

# PERCEPTUAL BANDWIDTH

In the study of human-computer interaction, one of the two words surrounding the hyphen usually leads. Perceptual interfaces from the

*What happens to people when computers become perceptually complex?*

perspective of computers signal an interest in machines that can accomplish human-like sensory tasks. But it's the *machines* that do the perceiving. There may be clues in the science of human perception about how to automate perceptual tasks, but it's computer code, not neural circuitry, that will ultimately allow machines to simulate human performance.

Perceptual interfaces defined from a *human* perspective suggest similar topics (for example, vision, hearing, speech, touch), but radically different ideas about how sensory experience works and about how machines might change or facilitate human perception. The goal of a human-centered look at perceptual interfaces is to transform the world of computers into features of sensory experience that determine how the interfaces work with people.

We have emphasized the human-centered approach in our lab at Stanford University [10]. The results of our work, taken from the psychology of sensation and perception, offer some surprising insights when applied to the design and evaluation of interactive media, especially media that seek to increase *perceptual bandwidth*.

We have used these insights to help with four important questions about perceptual interfaces:

- How should perceptual interfaces be defined?
- How do perceptual interfaces work?
- How can psychological research help create better perceptual interfaces?
- Is perceptual bandwidth in interfaces a good thing?

## What Are Perceptual Interfaces?

One common definition of perceptual interfaces is they describe features that increase perceptual bandwidth. Perceptual interfaces can interact with users via more and different sensory channels than are possible with traditional

interfaces. If a traditional interface has only typing and mousing in response to pictures and words, a perceptual one adds speaking, touching, gesturing, emoting, and gazing.

This notion of perceptual interfaces illustrates the enthusiasm for more elaborate sensory exchanges between humans and computers. But other elements of a definition must be added to go beyond a simple notion that the number of senses involved in the exchanges is greater. First, definitions need to specify whose perceptions and of what. Second, definitions should provide categories that help organize the different elements of perception. We'll try to answer the first issue with a model of communication exchanges, and the second with a brief review of the domains of perception research.

This figure shows a model of an interaction. Two people (A and B in the model) are interacting via a computer (C) about objects in the real world. We could limit the model to an exchange between one of the people and the computer in the case of human-computer interaction, or keep the other person in the equation in the case of computer-mediated communication. In either case, however, this simple model shows at least four different perceptions that are relevant to any exchange.

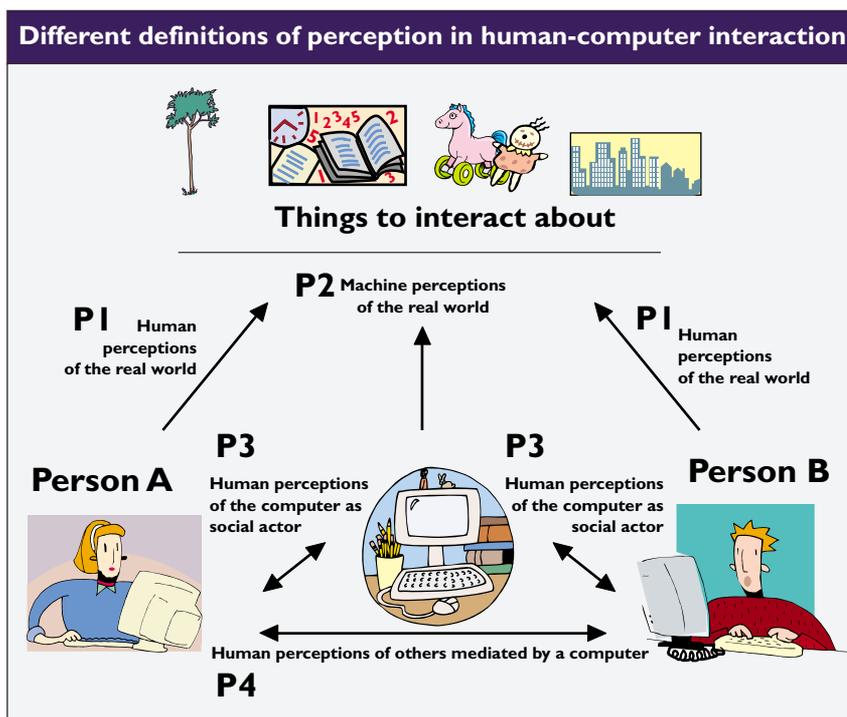
First, there are perceptions that people have of the outside world (P1). This constitutes all of the considerations that are part of any psychological treatment of perception. During any human-computer exchange, people perceive the real world using all of the senses available to them.

Second, there is the perception of the outside world sensed by the computer (P2). This perception includes the ability of computers to recognize people and objects, sense the emotions of interactants, or identify the personality or gender of a user. This perception, however, is virtual, and hence not regulated by human psychology. Indeed, successful perceptions in the machine domain need not follow the rules of human perception at all.

The other two perceptual possibilities in the model (P3 and P4) are the most pertinent to a psychological consideration of perceptual interfaces. Both involve the possibility that a computer can alter perception. In these cases, the computer is a stimulus machine,

reconfiguring reality in different ways, and making it available to more, fewer, or different perceptual systems than might otherwise be engaged.

There are two different sources of information that a computer could present as a perceptual stimulus. The computer could be nothing more than a conduit to another person or object (P4). This would be the case in all computer-mediated communication, including teleconferencing and other technologies that enable two or more people to sense each other



using technology as an intermediary. In these cases, the perceptions of other people and objects may change because stimulus information about them is filtered through the representational capabilities of a machine. In a teleconference, for example, the perception of people may be changed because a computer alters the way people speak, move, or gesture.

In addition to computer-mediated communication, a computer could also present virtual social actors that are automated interactants (P3). This would be the case for human-computer exchanges with avatars or other manufactured interactants. New perceptual considerations are added in these cases. For example, there is a greater range of representational choice (for example, photorealistic faces, recorded voices) as well as representations that could not exist in the real world (rapidly morphing images, text-to-speech synthesis).

The perceptual issues we emphasize in our lab are those that arise when computers mediate person-to-person exchanges, and when they automate the pre-

sensation of another social actor. We have not focused on the most computing-intensive and information-rich contexts, such as VR systems and immersive games, which often create veridical social actors and natural environments. Instead, we have focused on simpler and more ubiquitous contexts. Our goal has been to apply what is known about unmediated perception to the interactions enabled by computers. Essentially, this is a process of stealing the best of psychology, substituting a computer in the interaction, and then applying the conclusions—usually straightforwardly—to the resulting human-computer relationship.

The application of psychological research depends critically on defining interactions in terms of human perceptual capabilities. What perceptual stimuli can computers produce that will affect how people experience information that computers deliver? One useful way to answer this question is to adopt the same distinctions that social scientists use to categorize the world of perceptual stimuli.

The psychology of perception considers human experiences caused by stimulation of the senses [2]. This includes the chemical senses (taste and olfaction), the cutaneous senses (the skin and its receptors), and most importantly, vision and hearing. The latter two categories encompass the most extensive literatures. In the case of vision, there are visual mechanics, color, brightness and contrast, objects and forms, depth, size, and movement. Hearing includes psychophysics (loudness, pitch, timbre, sound localization), physiological mechanisms (the auditory components of the ear, and the neural activity associated with hearing), and the perception of speech (units of speech such as phonemes and the mechanics of word recognition).

This list—a typical map of psychological research about perception—has some critical implications for a definition of perceptual interfaces. First, it is important to note what is not on this list—most notably, social issues. The goal of humanizing interactions with a computer involves a significant amount of literature in psychology. Social issues include how interfaces should present consistent personalities, the various social roles that an interface might adopt (such as that of teammate, occupational roles, gender roles), or how an interface should express criticism, repair errors or give praise. These issues are critical, but not necessarily perceptual. In social sciences, the differences between perceptual and social issues are distinguished by how fundamental they are—sensing and physiology are primitive—while social interaction relies on more thoughtful processes.

A second important implication is the prominence of vision and hearing. Since the earliest days of computing, these two elements have been the fundamental forms of

output (although sound was limited to a simple warning beep). The first pictures (and even icons in a graphical user interface) and the first rich sounds were likely more important perceptual thresholds than might be true for the addition of other senses like touch or smell. As pictures and sounds increase in realism, there are certainly thresholds left to cross. But computing, however sophisticated it may become, will still be primarily about the senses of sight and sound, not because other senses won't be added to machines, but because sight and sound dominate human perception [2].

Third, the psychology of perception emphasizes human speech [3], and on this feature, there is similarity in the psychological emphasis and commercial excitement. We already know that when machines can speak and recognize speech, the social responses to computers change. For example, people are better able to recognize personalities in an interface, and they find computers more engaging. The addition of speech capabilities may well be the single most significant new perceptual interface feature in the near term.

Fourth, there are interesting possibilities for more elaborate interactions using the sense of touch. The skin senses are divided into three categories. The mechanoreceptors respond to indentations of the skin; the thermoreceptors to specific temperatures and changes in temperature; and the nociceptors to intense pressure or high heat. The ability to sense with skin is critical for fine motor coordination (for example, manipulating a joystick), and it is critical for survival because changes in skin sensitivity, for example, can warn people about potential injury. The primitive significance of touch suggests people will also find these cues useful when interacting with a computer, and that people will feel comfortable when they appear because of their ubiquity and importance.

There is a final implication of this list for how we might think about the differences between computers and other media. The boundaries between computing, and TV and film, for example, are not substantial, even if the industries that support them continue on separate paths. Attention to distinctions between old and new media will allow us to separate features of new media that work like they have in the past from features that are perceptually novel and that open new areas of study. This is one reason why the experiments in our lab move freely between computers and other media as stimuli in the research [10].

### **How Do Perceptual Interfaces Work?**

Theories about how perception works come from research about the psychology of objects and people in the real world. This leaves a question, quite significant for some, about whether mediated information will

work any differently. We have a simple and well-tested response: Human-computer interaction is fundamentally social and perceptual in exactly the same ways all other interactions with people and the physical world are social and perceptual. This means the lessons of psychological research about perception can be applied to media with few considerations for the special status of technology.

Consider the example of how people understand the contents of a simple 2D picture; for example, the tree at the top left of the figure. There are plenty of reasons to believe the perceptual experience of this picture should be different from the experience of the real tree that it depicts. This picture doesn't reflect the same light as a real tree. Depth cues are missing and we can't tell how large it is. And the picture looks the same regardless of how the reader moves relative to the picture.

A conclusion that perception of the picture of the tree was significantly different from the perception of the actual tree, however, would be a mistake. In the world of perceptions, close counts. There is no special mental calculation that humans perform to translate the incomplete cues of pictures into images more consistent with real life. Pictures were never important for humans during evolution, and consequently, the human brain is not specialized to separate them from other experience. Pictures appeal to us because they engage neural machinery that had previously evolved for other purposes. The consequence is that everyone can easily understand pictures and they will evaluate their significance as they would the same objects and events in the 3D world.

The same logic can be applied to other important dimensions of social and perceptual life. If the cues available to our perceptual systems are close enough to those we are evolved to notice, then all of the same evaluations that we make in real life will be true with media as well. This has been the central conclusion in the summary of work in our lab, a conclusion that applies equally well in the domains of personality, emotions, and social roles. So fundamental is this conclusion that the title of our research summary, *The Media Equation* [10], was chosen as a recognition that mediated life equals real life. New media engage old brains, and to the extent that new interactions mimic real life, then the principles that explain perception in real life can be applied straightforwardly to computers and other media.

### **How Can Research Aid the Design of Perceptual Interfaces?**

Our interest has been to understand the different ways that perceptual systems of humans are important determinants of human-computer interaction. We

have taken cues in this research from psychology, applying where possible the most useful theories about human sensory experience to the study of mediated interactions. Here are examples of our research, and some of the conclusions about how the research might influence the design of perceptual interfaces.

***The perception of motion.*** The onset of motion is a fundamental perceptual cue. Things that move demand attention, especially if the motion is toward us. There are several parts of the visual system that are extremely motion-sensitive, and there is specialized neural circuitry for motion perception. Orienting responses to motion occur automatically; a perceptual awakening that prepares people for possible action.

We have tested the application of this perceptual law to the study of media, measuring the levels of brain activity and changes in other perceptual responses that might signal action readiness, even though the stimulus is on a screen. We have found that objects moving toward viewers cause perceptual orientations at the onset of their movement. All available mental resources are directed at the moving object. Furthermore, this orientation is not merely a primitive response with no subsequent effect on how people think. Rather, the orientations begin periods of maximum attention, and they mark those parts of a media presentation most likely to be remembered [11].

The implications for computer presentations rich in perceptual cues are numerous. Motion can guide attention; use it to initiate a task sequence. But the processing of motion shuts down other thinking; don't let things move in an interface when people need to read text or concentrate on a task. Motion is even more distracting when it occurs in peripheral vision; try to keep characters and icons still on the boundaries of attention. Constant motion can cause perception to shut down; give people visual breaks and opportunities to escape the demands of constant perceptual changes.

***The perception of novelty.*** Novel people and places are perceptually more interesting than familiar ones. In real life, it's usually only specific features of the environment that change while everything else stays constant. But with media, much more is possible. In milliseconds, an entire visual display can change (in the case of film and television through the use of scene changes and cuts; in the case of computers through documents and displays stacked one on the other.)

The perceptual salience of novel media presentations is clear—they get attention. Visual discontinuities increase cortical arousal (as indicated by brain wave patterns during viewing) [11]. When the different displays are unrelated, the responses are even more jarring than when the sequences appear related to the same story or task [1]. The upside of changing visuals

is they can make presentations more dynamic by increasing interest. However, when there are a large number changes in a short sequence, people pay less attention because the sequence becomes so complex that they tune out [12].

Some examples of design rules for changing visuals include: visual changes cause disruption, so insure they connect related material; changes join unrelated material, so allow people time to adjust; relatedness of material can be signaled by showing similar visuals on both sides of a visual interruption, so change as few features as possible when switching from one visual to another; and cuts mark important steps in a process, so use them to help people remember a sequence of events.

***Display size and the perception of media content.***

The perception of size and distance is a significant psychological issue, largely because size matters a lot in the world of perceptions. It's important to know how far you are from danger or opportunity (by knowing how to judge distance cues from size), and it's important to keep track of the physical attributes of those with whom you interact. Size is a benefit in everything from job interviews to presidential elections (the taller candidate almost always wins).

The perceptual importance of size in mediated interactions is clear in the case of displays. New media enable the same digital information to be played on devices held in your hand or on screens that cover a wall. Our research shows larger displays are preferred and they create a greater sense of presence [10]. Of more importance, however, larger displays are more arousing, as measured by skin conductance levels and heart deceleration when visual material first appears [9]. The arousal results are particularly interesting because of the solid relationship between arousal and memory for media experiences: The higher the arousal, the better the memory [10].

Pictures should be produced with as much thought about size as possible. A face to appear in a window on a desktop computer could be framed to fill available space. The same picture on a wall, however, could be overpowering because the face would appear to be much larger than life. The arousal that accompanies large-screen displays competes for the same mental effort that could otherwise be given to thinking hard about information. Hence, it's best not to overdo size when learning and memory for information (as opposed to memory of the experience) are the important goals.

***The perception of faces and voices.*** Humans are biased to see other humans everywhere. People can see a face in an electric socket, in shadows on the wall, and even with two dots and a line. Similarly, people hear voices everywhere, from oboes to bird songs to the beeps of R2D2.

Given this remarkable acceptance of nonhuman faces and voices as essentially human, it's tempting to assume that as long as a computer had a remotely human representation, no further thought would be necessary. Unfortunately, this is not the case. In one study, we had people interact with stick figures or richer animated representations. Even though the content in the interaction was identical, the more compelling and life-like figures were seen as more intelligent, obtained more conformity from the user, and were more likable [3]. This is consistent with the social psychological finding that attractive people are perceived more positively on a wide variety of dimensions.

We have also explored what happens when designers mix modalities of varying quality. In one study, we presented users with either a computer-synthesized face (which did not quite look human) or no face, or synthesized speech (which was deficient in prosody and clarity) or recorded speech. We might assume the combination of the better representations (that is, synthesized face combined with recorded speech) would be the most desirable and comfortable. Instead, we found people were more willing to disclose personal and undesirable information about themselves, and the characters were more socially present when there was consistency between perceptual modalities [5].

The principle of consistency seems to be a general and powerful one in responses to many different interface features. For example, we have shown when the posture of a character (even a stick figure) on the screen is inconsistent with the language the character uses, its credibility is undermined. People perceive the character as less intelligent, less trustworthy, and less persuasive than when the character's posture is matched with its words [6]. Similarly, when a Web-based auction site had a text-to-speech voice whose volume, pitch, pitch range, and speed was inconsistent with the description of the items, the mismatch undermined purchase behavior, trust in the descriptions, and perceived quality of the items [7].

***Perceptions and the bias of stereotypes.*** When one person encounters another, the first salient perceptions are of gender and ethnicity. In the lab, we have attempted to determine whether people focus on the same characteristics when humans are presented on a computer. In one study, we presented users with a male or female-recorded voice that taught people about various topics. Another computer with a different male or female voice then commented on the performance.

Because praise from a male in real life is often taken more seriously than praise from a female, both male and female participants found the female-voiced computer to be significantly less friendly than evaluations from the male-voiced computer, even though the com-

ments made by each were identical. In addition, the generally positive praise from a male-voiced computer was more compelling than the same comments from a female-voiced computer. Participants thought the tutor computer was significantly more competent (and friendlier) when it was praised by a male-voiced computer, compared to praise by a female-voiced computer. And finally, the female-voiced computer was rated as significantly more informative about love and relationships compared to the male-voiced tutor, while the male-voiced tutor was rated as more informative about computers [8]. Once people perceive a gendered voice, they invoke the entire range of stereotypical beliefs that are perceptually associated with the beliefs (even though all of the subjects denied having any gender stereotypes).

An even stronger demonstration of the power of stereotyping came from a recent study on ethnicity. In this study, Koreans interacted with a Korean or Caucasian on a computer screen. We told half of the participants in the experiment they were interacting with another person via videoconference; the other half were told they were interacting with an advanced computer agent. In fact, both were simply videotapes that simulated an interaction.

It is perhaps not surprising that individuals showed a wide range of in-group favoritism toward the (ostensible) videoconference participants. More interesting, however, was the strong effect exhibited toward the computer agents. In-group participants (that is, Koreans working with Koreans) perceived the computer agent to be more socially attractive and trustworthy, and they perceived the agent's arguments to be better and more similar to the suggestions of their in-group partner. Perhaps most remarkable is that ethnicity made no more difference in human-computer interaction than it did in human-human interaction [6].

### Is More Perceptual Bandwidth Good?

Unfortunately, all of these studies do not allow a conclusion that increases in perceptual bandwidth will be universally good or bad. The most important conclusion is that more and different perceptual experiences turn up the volume on perceptual responses, an outcome that increases the importance of successes and failures in interface design. Getting a perceptual experience right will be better than an experience that is less perceptually rich; getting it wrong could easily make things worse.

This lesson is similar to most other perceptual additions to traditional media. One of the most prevalent hypotheses about media this century is that more and richer perceptions make for better experiences. The particular inventions this century that have enabled these

perceptions vary considerably: moving pictures, wide screens, stereo audio, color screens, CinemaScope, Sensorama, color television, high-definition images, and virtual reality goggles. But the hypothesis was always similar: If you increase the range of sensory experiences available to people, then interactions with media will be better. Indeed, they will be more engaging, more memorable, and more commercially valuable.

However, the assumption that more is always better is misguided. An increase in the breadth and depth of media representations certainly turns up the volume knob on perceptual responses, but greater presence does not translate into greater efficacy or desirability; intensity does not equal quality. Indeed, each increment in perceptual response necessitates a much more thoughtful and careful concern with design principles and strategies, many of which can be derived from the literature in perceptual psychology. ■

### REFERENCES

1. Geiger, S. and Reeves, B. The effects of scene changes and semantic relatedness on attention to television. *Commun. Research* 20, (1993); 155-175.
2. Goldstein, E.B. *Sensation and Perception*. Wadsworth, Belmont, CA, 1989.
3. Lee, E.J. and Nass, C. (1999). Effects of the form of representation and number of computer agents on conformity. In *Proceedings of the CHI 99 Conference*, (Pittsburgh, 1999). ACM, New York, N.Y.
4. Massaro, D.M. *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. MIT Press, Cambridge, MA, 1997.
5. Nass, C. and Gong, L. Maximized modality or constrained consistency? In *Proceedings of the AVSP 99 Conference*, (Santa Cruz, Calif., 1999).
6. Nass, C., Isbister, K., and Lee, E.J. Truth is beauty: Researching conversational agents. J. Castells (Ed.). *Embodied Conversational Agents*. MIT Press, Cambridge, Mass. In press.
7. Nass, C. and Lee, K. M. Does computer-generated speech manifest personality? An experimental test of similarity-attraction. In *Proceedings of the CHI 2000 Conference*. (April 1-6, 2000, The Hague, The Netherlands). In press.
8. Nass, C., Moon, Y., and Green, N. Are computers gender-neutral? Gender stereotypic responses to computers. *J. Applied Social Psych.* 27, 10 (1997), 864-876.
9. Reeves, B., Lang, A., Kim, E.Y., and Tatar, D. (1999). The effects of screen size and message content on attention and arousal. *Media Psychology* 1, (1999), 49-67.
10. Reeves, B. and Nass, C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press/CSLI, New York, N.Y., 1996.
11. Reeves, B., Thorson, E., Rothschild, M., McDonald, Hirsch, J., and Goldstein, B. Attention to television: Intra-stimulus effects of movement and scene changes in alpha variation over time. *Intern. J. Neurosciences* 27, (1985), 241-255.
12. Thorson, E., Reeves, B., and Schleuder, J. Message complexity and attention to television. *Commun. Research* 20 (1985), 155-175.

---

**BYRON REEVES** (reeves@leland.stanford.edu) is the Paul C. Edwards Professor of Communication at Stanford University. He is codirector (with Nass) of the Social Responses to Communication Technologies project at the Center for the Study of Language and Information.

**CLIFFORD NASS** (nass@leland.stanford.edu) is Associate Professor of Communication at Stanford University. He is codirector (with Reeves) of the Social Responses to Communication Technologies project at the Center for the Study of Language and Information.

---