Human Visual Performance Model for Crewstation Design

James Larimer & Michael Prevost
NASA Ames Research Center
Moffett Field, CA

Aries Arditi & Steven Azueta
The Lighthouse, Inc.
New York, NY

James Bergen & Jeffrey Lubin
David Sarnoff Research Laboratories
Princeton, NJ

ABSTRACT

In a cockpit, the crewstation of an airplane, the ability of the pilot to unambiguously perceive rapidly changing information both internal and external to the crewstation is critical. To assess the impact of crewstation design decisions on the pilot's ability to perceive information, the designer needs a means of evaluating the trade-offs that result from different designs. The Visibility Modelling Tool (VMT) provides the designer with a CAD tool for assessing these trade-offs. It combines the technologies of computer graphics, computational geometry, human performance modelling and equipment modelling into a computer based interactive design tool. Thru a simple interactive interface, a designer can manipulate design parameters such as the geometry of the cockpit, environmental factors such as ambient lighting, pilot parameters such as point of regard and adaptation state, and equipment parameters such as the location of displays, their size and the contrast of displayed symbology. VMT provides an end to end analysis that answers questions such as "Will the pilot be able to read the display?" Performance data can be projected, in the form of 3D contours, into the crewstation graphic model providing the designer with a footprint of the operator's visual capabilities, defining, for example, the regions in which fonts of a particular type, size and contrast can be read without error.

Geometrical data such as the pilot's volume field of view, occlusions caused by facial geometry, helmet margins, and objects in the crewstation can also be projected into the crewstation graphic model with respect to the coordinates of the aviator's eyes and fixation point. The intersections of the projections with objects in the crewstation, delineate the area of coverage, masking, or occlusion associated with the objects.

Objects in the crewstation space can be projected onto models of the operator's retinas. These projections can be used to provide the designer with the retinal coordinates and the visual angles subtended by objects in the crewstation space. Both the right and left eye retinal projections are mapped. The retinal map is yoked to the fixation point and changes as the fixation point is interactively manipulated. Performance contours on the retinas can also be indicated thus aiding the designer in understanding the limitations to visibility imposed by retinotopic processing.

1.0 INTRODUCTION

This paper describes a software simulation system designed to assist in the design and analysis of aircraft cockpits. The intended function of the system is to allow a designer to evaluate the visual performance properties of a proposed design. This function can be used in several ways. For example, a particular design can be analyzed and assigned a figure of merit with respect to some visual task. Then other designs can be evaluated relative to this figure. Alternatively, an analysis can be performed to determine the placement or characteristics of a display device in order to maximize such performance.

The purpose of this system is to determine whether displayed information will be available to the pilot. In order to fulfill this function, the system must integrate three major kinds of information:

1. Information about the physical cockpit environment ("THE WORLD"). This includes geometric descriptions of structures, photometric properties of display devices and other objects, and illumination conditions.

2. Information about the pilot's eye-head system ("THE EYES"). This includes the effects of head position, direction of gaze, locus of accommodation and vergence. This is the information necessary to determine a retinal image given the physical description.

3. Information about the pilot's visual neural processing ("THE BRAIN"). This includes the effects of light adaptation, position in the visual field, spatial, temporal and chromatic factors on the extraction of visual information.

To allow analysis of the availability of visual information, the software system must contain modules embodying each of these three types of information. Before these can be constructed, however, we must have a model of each subsystem. These models must specify the representation of information within each subsystem, and the information that it exchanges with the other systems. The "world" subsystem must represent information about objects positions and properties, and supply to the "eye" subsystem the information necessary to determine the retinal image. The "eye" subsystem must represent information about the state of the eye-head system and supply the "brain" subsystem a description of the retinal image. The "brain" subsystem must represent information about the pilot state and about the visual task and determine whether the information required for the task is accessible in the retinal image. These relationships must be incorporated in a computational module or set of modules in order to produce an actual software tool.

The following sections describe individually, the models of the cockpit environment, eye-head system, and visual neural processing that form the basis of the visibility analysis tool. Following this we describe the embodiment of these models in computational modules, and the integration of these into a software environment.

2.0 THEORETICAL BASIS FOR THE MODEL

2.1 The World/Eyes Subsystem

The World/Eyes subsystem is derived from a method of visual field analysis useful in understanding the effects of certain visual pathologies [1]. It can be viewed as a way of studying specific geometric relationships between objects in the world and images on the retinas. With impaired vision this technique, called volume perimetry, is used to analyze how areas of dysfunctional retina affect perception of the environment. In particular, we are interested in understanding how such areas constrain the visual field of view, and hence visual function. Given a map of functional and dysfunctional retina for each eye, the analysis produces a representation of the volumes of environmental space whose locations are visible to one or both eyes. In normal vision some properties of visual processing are retinatopic. For example, the peripheral retina is highly sensitive to image motion, whereas the central retinal field is most sensitive to spatial detail in an image. In VMT we use retinal maps that indicate regions of functional significance in normal vision, such as maps of visual acuity, of color discrimination, of motion sensitivity, etc., to construct a representation of the volumes in world space that are served by such visual functions.

2.1.1 World Space Representation

An environment of sufficient detail must be created before any analysis can be performed. There is a tradeoff between detail and performance that behooves the user to keep the representation as simple as possible. Many of VMT features depend on an interactive environment to be of maximum use and performance is adversely effected by very complex objects.

Properly scaled, 3 dimensional objects must be created and positioned accurately with respect to other objects and the human model. An object in the environment is a boundary model that consists of a set of scaled vertices, faces and linkages that describe its geometry. Surface properties can be defined for these objects but they are not used (with the exception of the CRT) in the analysis. These objects can be built by many commercial CAD packages. The output of the CAD package is then translated to the internal representation of VTM and imported into the environment. In our example, this is an AH-64 Apache helicopter complete with a detailed cockpit.

In addition to geometric objects it is often useful to define sites that are of particular importance in

the environment. Sites are 3 dimensional locations that allow the user of the system to have a way to reference that point in space. For example, the user can define a site on the surface of a multi-function display in the crewstation. He/she could then command the human model (to be explained later) to fixate on that point. A sequential collection of sites can specify a path, such as a scan pattern, in three space.

2.1.2 Three Dimensionality of Retinal Images

In the World/Eyes subsystem, visual space is recognized as being three-dimensional, even though each of the retinal images is only two-dimensional. In addition to the horizontal (x) and vertical (y) coordinates of the traditional visual field map, a composite of the two retinal images is used to compute the third dimension of visual space, the additional coordinate denoting the horizontal difference between positions of retinal object points in the two eyes. These three coordinates are used to compute locations in visual space that correspond to, or are constrained by, retinally-mapped functions, such as visual acuity. The geometry underlying this is largely the same as that underlying stereoscopic depth perception in biological vision, although the analysis is otherwise unrelated to depth perception.

2.1.3 Concurrent Retinal Images

A key feature of the World/Eyes subsystem is the concurrent viewing of retinal information in one or more RETINA windows with viewing of one or more WORLD representations, including the prompt updating of all such views. In this way, the impact of transformations in one domain can be appreciated in the other. Motion of world objects, for example, will result in retinal image motion, but often in nonintuitive directions and speeds -- viewing yoked displays of visual space and retinal space simultaneously is a powerful aid to the understanding of such motion. Similarly, changes in global and local retinal sensitivity over time can drastically alter the visibility of objects in space. The ability to view a rendition of visual space that shows changes in the visibility of environmental objects would be of obvious value.

2.1.4 Retinal Projections and Retrojections

Retinal images in the WORLD/EYES subsystem are projections of environmental object representations onto the retina(s) of a RETINA window. A kind of converse operation that we call here "retrojection" is the construction of the locus of possible points in the environment which may stimulate, through projection, a specific location on the retina(s). Retrojection of a retinal map shows the footprint of that map in the environment. For example, if a circular region on the retina of 5 deg radius represents the boundary beyond which a normal observer may not reliably identify a letter of, say, 20 min arc height, then the retrojection of that 5 deg circle, which will be a cone whose apex lies at the optic node of the eye, will represent the volume of space within which, assuming proper focus of the eye, the letter can be identified. If we intersect this cone with the corresponding cone of the other eye, this retrojection will represent the volume of space within which either eye can identify the letter. Thus projection and retrojection are the operations which relate visual world geometry to retinal geometry, and may be used to analyze visual functions in the space of the environment which they serve.

2.1.5 Volume Visual Field

The volume visual field (VVF) is a global construct that embodies the main features of the WORLD/EYES subsystem. It is the locus of points in the environment that fall on light-sensitive, functioning retina of one or both eyes. It should be thought of as a geometrical construct that delineates boundaries of visibility in the coordinates of environmental space, rather than in retinal coordinates. Its characteristics will vary with the visual task being performed, with eye and head position, and with sensitivity of the retina and visual apparatus to light, color, motion, and other relevant stimulus attributes.

2.2 The Brain Subsystem

The final subsystem models the processing of the image formed on the retina by the visual pathway of the brain. Because the goal of VMT is to provide the designer with a tool to evaluate design parameters such as the placement of displays, and the size and contrast of symbology, the model need only be accurate with respect to these human visual system behaviors. It does not need to be an accurate model of the actual physiological mechanisms underlying these behaviors.

The purpose of the model is to provide visibility data to display designers, who need quantitative information on the effect that their design

choices will have on crew performance over a broad range of mission scenarios. Given this range, the visibility of alphanumeric and other information depends not only on the spatial configuration of the display, but also importantly on lighting parameters such as light level in the cockpit and the observer's state of adaptation. The model must therefore accurately assess the effect of such parameters, and convey this information to the display designer in an easily understandable form.

As in any model of a physical or biological process, simplifying assumptions and approximations are sometimes required. Simplification and approximation can lead to error in the predictions of the model. We have tried to bias the error in the direction of underestimating real behavior rather than overestimating it. The rationale behind this is that it is better for a CAD tool to detect design faults with some false alarms than it would be for it to falsely convey a sense that the design is adequate when it is not. The size of this error, of course, determines the usefulness of the design tool. Therefore we have attempted to validate the predictions of the model by using it to predict known results as well as conducting empirical tests of its predictions.

2.2.1 Background

A number of visibility models have already been developed by workers in both the applied vision and basic psychophysics communities. Each can successfully predict human performance within a restricted range of stimulus and task domains.

2.2.1.1 Applied Psychophysics

An early success in applied vision was the JND Model of Carlson and his associates. [2,3] In this model, an input image is decomposed via a one dimensional fourier transform into a number of spatial frequency bands. These filtered bands are then perturbed by various noise sources, squared, and spatially integrated. Changes in the output of this process from one member of a pair of images to the other provides a simple perceptual measure of the visibility of differences between the two images. This model has successfully predicted the visibility of changes in edge sharpness and of various display artifacts, among other things. The disadvantages of the model are that it is spatially one-dimensional, and is somewhat complicated to compute, since a noise parameter must be adjusted for each change in display parameters such as luminance and display size. A more recent variant of this model, the Barten SQRI Model [4], solves some of the complexity problems by introducing polynomial approximations for the changes in human sensitivity to changes in display parameters.

2.2.1.2 Basic Psychophysics

Similar models have been introduced into the basic psychophysics literature by Wilson and his colleagues [5,6] based on the threshold model of Wilson and Bergen[7], and by Legge and Foley [8,9]. These models successfully predict human performance in simple psychophysical tasks such as grating contrast detection and discrimination. In all of these models, the input image is first decomposed into independent spatial frequency channels by a set of linear filters. The output of each filter is then put through a sigmoid non-linearity, the shape of which matches very closely that of the non-linearity imposed by the noise and squaring steps of the JND Model.

2.2.1.3 Two-Dimensional Models

All the models described above are spatially one-dimensional; that is, they predict sensitivity to spatial variation in one dimension only. Watson and his colleagues [10,11,12] have implemented a model which generalizes the linear filtering stage of these models to two dimensions. Each filter is a two-dimensional gabor function, with a number of different scales, orientations, and phases of filtering at each point in the two-dimensional visual field, and an increase in the overall scale of filtering as a function of eccentricity. The model has been validated on some detection and discrimination data. One limitation is that it is only accurate at stimulus levels near detection threshold since, unlike the other models described above, there is no point non-linearity after the linear filtering stage.

2.2.2 Combining Information Across Channels

One problem all these models must face is how to combine information across a large number of different filtering channels, so that a unidimensional value for human performance on a discrimination task can be obtained. In other words, from the large dimensional space represented by the channel outputs, the models must derive something like a single value for the probability of detecting a difference between a pair of stimuli.

One way to perform this reduction of dimensionality is to base the performance measure only on the single channel which shows the maximum change in output from one member of the stimulus pair to the other. This "maximum-of" decision rule is implicit in most of the basic psychophysics modelling, and is motivated by the hope that the simple stimuli usually used in that paradigm are sufficiently localized in the space of channel outputs so that only one channel governs performance, regardless of the degree to which channel outputs are combined.

Other models, like the Watson model, use variants of an optimal Bayesian classifier at the channel combination stage. Given the assumption that each channel output is perturbed by zero mean, unit variance Gaussian noise, the detectability of a pattern by an optimal observer is directly proportional to the euclidean distance of the pattern's feature vector from the origin of the channel output space. Under a more general model, detectability can be modeled in the same way, but with an exponent different than the euclidean 2.

2.2.3 Basic Strategy

Our approach to the legibility modeling task has been to augment the one-dimensional discrimination models of Carlson, Barten, and the basic psychophysics community to include the two-dimensional spatial processing used in the Watson detection model. In addition, because the task requires performance prediction over a wide range of lighting conditions and observer perceptual states, the model includes a front end that pre-processes the input images to model the effects of changes in illuminance, screen luminance, and observer fixation location in three-dimensional space.

2.2.4 Choice of Performance Measure

The performance measure chosen for the model is probability of a correct discrimination between two input images; e.g., between two symbols or alpha-numeric characters. Other performance measures are of course possible; one common measure from the applied psychophysics community is image quality, expressed in units of just-noticeable differences (JNDs) between the two images. In fact, the probability measure used in the model is derived from a JND-like measure, as will be described below.

Another important consideration in the model is that the performance measure should be conservative. To estimate the legibility of letters we decided to model a human's ability to discriminate between the letters O and Q. This is one of the most difficult discriminations to make and generally when these letters can be discriminated most of the time other letter pairs are equally or more discriminable.

2.2.5 Incremental Improvement

Finally, the model has been constructed in such a way as to allow incremental improvement in its ability to predict performance among complex stimuli. For example, the model as currently formulated can accurately predict performance only among static, monochromatic images. However, by replacing the model's spatial filters with a set of spatio-temporal filters, performance measurement among moving stimuli would be possible. The strategy in model development is therefore to validate the simple model on simple stimuli, and then incrementally build in model complexity to handle the more complex stimuli.

3.0 COMPUTATIONAL MODELS

3.1 Computational Environment

The visibility modelling tool has about 100,000 lines of code, completely written in the C programming language. It is comprised of three major software components. The first, VP, which stands for volume perimetry, is the portion of the software that represents the VVF and retinal images as distinct graphical constructs in separate windows that are yoked to illustrate the effects of making changes in the VVF or on the retinas.

The second software component, the legibility model, computes the probability for correct discrimination between two symbols based on stimulus characteristics, environmental factors and observer state. Due to the intensive computational requirements of the model, and the desirability of building an interactive visibility tool, legibility data is precomputed and stored in files. During operation of the VMT, the appropriate files are read in and displayed based on the current vision relevant parameters.

The third software component is an anthropometric modelling tool called Jack. This software provides the kinematically correct models of stereotypical male/female body types.[13]

The three software components can be run as an integrated program on the Silicon Graphic 4D series of workstations. These workstations have 24 bit color, 8 bit alpha and 24 bit Z buffered double frame buffers. The workstations have hardware support for light models and viewing transformations. While this workstation provides the performance we require for the integrated VMT tool, VP and the legibility model can be run independently on less costly workstations.

3.2 User Interface

VMT is an easy to use, mouse driven, multi-windowed program. The anthropometric model in Jack was extended to include eye coordinates and fixation point. The normal Jack interface [13] was extended to include manipulation of vision relevant parameters. The designer can interactively set the fixation point, via the mouse, anywhere in 3D space as long as it is within the normal eye joint limits. The fixation point can also be set to predefined points. As the fixation point is varied the designer can watch the vision plots and the field of view cones reflect those changes in real time. The user can set model parameters, such as, ambient light, font size, etc. via slider bars.

The designer must be able to relate dynamic vision characteristic to the objects within the design environment and based on those relationships make design decisions. Data visualization techniques are employed to illuminate these relations.

3.3 The VVF Display

The VVF or display of the "world" is a three-dimensional graphical construct. In the VVF display, simple three-dimensional objects can be introduced directly from the menu. More complicated objects can be created by instructing the program to read a description file that specifies the structure, location and orientation of the object. Newly created objects are assigned a default position in the VVF if no position information was specified in the description file, but can be moved to any location in the VVF. They can also be connected to build compound objects out of simple objects. Each object has a dynamically allocated database that contains information about the vertices, faces, and location of the object.

The VVF is displayed as a perspective projection, the parameters of which are stored in the window's database. The view reference point and center of projection are also stored in the window's database, and can be changed interactively to look at any point in the VVF from any location. Each time a new VVF window is created, the window is assigned default values for the projection and view, but it inherits the objects that already exist in the VVF. Essentially, opening a new VVF window gives the user the ability to view the VVF in a different way.

3.4 The Retina Display

The retinal display illustrates the dynamic relationships of three types of objects with respect to the visual axis. By default, it shows the retinal images that are formed by objects in the VVF. The orientation of these images varies with changes to the fixation point and with orientation and position changes of the head. A subset of retinal images that form the second object type, are formed from objects that are fixed with respect to the head. Since helmet mounted devices, nose bridge, glasses, etc., are stationary to the head coordinate system, this object type remains stationary in the retina display unless the fixation point is changed. The retina display can also illustrate areas that are characteristic of the retina, and are therefore fixed in location and orientation even when fixation and the head position is varied. The fovea and the natural blind spot are examples of fixed data. (We'll refer to fixed data as retinal objects to distinguish them from the variable retinal images discussed above.) The proper dynamic relationships between the data types are maintained as the user manipulates the fixation point.

The user can open additional retina windows as needed, and the retinal displays can be interactively customized to make it easier to see data of interest. For example, each retinal window can display either the right or left retina, or the two superimposed on each other. Left retinal data is currently displayed in shades of green, right retinal data is displayed in shades of red, and areas of overlap are shown in yellow.

The user can build up a database of retinal objects e.g. cone density data, scotomas, etc. This is the means by which the legibility data is brought into VMT. Retinal objects can also be interactively drawn directly onto the right, left or both retinas.

An important difference between retinal images and retinal objects, is that the objects can be retrojected (discussed below) into the VVF. There is no need to

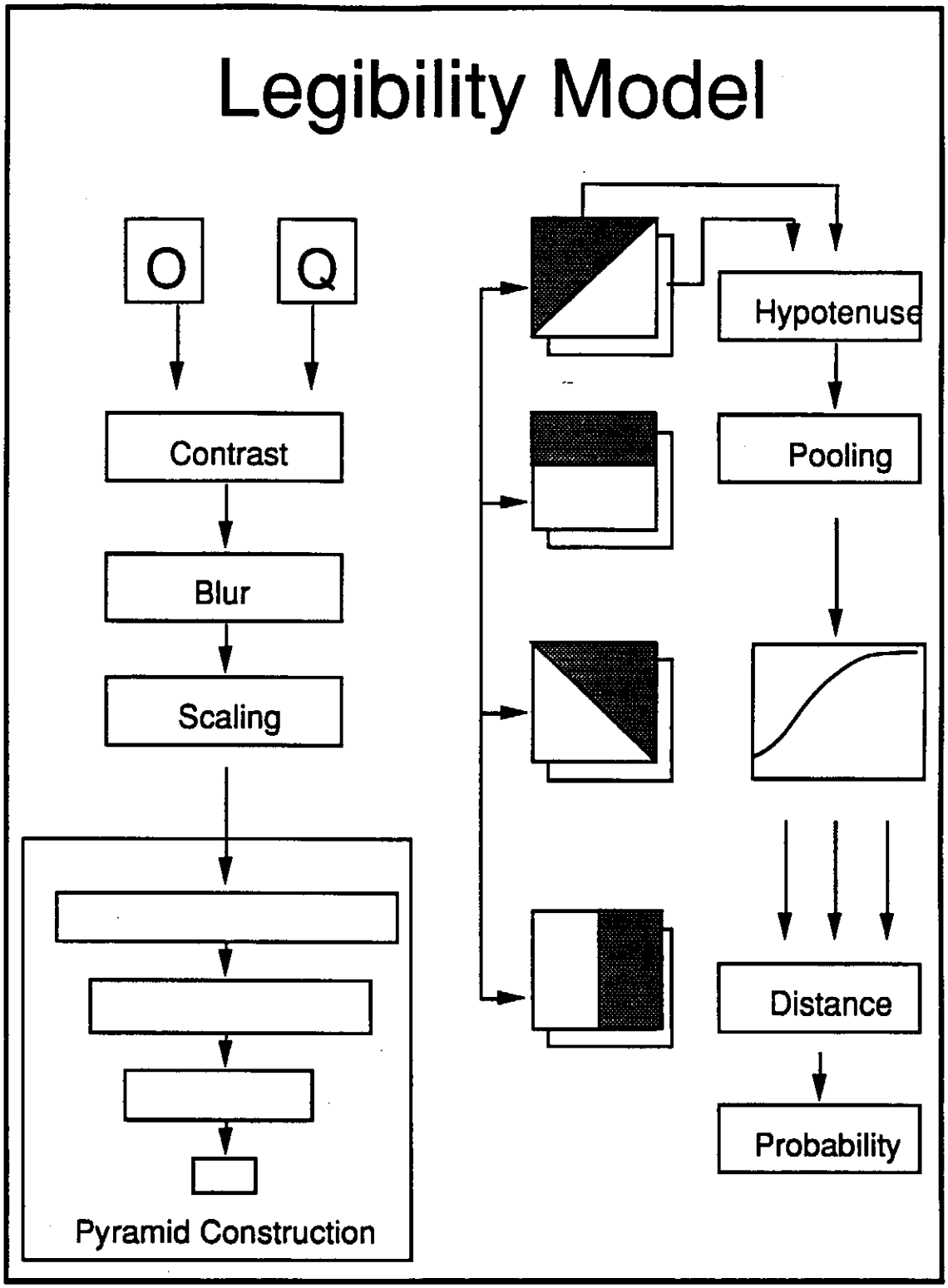
Legibility Model
O
Q
Contrast
Blur
Scaling
Pyramid Construction
Hypotenuse
Pooling
Distance
Probability

Figure 1

retroject retina images because the object that generated the image already exists in the world data base and its prepresentation.

3.5 Field of View Cones

Designers often have requirements to place instrumentation within some specified field of view. The designer using VMT can set the solid angle for any field of view of interest. Semi-transparent view cones are then projected along the current visual axis for the right and left eyes into the crewstation. The apex of the the view cones originates at the nodal point of the left/right eye.The convergence of the cones is at the fixation point. The intersection of the cones and the crewstation, delineates the objects in the world that fall within the current field of view settings.

3.6 Total Field of View Plots

Frequently, of major importance in vehicle design is the area visible out the window. Total field of view plots provide the designer with a 360 degree plot of external visibility. It is possible to read information such as over the nose visibility directly from these plots. The results can be seen immediately as the designer explores effects of various parameters such as the pilot's body size, the location and size of windows, the placement of seats in the cockpit,etc.

3.7 Retrojections

Retinal objects such as scotomas or cone density maps can be selectively retrojected into the VVF or "world." By retrojecting a retinal object, the user can see where it intersects the VVF. The retrojection is drawn as a semi-transparent volume in the VVF. A retinal object can be defined to be the retinal locus wherein letters of a particular size and image contrast can be discriminated with 100% accuracy. This object can be retrojected for a given head/eye position onto surfaces in the VVF or "world." If two letters of that size and contrast are surface attributes of a surface at the fixation distance within the retrojection volume, then they will be accurately discriminated from each other. This can be a useful tool for setting contrast and size specifications for symbolic displays in a cockpit.

3.8 The Legibility Model

Figure 1 is a schematic diagram showing the different stages of the legibility model. A bit map of the input is adjusted for contrast, blur, and retinal locus. This bit map is filtered by a set of oriented filters of different spatial size as in a pyramid image processing scheme. For each filter pair, i.e. sine and cosine phase, are then combined by a hypotenuse rule to yield a single scalar value. These values are subjected to a monotone nonlinear transformation. The input is then represented as an n-tuple which is used in a distance calculation when comparing two inputs. These distances are finally transformed into probabilities. The details of these steps is discussed below.

3.8.1 Input Parameters and Stimulus Format

The model, as currently formulated, takes as input one or two image files, and a number of optional lighting and observer state parameters. Since the legibility model, as we are currently applying it, predicts correct character discriminations as displayed on an Apache multi-function display, the images pass into the model are consistant with the font size and pixel representation of the characters of that display. These parameters are listed here with the default value and units in parentheses:

Screen luminance (10.0 foot-lamberts),
Illuminance (0.0 foot-candles),
Eccentricity of displayed stimulus (0 degrees),
Fixation depth (741.12 mm),
Stimulus depth (741.12 mm).

The image files start with two four-byte integers indicating the width and height of the image in pixels, followed by rows of pixel values in floats. The images can be any size, although they should be at least 256x256 to allow filtering within a large enough range of different frequency bands. The conversion factor from pixels to mm is currently fixed in the software as 13.21 pix/mm. With only one image as input, the software calculates the probability of detecting that stimulus; with two images, the discrimination probability is calculated.

In the current version of the model, pixel values are required to range from -1.0 to 1.0, with the maximum absolute value indicating the contrast of the stimulus. For example, a sine grating stimulus with peaks at ±0.5 would have a contrast of 0.5. To remove the need for this convention, a model stage which computes contrast from arbitrarily scaled input images is needed, but has not yet been implemented.

3.8.2 Front End Calculations

Several initial transformations on the input images are performed prior to the linear filtering stage of the model. These transformations model the effect of changes in fixation depth, veiling luminance, and fixation eccentricity.

3.8.3 Fixation Depth

In order to account for changes in effective image resolution with changes in the difference between image depth and fixation depth, we used geometrical optics to calculate the size of the blur circle, and then pre-filtered each input image with this disk-shaped convolution kernel. This calculation requires knowledge of the distance from the exit pupil to the imaging surface (i.e., the retina), which we took as 20.3 mm from Westheimer[14]. It also requires an estimate of pupil size. For this, we wrote a simple interpolation routine to estimate pupil diameter at any light level from a table published in Hood and Finkelstein [15].

3.8.4 Contrast Reduction

Veiling luminance, caused by the reflection of ambient light by the display screen, reduces the effective contrast of displayed information. We are modelling the screen face as a perfectly lambertian surface with a reflectivity of 10%. This assumption implies that an illuminance of 10 fcd will result in a veiling luminance of 1 fL. We are defining contrast as

$$c = [L_{max} - L_{min}] / [L_{max} + L_{min}]$$

where L_{max} and L_{min} are the maxmimum and minimum displayed luminances. Given this definition, the addition of a veiling luminance v to both L_{max} and L_{min} changes the contrast by a factor

$$[L_{max} + L_{min}] / [L_{max} + L_{min} + 2v].$$

3.8.5 Eccentricity Scaling

Psychophysical evidence shows that contrast sensitivity remains roughly constant across the visual field, if the grating patch is scaled up by a linear function of eccentricity. This strongly suggests that processing is similar across the visual field, except for a linear scaling up of sensor size towards the periphery. In order to model the effect of this change in sensor size, we found it more convenient to scale down the size of the input images as a function of eccentricity, rather than to scale up the size of the sensors. We used the scale factor (k) of 0.4 quoted by Watson [10], so that the scaling of input images as a function of eccentricity (e) is

$$1.0 / (1.0 + ke).$$

3.8.6 Linear filtering

3.8.6.1 Pyramid Decomposition

In order to filter within a range of different frequency channels, the input image is first decomposed with a gaussian pyramid into channels separated from each other by one octave. The frequencies we chose for these channels are identical to those used by Watson [10]; i.e., 32 through 0.5 cycles/degree, corresponding to seven octaves or equivalently, seven pyramid levels.

3.8.6.2 Computing Filter Gains

The human visual system is not equally sensitive at all frequencies. A plot of contrast detection threshold as a function of spatial frequency shows roughly an inverted-U shape, with a peak at roughly 2 c/d, and complete loss of sensitivity by approximately 60 c/d. Moreover, as shown by van Nes and Bouman [16], the shape of this contrast sensitivity function changes with retinal illuminance. To model these dependencies, the image component in each frequency channel is weighted by a gain factor appropriate for the retinal illuminance, before any oriented filtering is performed.

Retinal illuminance (in photopic trolands) is calculated as the amount of light incident on the cornea (in cd m^{-2}) times the pupil area (in mm^2), where the light incident on the cornea is assumed to be the screen luminance plus the veiling luminance, appropriately converted from fL to cd m^{-2}. The gain at each frequency is then calculated directly from the van Nes and Bouman data with a simple log interpolation function to return sensitivities at retinal illuminances other than those reported in the data. For example, if the threshold modulation for a 1 c/d grating were 1% at 10 tds and 5% at 1 tds, then we interpolate the threshold at 3.16 tds (half the log distance from 1 to 10) to be 2.23% (half the log distance from 1% to 5%). This direct calculation is possible only under the assumption of no summation among different frequency channels; any assumed

summation would require a more complicated relationship between the contrast sensitivity function and the gain of each channel.

3.8.6.3 Steerable Filtering

For convenience and speed of oriented filter operation, we use the steerable filters of Freeman and Adelson [17], which allow separable calculation of linear filter responses at any orientation and phase. The filters implemented here, a second derivative of a gaussian and its Hilbert transform, have a log bandwidth at half height of approximately 0.7 octaves. This is within the range of bandwidths inferred psychophysically.[18] We are using four orientations (0, 45, 90, and 135 degrees) and two phases (sine and cosine), for a total of eight oriented filter responses per pyramid level. The orientation bandwidth of these filters (i.e., the range of angles over which the filter output is greater than one half the maximum) is approximately 65 degrees. This figure is slightly larger than the 40 degree tuning of monkey simple cells reported by Devalois et al [19], and the 30 to 60 degree range reported psychophysically by Phillips and Wilson.[20]

3.8.6.4 Energy Calculation

During the early stages of model testing, we found that the detectability of a simple edge could change dramatically with small changes in edge position. To combat this problem, a small amount of spatial summation was added by computing energy after the linear filtering stage. That is, corresponding sine and cosine filter responses were combined as

$$e(x_i) = \sin^2(x_i) + \cos^2(x_i)$$

where x_i is a linear filter response, indexed over filter position, orientation, and frequency band.

3.8.7 Point Non-Linearity

Nachmias and Sansbury [21] showed that the results of a grating contrast discrimination experiment, when plotted with threshold contrast increment as a function of the base contrast from which the increment threshold is being measured, produce a dipper-shaped curve. These authors argued that the results can be modelled by assuming a sigmoid non-linearity following a linear detection mechanism. The decision mechanism has available to it only the output of this non-linearity, and reliably discriminates between inputs of two different contrasts when the difference in outputs is greater than some threshold.

To quantitatively model the dipper-shaped contrast discrimination curve, we follow Legge and Foley [8] in using a non-linear transducer of the form

$$T(L_i) = r|L_i|^n / [|L_i|^2 + s^2]$$

where T is the non-linear transducer output, L_i is the linear filter response (indexed as above), r is an overall gain-setting parameter, n is a real number greater than 2 (2.4 here), and s is a semi-saturation constant (0.0075).

3.8.8 Distance Calculation to Probability

We are assuming no summation among transducer outputs, and a decision mechanism in which discrimination is governed by the pathway whose transducer output shows the maximum change between the two stimulus presentations. Although these assumptions are probably not correct in detail, they dramatically simplify model theory and calculations, and thus are useful as a first pass at the truth.

One way of expressing this maximum change assumption is that the discriminability between two images is assumed to be proportional to a non-Euclidean distance between the points representing the two images in the space of transducer outputs. That is,

$$D(s_1,s_2) = \left[\sum_{i}^{n} [T(e_i(s_1)) - T(e_i(s_2))]^Q\right]^{(1/Q)}$$

where s_1 and s_2 are the two images, and n is the number of different transducer channels. If Q=2, this expression returns the Euclidean distance between points in the transducer output space. As Q goes to ∞, the distance metric gets closer and closer to the maximum change model described above.

3.8.9 Distance to Probability

This distance measure can be used in the following generalization of the Nachmias and Sansbury model: Two images are reliably discriminated whenever the result of the distance calculation for the two images is greater than some threshold.

This simple generalization is sufficient to model results in which discrimination thresholds are

measured as a function of a change in a stimulus parameter (e.g. contrast, as in the Nachmias and Sansbury results, or spatial frequency, as in the edge transition results to be discussed below.) For these tasks, the threshold distance can be understood as the distance which gives a 75% probability of discrimination among the two stimuli. However, for the cockpit display visibility modelling, the relevant performance measure is the probability of discrimination given two images, not the required change between two images given a fixed probability. Therefore, a mapping between distance and probability is required, not for just a single value of distance and probability, but along the entire range of each.

We have generated this complete mapping from distance to probability using two sets of data: (1) contrast detection psychometric functions from Foley and Legge[9], and (2) contrast discrimination functions from Legge and Foley[8]. In the former data set, probability of detecting a sine grating is plotted against the contrast of that grating. Foley and Legge showed that these data are well fit by an expression

$$p(C) = 100 - 50\exp(-aC^b)$$

where C is contrast, and a and b are parameters, fitted to one observer as a=52.60 and b=3.0.

For the latter data set, a good fit is obtained using the transducer function described above. This transducer expresses distance as a function of linear filter output, but can be expressed equally well as a function of grating contrast, since linear filter output is proportional to contrast. This means we can express both distance (i.e., transducer output) and probability as a function of contrast, and have only to invert the transducer function to obtain an expression for probability as a function of distance. We were not able to solve this problem analytically, but instead generated an accurate computational solution that was incorporated into the visibility model software as a lookup table.

To test this mapping, we applied the model to predict psychometric functions for contrast discrimination, also published in Legge and Foley [8]. In these functions, probability for detecting a change in contrast of a sine grating at threshold is plotted against that change in contrast. The model predicted these functions very accurately.

3.8.10 Position Uncertainty

One important task for performance measurement in a cockpit display environment is character discrimination as a function of eccentricity. However, predictions on O vs. Q discrimination from the model as described so far incorrectly assert near perfect discrimination out to as much as 16 degrees, whereas in reality, probability of successful discrimination among these two characters falls to chance by about 5 degrees.

Because of the lack of spatial pooling in the model, the task of discriminating between O and Q becomes, for the model, the task of simply detecting the diagonal segment in the Q, a task which could in fact be performed reliably out to 16 degrees. But for letter discrimination, it is not sufficient to simply detect all the component features; accurate spatial relationships among these features must also be recovered. This led us to consider other psychophysical tasks for which accurate localization is important, most notably the three dot bisection task of Yap, Levi and Klein.[22] Here, as in contrast detection, a linear scaling of image area with eccentricity leads to constant performance across the visual field. However, the scaling factor for three dot bisection is larger by about a factor of three than the scaling factor for contrast detection.

The fact that tasks requiring accurate localization scale by a larger factor than that of contrast sensitivity suggested to us the need for a stage of eccentricity dependent spatial pooling of sensor responses following the linear filtering stage. If filter responses were pooled over a progressively larger area towards the periphery, then the central visual system would become increasingly uncertain of the spatial position of a given image feature. Furthermore, if the gain of the pooling filters were scaled so that the magnitude of the pooled response of a uniform input were unchanged with the size of the input area, then the model would continue to accurately predict contrast sensitivity in the periphery.

When we incorporated these changes into the model, it was able to qualitatively predict published results in letter discrimination, three-dot bisection, and contrast sensitivity in the periphery. After making these changes, we found further verification for our letter discrimination predictions in a recent report by Farrell

and Desmarais.[23]

3.8.11 Output Format

Model predictions are generated in files for which all parameters are fixed, except for eccentricity (i.e., degrees of visual angle off of the fixation point.) The output files thus contain a simple two column table of eccentricity - probability pairs, one such file for each combination of image pair, lighting, and fixation depth parameters.

The initial set of model results included predictions for two different fonts of Q and O produced on a CRT planned as a primary display device for the modified AH64 Apache helicopter.[24] A screen depth of 30" was assumed, coupled with four different fixation depths: 15", 30", 60", and ∞. For each of the two fonts and four fixation depths, we generated model predictions at several luminance/illuminance combinations ranging from 10 fL to 350 fL for screen luminance and 10 to 10,000 fcd for ambient illumination. These illuminances range from bright sunlight (10,000 fcd) to late dusk (10 fcd) and the luminances range from typical high resolution color workstation moniters (10 fL) to bright monochrome avionic CRT displays (350 fL).

In order to generate complete discrimination contours from the model output files, the package of model software also contains a routine which generates probability contours from 55% to 95%, and the 99% contour. This routine performs a linear interpolation on the probability values listed in the model output file, to determine the eccentricity at which each contour's probability value would occur. It then computes a set of x,y points (in degrees of visual angle) for a complete circle at that eccentricity. Additional data may, in the future, require refinement of this module to produce contours which deviate from a purely circular shape.

3.9 Model Validation

We will now briefly describe two methods that were employed to validate the legibility model predictions.

3.9.1 Edge Transition Data

The discrimination model has been successfully tested against the Carlson and Cohen[2] edge transition data. In the Carlson and Cohen edge transition task, observers were asked to discriminate a change in sharpness of a vertical intensity edge, as a function of the base sharpness of that edge. The sharpness of the edge was controlled by a parameter which represented the frequency at which the modulation transfer function for the edge has fallen to one half of its maximum. The model fit to this data was better than the fit of Carlson and Cohen's original (JND) model.

3.9.2 Character Discriminability Data

We have begun collecting human data to test the model's predictions in a character discrimination task and these will be reported at the 1991 annual meeting of the Association for Research in Vision and Ophthalmology.[25,26] The stimuli, as in the model runs, were uppercase O and Q. The screen luminance was variable from between 0 and 13.5 ftL. The subject's distance from the screen is 30 inches. Veiling illuminance for the data collected so far was approximately 10 fcd.

We found that the falloff in performance with eccentricity is in good agreement with model predictions although shifted by a constant amount on the eccentricity axis. One possible reason for this discrepancy is that the short stimulus duration used in the experiment (to prevent eye movements) reduces the effective contrast of the stimuli, thereby significantly affecting performance. Since the model as it now stands is blind to changes in stimulus duration, it would not be expected to accurately predict performance for arbitrarily short duration stimuli. These results suggest that some additional model development is needed to accurately measure the effective contrast of a stimulus, given its time course and other relevant stimulus and observer parameters. We believe that the model is sufficiently accurate at this time to be a valuable design tool when making predictions about the legibility of fonts similar to the one we employed.

4.0 LIMITATIONS AND FUTURE DIRECTIONS

The Visual Modeling Tool that we have created is a useful design tool in its current form. It addresses virtually all of the field of view questions that are required by the standard design specifications for cockpits in commerical and military aircraft manufacturing. Moreover, it embodies many features such as the volume field of view, dynamic graphics for

visualization, retrojections, and legibility analysis that are extremely useful during the design phase. For example, with our system it is easy to see the limitations on visibility in terms of field of view that result from physically small pilots, seat heights, and window post placements. The system can also aid the designer in determining font sizes, contrast requirements, illumination levels during all ambient lighting conditions, etc. in setting specifications for displays.

Many of the predictions of the legibility module or "brain" subsystem are limited by the domain of stimulus variation that we have attempted to model to date. Thus a broad spectrum of font types, and spectral characteristics remain to be modeled. Currently the model is moot on issues related to display color, temporal dynamics such as flicker and motion, target localization or acquisition, reaction times, pilot's visual system adaptation state, and resource loading or work load. This of course is not an exhaustive list of desired features that are not currently part of the system. Additional feature could easily be added to the list.

Phenomena such as human light adaptation, flash blindness, or color perception are good candidates for future development. Efforts also need to be expended to verify the performance models as they are developed. The system was engineered as a demonstration of this CAD concept, therefore considerable effort could be expended to optimize the software and to improve the user interface. All of these options for future development are under consideration.

5.0 ACKNOWLEDGMENTS

This work was was sponsored as part of the Army-NASA Aircrew/Aircraft Integration (A3I) program under cooperative agreements NCC2-541 and NAG2-426, and contracts NAS2-12852 and NAS2-13210.

6.0 REFERENCES

1. Arditi, A. The Volume Visual Field: A basis for functional perimetry. Clinical Vision Sciences, 3:173-183, 1988.

2. Carlson, C. and Cohen, R. (1980) A simple psychophysical model for predicting the visibility of displayed information. Proceedings of the Society for Information Display, 21:229-245.

3. Carlson, C. and Klopfenstein, R. (1985) Spatial frequency model forhyperacuity. Journal of the Optical Society of America A, 2:1747-1751.

4. Barten, P.G.J. (1987) The SQRI method: a new method for the evaluation of visible resolution on a display. Proceedings of the Society for Information Display, 30:253-262.

5. Wilson, H.R., McFarlane, D.K., and Phillips, G.C. (1983) Spatial tuning of orientation selective units estimated by oblique masking. Vision Research, 23:873-882.

6. Wilson, H.R. and Regan, D. (1984) Spatial-frequency adaptation and grating discrimination: predictions of a line element model. Journalof the Optical Society of America A, 1:1091-1096.

7. Wilson, H.R. and Bergen, J.R. (1979) A four mechanism model for threshold spatial vision. Vision Research, 19:19-32.

8. Legge, G.E. and Foley, J.M. (1980) Contrast masking in human vision. Journal of the Optical Society of America, 70:1458-1470.

9. Foley, J.M. and Legge, G.E. (1981) Contrast detection and near-threshold discrimination in human vision. Vision Research, 21:1041-1053.

10. Watson, A. (1983) Detection and recognition of simple spatial forms. In O. Braddick and A. Sleigh, eds., Physical and Biological Processing of Images. Berlin: Springer-Verlag.

11. Ahumada, Jr., A. and Watson, A. (1985) Equivalent noise model for contrast detection and discrimination. Journal of the Optical Society of America A, 2:1133-1139.

12. Nielsen, K., Watson, A., and Ahumada, Jr., A. (1985) Application of a computable model of human spatial vision to phase discrimination. Journal of the Optical Society of America A, 2:1600-1606.

13. Badler, N. Jack 4.8 User's Manual. Computer Graphic Research Laboratory, University of Pennsylvainia, 1990

14. Westheimer, G. (1986) The eye as an optical instrument. in K. Boff, L. Kaufman, and J. Thomas, eds., Handbook of Perception and Human Performance. New York: Wiley.

15. Hood, D.C., and Finkelstein, M.A. (1986) Sensitivity to light. in K. Boff, L. Kaufman, and J. Thomas, eds., Handbook of Perception and Human Performance. New York: Wiley.

16. Van Nes, F.L. and Bouman, M.A. (1967) Spatial modulation transfer in the human eye. Journal of the Optical Society of America, 57:401-406.

17. Freeman, W.T. and Adelson, E.H. (1990) The design and use of steerable filters in image analysis, enhancement, and wavelet representation. MIT Media Laboratory Technical Report 126a.

18. Watson, A. and Robson, J.G. (1981) Discrimination at threshold: Labelled detectors in human vision. Vision Research, 21:1115-1122.

19. DeValois, R.L., Yund, E.W., and Hepler, N. (1982) The orientation and direction selectivity of cells in macaque visual cortex. Vision Research, 22:531-544.

20. Phillips, G.C. and Wilson, H.R. (1984)

Orientation bandwidths of spatial mechanisms measured by masking. Journal of the Optical Society of America A, 1:226-232.

21. Nachmias, J. and Sansbury, R.V. (1974) Grating contrast: Discrimination may be better than detection. Vision Research, 14:1039-1042.

22. Yap, Y.L., Levi, D.M., and Klein, S.A. (1987) Peripheral hyperacuity: three dot bisection scales to a single factor from 0 to 10 degrees. Journal of the Optical Society of America A, 4:1554-1561.

23. Farrell, J.E. and Desmarais, M. (1990) Equating character-identification performance across the visual field. Journal of the Optical Society of America A, 7:152-159.

24. Crew Systems Interface Document for the Apache Longbow, Volume 5, Revision A, Controls and Displays Management, 22 December 1989, McDonnell Douglas Helicopter Company, Mesa, AZ.

25. Lubin, J. (1989) Discrimination contours in an opponent motion stimulus space. Investigative Ophthalmology and Visual Science Supplement, 30:426.

26. Lubin, J. and Nachmias, J (1990) Discrimination contours in an f/3f stimulus space. Investigative Ophthalmology and Visual Science Supplement, 31:409.