PSYCHOPHYSICAL APPROACH TO VISUAL DISPLAY ACCEPTANCE

LCDR Michael G. Lilienthal
Naval Air Systems Command (PMA205-42)
Washington, DC 20362-1205

ABSTRACT

There has been a rapid development in photo-based generated imagery for flight simulators without an accompanying development of knowledge, test, and acceptance criteria. Trainer design engineers have developed visual displays based on years of previous experience rather than upon aircrew visual psychophysical requirements. The criterion for the merit of a display has relied heavily on the acceptance of the visual system by a few experienced aviators and program managers. Visual scientists and psychophysicists have played a minor role in deciding how and what visual information must be displayed in a simulator to ensure that the scene provides the proper cues to accomplish the training tasks.

This paper presents a review of several Navy performance specifications for visual flight simulators and proposes a psychophysical scaling test and acceptance approach for visual cue requirements. The move to photo-based systems with increased texturing fulfills part of the requirement for visual scene cues. However, the visual systems must not only generate the proper number of leaves on trees, but they must give the aircrew sufficient dynamic visual cues. The aircrew should receive the same psychophysical cues that are needed in actual aircraft flights. These include, for example, the same depth cues, vection, velocity cues, perceptual experience, and closure cues as experienced in flight operations. Highly reliable direct psychophysical measurement techniques are proposed as part of the test and acceptance protocol for such visual flight simulators.

INTRODUCTION

In recent years, flight simulators have become such a major factor in pilot training that the current focus is not so much on whether flight simulators will be used as training tools but to what extent they will be incorporated into the training process. The success of flight simulators for training has been demonstrated by both civil and military aviation increasing reliance on these systems (9). The overwhelming savings to the Department of Defense from the use of flight simulators (e.g., 6,7,8) will increase as training device technology continues to improve. This rapidly expanding technology mandates that specification language and test/acceptance protocols keep pace with technology to ensure that emerging simulator designs answer the training requirements of the customer.

There has been a rapid development in computer generated imagery for Navy/Marine Corps flight simulators without an accompanying development of quantified engineering specifications based upon aircrew perceptual systems/processes. Simulator design engineers have used as their criterion for the "goodness" of the display, the acceptance of the system by experienced aviators. Acceptance by the end user is a vital goal until something is considered unsuitable. When this occurs, if we lack understanding of the psychophysics of the problem, we are at a loss for offering solutions. Systems specifications do not contain enough detail about the visual information required for the pilot to receive optimal visual perceptual cues for a flight simulator in a cost-effective manner.

Because visual scientists and psychophysicists have played a minor role in deciding how and what visual information the simulator must display, design engineers have developed visual systems based on common sense and previous experience. There is still much "art" in the science of simulator visual display technology. System specifications do not contain enough detail about the visual information required by aircrew. For example, the depth cues in most helicopter simulators are not sufficient, especially at or below the simulated height of 50 feet. However, because the engineers are uncertain about the visual perceptual requirements of the aircrew, they have relied upon maximizing scene detail and display output (i.e. realism and fidelity) as the criteria.

It is necessary to develop specifications that can be quantified and verified to meet training requirements. Engineers have designed measurement equipment to test the specifications for "objective" measures such as contrast and surface resolution; however, "subjective" measures such as depth cues, vection, optical flow rates, and velocity cues necessary for the pilots to carry out their visually-oriented tasks have been ignored. A means is needed to easily determine, with the pilot-in-the-loop, the magnitude and changes in magnitude of visual parameters necessary for proper fidelity and realism required for training.

SPECIFICATIONS

In writing of statements of work and specifications, Navy engineers have focused on the criteria with which they have the most familiarity and can measure without having to worry about variation in pilot experience, perception, and performance. New visual systems are better because they have a higher refresh rate, fewer arc-minutes per optical line pair, and more polygons per update rate. The criteria have been more computer power, more field of view, more fidelity, and more realism. This, by default, has been the engineer's only choice. The communication between the aircrew and the government engineer about visual criteria seldom revolves around the perceptual requirements. It often devolves to such data base requirements as the number of landing fields, number of friendly and hostile ships, and what the instructor display page will contain.

Over 40 years ago, Thorndike ([15]) wrote "Certainly the most fundamental and probably also the most difficult problem in the Aviation Psychology Program was that of obtaining satisfactory criterion measures against which to validate tests and evaluate variations of training methods" (pg 29). In the comparison of training simulator and actual combat operations of a bombardier, Thorndike stated that in combat: (a) target identification was much more difficult, (b) there were more demands on bombardier flexibility, (c) there were more demands on resourcefulness, (d) there were more distractions, and (e) there were important differences in the relationships among the crew members. Flannagan ([4]) added that with these and other differences between the simulated and the actual bombing mission that correlation of performance under the two conditions is generally fruitless. We have come a long way since that time technologically but still the articulation of the requirements remains a difficult task.

Quantifiable perceptual criteria are needed to bridge the gap between the aircrew and the specifications. This will help avoid discrepancy reports during test and acceptance such as "Insufficient visual cues for height above ground and longitudinal speed during transition to hover." which do not provide diagnostic feedback to the contractor.

Table 1 has examples of specifications for various visual scene requirements for Navy flight simulators that attempt to include some perceptual requirements in the contract. What the engineer who wrote the specification meant by "sufficient," "adequate," and "ensuring a maximum of useful visual cueing information" in this table is not spelled out for the contractor. In contrast, Table 2 has examples of visual generation requirements for flight simulators where the government included operational definitions for each term. That is, a requirement is defined by how it will be tested. This makes it much easier for a contractor to design, develop, produce, and test a visual system to these specifications than to those included in Table 1.

However, neither contractors nor the government have validated the differences in training effectiveness delivered for different performance capabilities of a visual system. That is, engineering specifications are not related to a training criterion but rather a system output criterion. The surface resolution and contrast capability of a simulator is only a means to an end. The "true" criteria for a successful visual system is whether or not it enables the aircrew to train to perform perceptual, mediational, communication, and motor processes ([1]) necessary for aircraft mission accomplishment.

Along with buying theodolites, photometers, and strip charts to test a flight simulator, psychophysical measurement techniques should be incorporated into the government acceptance protocol. That is, in addition to the physical measurement of hardware that the engineers use, a psychological measurement of the magnitude or intensity of sensation is required. Psychophysical scaling and measurement techniques are the first step towards improved specifications and test protocols which will better translate fleet requirements into a training system.

PSYCHOPHYSICS

Psychophysics is the scientific study of the relation between perceptions in the psychological domain and stimuli in the physical domain. A physical scale such as dynes/cm squared, frequency, footlamberts, wavelength, and temperature is measured through techniques developed in physics. The corresponding sensory scale of loudness, pitch, brightness, hue, and warmth is measured through techniques of psychophysics. Proponents of these techniques assume that people can assign numbers or adjust a continuum (e.g. the loudness of a tone) in a "ratio-like manner" in order to reflect the sensory continuum.

For example, the direct scaling technique called magnitude estimation can enable the development of a psychophysical function which plots the magnitude of the sensory attribute against corresponding physical values of the stimulus. Table 3 is a typical set of instructions for the magnitude estimation of loudness. Stevens ([10,12,13]) has used various direct scaling techniques for scaling different continua such as loudness, brightness, finger span, vocal effort, and handgrip force. Ratio scaling techniques have been subjected to intensive validation testing ([10,11]). The scales constructed by magnitude estimation, magnitude production, fractionation, and ratio estimation have consistently been found to be linearly related to each other ([3]). The application of validation procedures such as cross modality matching, comparing fractionation scales on different fractions, and comparing results on scaling two overlapping sets of stimuli support the validity of scales based on these techniques ([2,14]). The validity and consistency of these methods for a broad spectrum of continua have importance for the acceptance of simulators.

TABLE 1. SIMULATOR VISUAL PERCEPTUAL SPECIFICATIONS

1. The visual system shall produce successive images at a rate sufficient to give the impression of smooth motion to the observer.

2. The CGI visual system shall provide full color real-time, flicker free, visual displays of the visual scene.

3. Night vision systems shall generate and display true perspective images of the three-dimensional visual scene.

4. The scenes shall be adequate for training pilots and copilots in the FRS.

5. The system shall have velocity and altitude cues to allow flight at 50 - 200 ft.

6. Scene content management shall optimize the training value of the system by ensuring a maximum of useful visual cueing information.

TABLE 2. SIMULATOR VISUAL HARDWARE GENERATION SPECIFICATIONS

1. The resolution for each display zone shall be achieved as follows based on measurements at criterion points throughout the display. The spacing between criterion points in each channel shall be such that nine equally spaced points are provided in each channel with a maximum spacing between criterion points and channel boundaries of five degrees.

2. Surface resolution shall be determined based on a test pattern of alternate, equal width, dark and light bars (50 percent duty cycle). Resolution is the test pattern spacing at which the Modulation Transfer Function is 10 percent measured at the display and including all system elements including image generator, display device and intervening optics. Two sets of ten light bars and nine dark bars each and displayed on a gray background shall be included in the pattern with the sets perpendicular to each other as in the standard USAF resolution chart.

3. Contrast ratio for visual imagery shall be as specified for each type of display and throughout the viewing volume and field-of-view using a checkerboard test pattern (all displays illuminated with the test pattern simultaneously). The pattern must provide at least 16 squares per display channel with 50 percent of the squares at the specified maximum luminance.

4. The total geometric distortion from all causes for each image channel shall not exceed 1.5 percent within a circle whose diameter is 0.6X the display channel diagonal, centered at the channel center as measured from the trainee eyepoints.

   The lack of quantified measures for the fidelity and realism requirements in the specification phase of flight simulators has made it difficult to develop adequate objective test measurement tools. Without standardized pilot-in-the-loop operating the system. Without pilot-in-the-loop measurement, the Navy cannot properly evaluate the simulator in the manner in which it would be used in the field.

TABLE 3. MAGNITUDE ESTIMATION (LOUDNESS)

You will be presented with a series of tones through the headphones that you will be wearing. Your task is to tell me how <u>loud</u> they seem by assigning numbers to them. The first will be the standard loudness, which we will call 10. Your task is to assign numbers proportionally to your subjective impressions. For instance, if a loudness is 5 times as loud as the first, assign a number 5 times as large as the first, that is 50. If it seems 1/10th as loud, assign a number 1/10th as large, that is 1. Use whatever numbers seem appropriate--fractions, decimals, or whole numbers. Try not to worry about being consistent, just make your matches agree with your casual impression of the loudness. In this experiment, there are no right or wrong answers.

Psychophysical scaling techniques have an advantage in that they are capable of being rapidly taught to the subject matter experts who will be part of the acceptance of the trainer. They also produce a sensory ratio scale. That is, the scale of measurement possesses magnitude, equal intervals, and an absolute zero point. The variance associated with magnitude estimation measurement also follows a normal distribution. Parametric statistics can then evaluate the data.

Thus, a psychophysical test protocol can define the requirement in Table 1, specification 5). Aircrew could fly through a prescribed low level route at different altitudes and airspeeds. The aircrew will not be allowed to view the actual air speed indicator but will give magnitude estimations of the aircraft velocity at various points along the route. The aircrew could then be flown in the same protocol in the simulator. The mapping of the magnitude estimations of velocity in the aircraft plotted against those made in the simulator will provide a measure of the degree of velocity perception fidelity for that simulator. The closer the mapping is to a linear function with a slope of 1.0, the closer the simulated visual cues are to the real ones. Any nonlinearity will show that at certain altitudes and airspeeds there is a lack of perceptual fidelity.

There are laboratory data of individuals who made magnitude estimations of the angular velocity of a stimulus moving across a 12 degree field of view display (5). Individual velocity estimation in degrees per second was very accurate with a standard error of approximately 1 degree/second. In this case, the relationship between physical velocity and perception of velocity is linear. That is, individuals can estimate movement accurately and consistently. This gives added confidence for using subjective measurement.

Figure 1A-D provides a series of hypothetical plots of the aircrew perceived velocity magnitude estimations for a simulator flight plotted against those estimates made while flying the same profile in the actual aircraft. Figure 1A indicates that the visual system has a high degree of agreement between velocity cues in both the real and simulated "world." Figure 1B demonstrates that aircrew receive cues that make them overestimate velocity at low simulated airspeeds; cues that are accurate for a small range at medium airspeeds; and cues that make them underestimate velocity at higher simulated airspeeds. Figure 1C shows that after a certain airspeed any increment in simulated velocity results in no perceptual change. That is, the aircrew do not receive cues in the simulator that tells them that they were flying any faster. Figure 1D provides diagnostic information that after a certain simulated airspeed, the perception of velocity disappears.

Psychophysical scaling can thus be more diagnostic than current protocols where an aircrew writes up a discrepancy report telling the contractor that the visual cues are inadequate. It shows in a systematic, reliable, and replicable manner under what conditions the fidelity for a particular visual cue breaks down. It will help enable the mapping of system hardware capabilities to perceptual requirements. We may find that the total geometric distortion specification (Table 2, Specification 4) should become more or less stringent because of how it relates to visual perception requirements. The government can then move towards specifying, testing and paying for visual systems that respond to training requirement perceptual criteria.

REFERENCES

(1) Christensen, J. M., & Mills, R. B. (1967). What does the operator do in complex systems. Human Factors, 9, 329-340.

(2) Ekman, G., & Sjoberg, L. (1965). Scaling. Annual Review of Psychology, 16, 451-474.

(3) Engen, T. (1971). Psychophysics II. Scaling Methods. In Kling, J. W. & Riggs, L. S. (Eds.), Experimental Psychology. (3rd edition). New York: Holt, Rinehart, and Winston.

(4) Flanagan, J. C. (Ed.) (1948). The aviation psychology program in the Army Air Forces. Washington, DC: U.S. Government Printing Office, Army Air Forces Aviation Psychology program Research Report No. 1.

(5) Kennedy, R. S., Yessenow, M. D., Wendt, G. R. (1970). Magnitude estimation of visual velocity. NAMI-1051, Naval Aerospace Medical Institute, Pensacola, Florida

(6) Orlansky, J., & String, J. (1977a). Cost effectiveness of flight simulators for military training: Volume I. Use and effectiveness of flight simulators (IDA Paper P-1275). Alexandria, VA: Institute for Defense Analyses.

(7) Orlansky, J., & String, J. (1977b). Cost effectiveness of flight simulators for military training. Vol. II: Estimating costs of training in simulators and aircraft (IDA Paper P-1275). Alexandria, VA: Institute for Defense Analyses. (NTIS No. AD A052801)

(8) Orlansky, J., String, J., & Chatalier, P. R. (1982). The cost effectiveness of military training. Orlando, FL: Proceedings of the 4th Interservice Industry Training Equipment Conference (Vol. 1). (NTIS No. AD A122155)

(9) Skans, N. S., & Barnes, A. G. (1979). Fifty years of success and failure in flight simulation. Paper presented at the Royal Aeronautical Society Conference, Fifty Years of Flight Simulation, London, England.

(10) Stevens, S. S. (1957). On the psychophysical law. Psychological Review, 64, 153-181.

(11) Stevens, S. S. (1958). Problems and methods of psychophysics. Psychological Bulletin, 55, 177-196.

(12) Stevens, S. S. (1960). On the new psychophysics. Scandinavian Journal of Psychology, 1, 27-35.

(13) Stevens, S. S. (1961). The psychophysics of sensory function. In W. A. Rosenblith (Ed.), Sensory Communication. Boston: MIT Press, 1-33.

FIGURE 1A
HIGH CUING FIDELITY

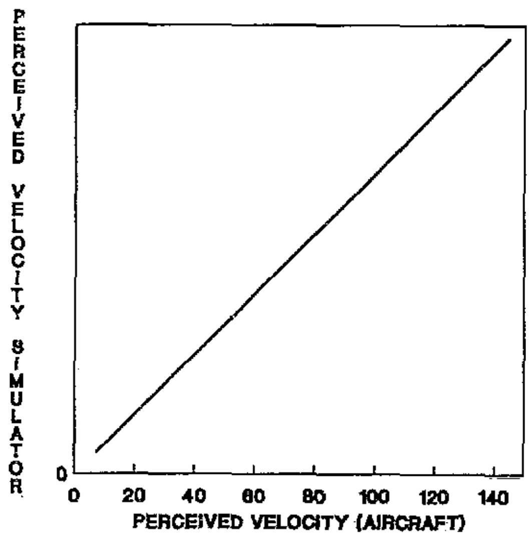
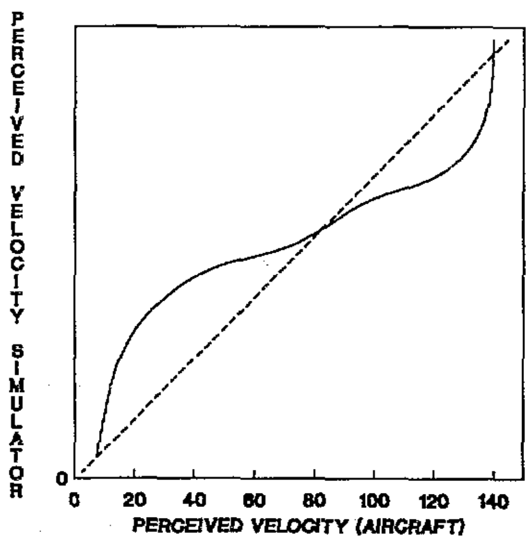


FIGURE 1B
OVER/UNDERESTIMATING CUING



(14) Stevens, S. S. (1969). On prediction of exponents for cross-modality matches. Perception & Psychophysics, 6, 251-256.

(15) Thorndike, R. C. (1947). Research problems and techniques. Washington, DC: U.S. Government Printing Office, Army Air Forces Aviation Psychology Program Research Report No. 3.

ABOUT THE AUTHOR

Michael Lilienthal has a PhD in Experimental Psychology from the University of Notre Dame. He has been an Aerospace Experimental Psychologist for eleven years in the Navy and is a Lieutenant Commander. Before his current tour at the Naval Air Systems Command, he was stationed at the Naval Training Systems Center.

FIGURE 1C
LIMITED VELOCITY CUING

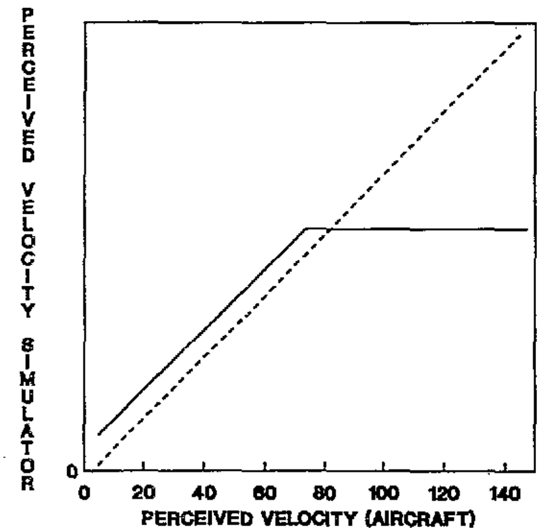


FIGURE 1D
LOSS OF CUING