Behavior Research Methods, Instruments, & Computers
1986, 18 (6), 487-492

SOFTWARE TECHNIQUES

The use of computer graphics animation in motion perception research

DENNIS R. PROFFITT
University of Virginia, Charlottesville, Virginia

and

MARY K. KAISER
NASA Ames Research Center, Moffett Field, California

The advantages and limitations of using computer-animated stimuli in studying motion perception are discussed. Most current programs of motion perception research could not be pursued without the use of computer graphics animation. Computer-generated displays afford latitudes of freedom and control that are almost impossible to attain through conventional methods. There are, however, limitations to this presentational medium. At present, computer-generated displays present simplified approximations of the dynamics in natural events. We know very little about how the differences between natural events and computer simulations influence perceptual processing. In practice, we tend to assume that the differences are irrelevant to the questions under study and that findings with computer-generated stimuli will generalize to natural events.

This paper is divided into two parts. In the first part, we discuss some of the advantages of employing computer-animated graphics in motion perception research. In the second part, we discuss some of the limitations inherent to this presentational medium. Although many research programs could not possibly be pursued today without computer animation, we suggest that existing graphics displays never model perfectly the environmental dynamics that they are intended to simulate. Little is known about how these differences may influence perceptual processing.

ADVANTAGES OF COMPUTER-GENERATED STIMULI

Much of the progress in motion perception research was made possible by the development of computer graphics animation technologies. Johansson (1950), in studies employing dynamic point-light stimuli displayed on an oscilloscope, not only influenced the conceptual understanding of the role that motion plays in organizing the visual world, but also established the directions of future research methodology.

This research was supported by Grant NCA2-87 from NASA, Grant HD-16195 from NICHD, and Grant INF-85-014 from the Virginia Center for Innovative Technology. Stephen Ellis, Scott Fisher, David Gilden, Jeffrey Lande, and Susan Whelan provided valuable criticisms on an earlier version of this paper. Reprint requests should be addressed to Dennis Proffitt, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, VA 22903-2477

Much of the work in motion perception has focused on identifying minimal conditions for perceiving various environmental properties. For example, researchers have investigated the perception of form (three-dimensional structure) from motion information. Probably the most dramatic demonstrations that motion is a minimal condition for perceiving form are the kinetic depth effect (Wallach & O'Connell, 1953/1976) and point-light walker displays (Johansson, 1973). In the original kinetic depth effect demonstrations, shadows of unfamiliar wire forms were projected onto a screen. Without motion, these shadows appeared as two-dimensional configurations of lines; however, when the wire forms were rotated, their three-dimensional structures were immediately evident. Johansson's point-light walkers were made by attaching small lights to the joints of people and then filming the people as they walked in the dark. As with the forms from the kinetic depth effect study, static frames from these displays appeared as meaningless two-dimensional arrays of dots; however, a very brief viewing of a moving sequence allowed the observer to identify the projection as a locomoting person.

More recently, researchers have increasingly turned to the use of computer-generated dynamic displays. Braunstein (1976) developed a computer-based methodology for creating complex kinetic depth effect displays consisting of point-lights. Cutting (1978) created a general program for generating point-light walkers. Bertenthal, Proffitt, and Keller (1985) wrote a very general animation program for the Apple microcomputer, interfaced to a Texas

Instruments TMS 9918A video display processor; the program allows a user to create point-light projections (limit is 32 points) of rigid or jointed objects capable of all six degrees of freedom movement. (All of the demonstrations in Johansson's, 1950, book can be easily created with this program.)

There is now a long list of research topics in which dynamic computer-generated displays are used. This list includes studies on perceiving: ego motion, texture segregation, form, form change, object displacement, and dynamics (the recovery of mass and force information from kinematics). It has also been found that infants as young as 3 months of age will attend with interest to computer-generated point-light displays and can extract some structure from them (Bertenthal, Proffitt, & Cutting, 1984). A recent issue of *Perception* (Ramachandran, 1985), devoted to motion perception, contained a wide variety of research reports employing computer animation. This issue even provided an Apple disk that allowed the reader to view many of the dynamic stimuli that were discussed.

Computer animation has numerous advantages over other techniques for creating minimal motion displays. Programmed displays are far more flexible and easier to create than are physical mechanisms constructed to produce desired motions. Computer displays can be easily programmed to display violations in natural dynamics (events in which the laws of physics are violated). This capability proves to be very useful in assessing a person's visual sensitivities to natural dynamics and is, of course, extremely difficult to achieve using real objects. Finally, computer-generated stimuli provide the researcher with exact knowledge of the display's parameters. An example of the importance of such tractability can be seen in research on perceiving point-light walkers. Proffitt, Bertenthal, and Roberts (1984) created a computer display that contained all of the information previously thought to be effective in the perception of Johansson's (1973) naturally produced point-light walker. After 1½ min of viewing, only about a third of their subjects recognized that this display could be seen as a projection of a person walking. We still do not know all of the parameters of information that people use when extracting the human form from Johansson's original, naturally produced point-light walker displays. Computer simulations can, in cases such as this, serve as empirical tests of processing models.

LIMITATIONS OF COMPUTER-GENERATED STIMULI

Dual Awareness

As with viewing static pictures, viewing computer-animated displays gives the observer a dual awareness: (1) an awareness of a transforming two-dimensional pattern appearing on the terminal screen, and (2) an awareness of the three-dimensional event that is being simulated. Gibson (1979) argued that this dual awareness was one of the aspects of picture perception that made it difficult to generalize from research employing pictures to the perception of real objects and events. One of the important environmental properties that is absent in pictures is, of course, motion; however, the substitution of motion for pictorial cues does not necessarily make dynamic computer displays more ecologically valid.

Whenever people look at a computer-animated display, they are presented with conflicting information about depth relationships. All of the primary depth cues specify that the transforming projections are two-dimensional. Moreover, unless the display's motions are yoked to the observer's head movements (Rogers & Graham, 1979), the absence of motion parallax will further define two-dimensional aspects of the display. At odds with this information are the display's dynamics specifying three-dimensional structures.

What is the influence of this dual awareness on the perceptual processing of computer-animated displays? Our research suggests that a person's ability to extract dynamic information from motion displays is related to the degree of naturalness found in the simulation (Kaiser, Proffitt, & Anderson, 1985). We suspect that too great a reliance on dynamic computer displays may result in an underestimation of people's sensitivities to motion-specified information.

Scaling Perspective, Depth, and Size

A perceived object in a computer-animated display does not appear to be located on a monitor's screen; it appears to be somewhere behind the screen. This indeterminacy of absolute depth (observer-relative depth) creates a set of difficult problems in programming a natural-appearing simulation.

Suppose that the researcher wants to generate a perspective projection of a rigid object undergoing rotation. To compute the perspective transformation, the researcher must specify, as one of the parameters, the distance between the geometric eyepoint and the simulated object. (Hagen, 1980, is an excellent source of articles on the geometry of perspective projection.) Should the viewer assume that this distance is equal to the distance between the eye and the screen, the viewer will be in for quite a surprise, since the simulated object will appear to deform drastically as it rotates.

A rigid object will appear to deform as it rotates unless either the perceived viewing distance is specified accurately or the perspective transformation is not salient. Determining appropriate viewing distances, thus, becomes an empirical, rather than a purely geometrical problem. Moreover, perceived viewing distance may vary across individuals and may not remain stable over time within an individual. With regard to the salience of perspective transformations, many researchers present displays in parallel projection and obtain few reports of nonrigid motions.

If perspective information is not given, or if it is not sufficiently salient, then depth-order ambiguities arise unless the animation program is capable of hidden surface removal. (*Depth order* refers to whether elements are in front of or behind other elements in the display.) Depth-

order ambiguities, in turn, can affect the motions that are seen. In kinetic depth effect displays, for example, the perceived direction of motion will spontaneously reverse unless perspective information or occlusion is provided (Braunstein, Andersen, & Riefer, 1982). For point-light displays, depth order can be specified by causing some points to disappear and other points to remain visible whenever they pass through a particular location. In such cases, the points that disappear are seen as being behind those that remain visible, and an invisible intermediate occluding surface is perceptually specified (Proffitt et al., 1984).

Since the perceived distance from observer to simulated object cannot be determined geometrically, neither can the object's absolute size. If apparent distance and size are different from the values intended by the programmer, then the display's dynamics may also appear inappropriate or out of scale. Consider, for example, a simulation of a falling object. If the object's perceived size and distance are different from the assumed parameters of the program, then the object will appear to fall either too fast or too slow. We have successfully specified size and distance in such a display by placing a simulated familiar object, such as a person, in the same depth plane as the falling object.

Size of Visual Field

Unlike natural scenes, computer displays subtend a limited area of an observer's field of view. The better computer graphics systems in today's market employ 1024×1024 pixel RGB monitors. If an observer views such a monitor from a distance of about 80 cm, the display has fairly good resolution (approximately 40 pixels per degree), but subtends less than 26° of visual angle. The observer has the clear sensation of viewing a window on a scene rather than the scene itself.

Attempts to enlarge the visual display entail significant compromises in terms of cost, resolution, color capability, or update rate. Given present technology, it is extremely costly to employ larger CRT displays; it is unlikely that CRT displays will ever become cost-effective for wide field-of-view displays. We are currently using two 45-in. rear-projection screens in studies of peripheral motion information processing; however, all currently available rear- or front-projection systems have several drawbacks, notably lack of resolution, contrast, and brightness. In addition, projection systems tend to be cumbersome and difficult to adapt to all experimental situations.

Recently, effort has been directed toward the development of head-mounted display systems. There are several advantages of such systems: they are capable of presenting binocular displays; a wide variety of apparent display sizes can be produced on a single, small screen (although fairly sophisticated optics are needed to produce appropriate geometries); and their displays can be yoked to the observer's head motions or other monitored activity. In fact, by employing head-tracking technologies, such a head-mounted display can create a 360° stereoscopic visual environment (Fisher, 1985). The disadvantages of such a system include the cost and awkwardness of high-resolution color displays and the relatively high hardware and software costs. In addition, head-mounted displays may prove inappropriate for some subject populations (e.g., infants) or experimental tasks.

At some point, the size of the visual field becomes constrained not only by display technology, but also by limitations in the computational hardware; the system simply cannot compute values for all of the pixels in the scene, given the required update rate. When this happens, several remedies are possible. First, the system can convert from real-time displays to storage of display sequences on some random-access storage device (e.g., a laser disc). Alternatively, one can create variable resolution displays that compute high-resolution displays only for the area surrounding the observer's current fixation point. The latter solution is fairly elaborate, but is being pursued in contexts requiring high-resolution real-time graphic-animation displays (e.g., flight simulators). A third solution that reduces the computational demand is to reduce the update rate slightly, yet retain the real-time nature of the display. This solution creates a new set of problems that are the next topic of consideration: The quality of motion in computer-generated displays.

Motion Quality

The quality of simulated motions depends primarily upon two things: the particular motion algorithms employed and whether or not the system possesses sufficient computational power to execute the algorithms in time for the next screen update. We discuss the issue of motion algorithm adequacy below and again under the topic of Simulation Dynamics. First, we consider the computational power issue.

As mentioned above, update rate can be increased if scene resolution (or complexity) is reduced. Thus, most researchers are left with a direct trade-off between update rate and computational complexity. If there is no need for real-time animation, the researcher gains a huge advantage. Complex events can be generated one frame at a time, with the ensemble of frames later shown at the desired update rate. Braunstein (1976) used this method to generate many of his depth-from-motion stimuli. Static images from a motion sequence were created on a computer terminal and recorded frame-by-frame on 16-mm movie film. Many expensive systems on the market today store the image on videotape or laser disc (the latter having the advantage of rapid random-access capabilities). However, all of these techniques require that the researcher have a well-defined, limited number of sequences to be computed. Furthermore, all but the laser disc make it extremely difficult to alter the order of sequence presentation, as one might in any response-dependent experimental design (e.g., staircase methodologies).

If one wishes to generate stimuli in real time, the trade-off between complexity and update rate remains. The third factor in this trade-off is cost: the more expensive com-

puter systems have greater computational capabilities and software enhancements. Fortunately, the power-cost trade-off continues to become more favorable to the consumer. Today's microcomputers are more powerful graphics systems than are minicomputers of a decade ago. In particular, processors with the Motorola 68000 chip (and their 68010 and 68020 successors) provide impressive performance. In addition, some microcomputers (e.g., the Commodore Amiga) have dedicated graphics-processing chips. Unfortunately, performance always seems to lag expectations, and researchers are likely to lag state of the art performance due to economic and procurement constraints.

Researchers disagree on what update rate is acceptable for dynamic stimuli, and, of course, the rate depends on characteristics of the event. The update rate at which time-sampled motion becomes indistinguishable from smooth motion depends both on the velocity and spatial frequency content of the image and on observer factors (Watson, Ahumada, & Farrell, 1983). Ultimately, update rates will be limited by display hardware constraints. Researchers using raster display systems are limited by the refresh rate of the monitor. This rate is generally 60 Hz for systems in North America and 50 Hz for systems marketed overseas (although some manufacturers use nonstandard rates; e.g., Sun Microsystems has 66-Hz monitors).

When depicted objects move at high velocities, computer-generated displays possess a notable artifact, most noticeable when comparing a computer-generated "frame" with a frame from a movie film of the event. In the film frame, quickly moving objects will be blurred. In the computer-generated frame, however, all objects are clearly defined, which leads to a strobing (or temporal aliasing) that appears quite unnatural. Effective motion-blur algorithms have been developed to solve this problem (e.g., Max & Lerner, 1985). In brief, these algorithms sample object location during each frame interval and distribute object density accordingly. Of course, the implementation of such algorithms increases the computational complexity of the sequence, and, thus far, cannot be performed in real time.

Object Realism

The problems of object realism are similar to those of motion quality: there are both computational power constraints and adequacy of algorithm limitations. The two aspects of object realism that we discuss are surface properties and shading. We also limit our discussion to computer-generated objects, excluding those impressive computer displays that are simply digitized photographs.

Real objects in the environment possess complex visual surface qualities. Texture and reflectance properties are difficult to model realistically for several reasons. First, reflectance properties have been studied for only a limited class of materials, with adequate mathematical description developed for fewer still. Second, few natural objects have smooth surfaces with constant reflective properties; most surfaces in the environment are anisotropic, meaning that reflection is a function of orientation. For example, the threads in a weave of cloth will scatter light more narrowly in the direction of the thread than perpendicularly. Although some anisotropic models are being developed (e.g., Kajiya, 1985), such surfaces are still quite difficult to simulate. Thus, we find a plethora of smooth, regular objects in computer graphics demonstrations. Finally, texture presents a challenge to efficient modeling. One wants to retain the stochastic nature of the texture while utilizing a consistent, efficient algorithm. Fractal geometry has been employed to this end (Mandelbrot, 1983) and has proved effective for a wide class of natural objects (e.g., mountains, trees, clouds). However, fractal models are not appropriate for all object classes, and most of the current fractal algorithms are computationally expensive.

Realistic shading is difficult to achieve for similar reasons. In fact, the two issues are related: in order to specify ambient light conditions for shading, reflectance properties must be known (Nishita & Nakamae, 1985). Consider, for example, the ray-tracing method of scene generation. In this method, a number of rays originate from each pixel and are allowed to propagate through the scene, bouncing from surface to surface in accordance with each object's reflectivity (Cook, Porter, & Carpenter, 1984). Thus, ray-tracing algorithms must adequately model interreflection as well as primary lighting sources in order to achieve realistic-looking continuous tone representations.

A final complexity is introduced when dynamic events rather than static scenes are generated. Since realistic surface and shadowing algorithms are computationally complex, it becomes extremely expensive to generate 20 to 60 frames for each second of the event. Ray-tracing techniques, for example, are beyond the capabilities of all but the most complex computer graphics systems, and a single ray-traced frame can require hours of CPU time to generate. Nonetheless, these complex algorithms are often employed, because adequate interpolation algorithms have not yet been developed.

The Adequacy of Simulated Dynamics

The final issue we consider is the adequacy of simulated dynamics in computer-generated animation. Presently, only the most simple dynamic events (e.g., colliding balls, rotating objects) are generated from mathematical motion models. Even these models often make simplifying assumptions, such as the absence of friction or the use of particle, as opposed to solid body, mechanics.

Consider, for example, biomechanical motions, such as those presented in point-light walker displays. Fully adequate mathematical models of biomechanical motion have yet to be developed, although progress is being made (Girard & Maciejewski, 1985). At present, the most impressive examples of computer-animated biomechanical forms were created by techniques borrowed from the traditional animation arts. For example, rotoscoping, an animation technique developed at the Walt Disney Studio, has been employed to capture the dynamics of human motion. In rotoscoping, one first films a person per-

forming the desired actions, then each film frame is used as a template to specify body and limb coordinate locations for each animation frame. Although such a technique produces impressive results for the cartoonist's character (e.g., Snow White) or for the computer animator's special effects (e.g., Abel Graphic's metallic woman), it affords few advantages to the perceptual psychologist interested in the specification of and the observers' sensitivity to biomechanical kinematics.

An approach midway between rotoscope techniques and true mathematical motion models is the keyframe technique. With this technique, critical points of the event sequence are sampled, and intermediate coordinate positions are calculated based on assumed motion properties and constraints (Steketee & Badler, 1985). As yet, keyframe techniques cannot provide motion parameters that are sufficiently precise to be used in perceptual research.

As indicated above, the algorithms used by perceptual psychologists in their dynamic simulations have been very reductionistic, even for relatively simple physical events (e.g., two objects colliding). There are good reasons for employing extreme simplifying assumptions: Precise motion modeling of complex physical systems is a huge computational problem. The difficulty of developing adequate models of such systems may be better understood by examining the problems confronted by other disciplines, such as computational fluid dynamics, which have attempted similar modeling.

Computational fluid dynamics (CFD) attempts to numerically model fluid and gas flow. This work has been most strongly pursued in studies of aerodynamics, but also has applications in streamlining, weather prediction, wake dynamics, and wind loading studies (Kutler, 1983). At institutions such as the NASA Ames Research Center, the ultimate goal of CFD is to allow aerospace engineers to optimize designs solely through computational models. At present, CFD modeling is complemented with wind tunnel and flight test evaluations. Convergence of CFD models with these laboratory and field results provides a basis for evaluating model adequacy. The advancement of CFD toward adequate models is seen as being paced by several constraints, notably the development of appropriate turbulence models, the ability to model dynamics in three dimensions instead of reducing the problem to a two-dimensional representation, and the development of more powerful computer architectures.

Despite the computational complexity confronting CFD modeling (almost all three-dimensional models exceed the computational power of the Cray X-MP computer system), there is the advantage that the criteria for model adequacy are well defined: a CFD model's performance should be functionally equivalent to that found for (1) a physical aircraft in the wind tunnel and in the field, and (2) explicit analytic solutions where they exist. As might be expected, however, there is a notable lack of consensus among experts as to when equivalence is reached.

The transportation of the CFD criteria for simulation adequacy to perceptual research would require us to demonstrate that our simulations and the corresponding natural events are perceptually equivalent. We shudder at the thought of perceptual psychologists' attempting to agree on criteria concerning when such perceptual equivalence is achieved. Ideally, we should "flight test" computer-generated stimuli to determine whether their dynamics are discriminable from those seen in natural events. In practice, we tend to assume that the differences are irrelevant to the questions under study and that findings with computer-generated stimuli will generalize to natural events.

CONCLUSION

Most current programs of motion perception research could not be pursued without the use of computer graphics animation. Computer-generated displays afford latitudes of freedom and control that are almost impossible to attain through conventional methods. We think it is important, however, to be ever mindful that computer simulations rarely, if ever, achieve a level of verisimilitude capable of causing an observer to confuse the simulation with reality. All of the limitations discussed above place constraints on the apparent realism of computer-animated displays.

When viewing a computer-generated display, the viewer always experiences a dual awareness: an awareness of both a two-dimensional contrived pattern and a projected three-dimensional event. Sensitivity to the dynamics manifest in the latter is almost surely influenced by the awareness of the former. We recommend caution in making unqualified generalizations about human sensitivities to natural events from studies on perceiving computer-animated displays. Convergent investigations employing natural objects are always desirable, although, in practice, such studies are often extremely difficult to conduct.

REFERENCES

BERTENTHAL, B. I., PROFFITT, D. R., & CUTTING, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of Experimental Child Psychology*, **37**, 213-230.

BERTENTHAL, B. I., PROFFITT, D. R., & KELLER, S. E. (1985). 3-D graphics animation program for the Apple II. *Behavior Research Methods, Instruments, & Computers*, **17**, 195-202.

BRAUNSTEIN, M. L. (1976). *Depth perception through motion*. New York: Academic Press.

BRAUNSTEIN, M. L., ANDERSEN, G. L., & RIEFER, D. M. (1982). The use of occlusion to resolve ambiguity in parallel projections. *Perception & Psychophysics*, **31**, 261-267.

COOK, R. L., PORTER, T., & CARPENTER, L. (1984). Distributed ray tracing. *Computer Graphics*, **18**, 137-145.

CUTTING, J. E. (1978). A program to generate synthetic walkers as dynamic point-light displays. *Behavior Research Methods & Instrumentation*, **10**, 91-94.

FISHER, S. S. (1985, December). *Virtual interface environment*. Paper presented at the Space Station Human Factors Research Review, NASA Ames Research Center, Moffett Field, CA.

GIBSON, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

GIRARD, M., & MARCIEJEWSKI, A. A. (1985). Computational modeling for the computer animation of legged figures. *Computer Graphics*, **19**, 263-270.

HAGEN, M. A. (Ed.) (1980). *The perception of pictures* (Vol. 1). New York: Academic Press.

JOHANSSON, G. (1950). *Configuration in event perception*. Uppsala, Sweden: Almqvist & Wiksell.

JOHANSSON, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, **14**, 201-211.

KAISER, M. K., PROFFITT, D. R., & ANDERSON, K. (1985). Judgments of natural and anomalous trajectories in the presence and absence of motion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 795-803.

KAJIYA, J. T. (1985). Anisotropic reflection models. *Computer Graphics*, **19**, 15-24.

KUTLER, P. (1983). *A perspective of theoretical and applied computational fluid dynamics* (AIAA Paper No. 83-0037). Reno, NV: AIAA.

MANDELBROT, B. (1983). *The fractal geometry of nature*. San Francisco: W. H. Freeman.

MAX, N. L., & LERNER, D. M. (1985). A two-and-a-half-D motion-blur algorithm. *Computer Graphics*, **19**, 85-94.

NISHITA, T., & NAKAMAE, E. (1985). Continuous tone representation of three-dimensional objects: Taking account of shadows and interreflection. *Computer Graphics*, **19**, 23-30.

PROFFITT, D. R., BERTENTHAL, B. I., & ROBERTS, R. J., JR. (1984). The role of occlusion in reducing multistability in moving point-light displays. *Perception & Psychophysics*, **36**, 315-323.

RAMACHANDRAN, V. S. (Ed.). (1985). Human motion perception [Special Issue]. *Perception*, **14**(2).

ROGERS, B., & GRAHAM, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, **8**, 125-134.

STEKETEE, S. N., & BADLER, N. I. (1985). Parametric keyframe interpolation incorporating kinetic adjustment and phrasing control. *Computer Graphics*, **19**, 255-262.

WALLACH, H., & O'CONNELL, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, **45**, 205-217. (Also in H. Wallach, (1976). *On perception*. New York: Quadrangle.)

WATSON, A. B., AHUMADA, A., & FARRELL, J. E. (1983). *The window of visibility: A psychophysical theory of fidelity in time-sampled visual motion displays* (NASA Technical Paper No. 2211). Moffett Field, CA: NASA.