Chapter 6

Discussion

‘ “Begin at the beginning,” the King said, gravely,
“and go on till you come to the end: then stop.” ’

(Lewis Carroll, *Alice in Wonderland*)

This chapter provides a focal point to discuss many of the issues pertaining to the subject matter of this dissertation. This will be done from several perspectives. We will begin by examining the specific issues associated with our particular model and its implementation. Following this, we will discuss the more general matters relating to any computer graphics system which attempts to modulate detail using perceptual criteria. Finally, we will offer some comment on the wider issues of applying knowledge from the domain of visual perception to that of computer graphics.

6.1 Discussion of our Model and Implementation

This first section deals with the specific model and implementation that we have developed as a result of this research effort. Through this examination we will review how successful the work has been, revise our visual acuity model based upon our experimental results, contrast a few of the major design decisions that were taken, and finally, we will consider the extension of our model to support fixed frame rate LOD.

6.1.1 Reviewing the Results from our Prototype Evaluation

The results from the previous chapter offer a proof of concept for the notion of perceptually modulated LOD in general, and our implementation in particular. They show quite plainly that such a system can substantially improve the performance of a VE, resulting in increased user performance; and that this can be done with no detrimental effect on the perceived fidelity of the VE. These results provide an indication of the amount of visual detail that a standard graphics renderer produces which the user may never see.

During the Introduction chapter, we listed a number of problems with the typical implementation of LOD and stated that we would address these in this thesis (see Section 1.3.3, page 11). Let us therefore review the success of our work in terms of these four criteria:

1. **Principled Selection Mechanism** : our model has been developed with close reference to contemporary theories and models of visual perception. This has resulted in a principled mechanism to select level of detail. There is no need for the user to provide any subjective input, or to use any trial and error judgements to find the best LOD switching conditions.

2. **Eliminate Popping Effects** : our model attempts to locate the optimal LOD to use at any point, i.e. the lowest model such that the user cannot perceive any difference between this and the highest LOD. As a result, the user will not experience any noticeable or distracting anomalies when different LODs are selected. The results from the previous chapter confirm this achievement.

3. **General LOD Framework** : our model is general, in that it encompasses all of the principal factors which can affect an object's visibility (i.e. size, eccentricity, and velocity). It is also orthogonal, in that any combination of these factors can be used selectively without compromising the accuracy or function of those parts of the model.

4. **LOD Generation** : we acknowledge that LOD generation techniques are integral to the topic of LOD. We have therefore developed a perceptually driven polygon simplification framework to complement the implementation of a perceptually modulated LOD system. This attempts to ensure that the extent of a simplification is restricted to a prescribed spatial threshold.

6.1.2 Re-implementing our Visibility Model

Using the results from our eccentricity and velocity psychophysical studies, we can now present a re-implementation of our visibility model, i.e. the phase that calculates α_{vis} for an object. Previously, we had to employ an iterative process in order to locate a user's threshold spatial frequency for a moving stimulus (see Section 4.4.4.1, page 123). However, using our empirically derived formula from Equation 5.8 we can now compute this value directly. The resulting frequency can then be scaled by our revised definition for cortical magnification (Equation 5.5) to produce a value for the highest visible spatial frequency at any eccentricity and velocity. This is described by the following formula for visual acuity (c/deg):

$$H(v, E) = G(v) \times M, \tag{6.1}$$

where,

$$G(v) = \begin{cases} 60.0, & \text{when } v \leq 0.825 \\ -27.78 \log_{10}(v) + 57.69, & \text{when } v > 0.825 \end{cases} \tag{6.2}$$

$$M = \begin{cases} 1.0, & \text{when } E \leq 5.79 \\ 7.49/(0.3E + 1)^2, & \text{when } E > 5.79. \end{cases} \tag{6.3}$$

Figure 6.1 presents a 3D graph of the function $H(v, E)$. This spatiotemporal threshold surface shows the visual acuity of a standard observer at any point in their peripheral field and at any angular velocity. This surface has been specifically tailored (by a process of empirical evaluation) to the genus of complex stimuli that occur within computer-generated imagery; rather than the simple, harmonic stimuli used by vision scientists.

The surface provides a description of the general ability of a user to perceive detail. We can apply this to a particular display device by thresholding the spatiotemporal surface with the appropriate value of highest displayable spatial frequency (ξ). To illustrate this process more clearly, the following source code segment presents a simple C++ implementation of the entire phase to calculate the value of α_{vis} for an object.

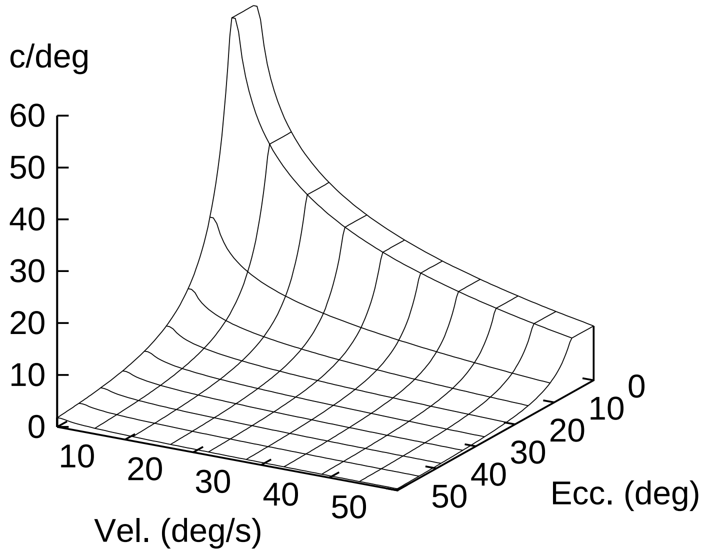

Figure 6.1: *The spatiotemporal threshold surface defined by our new visibility model (Equation 6.1).*

```
// Given values for eccentricity (deg), velocity (deg/s), and
// the display device's highest displayable frequency (c/deg),
// return the highest visible spatial frequency (c/deg).

float spatFreqLimit( float ecc, float vel, float maxDispSF )
{
   float motionSF = 60.0, cortMag = 1.0, result, temp;

   if ( vel > 0.825 )
      motionSF = -27.78 * log10(vel) + 57.69;

   if ( ecc > 5.79 ) {
      temp = ( 0.3 * ecc ) + 1.0;
      cortMag = 7.49 / ( temp * temp );
   }

   result = motionSF * cortMag;
   return ( result < maxDispSF ) ? result : maxDispSF;
}
```

This computation is of course only performed once for each degradable object during a frame. However, if we wished to optimise this routine in order to avoid the potentially expensive log operation and floating point arithmetic, then we could precompute the results of this function beforehand and store these in a look up table.

6.1.3 Contrasting our Design Decisions

Now that we have looked at the details surrounding our particular model, it may be worth discussing some of the major design decisions that were taken and contrast these against any alternative approaches which could have been adopted. This should also help to place our work into context with similar components of other systems.

6.1.3.1 Image versus Geometry for Measuring Detail

The pivotal design decision of our model is that we assess detail within a scene based upon the rendered image of an object, rather than its geometry. This is not a novel approach in itself (e.g. Maciel and Shirley, 1995, used an arbitrary image based metric to assess whether the modulation between two object representations might be perceivable from a certain viewpoint). However, there are a number of implications arising from this decision.

We have already examined many of the pros and cons for both the image based and geometry based approaches (see Section 3.1.1, page 67). We will therefore not replicate that discussion here. Suffice to say that any system which purports to utilise models of visual perception, must avail itself of the actual information which an observer perceives. For example, a purely geometry based approach could not describe the perceived effect of a texture mapped polygon. Also, a collection of adjacent polygons may be perceived as a single visual feature; however a geometry approach would simply use the sizes of the component polygons to decide the visibility threshold and hence underestimate the perceived size of that region.

The major limitation of our image based approach is that it cannot account for changes in environmental conditions (e.g. lighting, fog, etc.) because the spatial frequencies are collected off-line. However, it is worth noting that a geometry based approach would most probably suffer from this same limitation. We therefore conclude that an image based system is the correct approach to adopt for our particular application. Unfortunately, a complete model of perceptual image segmentation does not exist, so we were forced to develop an interim solution based upon colour difference theory. Given the results from the previous chapter, it would appear that this offers an adequate solution.

6.1.3.2 Traditional versus Hierarchical LOD

Most LOD systems operate by storing a collection of independent models for an object and switching between these representations during the simulation. An alternative method is to contain all of the various levels of detail for an object within a single data structure. For example, a quad tree structure can be used such that the deeper you progress through the tree, the more detail is added to the model. This technique is often used to provide continuous levels of detail for terrain databases (e.g. Lindstrom *et al.*, 1996; Economy *et al.*, 1990).

One advantage of hierarchical LOD over traditional techniques is that different parts of an object could be displayed at different levels of detail. From our perspective this would mean that if a large object is displayed on screen, then any parts of the object that are within the user's peripheral field could potentially be degraded whilst retaining high detail for those parts of the object which the user is focussing upon (Sen *et al.*, 1995). This would be most useful in the situation where the VE contains a small number of large objects: a situation in which traditionally implemented eccentricity LOD would not cope well.

There are however certain logistical difficulties associated with hierarchical LOD. Firstly, there may be a large development effort required to implement this technique: not all image generators may provide the ability to easily implement this; whereas most, if not all, renderers support a mechanism to implement traditional LOD (Reddy, 1995b). There is also the question of how to hierarchically decompose an arbitrary polygon mesh, and to do so in such a manner that there are no discontinuities when different parts of the model are displayed at a different detail level. This is a more simple problem for terrain databases because these data are composed of a uniform grid of height fields.

We have decided to adopt the more traditional approach to LOD because of its conceptual and computational simplicity, and also because of its generic applicability to any graphics system. This should make our work accessible to a wider audience, and in particular to users on low end machines.

6.1.3.3 Empirical versus Simulated Contrast Sensitivity

Sen *et al.* (1995) have proposed a model of the human retina in order to assess the contrast sensitivity of a user. This involves applying physiological data on the distribution and receptive field size of neurons in the retina, and then simplifying these to produce a manageable system. This model is then used in order to derive the expected contrast sensitivity for any region of the visual hemisphere.

This represents a fundamentally disparate approach to the one that we have taken, i.e. our contrast sensitivity model is based upon empirical studies of various observers' ability to resolve actual stimuli; whereas Sen *et al.*'s approach is to simulate users' perception using a model of the retina's structure.

There are a number of implied assumptions associated with Sen *et al.*'s approach which need to be resolved. For example, although it is clear that retinal ganglion cells have a bounding effect on the stimuli that we perceive, there are many other stages to perception which also affect the visual data, e.g. the retinal model approach discounts all processing that occurs in the visual pathways and visual cortex.

Having said this, the work of Sen *et al.* is most intriguing, and the author looks forward to discovering how their work progresses. It would be particularly interesting to compare the results from their simulated contrast sensitivity model with that of our empirically derived model.

6.1.4 Extending the Model for Fixed Frame Rate LOD

During our literature review of LOD techniques in Chapter 2, we stated that fixed frame rate LOD (degrading each object's LOD in order to achieve a desired constant update rate) is not directly relevant to our work (Page 38). However, we can suggest a way in which the model that we have developed may be extended to support fixed frame rate LOD in a perceptually principled manner.

If we were to add a global scaling factor, $s \in [0..1]$, to the computation of highest visible spatial frequency (α_{vis}) for all objects, then this would allow us to dynamically vary the magnitude of a user's threshold vision based upon the current system stress. When s is given the value 1.0, then the LOD system will act in the manner that we have already described; however if we choose a lower value of s then the system effectively believes that the user's visibility threshold is lower, and hence it will attempt to reduce the detail of objects that are of a larger size, a lower eccentricity, and a slower velocity.

For example, we could use this approach to produce a simple reactive fixed frame rate system. In this case, if the time for any frame exceeds the desired frame period, then we simply reduce the value of s accordingly. Alternatively, if we find that we have spare compute time after completing a frame, then we can increase the value of s until it reaches 1.0 (there is of course no need to exceed a scaling factor of 1.0). It is important to note that this enables us to develop a fixed frame rate system that will degrade resources in a perceptually linear manner, i.e. the less overloaded the system is (i.e. larger s), then the less perceivable the changes in scene complexity will be, and vice versa.

6.2 Discussion of Perceptually Modulated LOD

Up to this point we have discussed and contrasted various features of the specific model proffered by this thesis. However, there are a number of general considerations which relate to any system that implements perceptually modulated LOD. This section will cover a few of these issues and also undertake some examples to illustrate the potential merits of perceptually modulated LOD.

6.2.1 Considerations for Perceptually Modulated LOD

The following list presents various factors that apply to the general topic of optimising visual detail in a computer graphics system using perceptual criteria. These points are therefore applicable to the system that we have developed, as well as any future systems which may appear.

6.2.1.1 Desktop versus Immersive

A true perceptually-based LOD system should incorporate eye tracking, or at least head tracking, to know where the user is gazing. This means that a VE is optimised based upon the perspective of a single user, i.e. the system can only support one user per display device. This would imply that perceptually modulated LOD would be more suited to an immersive VR system where this situation is more generally true.

6.2.1.2 Rendering Consideration

Some researchers have contemplated using different rendering models to produce various LODs, such as flat-shading or smooth-shading (e.g. Funkhouser and Séquin, 1993). However we should note the existence of a fundamental dichotomy in this case. That is, a smooth-shaded object will generally have a lower spatial frequency content than the corresponding flat-shaded object (because the interpolation removes sharp edges from the image). However, smooth-shading is a more computationally intensive operation than flat-shading (because the surface normal is found for every pixel rather than for the entire surface). We therefore have the conflicting situation where smooth-shading is perceptually less complex, but computationally more expensive (see Figure B.4, Page 202). We therefore suggest that different rendering models should not be used to produce different levels of detail.

Another rendering consideration worth noting is that of antialiasing. The use of antialiasing techniques can improve the perceptual accuracy of an image. For example, they can be used to prevent the scintillating effect which is caused when near sub-pixel features continually appear and disappear from an image. We therefore advocate the use of antialiasing where this does not compromise the performance of a VE (e.g. if it is implemented in hardware).

6.2.1.3 Simplification Considerations

We made the recommendation in Section 4.3.1 (Page 111) that levels of detail for an object should be produced using a simplification algorithm which incorporates some means to limit the spatial extent of a degeneration. We should highlight here that this is not a specific requirement of our model, but a general requirement for perceptually-based LOD. In order to modulate detail using perceptual criteria, we require models with a range of visual as well as geometrical complexities.

One could of course use a standard polygon simplification algorithm with no framework for limiting the visual effects of the degeneration. However this may produce models which do not lend themselves well to selection based upon perceptual criteria; and as a result, the system would be unable to optimise the display as efficiently as might otherwise be possible.

6.2.1.4 Modeling Issues

We are not directly concerned with modeling issues in this thesis (beyond the generation of various LODs). However, we will attempt to bridge two important questions: how many levels of detail should be generated for an object, and at which complexities? These are notoriously difficult questions to answer because they will depend heavily upon the specific platform, the graphics renderer, the complexity and topology of the object, and the application in question. It is highly unlikely that a suitable metric could be developed to predict, *a priori*, the best suite of LODs to use in any situation.

However, we can suggest that any attempt towards a solution would be best achieved using some form of feedback loop; where the LODs for an application are adaptively refined through a pragmatic benchmarking process. We note that this might involve two distinct stages of feedback to the LOD generation phase, in order to improve the perceptual and the computational efficiency of the LOD suite, respectively. For example:

1. Firstly, we wish to ensure that we have LODs over a wide perceptual range; so that there is more chance that an object can be degraded under low visibility conditions. This could be done by comparing the spatial frequency profile of each LOD to see if they are suitably different. (N.B. our LOD generation framework is particularly suited to producing models with a range of perceptual profiles because we can vary the scale at which a reduction is performed.)

2. Secondly, based upon the usage of the LODs in a typical invocation of the application, we can find which models are used the most frequently and attempt to increase the range of LODs around this value. Also, if certain models are never used then we can alter the range of perceptual complexities accordingly.

6.2.2 Assessing the Value of Perceptual Optimisations

Throughout this thesis we have referred to the reduction of detail in an object contingent upon its eccentricity and velocity across a display. However, we have given little consideration to the physical merits of this process, i.e. how fast does an object have to travel, or how far does it have to progress into the periphery, before we can reduce its detail? And are these situations likely to occur regularly as a user navigates through a VE? These questions will be dealt with now.

6.2.2.1 A Desktop Example

To begin, let us take the example of a typical desktop display: 40×30 cm screen size, 1280×1024 pixel resolution, and viewed at a distance of 40 cm. Using Equations 3.2 and 3.3 (Page 77), we can calculate the display's field of view as: 53.1×41.1 deg. Finally, we can use Equation 4.4 to find the highest angular resolution of the display in terms of spatial frequency: 12.0×12.5 c/deg.

We can therefore see that for this example we can never be presented with a stimulus of a spatial frequency greater than 12.5 c/deg. This is about one fifth of the maximum spatial frequency that we can generally resolve (60 c/deg). So what effect does this have on our system?

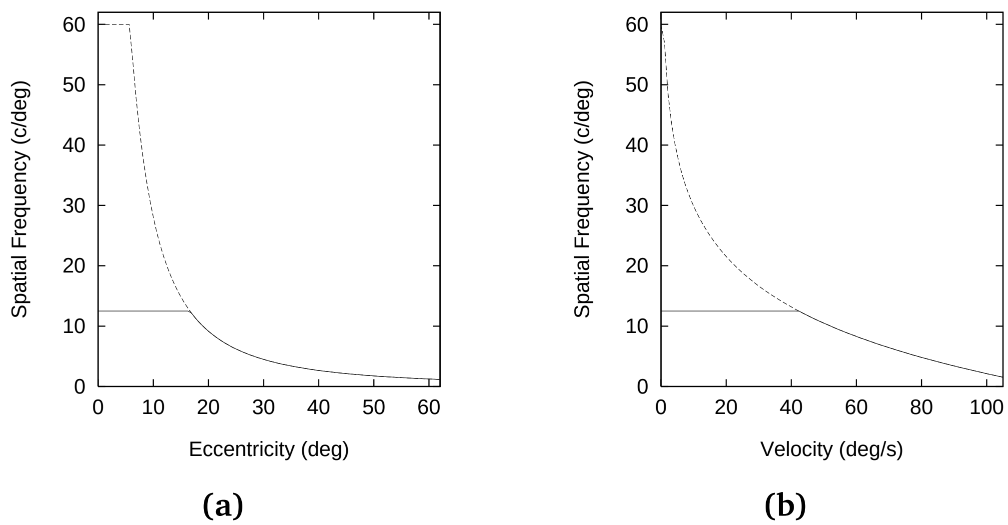

Figure 6.2: *The highest visible spatial frequencies for a user viewing a display where ξ = 12.5 c/deg. These are shown for (a) increasing eccentricity, and (b) increasing velocity. The eccentricity curve (broken line, a) is defined by Equation 5.5, and the velocity curve (broken line, b) is defined by Equation 5.8.*

Figure 6.2 presents graphs of our new models for cortical magnification and velocity. Both of these have been thresholded using a highest displayable spatial frequency of 12.5 c/deg to illustrate the potential stimuli in our desktop example. From these two graphs we can observe firstly that we do not begin to perceive less detail in our peripheral field until an eccentricity of $\sim$17 deg is reached. More striking than this is the result that our effective spatial perception will not degrade until an object exceeds a velocity of $\sim$42 deg/s.

In terms of our display example (which, recall, occupies 53.1 $\times$ 41.1 deg of arc) these results mean that an object would have to be displaced horizontally from the focus point by around one third of the display—or that it would have to travel from the left edge of the display to the right edge in around 1.26 seconds—before the system could even contemplate reducing the detail of an object.

Of course, these are extreme and isolated cases. When we combine the effect of both velocity and eccentricity, then the above figures will be substantially mitigated. For example, at an eccentricity of 10 deg, then an object need only travel above $\sim$13 deg/s before detail could become potentially invisible. It is therefore evident that, taken on their own, eccentricity and velocity based optimisations will probably provide a meager performance advantage; however, if the two are implemented together, then this will produce a synergic speedup.

6.2.2.2 An Immersive Example

We have already suggested that perceptually modulated LOD might be better suited to an immersive VR system rather than a desktop system. Let us therefore investigate the implications of such systems with reference to the previous desktop example.

Table 6.1 provides a cross-section of modern head-mounted displays. For each entry we have calculated the corresponding highest displayable spatial frequencies for that device[1]. These figures should be compared with the 12.0 $\times$ 12.5 c/deg resolution from our desktop example.

Let us examine further the affordances of the i-glasses! HMD. This LCD based unit is indicative of currently popular, cheap HMD systems; both in terms of resolution and FOV. From a casual inspection we can see that this device has a significantly lower angular resolution than our desktop example, almost three

[1]It is interesting to note an obvious technology divide emerging from Table 6.1: the three highest resolution units (MedView, Datavisor, and SIM-EYE) are all based on Cathode Ray Tube (CRT) technology; whereas all the other devices are Liquid Crystal Displays (LCDs).

HMD Model	Field of View	Pixel Resolution	Highest Displ. Spatial Freqs.
CE-100S	22.5 × 16.8 deg	208 × 155	4.6 × 4.6 c/deg
CyberFace 2	140 × 110 deg	385 × 119	1.4 × 0.5 c/deg
CyberMaxx	58 × 42 deg	267 × 225	2.3 × 2.7 c/deg
dVISOR	105 × 41 deg	294 × 141	1.4 × 1.7 c/deg
Datavisor	50 × 37 deg	1280 × 1024	12.8 × 13.8 c/deg
EyeGen3	32 × 24 deg	332 × 493	5.2 × 10.3 c/deg
i-glasses!	30 × 24 deg	263 × 230	4.4 × 4.8 c/deg
MedView	30 × 30 deg	1280 × 960	21.3 × 16.0 c/deg
MRG 2.2	84 × 65 deg	240 × 240	1.4 × 1.8 c/deg
MRG 4	61 × 46 deg	160 × 234	1.3 × 2.5 c/deg
SIM-EYE 60	100 × 60 deg	1280 × 1024	6.4 × 8.5 c/deg
VIM 1000pv	100 × 30 deg	710 × 225	3.6 × 3.8 c/deg
VR4	48 × 36 deg	247 × 230	2.6 × 3.2 c/deg

Table 6.1: *A comparison of the highest displayable spatial frequencies for a number of contemporary HMDs. The figures for field of view and pixel resolution are taken from VR News (1995). The corresponding figures for highest displayable spatial frequency were calculated using Equation 4.4.*

times so (compare 4.4 × 4.8 c/deg with 12.0 × 12.5 c/deg); also, the field of view is considerably smaller than the desktop case (30 × 24 deg versus 53 × 41 deg).

If we repeat the same analyses that we undertook for the desktop example, then we find that for $\xi = 4.8$ c/deg, an eccentricity of 29 deg, or a velocity of 80 deg/s, must be exceeded before it would be possible to degrade objects without the user being able to perceive the modulation (refer to Figure 6.2). As the horizontal FOV for the device is 30 deg, it is therefore apparent that employing only eccentricity LOD will have essentially no benefit. Also, when employing only velocity LOD, then an object must travel across the entire display in under 0.375 seconds before any optimisation can occur. Once again however, if we take these two components together, then the situation is less drastic: e.g. at an eccentricity of 15 deg, then an object need only travel at 24 deg/s before we may begin to reduce LOD. The exact relationship between eccentricity and velocity for this example can be more easily ascertained by examining the surface in Figure 6.3.

There is a further benefit in immersive VR systems because, whenever the user

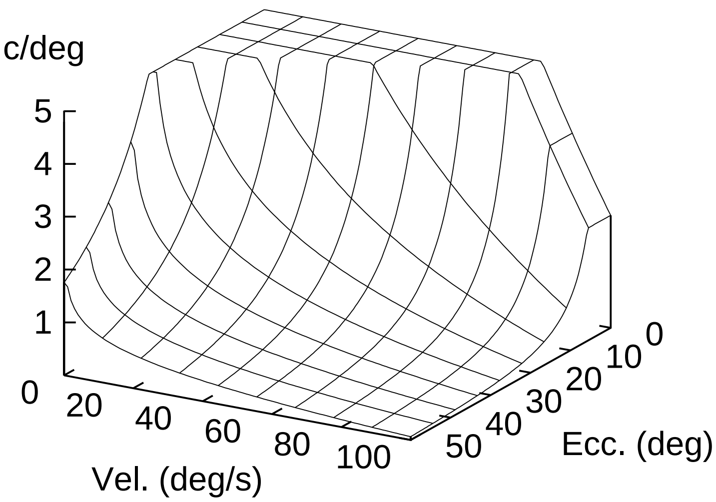

Figure 6.3: *The effective spatiotemporal threshold surface for the i-glasses! HMD, i.e. this represents the graph from Figure 6.1, thresholded with a highest displayable spatial frequency, ξ = 4.8 c/deg.*

moves their head, the whole world effectively moves (in relation to the user's viewpoint). A user's head rotation can often be in excess of around 180 deg/s; thus providing considerable opportunity for most objects to be degraded during exaggerated head movements. This is particularly desirable because many users tend to feel disoriented when they move their heads rapidly and the system does not keep up with their movements (Holloway, 1991). Therefore, by reducing detail during head movements we can improve the update rate of the system and thus make the simulation appear more smooth and interactive. Our model will of course automatically support this feature because we measure the velocity of objects across the display device, i.e. relative to the user's viewpoint.

In summary then, for many of the current generation of LCD HMDs, their lower angular resolution and smaller FOV means that there would be practically no benefit in simply using either eccentricity or velocity optimisations on their own (that is, if the requirement is to be maintained that the user should not be able to perceive the modulation of detail). However, by using the two optimisations in parallel, their contribution can become significant (refer to Figure 6.3). This will be particularly evident in situations when the user makes large head rotations. This latter optimisation could prove a major factor in making immersive VR systems appear less disorienting.

6.3 Discussion of General Perceptual Issues

The following sections will attempt to take a step back and reflect upon the more general issues of applying knowledge from the domain of visual perception to that of computer graphics. This includes fostering an appreciation for the subjective nature of many visual theories, and in particular for the principles upon which this thesis is founded. We will also delve into a number of relevant or interesting topics within the field of visual perception and illustrate how these may apply to our work.

6.3.1 Acknowledging the Extent of Perceptual Knowledge

As the reader has progressed through this thesis, they may have developed a feeling for the incompleteness of our understanding of visual perception. This is certainly true: there is no grand unified theory of perception that conveniently explains all perceived phenomena. This is not a criticism on the part of vision scientists. On the contrary, the discoveries and advancements which have been made in this field are estimable given the sheer complexity of the brain which must be fathomed. To illustrate this, the author Lyall Watson once made the perspicacious comment,

> '*If the brain were so simple that we could understand it, we would be so simple that we couldn't understand it.*'

Even Albert Einstein conceded,

> '*How difficult it is! How much more difficult psychology is than physics.*'

The difficulty of visual perception is beautifully encapsulated by Marr (1982) who develops the analogy of a computer system. Essentially, if we were simply to look at the fluctuation of ones and zeros in a computer memory, could we really decipher the underlying structure and design of a modern computer? This is a commensurate problem to that facing vision scientists who must try to understand perception by only analysing the state of millions of neurons in the retinae and visual cortex.

It is therefore important to appreciate that there are limits to our perceptual knowledge. For example, in the context of this thesis, we know about the contrast sensitivity of the eye, and how this can be measured and modeled; however we do not know how the brain segments the retinal image into component features so that we may accurately apply our contrast sensitivity models to predict the visibility of an arbitrary scene.

Additionally, even if we do understand the underlying principles for a small area of visual perception, we must also appreciate that our visual system is an extremely adaptive and tolerant machine. There will always be a degree of subjectivity when we talk about perception because this faculty is, by definition, a personal perspective and will vary marginally for different people. The following section will illustrate this latter point with specific reference to the subject matter of this thesis.

6.3.2 A Note on Perceptual Subjectivity

We made the point in Section 2.3.4.1 (Page 64) that each individual's visual system will exhibit subtle nuances, and that our contrast sensitivity model therefore only pertains to a standard, averaged observer. In general, the notion of a standard observer is valid because we find that most individuals tend to have a visual performance close to this ideal (the reader is directed to Figures 5.5 and 5.8 for evidence of this).

However, we will take this opportunity to present a list of factors which can affect a user's visual perception, and thus specify the various variables which can introduce a degree of subjectivity into our perceptual algorithms. We can classify these into two distinct classes: environmental (to do with the state of the environment which the user occupies), and individual (to do with the state of the user's specific visual system).

Environmental Considerations :

- **Background Illumination** : the background light intensity which is used to illuminate a stimulus can affect its visibility. Kelly (1975) presents theoretically derived contrast sensitivity functions for an observer under a range of background illuminations. These show a degradation of sensitivity in dim lighting conditions. Following from Lamming's (1991a) exposition, we can attempt to explain this effect from a quantum perspective: under dim conditions, there will be less light quanta entering your eyes and so less opportunity that sufficient information can be gathered to form a detailed image within the integration period of the eye.

- **Light Adaption** : the human eye is sensitive to light intensity over a remarkable range. This is due to the range of photoreceptors in the retina as well as optical factors such as pupil dilation. Sekuler and Blake (1994) offer the example that when entering a dark cinema your light sensitivity can improve by a factor of around 100,000. The level of an observer's light adaption is controlled by the degree of retinal illumination via a feedback system. However, different photoreceptors can take different lengths of time in order to adapt to a new light intensity, e.g. the dark adaption period for rods is around 40 minutes, but only around 10 minutes for cones.

- **Display Factors** : if we are viewing images on a display device then the brightness, contrast, colour, and gamma settings of that display will affect the appearance of any stimuli to which we are exposed.

Individual Considerations :

- **Age** : contrast sensitivity varies as a function of age. For example, an infant's CSF is significantly displaced from an adult's CSF. Consequently we know that infants can only see large, high contrast objects (Banks, 1982).

  Owsley *et al.* (1983) investigated the contrast sensitivity of adults over a range of ages (20 to 80 years). They found that contrast sensitivity degrades notably over this range, such that an 80 year old person will be unable to perceive many of the high spatial frequencies that a 20 year old person could.

- **Colour Vision** : colour is not perceived equally by everyone. Many people suffer from colour blindness of one type or another. For example, Gregory (1990) states that nearly 10% of men have red–green colour deficiency, although this is extremely rare in women. In fact, Hurvich (1981) notes that Caucasian women are 10 times less likely to have colour deficient vision than men.

- **<u>Stereoscopic Vision</u>** : there is a surprisingly large percentage of the population who cannot perceive depth stereoscopically. That is, they cannot perceive three dimensions as a result of the disparity between the images from each eye. It has been estimated that as many as 1 in 10 people suffer from stereoblindness.

- **<u>Lens Aberrations</u>** : defective vision can be caused by an eye's inability to correctly adjust the focal length of its lens in order to project a fixated object exactly onto the retina (a process referred to as **accommodation**). For example, the lens of a person with **myopic** vision will cause light to converge at a point before the retina. This can cause distortions or blurring of an image, often requiring corrective spectacles.

- **<u>Emotional State</u>** : the emotional state of the observer affects the dilation of their pupils. A smaller pupil size implies that less light can reach the retina. This reduction of retinal illumination will cause a drop in the observer's visual acuity (Campbell and Green, 1965).

- **<u>Experience</u>** : Gregory (1990) suggests that our perception of objects may be influenced by *a priori* knowledge and past experience. For example, he speculates that common objects such as oranges and lemons develop a more rich and natural colour once they have been recognised as such.

6.3.3 Other Relevant Perceptual Material

This final section will look at a number of relevant topics within the field of visual perception and discuss how these relate our work. We do not pretend to provide conclusive solutions in all of these cases: each of the selected topics might deserve an entire thesis in their own right. However we will offer tentative solutions were possible, and highlight areas of further research where appropriate.

6.3.3.1 Eye Tracking

We know that if we want to make judgements about a user's perception of a scene, then we need to know what they are looking at. We must therefore use some form of eye tracking technology in order to monitor the user's oculomotor activity. In addition, some form of 6 DOF (degrees of freedom) tracker will normally be required in order to monitor the position and orientation of the user's head, and thus resolve the combination of head and eye movements. Such head trackers are already commonly used in immersive VR systems, so we will concentrate on the area of eye trackers as these are less pervasive in the field of computer graphics.

Modern eye tracking devices generally offer angular resolutions of less than 1 deg of arc, with possible lags in the order of milliseconds. However, many eye tracking technologies are too restrictive to be practically incorporated into a VR system; imposing various constraints such as strict lighting conditions, restrained head movements, or implanted sensors. Nevertheless, some of the more appropriate techniques include the following[2]:

- **Limbus Tracking** : uses phototransistors and infra-red LEDs mounted on an eyeglass frame to monitor the boundary between the iris and the sclera. These sensors are small and cheap, but prone to noise.
- **Image Tracking** : involves training a camera on the eye and using a real-time video processor to determine the position of the pupil (the camera is normally mounted on the head to maximise performance). These systems are fairly accurate but tend to be rather expensive and cumbersome.
- **Electro-oculography (EOG)** : uses electrodes placed beside the eyes to measure the standing potential between the cornea and the retina. Although these systems offer large working ranges (around 170 deg), they are susceptible to noise and drift, and are of questionable worth for accurate gaze tracking.

At the current time, eye tracking technology presents a number of problems as far as the VR researcher is concerned. These include being relatively expensive, susceptible to inaccuracies, or requiring frequent recalibration. However, as the technology matures, these problems will hopefully diminish. Specifically, we suggest that the following criteria should be addressed in order to facilitate the effective integration of eye tracking technology into a VR system:

[2]The interested reader is referred to Ciuffreda and Tannen (1995), and Young and Sheena (1975) for a general study of eye tracking technologies, and Stampe *et al*. (1993) for a more specific discussion of the issues relevant to computer interfaces such as VR.

1. **Interoperability** : the eye tracker must be able to co-exist with the selected 6 DOF tracker(s), and it should be able to function as part of an immersive system, e.g. inside a head-mounted display.

2. **Freedom of Movement** : many eye tracking technologies require that the observer's head be restrained, e.g. using a bite bar. Obviously in a VR system the user should be able to move their head freely and so any eye tracker must be able to operate on a moving subject.

3. **Low Lag** : VR is obviously a time-critical technology. We therefore require an eye tracking technology which can operate at real-time update rates.

4. **Persistence** : the eye tracker should be resilient to drift and not require frequent recalibrations.

It would also be useful to investigate the actual merits of integrating eye tracking technology into a VR system. For example, Watson *et al*. (1995) suggest that for their search task experiment, eye tracking would not offer any appreciable user performance gain over the situation where head tracking alone is used. As a possible explanation of this, Hitchner and McGreevy (1993) note that whenever the user evokes a large change in their point of fixation, there will normally be an associated head movement. As a result their resting gaze will generally relate quite closely to their head orientation. There is some perceptual evidence to support this claim: e.g. Bahill *et al*. (1975) report that most individuals will make a combined head and eye movement to focus on objects which are outside the central 10–15 degrees of their fixation field. It is therefore likely that, for certain applications, eye tracking would be an unnecessary complication; particularly in an immersive system where head tracking is already available.

6.3.3.2 Saccadic Suppression

A **saccade** is a rapid movement of the eye which is made in order to fixate a target onto the fovea (the name comes from the French verb *saccader*, which means 'to jerk'). Saccades can occur at velocities of up to 800 deg/s and last for periods of many milliseconds[3]. The term **saccadic suppression** is used to

[3] A good rule of thumb for the duration (ms) of a saccade is: 20 plus twice the angular distance travelled (Robinson, 1964), e.g. a 10 deg saccade will last about 40 ms.

describe the phenomenon that during a saccade we do not experience blurred vision, even though our eyes are moving at very high velocities; that is, we do not appear to perceive detail during a saccade. Sekuler and Blake (1994) collate a number of the reasons which have been postulated for this effect.

During the months of February and March 1994, there was a discussion on the USENET newsgroup `sci.virtual-worlds` into the possibility of reducing the detail of a VE during a saccade; under the assumption that the visual system might not be able to detect this change. However, it is not clear exactly to what extent we should reduce detail. It is certainly not true that our visual system 'shuts down' during a saccade—if we replace a scene with a black backdrop during a saccade then we do notice this.

Interestingly, Ohshima *et al*. (1996) recently built a system that takes advantage of saccadic suppression. In their approach, the rendering process is simply suspended whilst the angular velocity of a user's gaze movement exceeds 180 deg/s. However, Ohshima *et al*. do not offer any comment on how visually effective this was, or why they chose the value of 180 deg/s. Further work and experimentation is needed in this area.

6.3.3.3 The Blind Spot

The **blind spot** is the area of the retina where all of the axons of the retinal ganglion cells meet to form the optic nerve (see Figure 1.4, Page 15). There are no photoreceptors in this region, so we cannot detect any light which falls on the blind spot. This therefore raises the question: could we reduce the detail of objects that fall onto a user's blind spot?

The angular size of the blind spot is quite large: around 5 deg (Andrews and Campbell, 1991). However, because LOD implies a per-object modulation, we would require all of an object to be projected onto the blind spot before we could degrade its detail (or perhaps even remove it completely from the scene). For an object to be small enough to project completely within the blind spot, it would have to be very small on the display device; in which case, the size-based component of our model will very probably have already reduced the object's LOD. Also, the probability of an object being small enough on the display and also in the correct location to project onto one of your retinae's blind spots will normally be low enough to make it computationally unviable.

Finally, all of this is irrelevant under normal stereoscopic vision anyway because the blind spots for both eyes are in different regions of our visual hemisphere—i.e. we actually have two blind spots. Therefore, any part of a scene which is within one eye's blind spot will always be visible to the other eye (this is why you are always instructed to close one eye when attempting to experience the loss of vision around the blind spot). We should therefore not attempt to reduce the detail of objects when they are within the user's 'blind spot'.

6.3.3.4 Maximum Spatial and Temporal Disparities: D_{max} and T_{max}

The term D_{max} is used to represent the largest spacing between two successively presented stimuli such that an observer will perceive the sequence as a single moving stimulus, rather than two discrete stimuli. The term T_{max} is the temporal equivalent of this spatial phenomenon. These values are therefore important in order to understand how we may preserve the illusion of apparent motion in a computer graphics system.

It is known that D_{max} is a function of stimulus size and density (Eagle and Rogers, 1991), and that this value increases with retinal eccentricity. However, little work has been done to assess the values of D_{max} and T_{max} for complex, real-world images: most of the work in this field has been restricted to simple random dot displays. In our model we therefore make no quantitative attempt to measure D_{max} and T_{max} for an object and use these values to ensure the perceptually continuous presentation of the object. However, qualitatively, the act of improving frame rate by reducing detail will increase the likelihood that these thresholds are not exceeded, and therefore that object motion is perceived less discontinuously.

6.3.3.5 Hyperacuity

The term **hyperacuity** is used to describe the paradoxical phenomenon that certain stimuli can be perceived which are smaller than the size of a single photoreceptor cell. Photoreceptors in the fovea subtend around 25–30 sec of arc (which corresponds to a maximum spatial frequency of around 60 c/deg). However, it has been shown that it is possible to discriminate the non co-linearity of two thick abutting lines to a resolution of 2–5 sec of arc (this is referred to as **vernier acuity**). We must therefore ask ourselves: are we being over simplistic by only using contrast sensitivity to measure perceived detail?

Hyperacuity is thought to be caused by differences in the mean distribution of light sampled over a number of photoreceptors (Morgan, 1991). The effect is therefore dependent upon the large spatial spread over which two adjacent features extends, i.e. any isolated feature which is smaller than a single receptor will still remain undetectable. We can therefore see that hyperacuity merely gives us a higher positional accuracy between adjacent features (discrimination), but it does not increase the fundamental resolution limit of our visual system (detection).

Our model is concerned with the *detection* of features—i.e. deciding whether an entire feature is no longer resolvable by a user—and is therefore unaffected by hyperacuity. Also, hyperacuity is largely confined to low velocity targets at the fovea. For example, Levi *et al*. (1985) and Schor and Badcock (1985) report that hyperacuity performance degrades markedly with eccentricity, more so than contrast sensitivity. In conclusion then, hyperacuity is an interesting curiosity, but not one which should affect the accuracy of our current contrast sensitivity based model.

6.3.3.6 Temporal Antialiasing

Just as spatial antialiasing attempts to smooth computer images over space, temporal antialiasing attempts to smooth images over time, i.e. to blur objects as a function of their velocity. This is done in order to simulate the effect that we perceive as fast moving objects race past our point of fixation. It would be fair to say that this a more technological rather than perceptual topic, however it does raise the question: what are the perceptual parallels between temporal antialiasing and the temporal optimisations effected by our model?

The most important distinction to make is that temporal antialiasing techniques attempt to *add* cues to an image in order to mimic characteristics of the visual system; whereas our velocity LOD model attempts to *remove* detail from a scene by taking advantage of limitations in the visual system. The similarity between the two techniques is that they both remove high spatial frequency components within an object; although temporal antialiasing does this by blurring the image of an object, and velocity LOD does this by using less complex geometry for the object.

Historically, most temporal antialiasing techniques have been developed for non real-time applications (such as ray tracing) and are therefore too computationally expensive for use in a real-time VR system (Williams, 1993); however, a number of real-time techniques are beginning to emerge. The following list presents a few of the techniques that have been developed thus far:

- **Super-Sampling** : for each frame, multiple images are generated over time and then merged together using a simple weighted sum averaging algorithm, e.g. a box filter (Korein and Badler, 1983).
- **Distributed Ray Tracing** : the various reflected and transmitted rays of a ray tracer are distributed in time as well as space, taking into account any changes in visibility and shading (Cook *et al.*, 1984).
- **Fourier Techniques** : a Fast Fourier Transform (FFT) is performed on the image and then convolved with a Point-Spread Function (PSF) which describes the image path of the projected motion. The inverse FFT is then calculated to produce the blurred image (Potmesil and Chakravarty, 1983).
- **Practically Frameless Rendering** : only a fraction (e.g. a quarter) of all pixels are updated each frame. By distributing these pixels uniformly, the entire image appears to update progressively over a number of frames; producing a 'dissolving' effect. Wloka *et al.* (1995) proffer this as a crude motion blur technique.

In general, temporal antialiasing is a desirable feature because it would allow rapidly moving objects to be perceived more faithfully (e.g. failing rain would appear as streaks rather than individual drops). However, as long as the implementation of such features could degrade the performance of a VR system, then they must be considered contrary to the goals of this thesis.

6.4 Summary

We have covered a wide range of topics in this chapter; with our perspective ranging from that of a computer scientist, to that of a vision scientist. As a result, we hope that we have addressed many of the questions which may have occured to the reader; and hopefully alerted them to further issues and considerations in the process.

We began our discussion with an examination of our model for perceptually modulated LOD. This was done in reference to the four major problems that were identified at the start of this thesis. We found that our model provides satisfactory solutions to all of these problems. Following from this we illustrated how the results from the various psychophysical studies of Chapter 5 could be used to produce a revised and more efficient model for highest visible spatial frequency (α_{vis}). We then discussed and contrasted a number of the more influential design decisions behind our model. Finally, we showed that our model could be easily extended to support fixed frame rate LOD. This generalises our model further such that it encompasses all forms of LOD modulation which are currently employed in real-time computer graphics systems.

Our next perspective was to look at the general issues facing a perceptually modulated LOD system. We noted that such a system would be more applicable to an immersive VR system; also that on the topic of level of detail, it would be inadvisable to mix various rendering models; and that when producing LOD, some mechanism should be employed to limit the spatial extent of a simplification. We also developed a couple of examples to illustrate the actual effects of a perceptually modulated LOD system in a real-world scenario. This showed that using either eccentricity or velocity optimisations on their own offers little advantage. However when used together, they represent a valuable tool for reducing detail.

In the final portion of this chapter we were concerned with looking at our model from a vision scientist's perspective. We drew the readers attention to the inherently subjective nature of perceptual theories and illustrated this in terms of our work by describing many of the factors affecting a user's visual perception. To complete our discussion we presented a cross-section of other relevant areas within the field of visual perception and illustrated how these relate to the topic of this thesis.

Upon reaching this stage, we have completed the main body of our work. We have described the background work and theories of our area, developed suitable tools to aid us in our task, produced a model and a subsequent implementation for perceptually modulated LOD, built a prototype system and evaluated its efficacy, and finally we have instigated a thorough discussion on the issues and implications of such systems. We are now in a position to draw conclusions on the merit and validity of this research programme.