

## “Put-That-There”: Voice and Gesture at the Graphics Interface

Richard A. Bolt

Over a weekend in mid-May 1979, I outlined an approach to creating and manipulating items on a graphics display using a magnetic space-sensing cube in concert with connected speech recognition technology. I presented my thoughts in a memo to my colleagues at the MIT Architecture Machine Group (precursor of today's MIT Media Lab). The scheme seemed plausible, and we decided to go ahead and build a demonstration prototype. While we already had a couple of space-sensing cubes in-house, we had to place a rush order for the speech recognizer—an item, in those days, by no means inexpensive.

Things moved fast after the recognizer arrived. In less than two weeks, Chris Schmandt (then a grad student, now leading speech research at the Media Lab) had designed, written, and debugged an initial version of the program—by then dubbed “Put-That-There.” Even in its first form (the items created and manipulated being simple squares, circles, triangles, and rectangles), the demo had undeniable impact. You saw immediately what was going on: a person addressing a computer display using everyday speech and gesture. Chris proceeded to create a second, more elaborate version of Put-That-There—one that let you create and direct a fleet of ships about the Caribbean; a variant of this second version involved icons of forts, campsites, and meetinghouses set against a map of colonial Boston.

In the years since, as part of an extended agenda in multimodal natural dialog, my students and I have developed prototype systems, demonstrating, for example, two-handed freehand gestural input with accompanying speech (Bolt and Herranz 1992) as well as concurrent speech, gesture, and gaze input (Koons et al. 1993). But even though these systems were vastly more sophisticated in both instrumentation and software intelligence than was Put-That-There, that earlier demo yet retains a certain *éclat*. Even after 17 years, looking at a video of the demo, you sense something special when Chris, seated before our media room's large screen, raises his hand, points, and says “Put that (pointing at a blue triangle) . . . there (pointing to a spot above and to the left),” and lo, the triangle moves where he told it to. I have yet to see an interface demo that makes its point as cleanly and succinctly as did that very first version of Put-That-There.

### REFERENCES

- Bolt, R. A., and Herranz, E. 1992. Two-Handed Gesture in Multimodal Natural Dialog. In *Proceedings of UIST '92, Fifth Annual Symposium on User Interface Software and Technology*, Monterey, CA, November 15–18, 7–14.
- Koons, D. B.; Sparrell, C. J.; and Thorisson, K. R. 1993. Integrating Simultaneous Input from Speech, Gaze, and Hand Gestures. In Maybury, M. (ed.), *Intelligent Multi-Media Interfaces*, 257–276. Menlo Park, CA: AAAI Press.