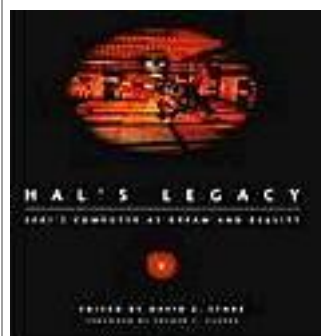


REVIEW OF HAL'S LEGACY: 2001'S COMPUTER AS DREAM AND REALITY

David. G. Stork, Ed.
*HAL's Legacy: 2001's Computer as
Dream and Reality*
1997
ISBN 0-262-19378-7
\$22.50
384 pp.

MIT Press
Cambridge, MA, USA

<http://www-mitpress.mit.edu/Hal>



Reviewed by Philip Hubbard, [Stanford University](#)

The central character of the landmark 1968 film *2001: A Space Odyssey* is a computer named HAL. As those who have seen the film may recall, HAL has a remarkably human set of qualities: along with his monitoring of the ship, he can see and hear, play chess, and more importantly, speak and understand English perfectly. Following the model of computers in the 1960s, HAL is a room-sized central processing unit in the bowels of the ship, but with eyes, ears, and mouth distributed throughout it. He is in a sense, an omnipresent intelligence, interacting in a warm and friendly though occasionally patronizing manner with Dave and Frank, the human members of the crew who stay awake during the long voyage to Jupiter while the rest of the crew hibernates. HAL is the vision of author Arthur C. Clarke and director Stanley Kubrick, an extrapolation some thirty years into the future. Given the painstaking attention to realism that characterizes the film as a whole, HAL is a convincing, even inspirational extrapolation.

In Clarke's novel based on the film, HAL tells us "I became operational ... in Urbana, Illinois on January 12, 1997." In honor of that date, David Stork has assembled a remarkable set of researchers in artificial intelligence and related areas to comment on the accuracy of Clarke and Kubrick's prophesy in *Hal's Legacy: 2001's Computer as Dream and Reality*. The sixteen chapters cover the full range of HAL's human capabilities and together paint a picture of how close we are to creating the kind of sense-endowed, intelligent personality that HAL symbolizes.

If we are as close as some of the contributors think to assembling significant elements of human intelligence inside a computer, then the ramifications for language teaching--and for education in general--are astonishing. This review will go through the articles most relevant to the future of language teaching, covering artificial intelligence, speech synthesis, speech recognition and understanding, and the display and recognition of emotion. It will conclude with a discussion of what these developments mean for the foreseeable future of language teaching.

After an introductory chapter, the book follows with Stork's interview with MIT's Marvin Minsky, one of the founders of the field of artificial intelligence (AI). Along with his reminiscences about the filming of *2001* (he was a science consultant on the set), Minsky comments on how AI researchers have largely

moved away from looking at general intelligence to focusing on specific areas, expert systems like IBM's chess-playing specialists. His view is that this focus on specialized AI has set the field back. Though Minsky acknowledges that we have "collections of dumb specialists in small domains" (p. 27), he believes we will not have the kind of generalized AI HAL represents until we develop a better set of tools for knowledge representation on multiple levels.

Minsky's position is echoed in a later chapter by Douglas Lenat, the founder of the AI research company Cycorp. Since the early 80s, Lenat and his team have been working on developing a machine with general intelligence and common sense. Among the interesting elements of Lenat's CYC program are that it structures knowledge in "microtheories" that appear somewhat analogous to the scripts and schemata in models of human cognition. And this knowledge comes not just from direct human programming: CYC is being programmed to learn on its own as well. This general learning approach is also being taken by Cog, a humanoid robot described in the chapter by Tuft University's Daniel Dennett. Cog, an inhabitant of MIT's Artificial Intelligence Lab, is being constructed with senses and limbs: the plan is for it "to go through an embodied infancy and childhood, reacting to people it sees with its video eyes, making friends, learning about the world by playing with real things with its real hands, and acquiring memory" (p. 358).

While the generalized AI described above is still a ways off, other chapters of the book show just how close we are to having programs with many of HAL's features in specific domains. These domains are built on three areas: speech synthesis, speech recognition, and computer vision.

Joseph Olive's contribution is in the area of speech synthesis. Olive, the head of Bell Lab's Text-to-Speech research division, describes the history of attempts to build talking machines and outlines the hurdles still left to overcome. These include the computer's "understanding" of what it's saying and the production of native-like stress, rhythm and intonation patterns. It is important to note, though, that these hurdles are incremental to a large degree. High-end speech synthesis today is moving increasingly closer to a native model. In describing what has already been accomplished, Olive notes that existing programs can reliably perform voice verification, so that a talking computer could "know" who it's talking to. He also describes Bell Lab's Talking Head: a 3D model of a human head which when linked to the lab's speech synthesizer moves its lips, jaws and tongue appropriately, giving the synthesized speech a personality.

Undoubtedly the richest chapter in the book in terms of breadth of information, prediction, and general optimism is the one by Raymond Kurzweil. Kurzweil, a leading researcher in speech recognition who runs his own company, describes the state of the art as well as the immediate and longer-term future of both speech recognition (the process of getting a reliable rendering of the words someone said) and speech understanding (the much more complex process of representing the meaning of what was said). He explains the different approaches used in current speech recognition systems, systems that work surprisingly well within specialized domains. For instance, in 1996 about 10% of doctors in emergency rooms in the US were dictating their medical reports using speech recognition systems.

Of all Kurzweil's statements, his short-term predictions have the most impact because they represent extrapolations from existing prototypes. These include:

- by 1997-98, commercially available, speech recognition systems in restricted domains (such as law);
- by early 1998, commercial speech recognition systems for business English that are speaker independent (require no training) with around a 97% accuracy rate initially;
- by 2001, systems that can produce a written transcription of a movie from a soundtrack.

And in the next ten years:

- unrestricted, speech to text dictation systems
- translating telephones (though he doesn't comment on how native-like the translations will be).

Given the current pace of development, he foresees a machine with the "raw computing power" of the human brain in about two decades. This is a prediction that appears in other chapters as well: over the past few decades, computer speeds and densities have doubled roughly every 18 months, and Kurzweil offers convincing evidence that this pace is likely to continue. In tandem with the development of increasingly more powerful and accurate medical scanning devices, such a system could conceivably image a human brain and reverse engineer it into silicon or even copy an individual brain, including memories, into a computer. He calls this a "modest extrapolation of current trends based on technologies and capabilities that we have today" (p. 166). In Kurzweil's world, HAL would simply be a copy of a human.

Roger Schanck, Director of the Institute for the Learning Sciences at Northwestern, also grapples with the issue of computer understanding. He addresses the question of whether we could build a computer today that is "not a fake, yet is fully capable of engaging in the dialogues HAL engages in" (p. 184) and answers in the affirmative. His approach--in marked contrast to that of Minsky, Lenat, and Dennett--is to create dedicated programs that are specialists, local experts who are very good at operating within a given context. Given the appropriate knowledge structures, such experts could engage in very human-like interactions within their designated domains. They could, for instance, use "pre-stored stories" in their conversations in much the same way humans do.

In the movie *2001*, a key turn of the plot occurs when HAL is able to read the lips of Dave and Frank as they discuss his behavior inside a soundproof pod on the ship. The current state of lipreading, or more accurately, speechreading, is described by the book's editor, Stork. While this seems a rather odd area for research at first glance, it turns out that existing speech recognition systems can have their error rates improved from 10% to 50% by integrating a visual speechreading component. In a separate chapter on computer vision in general, Azriel Rosenfeld, the Director of the Center for Automation Research at the University of Maryland, notes that computers already exist that can reliably recognize facial expressions of surprise, fear, disgust, anger, happiness, and sadness. He predicts that by 2001, computers will reliably recognize not only facial expressions but also certain body postures and gestures, along with a wide variety of objects and actions.

Rosalind Picard of MIT's Media Laboratory presents perhaps the most intriguing contribution with her chapter on affective computing. One area of affective computing centers on the ability of the machine to detect the user's emotional state and respond differentially to it. By recognizing emotion not only through visual cues but also through voice quality, timing, intonation, heart rate, and other inputs, the computer as a polygraph should be able to determine states like frustration and satisfaction. She also discusses the other end of affective computing: the programming of emotions within the computer itself. Given our current understanding of human intelligence, it seems unlikely that we could ever have AI without some form of emotional intelligence as a part of it.

HAL, though disembodied, was a computer that interacted with humans in a human way. By taking pieces from the various articles, it is possible to conclude that we have the technology available *today* to make a virtual teacher, tutor, or conversation partner that could be imaged with a human face and animated with some degree of realism, that could recognize and within a limited domain even understand some of what was said, that could talk back with synthesized or prerecorded speech, and that could recognize who it is talking to and at a rough level infer the user's emotional state. It is also clear that the

virtual teacher, tutor, or conversation partner we could make next year would be better--faster, more reliable, more generalized--than the one we could make this year.

Computer assisted language learning (CALL) has taken off in two directions recently. In one of these, centered on the world wide web and electronic mail, the computer is a tool for exploring, gathering information, and communicating and collaborating with others. In the other, built on increasingly sophisticated multimedia, the computer is a delivery system for interactive language experiences, often with dramatic or game elements in addition to more traditional practice activities. Given the state of the art reported in *HAL's Legacy*, we seem to be on the brink of a third direction for CALL, one based on interaction with virtual humans, teachers in the machine, electronic personalities.

We are not going to have HAL's artificial intelligence anytime soon (according to Minsky, somewhere between 4 and 400 years), but we are already beginning to see programs with renditions of people that learners can speak to and get differential responses from. At first, these are going to have a limited range of pre-programmed responses, but by drawing the learners into an engaging conversation, they will make CALL much more participatory than current multimedia. These programs will be followed by increasingly more complex ones which will handle a greater variety of input and be more adaptive to the individual learner. As there will be parallel developments in the entertainment and game industry and in other fields of education, it seems likely that within a few years people will be used to "talking" with such computer constructs. To be prepared for this future, it is necessary to start thinking now about how the field of language teaching is going to be impacted. In the short run, are such programs actually going to enhance language learning? Or are they merely going to give the illusion of enhancing it? If we expect them to actually enhance it, some of us at least are going to have to put some serious thought into how this is going to occur.

If you haven't seen *2001* recently and have access to the video, it's worth watching it before reading this book. And if you have the opportunity to read this book, do so. *HAL's Legacy* is important because it has brought together high-level practitioners in related fields and given them a clear and compelling common frame of reference. In addition to those already mentioned, there are chapters on supercomputer design, fault tolerance, IBM's chess programs, human computer interaction, and computer ethics, among others. The chapters discuss the technical concepts in predominantly non-specialist language, built around the metaphor of HAL. Perhaps most importantly, the contributors are not science reporters or pundits or outside observers seeking to sensationalize: they are insiders who have devoted substantial portions of their lives to making computers more like us, and in at least some realms they appear to be succeeding at a faster pace than most of us are aware.

ABOUT THE REVIEWER

Philip Hubbard is Senior Lecturer in Linguistics and Associate Director of the English for Foreign Students Program at Stanford University. Active in computer-assisted language learning since 1984, he is the author of a number of articles on CALL methodology and CALL programs for teaching ESL vocabulary, listening, and speaking.

E-mail: phil@turing.stanford.edu