple frames of reference. In R. Klatzky, M. Behrmann, & B. MacWhinney (Eds.), *Embodiment, ego-space and action* (pp.145–176). New York: Psychology Press. Discusses basic issues related to spatial visualization, particularly frames of reference.
- Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum. A classic article on interpreting medical images.
- Loomis, J.M., Klatzky, R., Avraamides, M., Lippa Y., & Golledge, R. (2007). Functional equivalence of spatial images produced by perception and spatial language. In F. Mast & L. Jäncke (Eds.), *Spatial processing in navigation, imagery, and perception* (pp. 29–48). New York: Springer. A discussion of amodal spatial images and how to test for them.
- 

**Acknowledgments**—This research was supported by National Institutes of Health Grant # R01EB00860-03 and National Science Foundation Grant # 0308096.

### REFERENCES

- Avraamides, M., Loomis, J., Klatzky, R.L., & Golledge, R.G. (2004). Functional equivalence of spatial representations derived from vision and language: Evidence from allocentric judgments. *Jour-*

- nal of Experimental Psychology: Human Learning, Memory & Cognition*, 30, 801–814.
- Blaisdell, A.P., & Cook, R.G. (2005). Integration of spatial maps in pigeons. *Animal Cognition*, 8, 7–16.
- Bryant, D.J. (1992). A spatial representation system in humans. *Psychology*, 3(16), 1.
- Golledge, R.G., Smith, T.R., Pellegrino, J.W., Doherty, S., & Marshall, S.P. (1985). A conceptual model and empirical analysis of children's acquisition of spatial knowledge. *Journal of Environmental Psychology*, 5, 125–152.
- Holding, C.S., & Holding, D.H. (1988). Acquisition of route network knowledge by males and females. *Journal of General Psychology*, 116, 29–41.
- Irwin, D.E., & Andrews, R.V. (1996). Information integration in perception and communication. In T. Inui & J. McClelland (Eds.), *Attention and performance* (Vol. 16, pp. 125–155). Cambridge, MA: MIT Press.
- Klatzky, R.L., Lippa, Y., Loomis, J.M., & Golledge, R.G. (2003). Encoding, learning and spatial updating of multiple object locations specified by 3-D sound, spatial language, and vision. *Experimental Brain Research*, 149, 48–61.
- Loomis, J.M., Lippa, Y., Klatzky, R.L., & Golledge, R.G. (2002). Spatial updating of locations specified by 3-D sound and spatial language. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition*, 28, 335–345.
- Moar, I., & Carleton, L.R. (1982). Memory for routes. *Quarterly Journal of Experimental Psychology A*, 34, 381–394.
- Richardson, A.E., Montello, D., & Hegarty, M. (1999). Spatial knowledge acquisition from maps, and from navigation in real and virtual environments. *Memory & Cognition*, 27, 741–750.
- Stetten, G., & Chib, V. (2001). Overlaying ultrasound images on direct vision. *Journal of Ultrasound in Medicine*, 20, 235–240.
- Sturz, B.R., Bodily, K.D., & Katz, J.S. (2006). Evidence against integration of spatial maps in humans. *Animal Cognition*, 9, 207–217.
- Wu, B., Klatzky, R.L., Shelton, D., & Stetten, G. (2005). Psychophysical evaluation of in-situ ultrasound visualization. *IEEE Transactions on Visualization and Computer Graphics*, 11, 684–693.
- Wu, B., Klatzky, R.L., Shelton, D., & Stetten, G. (2008). Mental concatenation of perceptually and cognitively specified depth to represent locations in near space. *Experimental Brain Research*, 184, 295–305.
- Wu, B., Ooi, T.L., & He, Z.J. (2004). Perceiving distance accurately by a directional process of integrating ground information. *Nature*, 428, 73–77.