Natural User Interfaces Are Not Natural

Donald A. Norman
Nielsen Norman Group, Northwestern University, KAIST Industrial Design | don@jnd.org

"I believe we will look back on 2010 as the year we expanded beyond the mouse and keyboard and started incorporating more natural forms of interaction such as touch, speech, gestures, handwriting, and vision—what computer scientists call the 'NUI' or natural user interface."

—Steve Ballmer, CEO Microsoft

Gestural interaction is the new excitement in the halls of industry. Advances in the size, power, and cost of microprocessors, memory, cameras, and other sensing devices now make it possible to control by wipes and flicks, hand gestures, and body movements. A new world of interaction is here: The rulebooks and guidelines are being rewritten, or at least, such is the claim. And the new interactions even have a new marketing name: natural, as in "Natural User Interface."

As usual, marketing rhetoric is ahead of reality.

Fundamental principles of knowledge of results, feedback, and a good conceptual model still rule. The strength of the graphical user interface (GUI) has little to do with its use of graphics: It has to do with the ease of remembering actions, both in what actions are possible and how to invoke them. Visible icons and visible menus are the mechanisms, and despite the well-known problems of scaling up to the demands of modern complex systems, they still allow one to explore and learn. The important design rule of a GUI is visibility: Through the menus, all possible actions can be made visible and, therefore, easily discoverable. The system can often be learned through exploration. Systems that avoid these well-known methods suffer.

Gestural interfaces are not new. Gestures have been part of the interface scene since the very early days. Brad Myers' 1998 review describes work in the 1960s and reminds us that they were first commercially deployed in systems for computer-aided design and with the Apple Newton of 1992. Myron Krueger's pioneering work on artificial reality in the early 1980s was my first introduction to gestural interaction with large, projected images. Multiple-touch systems have been around since the 1980s: Bill Buxton's review correlates the date of the first multi-touch system designed for human-computer interaction with the 1982 M.S. thesis of Nimish Mehta. Specialized sensors for detecting human location and movement have long played a role in game design. Musical instruments are both multi-touch and gestural, and electronic input devices such as drum pads and electric guitars extend these modes of mechanical interaction into the world of electronics. But even electronically mediated gestures are over a half-century old for musical instruments: The Theremin, a gesture-controlled electronic music synthesizer, was patented by its Russian inventor in 1928.

Most gestures are neither natural nor easy to learn or remember. Few are innate or readily predisposed to rapid and easy learning. Even the simple head-shake is puzzling when cultures intermix. Westerners who travel to India experience difficulty in interpreting the Indian head shake, which at first appears to be a diagonal blend of the Western vertical shake for "yes" and the horizontal shake for "no." Similarly, hand-waving gestures of hello, goodbye, and "come here" are performed differently in different cultures. To see a partial list of the range of gestures used across the world, look up "gestures" and "list of gestures" in Wikipedia.

More important, gestures lack critical clues deemed essential for successful human-computer interaction. Because gestures are ephemeral, they do not leave behind any record of their path, which means that if one makes a gesture and either gets no response or the wrong response, there is little information avail-

Photograph by Elisabeth Moore

I am reminded of those old comedies of people in formal clothing at auctions doing silent bidding. One person sneezes and thereby purchases an unwanted painting. A couple argues, and as they wave their hands at one another, the waving gets interpreted as ever-escalating bids.

able to help understand why. The requisite feedback is lacking. Moreover, a pure gestural system makes it difficult to discover the set of possibilities and the precise dynamics of execution. These problems can be overcome, of course, but only by adding conventional interface elements, such as menus, help systems, traces, tutorials, undo operations, and other forms of feedback and guides.

Are gestures a powerful mode of interaction? Yes, I have no doubt that gestures will find an appropriate place in the repertoire of interaction systems. The main difference between the systems of today and those developed over the past 50 years is the rise of powerful, inexpensive technologies for sensors and processing, which makes it now practical to deploy these systems on inexpensive, mass-produced items. We have already seen great advances in their use. Gestures will become standardized, either by a formal standards body or simply by convention—for example, the rapid zigzag stroke to indicate crossing out or the upward lift of the hands to indicate more (sound, action, amplitude, etc.). Shaking a device is starting to mean "provide another alternative." A horizontal wiping motion of the fingers means to go to a new page. Pinching or expanding the placement of two fingers contracts or expands a displayed image. Indeed, many of these were present in some of the earliest developments of gestural systems. Note that gestures already incorporate lessons learned from GUI development. Thus, dragging two fingers downward causes the screen image to move upward, keeping with the customary GUI metaphor that one is moving the viewing window, not the items themselves.

New conventions will be developed. Thus, although it was easy to realize that a flick of the fingers should cause an image to move, the addition of "momentum," making the motion continue after the flicking action has ceased, was not so obvious. (Some recent cell phones have neglected this aspect of the design, much to the distress of users and delight of reviewers, who were quick to point out the deficiency.) Momentum must be coupled with viscous friction, I might add, so that the motion not only moves with a speed governed by the flick and continues afterward, but that it also gradually and smoothly comes to a halt. Getting these parameters tuned just right is today an art; it has to be transformed into a science.

Once again, though, the concept of clicking coupled with momentum is old. I first saw this flicking gesture, complete with momentum (although that term was not yet in use) in work developed by Joy Mountford's Human-Interface Group at Apple in the late 1980s to early 1990s.

The timing and dynamics of gestural motions will no doubt be the topic of many dissertations and conference papers. Even today, different groups take different conventions. What should a flicked object do when it encounters the edge of its window, or the edge of the enclosing screen? What if there are multiple screens? If several people are jointly cooperating on a task, but each is using a differ-

ent computer, should a flicked object move from one computer to the other? And if so, how can the sender also retain a copy? (Note that systems that have faced—and created answers to—these issues have existed for quite some time.)

The problems faced by gesture developers remind me of similar issues that arose during the early days of development of the GUI. Thus, in the development of the early Xerox PARC systems, when one moved the icon of a file across the screen to a file folder, it was natural that the icon would disappear into the folder. Similarly, when a file was moved to the trash, it was natural that the icon—and the file—disappeared from sight. But this movement principle got into trouble with the printer: Moving the file to the image of the printer caused the item to be printed, but it also caused it to disappear from the screen. Much rethinking took place then. Much rethinking is required now.

The proper behavior for moving something to a printer is obvious: The object should remain in view. What if the movement is to an external storage device or a different computer? Today, the file stays on the home computer as well. This difference in end result depending upon the nature of the destination is the source of continual confusion for some. What gesture signifies copy rather than move?

Some systems are trying to develop a gestural language, sometimes with the number of touch points as a meta-signal about the scope of the movement. A single finger gesture means one thing, the same gesture with two fingers means another, yet another with three or four. But note the existing failure of attempts to use multiple mouse clicks in this way. A single mouse click points, a double mouse click selects a word, a triple mouse click selects a paragraph. But if each additional click moves up one level in the hierarchy, shouldn't three clicks select the sentence? How well known and followed is that triple mouse click? Note that the early developers of the Xerox Star computer spent considerable effort and time to develop a systematic clicking language; although some of their efforts survived, much was lost.

Physical gestures have other side effects. By their potential to engage the entire body, they can enhance the pleasure and engagement of participants. They can even be used as exercise machines. But they also can do damage.

When the Nintendo Wii introduced its bowling game, the "natural" interface was to swing the arm as if holding a bowling ball, and then, when the player's arm reached the point where the ball was to be released, to release the pressure on the hand-held controller's switch. Releasing the pressure on the switch was analogous to releasing the ball from the hand and it was readily learned and employed. Alas, in the heat of the game, players would also release their hand pressure on the controller which would fly thorough the air, sometimes with enough force to hit and break the television screen on which the bowling lane was being displayed. Nintendo had to issue warnings about the need to fasten a wrist strap, but when that didn't work, it redesigned the wrist strap. The problem remains. (This of course is reinforcement of yet another design dictum: Proper behavior comes about through careful design, not through instruction manuals and warnings.) Is it beneficial for gestures to be natural? Not in this case. Here, the gestural convention was too natural. It led to an unexpected, unfortunate side effect, one that is difficult to overcome.

Those who champion full-gesture systems are apt to respond that they do not need a controller, so there would be no physical object that could do damage. True, but what gesture would they then use to signal when the ball should be released? It is also unlikely that complex systems could be controlled solely by body gestures because the subtleties of action are too complex to be handled by actions—it is as if our spoken language consisted solely of verbs. We need ways of specifying scope, range, temporal order, and conditional dependencies. As a result, most complex systems for gesture also provide switches, hand-held devices, gloves, spoken command languages, or even good old-fashioned keyboards to add more specificity and precision to the commands.

Gestural systems are no different from any other form of interaction. They need to follow the basic rules of interaction design, which means well-defined modes of expression, a clear conceptual model of the way they interact with the system, their consequences, and means of navigating unintended

Ballmer, S. "CES 2010: A Transforming Trend—The Natural User Interface." The Huffington Post, Jan. 12, 2010, from http://www.huffingtonpost.com/steve-ballmer/ces-2010-a-transforming-t_b_416598.html/

Buxton, B. "Multi-Touch Systems that I Have Known and Loved," available from http://www.billbuxton.com/multitouchOverview.html/

Krueger, M.W. *Artificial Reality.* Reading, Mass.: Addison-Wesley, 1983.

Myers, B.A. "A Brief History of Human Computer Interaction Technology." *interactions* 5, 2 (1998): 44–54. http://www.cs.cmu.edu/~amulet/papers/uihistory.tr.html/

consequences. As a result, means of providing feedback, explicit hints as to possible actions, and guides for how they are to be conducted are required. Because gestures are unconstrained, they are apt to be performed in an ambiguous or uninterruptable manner, in which case constructive feedback is required to allow the person to learn the appropriate manner of performance and to understand what was wrong with their action. As with all systems, some undo mechanism will be required in situations where unintended actions or interpretations of gestures create undesirable states. And because gesturing is a natural, automatic behavior, the system must be tuned to avoid false responses to movements that were not intended to be system inputs. Solving this problem might accidentally cause more misses, movements that were intended to be interpreted, but were not. Neither of these situations is common with keyboard, touchpad, pens, or mouse actions.

What do I conclude? Gestures will form a valuable addition to our repertoire of interaction techniques, but they need time to be better developed, for us to understand how best to deploy them, and for standard conventions to develop so the same gestures mean the same things in different systems. And we need to develop the supporting infrastructure to handle guides, feedback, error correction, and the other consequences of gestures, some of which can use well-known procedures, some of which will be novel.

Gesture and touch-based systems are already so well accepted that I continually see people making gestures to systems that do not understand them: tapping the screens of non-touch-sensitive displays, pinching and expanding the fingers or sliding the finger across the screen on systems that do not support these actions, and for that matter, waving hands in front of sinks that use old-fashioned handles, not infrared sensors, to dispense water.

Gestural systems are indeed one of the important future paths for a more holistic, human interaction of people with technology. In many cases, they will enhance our control, our feeling of control and empowerment, our convenience, and even our delight. But like all technologies, gesture-based systems will come at a cost. Different systems will devise different conventions. There will be a learning curve. People with handicaps will have to be accommodated. And there will be an entirely new source of material for comedians. Imagine the problems when a system has a repertoire of dozens of gestures, all of which mean something, but not all of which may be known by the person near the device. I am reminded of those old movie comedies of people in formal clothing at auctions doing silent bidding. One person sneezes and thereby purchases an unwanted painting. A couple argues, and as they wave their hands at one another, the hand waving gets interpreted as ever-escalating bids.

Control of our systems through interactions that bypass the conventional mechanical switches, keyboards, and mice is a welcome addition to our arsenal. Whether it is speech, gesture, or the tapping of the body's electrical signals for "thought control," all have great potential for enhancing our interactions, especially where the traditional methods are inappropriate or inconvenient. But they are not a panacea. They come with new problems, new challenges, and the potential for massive mistakes and confusion even as they also come with great virtue and potential.

All new technologies have their proper place. All new technologies will take a while for us to figure out the best manner of interaction as well as the standardization that removes one source of potential confusion. None of these systems is inherently more natural than the others. The mouse and keyboard are not natural. Speech utterances will have to be learned and gestures carefully developed and standardized through time. The standards don't have to be the best of all possibilities. The keyboard has standardized upon variations of qwerty and azerty throughout the world even though neither is optimal—standards are more important than optimization.

Are natural user interfaces natural? No. But they will be useful.

ABOUT THE AUTHOR Don Norman wears many hats, including cofounder of the Nielsen Norman group, professor at Northwestern University, visiting professor at KAIST (South Korea), and author. His latest book, *Living with Complexity*, started out as a series of essays in this magazine. He lives at jnd.org.

DOI: 10.1145/1744161.1744163