9 Beyond the Desktop Metaphor

Nicholas Negroponte

Director, Media Laboratory
Professor, Department of Architecture

Abstract

This paper concludes that the desktop metaphor will be replaced with a theoretrical one, with actors to whom you delegate tasks. Speech is seen as the primary channel of communication, in a plural interface, characterized by extraordinary redundancy and concurrency. This picture of the future includes a gaggle of small, independent, but intercommunicating agents embedded in common objects.

9.1 Hospital Corners

When I learned to make my bed, I was artfully taught and dutifully learned to make hospital corners. This simple fold had both functional and esthetic advantages, and with practice, added no time to the tedium of bedmaking. Today, notwithstanding this skill, I cherish the opportunity of delegating the task and have little interest in the *direct manipulation* of my bedsheets.

Likewise, I feel no imperative to manage my computer files, route my telecommunications, or filter the onslaught of mail, messages, news and the like. I am fully prepared to delegate these tasks to agents I trust, as I tend to other matters (which could be as banal as getting dressed), while those other tasks are brought to a satisfactory conclusion. For the most part today, these agents are humans. Tomorrow they will be

An earlier version of this paper appeared as "An Iconoclastic View Beyond the Desktop Metaphor" in the *International Journal on Human–Computer Interaction*, 1(1), pages 109-113, 1989, and is reprinted with permission from Ablex Publishing Corporation, Norwood, NJ.

machines. But what cannot change in the equation, in the transition, is **trust**.

9.2 Dynadots

Stalking the future is a curious game. More often than not we project our current images, sometimes unwittingly, with mere technological changes which, in the case of computers, include speed, memory and the exotica of I/O. In the text which follows, I have a view of the future not generally held, which enormously affects the human interface and is not a mere technological advance, but a fundamental departure from commonly held views.

The common assumption is that computers will move in the direction of paper and/or clipboard-like paraphernalia. Alan Kay's early vision of the **dynabook** is now beginning to find implementations in the back rooms of many corporate laboratories. The general portrait of that future machine is: about the size of *Time Magazine*; maybe flexible; radio transmitting; high resolution, color and bright display; stylus input; massive memory and lightning speed. When not in transport, such a machine might be hung on the wall in front of a keyboard for more traditional use. Anyway, that is the flavor of most people's view of the future personal computers and is clearly embodied in John Sculley's excellent videotape, "The Knowledge Navigator."

My view includes a slightly different mix of form factors and physical embodiments of computers. Many or most will be small objects (hence *dynadots*, to be in keeping with the jargon) that intercommunicate with each other, serving special needs. I expect to carry much more computing power on my wrist tomorrow than is in my office today. Agents, great and small, will be distributed all over the place.

The reason this dynadot view is important is twofold. For one, it leads to a wider computer presence upon which truly personalized systems will be built. Computers will go underground. My refrigerator will know when it is out of milk, but take appropriate action only after a conversation with my calendar or travel planning agent.[1] Yes, small, highly interconnected objects which compute are a *society of objects.*

[1] "Some terminals of the future will be all knowing rooms without walls. Others will be flat, thin, flexible touch sensitive displays. And others will be wrist watches and cuff links with the right hand talking to the left by satellite." –*The Metaphysics of Television,* N. Negroponte, Methodology of Interaction, North-Holland Publishing Company, Amsterdam, 1980.

The second consequence of this view of the future is that the form factor of such *dynadots* suggests that the dominant mode of computer interaction will be speech. We can speak to small things.

9.3 Speech Works in the Dark

Speech has other values too often overlooked. For example, it allows us to deal with computers out of arms' reach, it works around corners, it carries the richness and information content of tone and prosody, and it is frequently the free channel (like: while driving, dressing or reading).

Our tendency is to couple speech recognition with natural language understanding. While this is surely a proper long term view, it misleads us into believing that it is too far off to consider its current worth and usability.

The "space" of speech recognition is characterized by three axes:

- user-dependent versus user independent,
- small vocabularies versus large vocabularies, and
- discrete versus connected utterances.

In this volume, the hardest corner is clearly:

- user-independent, large vocabularies, connected speech.

Let's examine them one by one.

User-independent speech is a goal driven by two desires: not to require training the machine for each word, and not to be limited to a single user. One is obvious and one is suspect. It does not follow that user-dependent speech systems need to be trained word-for-word, versus evolve over time. And, it certainly does not follow that a system which can recognize anybody's speech is an asset. Sure, it should be able to cope with family, friends and our dog's bark. But, there is no real value in the ability to have such systems cope with random accents or strangers.[2]

[2] The telephone company might argue that such systems would allow for *anybody* to call American Airlines from *any* telephone and hold a discussion about flight plans. But, I can call my personal computer or talk to my wrist watch, instead, and have either of them deal with the airline in ASCII. No. Speaker independence only has niche applications, like vending machines and security systems.

Large vocabularies are also a myth. The secret is not size, but rapid downloading of speech subsets from a very large word space, call these: word windows. At any one moment, the total number of utterances required in a machine at any single moment might be as small as 500. Context is the trigger of what needs to be folded into the system from instant to instant. And remember, we are not talking about discussing Thomas Mann, but delegating, which more or less means issuing commands, asking questions and passing judgments; more like drill sergeant talk.

Figure 9.1

To complete the "problem space" of speech recognition, consider the matter of *connected speech.* Surely we have little interest in talking to a machine in broken language. Much of this in fact can be achieved with multiword utterance recognition. This is not the right way to do connected speech, but addresses the issue by widening the acoustic duration. Alas, sometimes we just might have to use more discrete talk.[3]

The real problem in speech recognition is spatial independence, so that one can speak to computers from a distance, in the presence of noise and the like. Many people think that speech is primarily to liberate you from the keyboard in the sense of not having to type (which some people can not do or feel it is beneath their station). I think it is indeed to liberate us from the keyboard, but in the different sense: of not having to be near it or to try to use it while our hands are tending to something else. Also, right now for personal computers there is a real **captive audience** of people. One cannot deal with them *en passant.*

9.4 "About So Big"

Speech does not stand alone. "So big" might be the small distance between two of your fingers, the space between your cupped hands, the volume embraced by your arms or the size of some huge mountain or building to which you are pointing in the distance. In short, the utterance is meaningless without the parallel channel of communication (in this case) of gesture.

Such plurality has another, and in my mind, greater significance: concurrence and redundancy. My classic (read: often used) story is that of dinner talk in a foreign country. If your hosts ask you, say in French, "Do you want some more wine?" or "Please pass the bread," you will understand them with and even without your high school French. Whereas, if they talk among themselves about politics, you will be totally at a loss unless your command of French is almost fluent.

Most people conclude that this is obvious because the first example is "baby talk," whereas a discussion about political matters engages all kinds of sophisticated metaphors, complex terms, grandiose concepts and oblique referents. These differences are true, but not the important

[3] This is the weakest argument I have. It is a near-term solution. Added with the rest, I suggest that user-dependent, small vocabulary, discrete speech systems could be used today, if people really were interested in a plural human interface.

ones by far. The important difference is that in the baby talk, table talk scenario all the objects and subjects of discourse are in the same **space and time** as you (and your hosts). This suggests that when somebody says "Please pass the water," his or her arm can be stretched out in the direction of the water pitcher, and his or her eyes can gaze upon the empty glass. What this means is that the message is enormously **redundant**, and you can pull the signal out of any of many **concurrent** channels. That's the real difference.

For this reason, when you know a foreign language to a limited degree, a most painful task is to use that language over the telephone. In such situations you are at the mercy of a single channel. But that is the way we use computers today, with single channels.[4]

9.5 Why Winking Works

At a recent dinner party I winked at my wife,[5] and she knew all the *paragraphs* of information it would have taken me (otherwise) to explain the same to some stranger. The reason is quite obvious. A vast amount of shared experiences and robust models of each other make the epitome of communication **be the lack of it.**

In an agent-model of computers, similar familiarity is required in order to preclude relentless explicitness that would destroy the value of agencies. In daily life, it can often be the case that it is easier to do something one's self than to explain it to somebody "new." No, the computer must be an old friend with as many shared experiences (facts, at least) as the future *computer presence* will allow. This is absolutely critical to a new metaphor, beyond the desktop.

[4]Note that when you talk on the telephone your face and body still emit expression, even though you know full well that the person at the other end can see none of it. This suggests that our human system of expression is quite tightly wired with the very same (somewhat uncontrollable) parallelism discussed above.

[5]In some perverted measure of information theory, this could be construed as one bit.

Figure 9.2

9.6 The Theatrical Metaphor

If you are prepared to accept the promise of **delegation** and the viabilit of speech, the desktop metaphor is subject to serious change, soon.[6]

My view of the future of desktop computing (versus dynadots in th case) is one where the bezel becomes a proscenium and agents are en bodied to any degree of literalness you may desire. In the longer tern as holography prevails, little people will walk across your desk (if yc have one) dispatched to do what they know how.

The picture is simple. The stage is set with characters of your ow choice or creation whose scripts are drawn from the play of your lif Their expressiveness, character and propensity to speak out are drive by an event (external) and a style (yours). If you want your agents wear bow ties, they will. If you prefer talking to parallelpipeds, fine.

[6]I have not addressed animation because I think it goes without saying that th will be a dramatic change in three dimensional engines and low cost, real time, hi resolution animation will be as commonplace as *pull-down* menus.

This highly literal model of agents can be dismissed as a foolish scheme to replace serious **icons** with *Snow White and the Seven Dwarfs.* [7] But this begs the question about delegation and speech. In some form, we can expect surrogates who can execute complex functions, filter information and intercommunicate in our interest(s).

Figure 9.3

Direct manipulation has its place, and in many regards is part of the joys of life: sports, food, sex and, for some, driving. But, wouldn't you really prefer to run your home and office life with a gaggle of well trained butlers (to answer the telephone), maids (to make the hospital corners), secretaries (to filter the world), accountants or brokers (to manage your money)[8] and, on some occasions, cooks, gardeners and chauffeurs when there were too many guests, weeds or cars on the road?

[7] In 1977 I recall almost being laughed out of an auditorium when I suggested that a calculator icon would invoke that object on the screen and ease of use would naturally stem from familiarity.

[8] This one might make you nervous.

10 Computation as a Framework for Engineering Education

Harold Abelson

Associate Professor, Department of Electrical Engineering and Computer Science
Co-Leader, AI/LCS Project on Mathematics and Computation

Abstract

Twenty-five years ago, when Project MAC was launched, MIT located it not on campus, but rather in some rented office space across Main Street. Even though everyone recognized computation to be an important technology, computation did not play an important enough educational role to justify making room for it on campus. Twenty-five years later, computer science at MIT has grown substantially—but we are still located in the same rented office space. In all fairness, it is still uncertain whether computer science remains merely the study of an important technology, or whether it has a more central place in MIT's educational mission.

In this paper, I want to hint at that second possibility: that computer science can contribute something fundamental to our educational perspective, not only in teaching about computers, but in teaching about engineering more broadly, and perhaps in teaching about other disciplines as well.

10.1 Formalizing Imperative Knowledge

To begin with, "computer science" is a terrible name for this pursuit. It is not a science. Or at least it is not *yet* a science. It is also not fundamentally about computers—in the same sense that chemistry is

The ideas and examples in this paper are drawn from the MIT Department of Electrical Engineering and Computer Science undergraduate subjects "Structure and Interpretation of Computer Programs" and "Signals and Systems," both of whose development is joint work with Gerald Sussman and William Siebert.