EUROGRAPHICS '99 / P. Brunet and R. Scopigno
(Guest Editors)

Volume 18, (1999), Number 3

Perceptual Principles and Computer Graphics

Jon May

Department of Psychology, University of Sheffield, UK

EXTENDED ABSTRACT

Until comparatively recently, the major problems in computer display technology were caused by the difficulty of making anything recognisable at all. Eighty character-width displays, with eight or nine brilliant green lines per character, slow to respond and slow to decay, somehow enabled people to use their vast new computers with their kilobytes of memory. The pace of change should really astonish us, as we contemplate flat, bright and crisp LCD screens that require separate graphics processors and megabytes of video memory chips just to display our favourite desktop images. It now seems possible for our technological artefacts to display almost anything in as much detail as we would like, whether from a high resolution photographic image or, via skilfully implemented algorithms, by photorealistic rendering from data. In the course of this rapid development, the major problems have themselves changed: now we must ask ourselves what it means for our displays to be recognisable, and what is it in the display that needs to be recognised?

As I am a psychologist, it might not surprise you to hear that I think that there is a role for psychology in answering these questions. The ultimate purpose of any computer display is to be recognised by a human, and visual perception has been a cornerstone of the discipline since Wilhelm Wundt founded the first Laboratory of Experimental Psychology a hundred years ago. The involvement of psychologists in human-computer interaction, and specifically in computer graphics, is not unheard of, of course. In the early days, choices of phosphor and of screen refresh times were driven not just by technical and manufacturing constraints, but also by detailed studies into phenomena such as critical flicker fusion frequency and contrast sensitivity. The introduction of colour within displays was (sometimes) backed up by usability studies showing that (sometimes) it improved performance. Compression algorithms were designed to take into account the discriminability of different levels of hue and saturation by the human visual system.

Contributions such as these have played an important guiding role within the development of computer graphics by providing principled and empirically justifiable ground rules. If the human visual system cannot see something, then you know there is no point displaying it like that. Now that the basic visual properties of displays have been determined, we know how to make displays that are readily perceivable and which are, to all extents and purposes, veridical. Seeing something upon a computer screen is now no more difficult than seeing it in any other representational medium: so why should psychology have a role?

Psychology still has a role because it is about more than low level constraints upon the visual system. It is true that if you go into any psychology department and ask to speak to an expert in vision, you will find that they are concerned with low level problems such as the perception of optical flow, or of binocular disparity in depth perception, or of texture discrimination. The visual stimuli that participants in their psychophysical experiments observe are dots, crosses, and lines of pure colours, not photorealistic or veridical images. Their research is, after all, still directed towards understanding the way by which information gathered by the retina is perceived at all. It is true that computer graphics has gone beyond this stage. There are other psychologists, though, who research later stages in perception, and it is this work that should now be of interest to the computer graphics community.

In providing a photorealistic image, the problem of making something recognisable has been solved, or at least, overcome. We are now faced with the problem of ensuring that the viewer sees what we intend them to see, rather than something else in the image. We have to ensure that the image is not ambiguous, and that the viewer will not interpret it as something other than we intended, or too rich in information suchthat the viewer is unsure about the relevant aspect. We have to be sure that the veridicality of the image does not lead the viewer to treat it as if the imaged scene were really present, and to respond to some channel of information that we have unwittingly introduced. These are the problems that can be investigated by cognitive psychologists, who are interested in the interplay of our thoughts, ideas, memories, emotions and perceptions.

Consider the addition of a realistic face to an audio signal that has been generated computationally, not spoken by a real human. I have read wonderful papers describing techniques for optimally rendering lip and face movements so that the generated face speaks visually in exactly the way a real person would. But the result is a rather Vulcan appearance. The absence of emotional components in the facial expression does not mean that we fail to attribute any emotional content to the message. Instead, we actively infer "absence of emotion", which is a very definite and undesirable emotional state. To overcome this, cheek, eye and forehead components can be manipulated to form a basic grammar of affect, and these can be overlaid to add emotional tone: but now we need to know what emotional tones are appropriate, and how they interact with the emotional state of the viewer. Is an interrogative raised eyebrow always perceived as such, or does it appear condescending if the viewer is unsure of themselves?

Another problem that has attracted a large amount of research effort is that of the realistic rendering of motion through three dimensional space, and how to link it in a usable and "natural" fashion with user interface actions. It has been a long time since our ancestors swung through the trees, after all, and while swooping through abstract cyberspaces may become second nature to our descendants, at the moment the best CAD packages still risk making people nauseous. Leaving aside the buzzword of "intuitive", which often means no more than "usable by other people in my lab", it is worth stepping back and asking whether it really is in the interests of the viewer to have every frame of their trip from A to B animated in front of their very eyes. Especially if they tend to close them to preserve their lunch. Is it worth using all those processor intensive routines to interpolate and ray-trace and blur, when film directors find it just as convenient to cut directly between camera positions? Pans and zooms in cinematography are limited by extensive conventions, that have been developed by a century of experience, and so film should have something to say for motion in computer graphics.

There are two real reasons why this wider world of knowledge has been difficult to integrate with computer graphics. First, the very pace of change mitigates against the systematic application of interdisciplinary knowledge. No-one can be expected to know about all of the research that has been conducted outside their domain of expertise, and those who do know about it cannot be expected to drop their own research to keep an eye on your field, on the off-chance that they will be able to help. Secondly, and more importantly, it is very hard to map knowledge or principles from one domain to another without some common theoretical framework. Film makers express their craft skills in terms of film-making situations that do not occur in computer graphics. Psychologists who are researching vision, or emotion, or spatial navigation, are doing so with their own theoretical concepts, and it is no easier for them to map these onto a computer graphics problem than it is for a computer graphics researcher to understand the psychologists. The theories within a domain are often too detailed and require too much specific input to be applicable to problems outside their native empirical paradigms. In fact, this is as much a problem within psychology as it is between psychology and other disciplines. Human behaviour has been partitioned into so many areas, at so many levels of analysis, that the mutual ignorance between researchers of vision, memory and emotion is astounding.

Fortunately there is an ongoing effort to develop integrative approaches within psychology that enable different aspects of behaviour to be linked at a less detailed level. Because they are not tied to any particular domain, such approaches also provide a way to communicate psychological research to non-psychologists, particularly those working in applied domains. One such technique that is becoming known within computer graphics is the Interacting Cognitive Subsystems model (ICS) that has been developed by Barnard and his colleagues. ICS deals with cognition as a flow of information between nine different levels of mental representation, each with their own memory, or "image record". One level deals with "vision" at a low level, where sensory attributes such as hue, brightness and motion are represented; another with the "objects" that can be perceived within visually based scenes; another with "propositions", semantic facts about objects and their relationships; and a fourth with "implications", the real meanings that can be inferred from sets of propositions. Barnard's approach is not limited to this linear or bottom-up process of recognition and comprehension, though. Currently active Implications feed back to influence the formation of Propositional representations, and these feed back to influence the formation of Object representations.

The object level can be thought of as the "mind's eye", where our awareness of a visual scene resides, but ICS includes a second route by which visual information can affect cognition. The visual level of representation is used to produce implicational meaning directly, in addition to the interpretative object and propositional route. A flashing red light, to take an extreme example, has an implicational meaning that is directly inferred from the sensory level. This meaning is, paradoxically, available to influence in a top-down manner the bottom-up structural interpretation. The same is true for sensory attributes that might not even be represented at the object level, such as aspects of facial expression, or of co-variation in movement of scene elements. The impact of such features of a display can occur despite our lack of awareness of their presence, and their consequent unreportability. As such, it is clearly dangerous to rely upon introspection or self-report assessments of display adequacy.

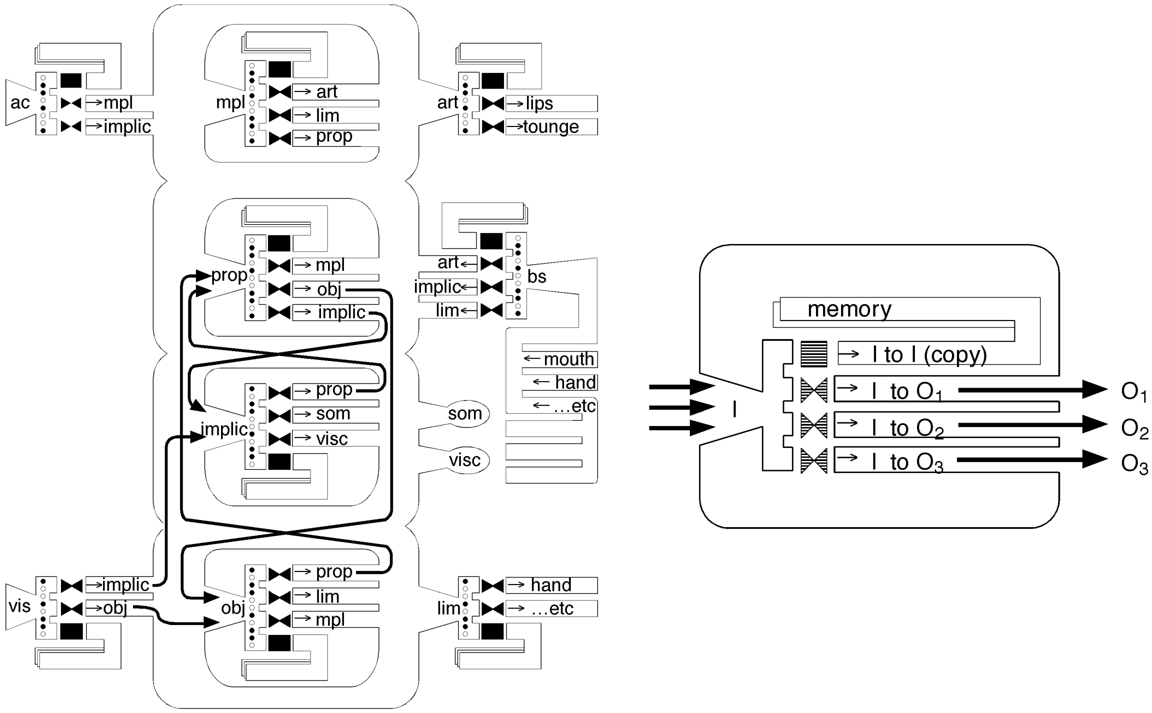

Figure 1: *The overall ICS architecture, with the two routes for visual perception indicated (left), and the generic subsystem architecture (right).*

The overall architecture is illustrated on the left of Figure 1, with the information flows for these two routes for visual perception shown by the arrows linking the visual, object, propositional and implicational subsystems. A similar and parallel process involving "acoustic" and "morphonolexical" levels of representation allow sound to be dealt with. The nine subsystems share the same generic architecture (shown on the right of Figure 1), copying incoming representations to the subsystem's own memory and transforming them into up to three different output forms. The consequence is that ICS allows "top-down" influences on perception to be modelled, and these include the effects of task context, message content, and emotional state, through the representations at the propositional and implicational levels.

This approach to cognition allows a new set of perceptual principles to be added to the low-level constraints upon visibility of displays and the mysterious Gestalt Laws of perception. These new principles govern the requirement of congruency between the arrangement of scene elements and the viewer's expectations about the scene; about changes within the scene and thematic transitions within the viewer's comprehension of the "narrative"; and about latent aspects of the interaction that can influence the viewer's interpretation of the scene. The traffic is not all one way, of course. By providing a framework for the modelling of cognition in complex tasks, ICS may enable psychologists to develop empirical paradigms that do not rely on highly reduced stimuli. The powerful graphics workstations in our laboratories that currently display red and green dots for hours on end may also be used to display photorealistic or rendered images, without the psychologists muttering about irrelevant complexity.

At the heart of ICS is its assertion that the meaning of an image can have as important a contribution to its perception as its physical structure. This is perhaps the holy grail of graphical rendering: to convey meaning as economically and accurately as is possible. Economy resides on both sides of the interaction: in terms of processing resources and hardware constraints on the system side, and in terms of attention, cognitive effort, and time on the user side. The solution will require an understanding of meaning, of the representation of meaning, and of the perception of meaning. The research path that is opening up requires the computer graphics community and cognitive psychologists to work together in a truly meaningful way.

Web page and email of the author

Web:
`http://www.shef.ac.uk/psychology/staff/may.shtml`
Email: `jon.may@shef.ac.uk`