Important Notice

"Audentis Fortunas Iuvat"

The basis for all that follows.

UNIVERSITY OF CALGARY

More than Meets the Eye – A Study in Seismic Visualization

by

Steven Lynch

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF GEOSCIENCE

CALGARY, ALBERTA

MAY, 2008

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "More than Meets the Eye – A Study in Seismic Visualization" submitted by Steven Lynch in partial fulfilment of the requirements of the degree of Doctor of Philosophy.

Supervisor, Dr. Laurence R. Lines, Department of Geoscience

Dr. Robert R. Stewart, Department of Geoscience

Dr. Donald C. Lawton, Department of Geoscience

Dr. John C. Bancroft, Department of Geoscience

Dr. Brij Maini, Department of Chemical and Petroleum Engineering

---

External Examiner, Dr. Mathew Yedlin, University of British Columbia, Department of Geophysics

---

Date

Abstract

This thesis is primarily concerned with examining the properties of SeisScape displays, which render seismic data as a three-dimensional surface. SeisScape displays are fundamentally different from conventional seismic displays in that they fully engage the visual system and produce sensations of perception. These perceptions are the goal of scientific visualization. Visualization itself is placed into context with respect to seismic data by discussing how the display acts as a filter upon seismic resolution. There are two levels of seismic resolution; absolute resolution, which is a product of spatial and temporal resolution; and apparent resolution, which is a product of the display. It is established that the apparent resolution of conventional displays is significantly lower than the absolute resolution of the data.

The primate visual system is the second, immutable, stage of the seismic display filter. It is not, however, a general purpose tool. To learn how to use it appropriately, the evolution and properties of the primate visual system are discussed in the context of determining how primates establish their perceptions of form and color.

Two terms that describe the structure of a seismic section are introduced. The first is macrostructure, which is the collection of strong amplitude events that are visible on any seismic section. The second is the microstructure, which is the collection of weak amplitude events that are often only observed as perturbations upon the macrostructure. Several techniques for tessellating the seismic surface are developed. Examples are presented to illustrate the effect that tessellation has on the ability to perceive both macrostructure and microstructure. Various techniques are developed to calculate the reflectance of the seismic surface and examples show how reflectance is primarily responsible for our ability to perceive microstructure.

The use of color on seismic data is examined from the perspective of the evolution of primate trichromacy. Conventional color palettes, which were developed for use in a perception free environment, are shown to be inappropriate for use on SeisScape displays. Two new color palettes, HA1 and HA2, are developed and examples show that they are more appropriate for use in a perceptive environment.

Preface

This report, by its very length, defends itself against the risk of being read.
Winston Churchill

In the late fall of 1999, my son convinced me to experiment with displaying seismic data as a three-dimensional surface. His motives were hardly scientific. He was fourteen and all he wanted was a new graphic card. This, of course, was so he could kill even more aliens in even more gruesome ways. His reasoning was that I would have to buy a new graphic card and if I got one, it was only fair that he got one as well. Crafty little man, he was right on with his logic and it worked out for both of us. He got his graphic card and I got this thesis.

At the time, it was a toss-up which of us cared less about seismic visualization. He had no idea what seismic data was or what it was used for. This put his interest at around zero. I, on the other hand, had worked with seismic for decades and had every idea what it was used for. This put my interest at … somewhere around zero!

Seismic data looked like seismic data. There was nothing else to see but what we could see. Graphic cards were for games and games were for children. With attitudes like that, it was a miracle that I developed SeisScape displays in the first place. I did not have any insight into what they would look like and I expected absolutely nothing from them. But that all changed as soon as I saw the very first display.

I produced my first SeisScape display in December of 1999 and from the very first moment that I saw it, I was taken by two things. The first was that SeisScape displays are ethereal. I lucked out and my very first image was one of the most beautiful things that I have ever seen. It was almost mesmerizing and it took me completely by surprise. The conventional displays that I had looked at for years had no such effect. They were cold, dispassionate and if I expected one thing from SeisScape displays, it would have been that I expected them to have the same dispassionate nature.

They did not though. They were the exact opposite and they engaged me on a level that seemed strange and inappropriate. If I had to describe that first experience one way, I would say that the display was speaking to me but it was saying things that I did not understand.

The second thing, I was taken with, was that I could see an unexpected amount of detail. I had looked at thousands of kilometers of seismic and I never realized that I was missing so much. My first display was of a very small section of data over a reef and I had used it as a test case for over a decade. I had looked at it thousands of times and I thought that I knew every sample by heart. Yet, when I looked at it as a SeisScape display, I immediately began to see things that I never suspected were there. I saw new features and new alignments and even though most of them were just noise, they were so obvious that I began to wonder why I had never seen them before.

The title of this thesis is “More than meets the eye”. This refers back to that first moment when I began to realize that there was more to seismic data and to visualization than I ever suspected. First and foremost, this thesis is about seismic resolution. It is about the effect that any display has upon resolution. It is about the resolution that we never suspected was there because our displays filtered it out. It is about discovering and quantifying the sciences of visualization and learning how to use them to get it all back.

Beyond that, however, this thesis is also about discovering why my early displays had such an emotional impact. The first SeisScape displays were so engaging that they were almost uninterpretable. Everyone loved them but no one could use them. It is about discovering the nature of art. It is about learning how to separate art from science and how to keep each in its appropriate place. You would think that would be easy, wouldn’t you? I certainly thought so when I started. I thought that the answer was just beyond my reach … just wait until you see where I found it and what I learnt along the way.

In January 1978, I sustained a serious head injury in a soccer game and its misdiagnosed consequences slowly, but inexorably, eroded my life. It destroyed me physically, it destroyed me emotionally, and it destroyed me mentally. It destroyed everything about me until in January 1990, its aftermath would come within hours of destroying me completely. By January 1991, when I reached my lowest point and started the long process of recovery, all of the promise that I held in 1977, had drained away and all that was left was an empty shell.

It was always my intention that when I had completely finished this work to go back, and re-examine everything that happened both on the way down and on the way up. To that end, for seven straight days, I lived through it all again. I wrote it all down, I edited it and then I proofread it. I intended to include it as an epilogue but once I had finished it completely, I realized that I could never show it to anyone. I decided that as long as people could read it, they would always ask questions and the questions would keep it alive.

By the time I saw the first SeisScape display, I had long since recovered physically and emotionally. However, the surgeon who repaired me warned that I would never get all the way back. By 1999, I knew what he meant. Mentally, I had only come back part of the way. What came back only came back by default and I knew that I had to earn the rest of it. I knew that the only way was to take on a challenge that was beyond me. I would have to push far beyond my limits.

As you read this, you may become aware that it is harder (and longer) than it needs to be. If you do, then you get the point. This was never about the degree; it was only ever about the difficulty of earning the degree. It may be harder than it needs to be for you but rest assured it is as hard as it needed to be for me. This was the only way that I could get completely back and getting completely back was the only reason why I did it. It was simply the hardest thing that I could think of to do at the time. If I could have thought of something harder, however, I would have done that instead.

As you can imagine, starting a degree at the age of 50 presents some unique challenges. Probably the greatest was scraping the barnacles from my code-encrusted brain. Love is wasted on the young. So, apparently, is memory because mine did not work anymore. I saw this as a potential problem and so, in an effort to restore it to its former glory, I went back to doing something that I did in my teens; I started to memorize poetry.

I have always loved poetry and I used to know dozens of poems by heart. It was only natural that I should go back and relearn the ones that I used to know. Unfortunately, I had an ill-spent youth and most of the poems that I knew were Rugby songs. Whereas my teenagers enjoyed them, for some strange reason my wife did not consider reciting "The Ballad of Eskimo Nell" to be appropriate at the dinner table. I switched to memorizing Kipling instead and my memory (and marital bliss) were somewhat restored.

You will find excerpts from some of the poems I learned scattered throughout what follows. I put them in to serve as a welcome respite from the dispassionate nature of the text. They are always in context to what immediately follows. I admit that sometimes the context is a little thin but it is always there if you look hard enough.

Unless otherwise noted, all of the poems are credited to Rudyard Kipling.

Acknowledgements

Starting a PhD at my age is definitely not for the faint of heart. In my case, that should read “not for the faint of hearts” because I have been married for almost 25 years and I have another heart to consider. Given how long we have been together, a project such as this could never be an individual effort because both of us had to sacrifice. Both of us had to understand why we were doing it and both of us had to be prepared to see it through. This has been, from start to finish, a lifetime experience and I would recommend it to anyone. I would caution you, however, not to attempt it yourself unless you are completely sure of the heart you share your life with.

In my case, there was never a doubt. I knew, before I even suggested it, that my wife would support me. I knew she would understand why I had to do it and I knew that she would be an active and willing part of it. And she was. She served as my editor in chief (*you are not going to leave that in are you?*). She served as my spell checker (*making up new words as needs be*). She served as my style editor (*are you going to rewrite this in English?*). She served as a sounding board when I was stuck on a concept (*you have been doing this for five years and that is the best you can come up with?*). She served as my main cheerleader (*but she refused to wear the skimpy outfit*). This was always a joint project and with all my heart, I can say that it is as much hers as it is mine. So, if you find any bits you don’t like (*they’re the bits she wrote*).

I am also deeply indebted to my children, Sean and Katherine, and I cannot say enough about what they have meant to me. What I can say is that without their help, I would never have gotten through this. Sometimes it was a challenge to keep going and sometimes my enthusiasm started to wane. When that happened, all I had to do was consider their spending habits and it gave me all the motivation any man could ever want.

I also owe a debt of thanks to my Mother who has been my lifelong friend and confidant. When I first thought of doing this, I discussed it with her at length. It was such a strange idea, to go back to university when I had been out for so long, that I wondered if I had lost my mind. I think if anyone had brought me back to sanity, I probably would never have started. I gave my mother the chance to do that but she never took it. Instead,

she saw this as my dream and she encouraged me to pursue it, regardless of how long it took. She told me to ignore all those voices that said I had lost touch with reality. “Don’t worry about them”, she said, “you were never in touch with reality to begin with, why change now.” Thanks Mom, you were a big help!

As much as you need support at home, you also need support at the office. I was very fortunate there because before I started, I discussed it with the CEO of Divestco, Stephen Popadynetz. He was great. He told me that the company would fully support me and that I could have all the time I needed. He was so supportive that he insisted I tell everyone I would be spending more time away from the office. “That”, he said, “will help us with employee retention.”

And where would I be without my supervisor, Dr. Larry Lines. We have been friends since graduate school (the 1975 rendition) and he was instrumental in convincing me that this was a good idea. Over the past five years, my association with him has brought many rewards, not the least of which is he cured me of a lifelong problem - I never learned to like Scotch. I am not sure where I would be without him, but I am pretty sure my head wouldn’t hurt quite as much!

On a serious note (and they are about to come a lot more often), this would definitely not have been possible if it had not been for two doctors. The first is Dr. R.C. Westbury, who kept me alive long enough to meet the “smart” doctors. The second is Dr. Wayne Tunis, the surgeon who turned out to be the smartest doctor of them all (*you were right on the one, it didn’t help me at all. But I am going to prove you wrong on the other*).

Finally, for being my only friends during the dark years, I am deeply indebted to “the boys”. I am sorry I am breaking my promise but I think you know why I have to.

Dedication

To my Grandfathers

Andrew Lynch

Thomas Hopley

Both of whom died, 7:00am, July 1st 1916, on the banks of the river Somme.

And 50 years later told me why.

To my Father

George Lynch

Who died hours before I could ask him the only question that ever matters.

Table of Contents

List of Tables

List of Figures

Epigraph

This is the ballad of Boh Da Thone,
Erst a Pretender to Theebaw's throne,
Who harried the district of Alalone:
How he met with his fate and the V.P.P.[1]
At the hands of Harenda Mukerji,
Senior Gomashta, G.B.T.[2]

"The Ballad of Boh Da Thone"

[1] Value Payable Post = Collect on Delivery.

[2] Head Clerk, Government Bullock Train.

What is the moral? Who rides may read.
When the night is thick and the tracks are blind
A friend at a pinch is a friend indeed,
But a fool to wait for the laggard behind.
Down to Gehenna or up to the Throne,
He travels the fastest who travels alone.
“The Winners”

Theme

The picture on the right is of the Chapel Street Junior School's first eleven (1963 edition). The school is in Hazel Grove, which is about 15 miles south of Manchester. I am the goalkeeper, back row center; you can't miss me I am the one in the wool sweater, the perfect attire for playing in the infamous Manchester rains.

1964 Chapel Street First Eleven. Author is the goalkeeper, back row center.

When this picture was taken I was 11 and had already been a goalkeeper for several years. I played my first game in 1961 at the age of 8. I remember it well because it was the last time I wasn't injured. I was the stand in keeper for my local Cub Scout team: we lost 12-0 and I let in all of them! Never one to be easily deterred I continued playing for the next 40 years and am proud to say that my first game was my worst ever defeat.

Between that first game and my last in the summer of 2001, I accumulated an impressive series of breaks, sprains, tears and concussions. Even my first date with my future wife was punctuated by a quick trip to an emergency room. I invited her to watch me play in an indoor game during which I broke my wrist. Most of the injuries that I accumulated were minor but several were more serious and one proved cataclysmic. By the time I finally retired, I could proudly claim that I had been treated in nearly all of the emergency rooms in Southern Ontario and Western Canada.

My final game was in Penticton B.C. and for soon to be obvious reasons I don't remember the score. It wasn't meant to be my final game but towards the end of the first half I continued my habit of stopping close range volleys with my face. I did not loose consciousness (for once) but I was badly stunned and had to leave the game at the half.

I drove home by myself, which, in my semi-conscious state, gave me lots of time to think. By the time I got home it had all finally all come clear and as I walked into the house I was struck by a revelation of tectonic proportions. I decided to reveal this great new understanding to my wife so I called out to her. "Jan", I said, "I have just realized something; this goalkeeping thing, it's dangerous - I could get seriously hurt".

And with that revelation, I promptly retired and I haven't played since.

The moral of the story (and the theme of this thesis) is that if you hit a man in the head often enough, hard enough and for long enough - he will eventually come to see even those things that are blatantly obvious!

Part I: Observation

Initial Observations and Conclusions

You think you had the last laugh
Now you know this can't be true
Even though the sun shines down upon you now
Sometimes you must feel blue
You make the best of each new day
You try not to be sad
Even though the sky falls down upon you
Call it midnight feelin' bad

When you wake up to the promise
Of your dream world comin' true
With one less friend to call on
Was it someone that I knew
Away you will go sailin'
In a race among the ruins
If you plan to face tomorrow
Do it soon
"Race Among the Ruins"
Gorden Lightfoot

Imagine if you lived in a world where geophysicists did not understood the theory of migration. A world where seismic processing was as sophisticated as it is today but with one exception – everyone believes that unmigrated seismic data is the state of the art. Now, imagine that on one of your sections, someone accidentally used a migration algorithm developed for an entirely different field of study. If that happened, what would you think? Since you do not know the theory behind migration (or even how waves propagate through the earth) would you immediately recognize that the migrated sections were better than the original ones? Would you instantly grasp what made them better, would you immediately understand the relationship between migration and resolution; or would you simply be confused?

I vote for confused because digital signal processing, the area of exploration geophysics that I am most familiar with, does not work that way. We rarely get the observations before we understand the theory and consequently practice never comes about by accident. Digital signal processing is primarily a theoretical science; theory leads to practice, which is tested by observation. The observation is empirical but we always begin with theory.

Nevertheless, in this thesis, observation came long before theory. It began with a series of observations brought about by the accidental use of someone else's technology. I observed things but did not understand what they were or the theories and technologies that my observations were based upon. In my case, accidental observation led to the discovery of theory, which in turn led to practice; and that is how the three main parts of this thesis are organized – Part 1: Observation, Part 2: Theory, Part 3: Practice.

Part 1, which follows, contains four chapters, each of which details one lesson that arose from my initial, accidental observations. The objective of Part 1 is to produce a conceptual understanding of what visualization is and to place the science of visualization into overall context with respect to seismic data.

CHAPTER ONE: INTRODUCTION

When the flush of a new-born sun fell first on Eden's green and gold,
Our father Adam sat under the Tree and scratched with a stick in the mould;
And the first rude sketch that the world had seen was joy to his mighty heart,
Till the Devil whispered behind the leaves "It's pretty, but is it Art?"
The Conundrum of the Workshops
Kipling

1.1 Introduction

Serendipity:

The process of finding what you weren't looking for.

Science:

The process of discovering what you didn't want to find.

Visualization:

The science of communicating information via the visual processing system.

This thesis is ostensibly about a new type of seismic display called a SeisScape[1] display. The SeisScape display, which I initially developed in the fall of 1999, considers that a seismic section is a three-dimensional surface and displays it as such complete with lighting effects. The three dimensions are (1) the spatial and (2) the temporal position of a sample, which form the mosaic of the surface, and (3) the amplitude of the sample, which provides the topography. A SeisScape display is thus fundamentally different from the conventional seismic displays, the wiggle trace display and the variable density display, both of which is strictly two-dimensional and has no concept of lighting or illumination.

From the outset, I want to make it clear that I do not claim to be the first person to look at seismic data as a three-dimensional surface. I know, from personal

[1] The term SeisScape is a registered trademark, originally of BirchTree Software and now of Divestco Inc.

communications, that other people before me had attempted to produce 3D views of seismic data. In all the cases I am familiar with, however, the people involved used software designed for other purposes and the sheer volume of data in a seismic section made the resultant displays too cumbersome to be useful. To the best of my knowledge, my development of SeisScape displays was the first time anyone had specifically designed software for the express purpose of viewing three-dimensional images of seismic data.

Figure 1.4 is the first SeisScape display that I produced and as soon as I saw it, I became aware of two things. The first and most obvious was that seismic data, displayed in this way, was surprisingly ethereal. This first display and the others that soon followed were the epitome of "beautiful science" and could be, if not produced carefully, mesmerizing. The second thing I became aware of almost as quickly was that somewhere behind the displays lay a science but a science that was frustratingly hard to discover. As much as this thesis is based upon the SeisScape display, it is really about the discovery of what that science is and how to use it correctly.

The primary purpose of this thesis is to discover the underlying principles of visualization and use them to improve our ability to communicate seismic information. I split it into three principal sections, (1) observation; (2) theory; and (3) practice and in that sense; it is similar to most other theses. Beyond that, however, it differs from the norm in two significant ways. The first is that this is a cross-disciplinary thesis because the sciences that dominate the last two sections are vastly different. As one might expect, the practice section remains true to its roots and is dominated by the sciences of geophysics and computer graphics. Nevertheless, whereas this thesis started out in geophysics, the theory section ultimately took me back to my undergraduate roots in biophysics because it became dominated by the sciences of psychophysics, evolutionary biology and the physiology of vision.

The second way in which this thesis deviates from the norm is that you find the most important conclusions of this work here, in this introductory chapter. It is those conclusions, which I state in 1.4.1, which lead to the cross-disciplinary nature of this

work. In addition, accepting and explaining those conclusions ultimately became the most challenging aspect of the entire thesis.

To the end of developing those conclusions, I split this introductory chapter into three parts, each of which has its appropriate definition at the beginning of the chapter. The first part discusses how I became interested in visualization and why I created the first SeisScape display. The second part discusses how I came to accept that although the displays were an improvement on conventional displays they were fundamentally flawed. The final part introduces the science of visualization, one I had to discover for myself, and sets the stage for how I would go about correcting the flaws.

1.2 The Background to SeisScape

This thesis dates back to the late the fall of 1999. At the time I was working out of a home office in Kelowna B.C. and regularly commuting back and forth to Calgary. I was working on three programs; the first was a structural modeling program called Outrider that I started work on in 1988; the second was a stratigraphic modeling program called InterpaLog and the third, was a geological cross-sectioning program called CrossLog (all three are still around at the time of writing). I had little, if any interest, in visualization.

This would all change due to a short conversation I had with my then 14-year-old son, Sean. He was very much into computer gaming and wanted me to buy him a new $249 graphic card. My initial reaction was to say no because I had purchased him a new card just months before. But that reaction changed when he told me that the card he wanted, an NVIDIA TNT2 Ultra (Figure 1.1), was rated at one gigaflop[2] (GF). With my background in seismic processing, the term gigaflop got my attention. Before he

Figure 1.1: An NVIDIA RIVA TNT2 chip, one of the earliest graphic processing units.

[2] A gigaflop is a measure of floating point performance, and refers to one billion floating point operations per second.

mentioned it I had no interest in visualization, afterwards I did.

To explain why I became suddenly interested I have to go back as far as 1986. At that time, I was working for Western Geophysical in Calgary. During the winter processing season we purchased four Star array processors that had a combined speed rating of the same one gigaflop but they cost us over five million dollars. Finding out that this same processing power was, in 1999, available for $249 got my attention. What also got my attention was that this explosion in price-performance ratio came from a very much unexpected source. Whereas the array processors that had driven the early development of processing power were developed for work, the graphic card which has been driving it since the mid-90's was developed for play. Before our discussion, I did not know of the revolution that was and still is taking place in the field of computer graphics. With my science and industry background, computers were tools of my trade and I assumed that they were developed for me in that role. However, I was wrong, from the mid-90's onwards, the most significant developments in computers occurred on the graphic card and those developments were driven by the gaming industry.

By chance, our discussion took place in my office during a time when I was developing an optimized wiggle trace seismic display (see Figure 1.2). This display was on my monitor during our conversation and this led us into a discussion on the nature of seismic and of the limitations of conventional seismic displays. It was Sean who first suggested that since seismic, the way I explained it, was actually a three-dimensional surface, that I try displaying it as such. As he told me, that is what the new graphic cards were all about – displaying large amounts of data in 3D, so why not take advantage of it?

My initial reaction was to believe that the sheer volume of data in a seismic line would be prohibitive and that the graphic card would be overwhelmed. He did get me thinking though and after researching it, I became convinced that the existing cards were not powerful enough to display large amounts of seismic data but they soon would be. I did not know what three-dimensional seismic displays would look like but I became intrigued by the possibility of using this new technology. As a result, over the next few

months I experimented with the concept of displaying seismic in 3D, with Figure 1.4 being the first result.

1.2.1 What Does Seismic Data Look Like?

It took me about three months of part-time effort to learn the basics of graphics programming and to produce the very first SeisScape display. During the period of this initial development, I had two questions in my mind. The first question was, "given the shear volume of samples in a seismic section, would it even be feasible to view it in 3D?" This was a purely technical question and the answer would be technology driven. However, the second question "what would it look like?" was more intriguing and the answer would prove to be more complex and elusive.

By the time I started the SeisScape project I had already looked at thousands of kilometers of seismic data using conventional seismic displays and I was intimately familiar with what seismic looked like, or so I thought. In 1999, I thought I knew what seismic data looked like but as I would soon learn, I only knew what seismic data looked like using archaic technology.

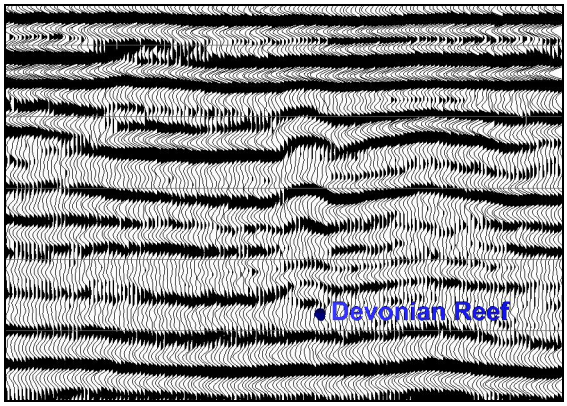


Figure 1.2: A conventional wiggle trace display of a small seismic section over a Devonian reef. Each wiggle represents the energy reflecting from a single common depth point. The deflection of the wiggle is proportional to the amplitude of signal. Positive amplitudes are displayed filled and to the right, negative amplitudes are displayed unfilled and to the left.

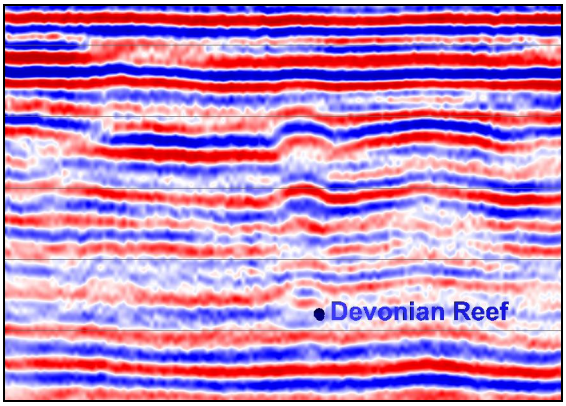


Figure 1.3: A conventional variable density display of a small seismic section over a Devonian reef. I plot the amplitudes of the seismic data in color. I display positive amplitudes in red, negative amplitudes in blue and zero amplitude in white.

The concept of using reflection seismology to infer geology is certainly not new. Reginald Fessenden first conceived it during World War 1 as an offshoot of his work on submarine detection. John Clarence Karcher, however, was the first to propose its use in the exploration for hydrocarbons in 1920. He was instrumental in developing the first practical acquisition system in the late 1920's and since that time, reflection seismology has been the principal tool used in the exploration for oil and gas. Thousands of explorationists have been looking at seismic data since that time and if you ask any explorationist today what seismic looks like, they would, in all probability, point to either Figure 1.2 or Figure 1.3 and say "it looks like that" and in 1999 I would have said the same thing.

However, is that answer correct? Is that what seismic really looks like? Consider Figure 1.2 and Figure 1.3, which are the "conventional" seismic displays. The seismic data that they show is a small section of a survey over a Devonian reef. The data is courtesy of a company that no longer exists and the reason I show it is that it was the first section that I showed as a SeisScape display. Historically, we did not develop these conventional displays because they were the best or most appropriate way to display seismic data; we developed them because they were the best way to display seismic data given the available technology. We are familiar with the displays because we have used them for decades but here is the point, they only take advantage of outmoded and archaic technology. In the case of wiggle trace displays that technology dates as far back as the 1930's. The technology behind the variable density display is more modern, it dates to the late 1970's, but it is still 20 years out of date.

Figure 1.3 is based upon more modern technology than Figure 1.2 and consequently it looks very different. This opens the possibility that if we use modern graphics technology that it will look very different again. In reality, seismic data forms a three-dimensional surface but neither conventional display conveys any hint of it and as a result, they cannot be showing seismic in its true form. Modern graphics technology, however, can show it any way we want, including in its "native" three-dimensional form.

With that in mind then my original simple question of “what does seismic look like” evolved into three, those being:

1) What does seismic data look like in its true form?

2) How different will it look from what we are used to seeing?

3) Will there be anything to see that we could not see before?

I answered the first two questions as soon as I produced Figure 1.4. Answering the third takes up a large part of this thesis.

1.2.2 The First SeisScape Display

Figure 1.4 shows the results of my initial efforts to display seismic data as a three-dimensional surface. In this display, I show positive seismic amplitudes as peaks and negative amplitudes as troughs. I used a simple color palette with negative amplitudes displayed in various shades of blue, zero amplitude as white and positive amplitudes as being shades of red. Because of its similarity to a topographical or landscape display, I initially called it a SeisScape display, a name that eventually stuck.

Usually when building a new object in software the results come together in pieces. That was not the case here. I did not know very much about graphics programming at the time and I struggled to make anything appear on the screen at all. All of my initial displays were simply black; I could not see anything at all. Trying to figure out why forced me to correct all of the small problems in the software. As a result, once I finally understood that I had to turn on some lights before I could see anything, Figure 1.4 appeared almost exactly as it does in the image. This is important because this initial image had a profound impact upon me. Before I could see it, I was literally in the dark about what it would look like. I had no idea if it would even make sense visually and I had no idea if it would ever be of any use.

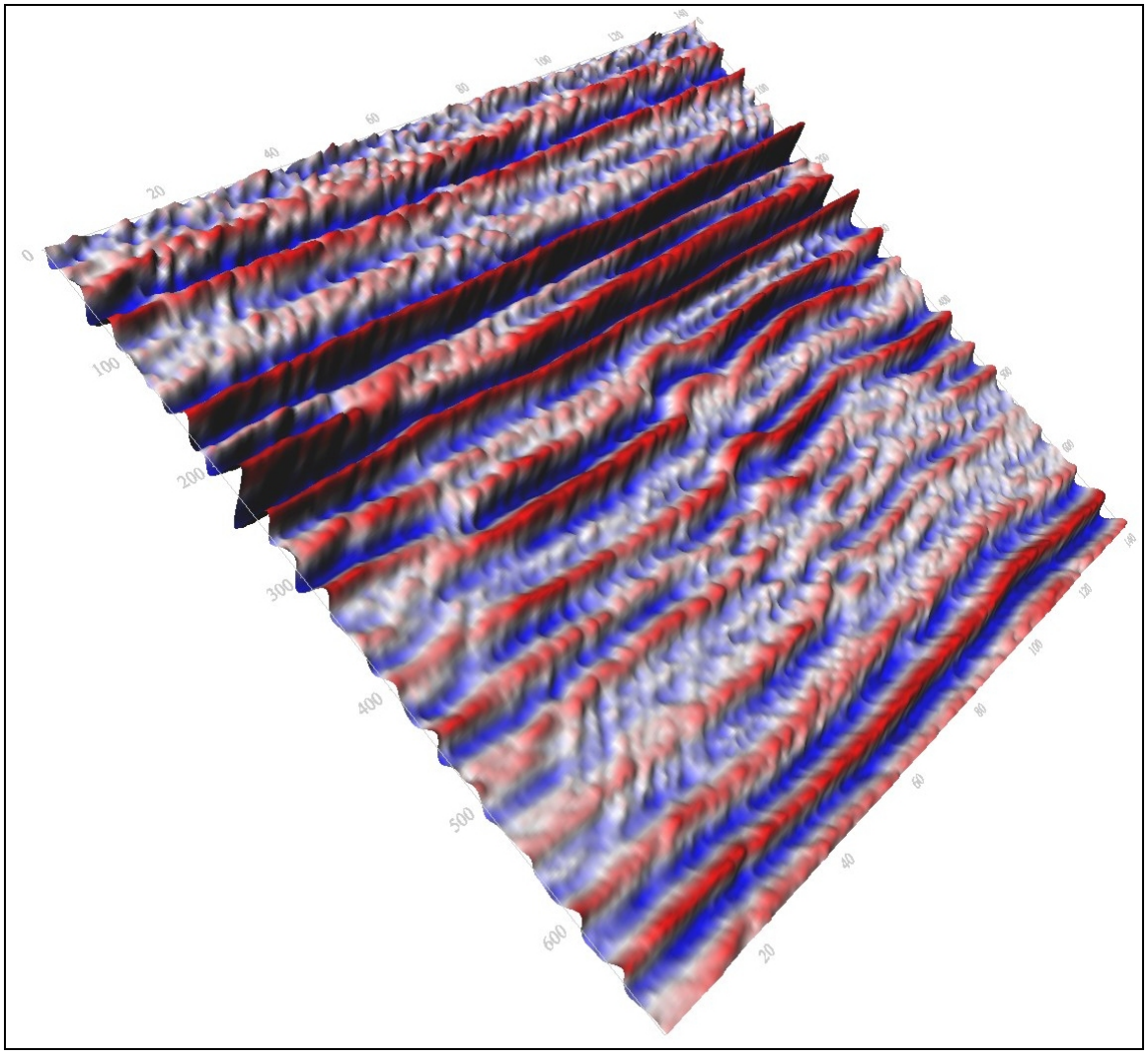


Figure 1.4: The first SeisScape™ display produced in December 1999 using an NVIDIA TNT2 Ultra graphics card. The seismic data shown is the same as in Figure 1.2 and Figure 1.3. I show the seismic data as a three-dimensional surface with positive seismic amplitudes displayed as peaks and negative amplitudes as troughs. I illuminate the seismic surface with diffuse lighting only.

However, that changed the moment that I looked at this first display. I am showing Figure 1.4 here at a large size so that with any luck it conveys some sense of the impact that it first had upon me. This first display answered my primary questions almost right away. The first question was, of course, “what would seismic data look like in 3D”. The answer was that it looked eerily realistic and that it had an almost ethereal quality to it. The best way to describe my initial reaction was that I felt almost a sense of déjà vu. It looked like something I had seen before, but I just could not put my finger on what it was. It could have been reminding me of a mountain range or it could have been reminding me of something else, I was not sure.

What I was sure of was that it was visually attractive and for want of a better term, I would have to call it beautiful. That was not what I expected at all but the answer to my first question of “what does seismic data look like in 3D”, was that it looked beautiful. However, all that aside, when I started developing SeisScape displays I had another, more important, question in mind. That question concerned the data itself and whether or not conventional displays showed all of its coherent signals. I expected that a three-dimensional seismic display would highlight amplitude changes better than conventional displays; it only made sense that it would. However, beyond that, there was the question of whether or not there were signals buried in the data that we were not able to see using conventional technology.

This is a more important question because in the Western Canada Sedimentary Basin, the area I am most familiar with, we have discovered the majority of the large and seismically prominent discoveries. The exploration targets are now smaller and more seismically subtle. There is also a switch from exploration to exploitation, which again requires more from the seismic data. What this means for the explorationist is that whenever they look at a seismic section it is most likely that someone has already either looked at it in detail or has at least looked at data very close to it. Being able to bring out more from the data, being able to show things that previous explorationists had not been able to see, I believed and still do believe, is extremely important for the future.

Therefore, the question then is “is there anything else there to see?” and to my relief I noticed almost immediately that there was; that there were coherent signals in this image that were not visible on the conventional displays of Figure 1.2 and Figure 1.3. To illustrate this, in Figure 1.4 there are a series of hyperbolic events sweeping upwards from the left edge of the section. The most prominent start at around 550 ms and sweep upwards from there but the reader can probably see more. These coherent events are artifacts produced by the migration and are not geologically meaningful. However, they are virtually invisible in both Figure 1.2 and Figure 1.3, which indicates that they lie below what I will term the “visual dynamic range” of the displays. The hyperbolic signals in Figure 1.4 may have been just noise but their presence proved that it was possible there could be hidden but meaningful signals present as well. What was clear from this first image is that its “visual dynamic range” was much higher than that of the conventional displays and that its ability to communicate information was much higher.

1.3 Beautiful Science

Oh thou who honerest both art and science …
Spoken by Dante to Virgil as they entered the gates of hell

Before the serendipitous conversation with my son, I was uninterested in visualization. Afterwards, I was somewhat intrigued. This mild interest kept me going through the frustrating months learning computer graphics in my spare time. However, once I started to produce SeisScape displays I became a convert to the new science of visualization and I started work immediately producing my first commercial version of the displays. I completed this program (which I confusingly called SeisScape) in the spring of 2000 and started using it to produce the SeisScape displays on a regular basis.

The more I produced of them the more I became aware of their duality. My early displays had a split personality; they had both a decidedly artistic nature and a decidedly scientific nature - something that at the start I thought was good. They were the epitome of beautiful science and I decided to present them as such to the industry at large.

1.3.1 Ancient Evenings: The first SeisScape presentation

I used the quote from Dante's "The Divine Comedy" that begins this section as the subtitle of the first talk that I gave on SeisScape displays, "Ancient Evenings: Seismic Visualization using very old techniques", a CSEG luncheon talk given in October 2000. By then I was struggling to reconcile the dual nature of the displays, the artistic and the scientific, and that struggle was reflected in both the title and the content of the talk. In the quote, Dante implies that Virgil was both an artist and a scientist, in this he was wrong, Virgil was primarily a poet. What Dante did get right was that in the ancient world science and art were intimately intertwined. They were also intertwined in SeisScape, but I was beginning to realize that this was more of a curse than blessing.

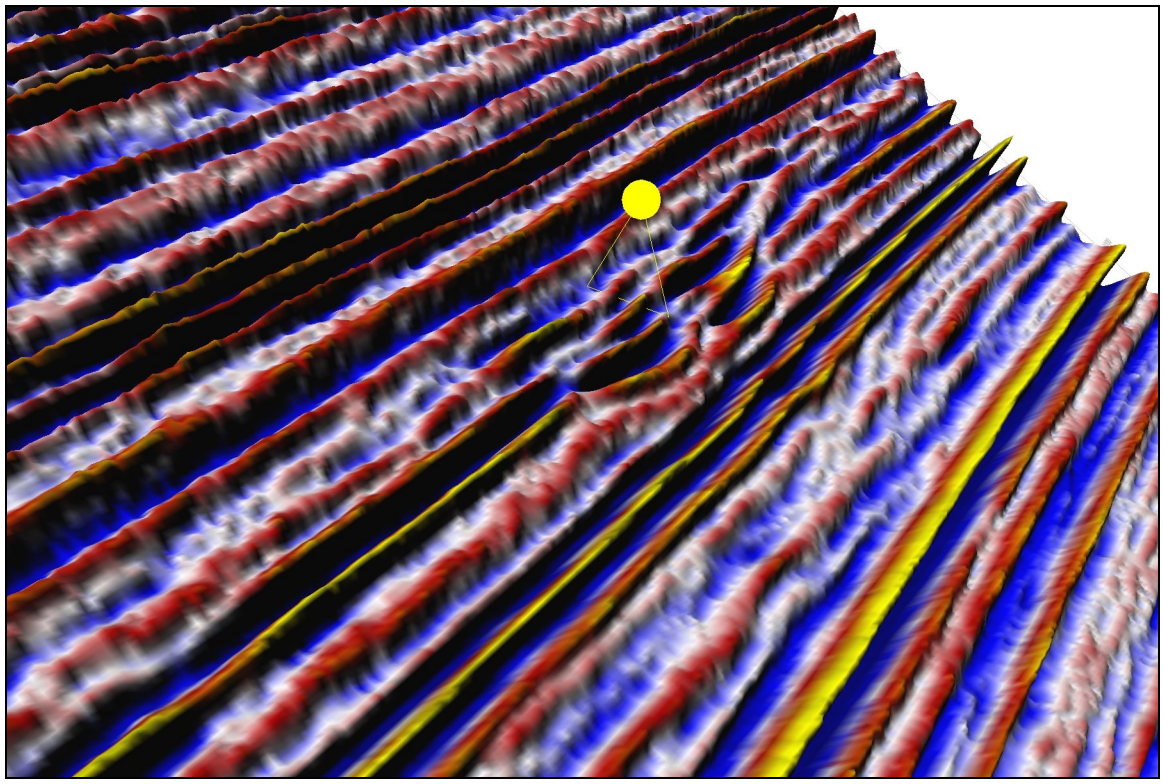


Figure 1.5: SeisScape display of a small channel, data courtesy unnamed source. The ball and triangle show the azimuth and angle above the horizon of the light source. The color palette used grades from dark blue (-‘ve) to white (zero) and then to deep red and yellow (+’ve). The effect I was trying to achieve was to make the display look mysterious. I did this by orienting the light source to show the darkened back faces of the channel.

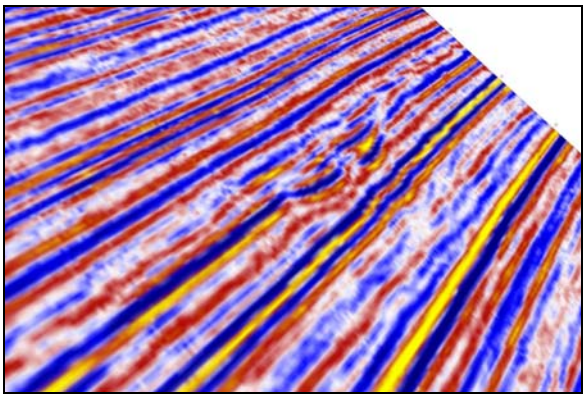


Figure 1.6: The conventional variable density equivalent of Figure 1.5, used for comparison purposes. The coloring and orientation of the displays are exactly the same. This was the first comparison that I showed between a SeisScape display and a variable density display. I used it to suggest how much more effective the SeisScape display was.

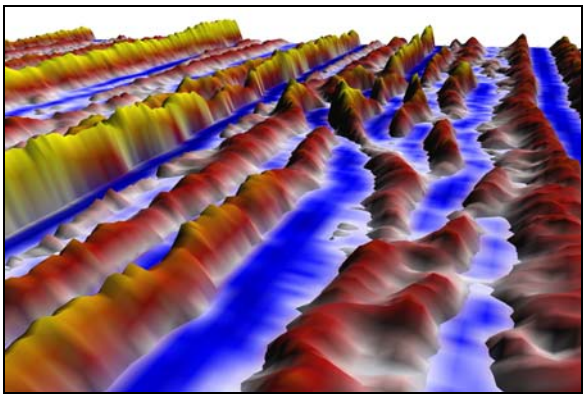


Figure 1.7: This is a combination of Figure 1.5 and Figure 1.6 viewed end on from the right. The variable density display cuts through the SeisScape display at the zero amplitude level. I was experimenting here trying to make the variable density display look like a sea level. The objective was to segregate the display and highlight positive amplitudes.

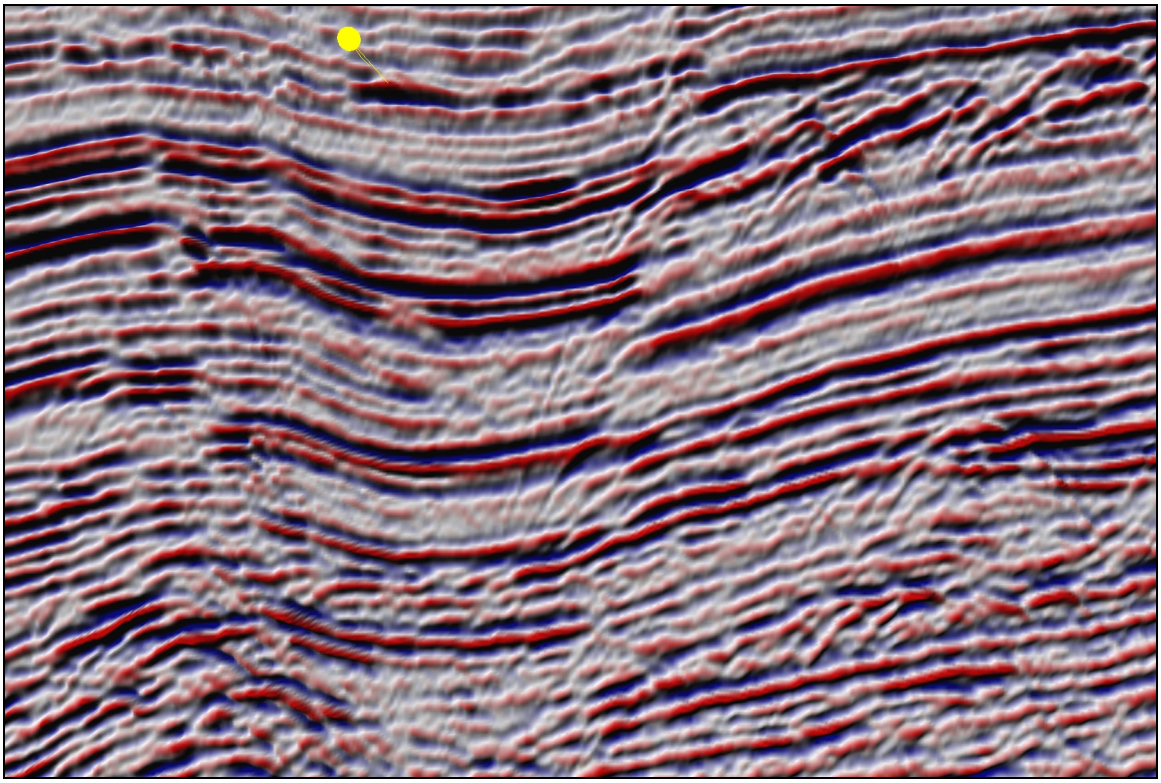


Figure 1.8: SeisScape display of the same section shown as variable density in Figure 1.9. I indicate the direction of the lighting by the yellow ball and triangle at the upper left. My objective in comparing this image with Figure 1.9 was to show that a seismic section contains pertinent coherent signals that are not visible on conventional displays. The lighting here enhances the ability to see coherent events that are oriented perpendicular to it. In this case, it brings out diffractions from the edges of faults and other terminations on the right half of the section that are not evident on the variable density image.

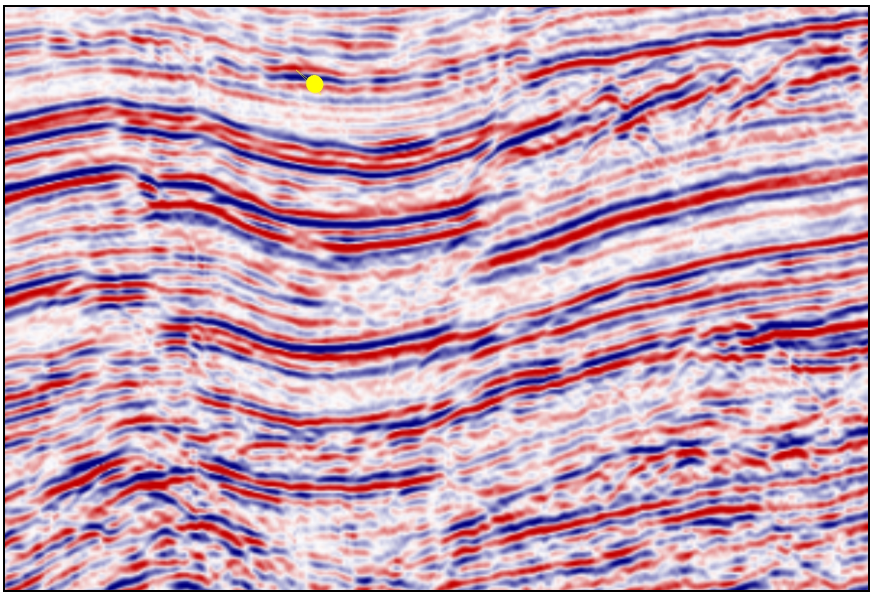


Figure 1.9: Variable density display of an unmigrated section showing both major and minor faulting. The coloring is the same as used in Figure 1.8. Since the section is unmigrated there should be diffractions coming off the edges of the faults but they are not visible in this type of display.

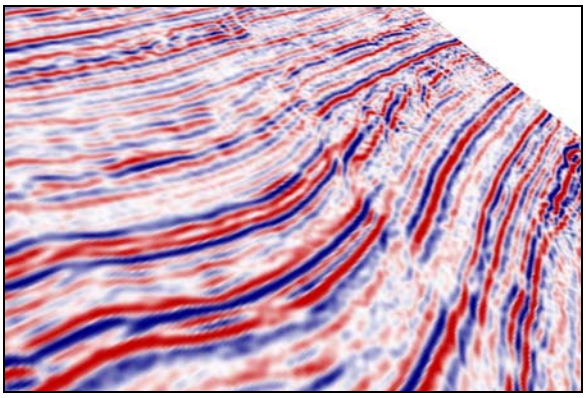


Figure 1.10: This is the same image as Figure 1.9 but rotated to see the section at an angle. My purpose in comparing this image to the following three was to show how important the direction of the light source was in enhancing high-angle events. It is a known phenomenon of lighting that it enhances coherent events perpendicular to it, I wanted to show that this principle applied to seismic data as well.

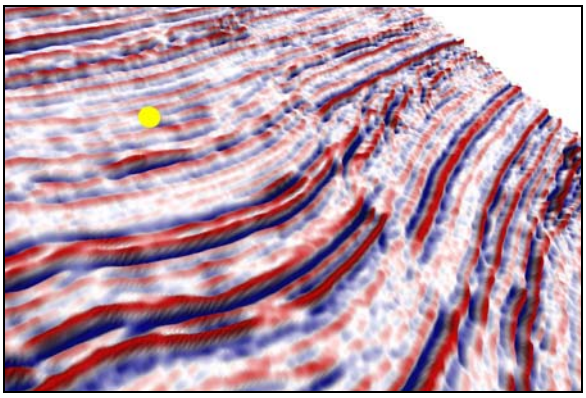


Figure 1.11: A SeisScape display of the data shown in Figure 1.10 using the same blue-white-red color palette. The yellow ball shows the direction of the lighting, which in this case is vertical. Since the light shines straight down it has little effect on our ability to perceive dipping coherent signals. The diffractions that are so evident on Figure 1.8 and Figure 1.13 are virtually invisible here.

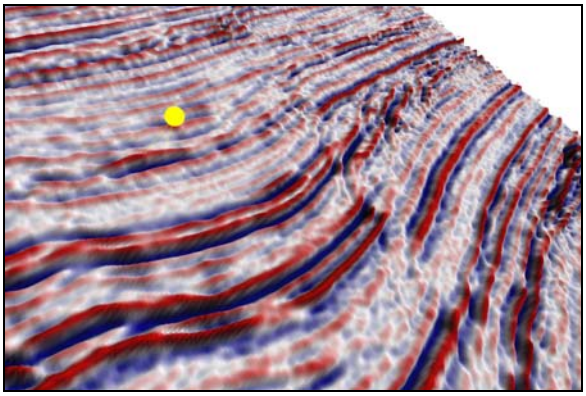


Figure 1.12: In this case, I have moved the light source so that I direct the lighting from the lower right and at an angle of 30 degrees to the vertical. Notice how diffractions from the edge of the major fault are now starting to appear.

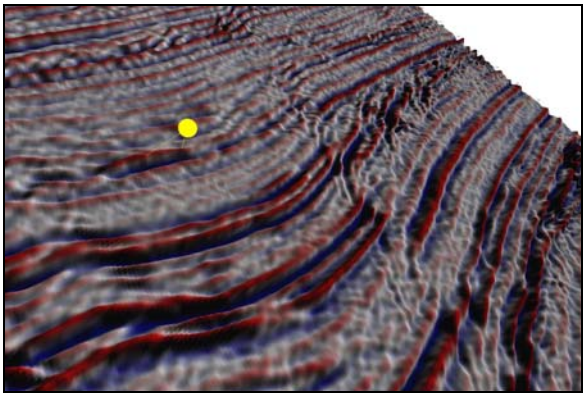


Figure 1.13: The lighting here is coming from the same direction as in Figure 1. but it is at an angle of 45 degrees to the vertical and is therefore lower to the horizon. Notice how diffractions from the edge of the major fault are now completely visible. Diffractions from minor faults further to the right are also now apparent.

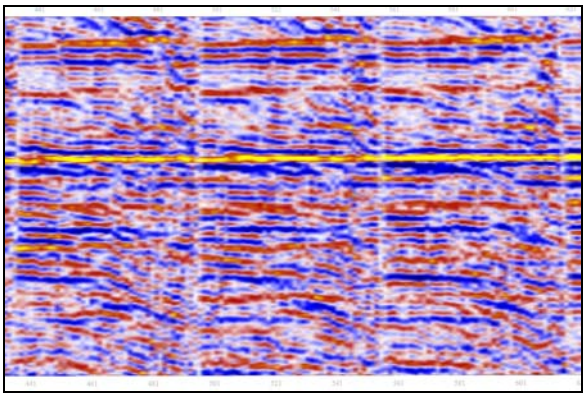


Figure 1.14: Variable density display of three common offset gathers. Gathers are delineated by the vertical white stripes with the offsets increasing from left to right. Note the multiple events that interfere with the main primary reflector (the yellow event).

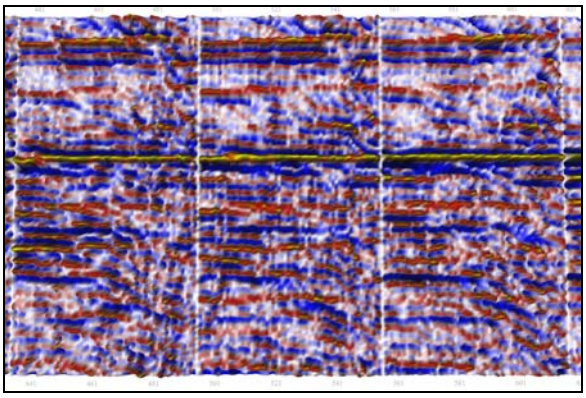


Figure 1.15: A SeisScape display of the same common offset records shown in Figure 1.14. The lighting here is vertical so beyond making the display more realistic it is not introducing any more information.

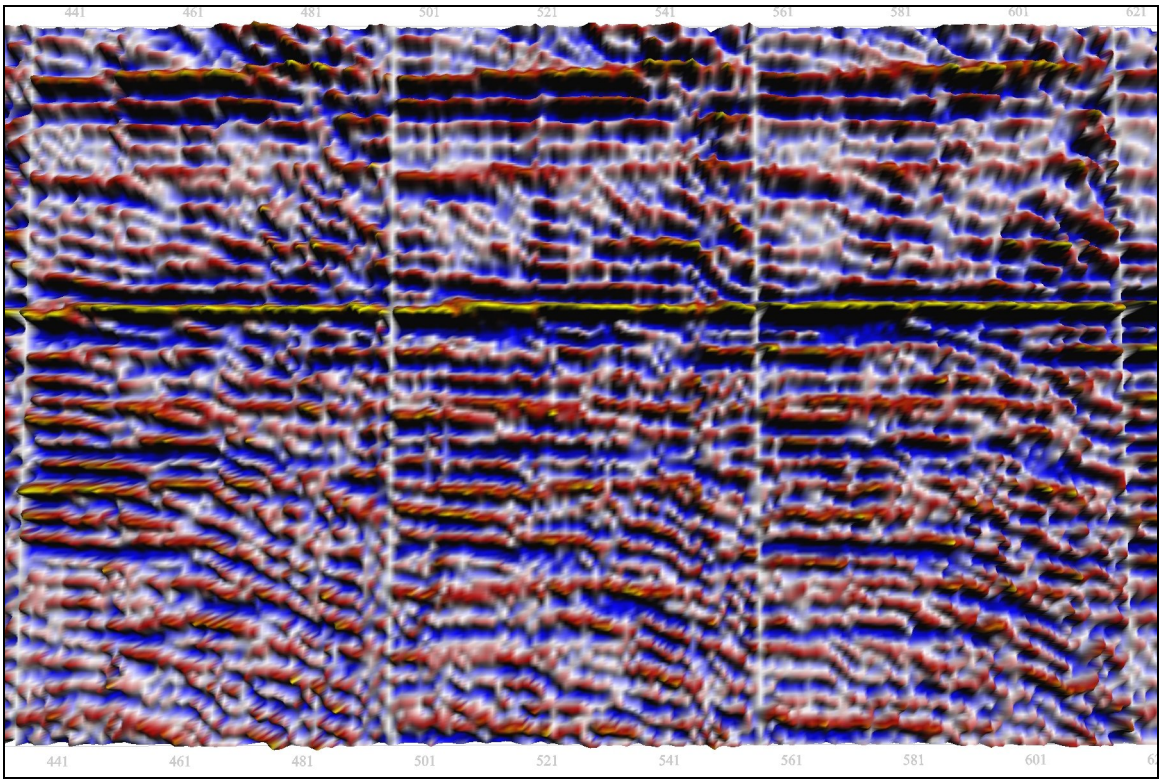


Figure 1.16: In this display, I direct the lighting from the lower left edge of the display at an angle of 40 degrees from the vertical. Note the multiple events that interfere with the primary reflector (yellow) are now much more visible than on either Figure 1.14 or Figure 1.15. I called this ability of the lighting to enhance the visibility of high-angle events "The Corrigan Effect" after Mike Corrigan, a consultant who first brought it to my attention.

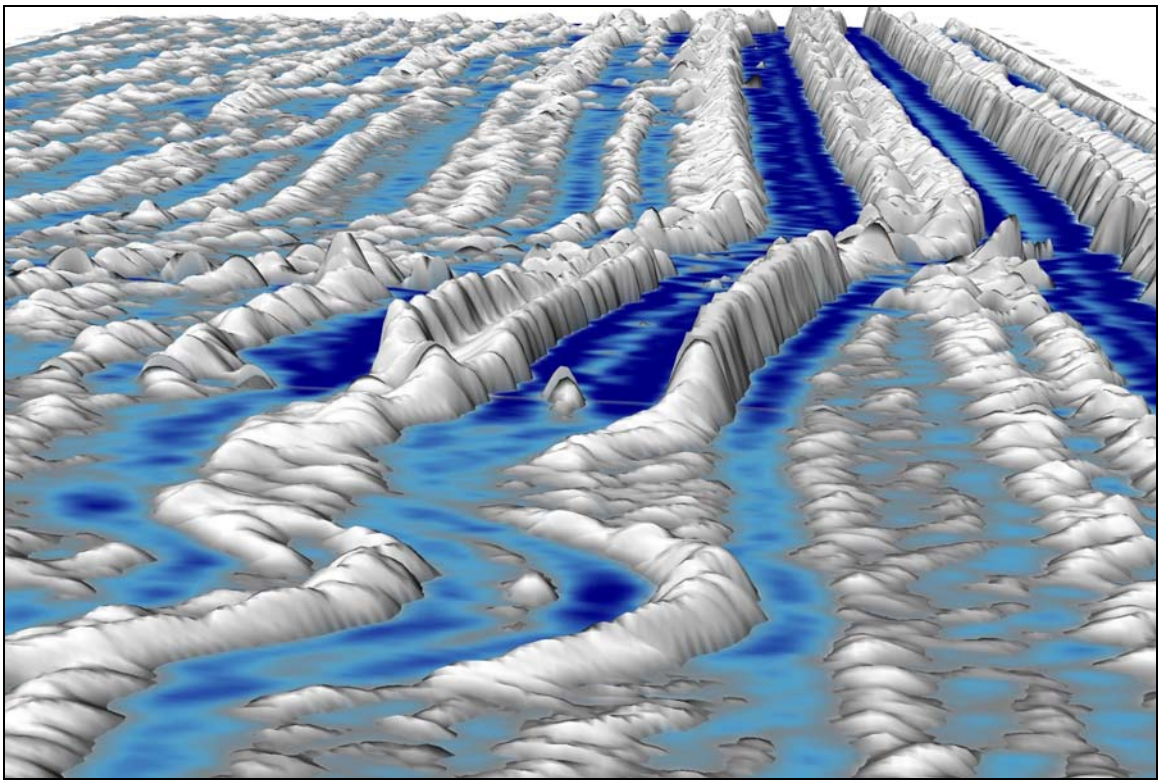


Figure 1.17: SeisScape display of the same section shown below in Figure 1.18, viewed end on and from the right. This was the final SeisScape image in the talk and I decided to experiment with the color palette and produce an image that was purely artistic. I wanted to get away from the stock blue-white-red color scheme and produce an image that was vaguely reminiscent of an arctic scene. To that end, I used a palette that represented low amplitudes as deep blue grading to grey and high amplitudes grading from gray to white. I also used the sea level concept by slicing the display in the middle with a variable density display that used the same color palette. This final image was quite dramatic and very well received.

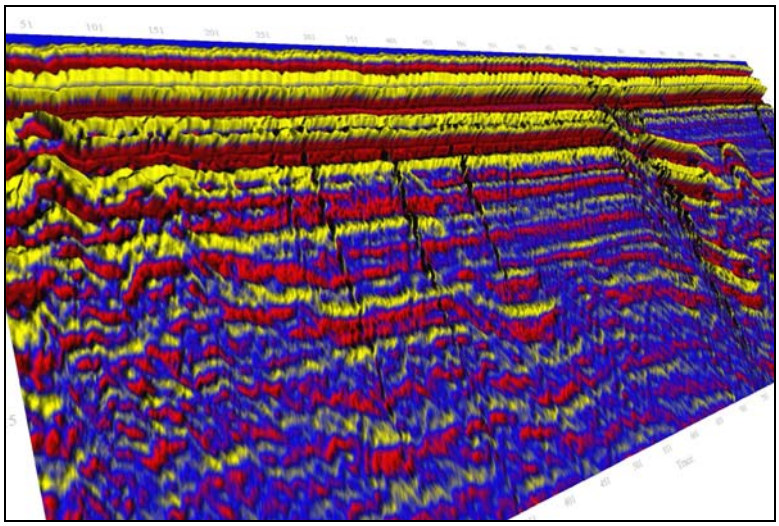


Figure 1.18: SeisScape display of CODA GeoSurvey Sonar data over a small lake in New Brunswick. Data courtesy GeoSurvey. CODA data is much higher frequency than conventional seismic data; the section shown being only approximately 6-8 meters in depth. Note the three vertical features in the middle of the image; these are likely buried pipes. The color palette here uses red for low amplitudes, blue for mid-range amplitudes and yellow for high amplitudes.

Figure 1.5 to Figure 1.18 inclusive is a collection of some of the actual images that I used in that first talk. I have included many of them at large scale so that the reader may get some understanding of the visual impact that I was trying to achieve. When I started preparing the talk, I initially wanted to impress the viewers with both how visual the displays could be and how they could bring out subtleties in the data. As the readers can judge for themselves, I succeeded in the first part by choosing color palettes that were vibrant, that in essence just looked good. Most of the images use color palettes that are very heavy on deep blues, reds and yellows and the effect when combined with the lighting of a SeisScape display can be very striking. I succeeded in convincing the audience that seismic data could look beautiful and to this day people remember the talk for just that, its visually stunning images.

However, that is not all the talk was about. I also wanted to show that seismic data contained significant coherent signals that were not evident on conventional displays. To that end, I showed a number of the images (Figure 1.8 - Figure 1.13) that compared a conventional variable density displays with SeisScape displays using various orientations of lighting. The seismic used in these examples was from an unmigrated seismic line shot over an area of significant faulting. Figure 1.9 and Figure 1.10 are the variable density images of this data set and since the data set is unmigrated, they should show diffractions coming off the faults and other terminations. However, they do not, there are no diffractions evident and the point I made is that even though we cannot see them they should still, in theory, be there.

The SeisScape displays proved conclusively that the diffractions were there. I showed that not only did lighting enhance faults but also that the direction and elevation of the light source were critical components in determining what coherent events became visible. I meant these images to be the most important images of the talk because they showed conclusively that there were important coherent signals (not just noise) hidden in the seismic and that SeisScape displays were capable of bringing them out. It is true that this was still unmigrated data and that the diffractions would be collapsed by migration prior to interpretation. Even so, they were distinct on the SeisScape images and invisible

on the variable density images, which was clear proof of the low visual dynamic range of the conventional displays.

Unfortunately, whereas the faulted images were the most significant to me, not a soul from the audience remembers them. I did a good job showing the dual nature of the displays, the artistic and the scientific but I overdid it. When I wrote the talk, the science was primary and the art secondary. This is not how the audience received it, they were more attracted to the visually stunning but less visually informative images; beauty won out over science. To this day, people still tell me how much they enjoyed the images but nobody talks about the importance of the dull gray diffractions.

Ultimately, I judged the talk to be a failure because I allowed the visual effect of the displays to override their importance to exploration. It was at this point that I realized that I did not understand the nature of visualization at all. I understood somewhat how to make the displays look "better" by using artistic techniques but this did not translate into making them better visualization displays. The artistic nature of the displays seriously detracted from the cold hard science that I was trying to get at.

1.3.2 The Purposes of Art and Science

As a high school student, I attended Nelson High School in Burlington Ontario. At the time (late 60's) Nelson High School was the home of a popular local artist and art and history teacher by the name of Robert Bateman (the same Robert Bateman who is now one of the world's leading wildlife artists). This is not to imply that I ever took a course from him, far from it. As a young man, I was schooled in the hard disciplines of science and logic, I had little use for the arts. This was made apparent one day in the study hall when I engaged an entire art class in debate with Mr. Bateman as the inadvertent moderator. I do not remember most of what I said but I do remember, in a particularly heated moment, telling Mr. Bateman "I don't think much of your paintings; any idiot with a camera can beat you any day of the week". Mr. Bateman's reply showed a high degree of tolerance for youthful arrogance, he simply told me that he had hung one of his

paintings in the hall outside of the main office and he suggested I go look at it and see if I still felt the same way.

I have always regretted my comment because it was not only foolish but also rude. In my own defense, I was young and I did do as he suggested, I went and studied the painting, which I have included here, with permission, as Figure 1.19. The painting was titled “Blown Grasses” and, although I cannot be sure, I believe it is of a field behind Nelson High School, one that I walked through every day on my way to school.

This painting is important because viewing it made me aware, for the first time, of the duality of vision that plagued my early SeisScape displays. In humans, the visual sense is not just functionary; it does not just inform it also provides us with an experience. When I looked at Figure 1.19 I realized that I was partially correct. I could most likely produce a sharper and better-resolved image of the grass with a camera; I could produce a more informative image. However, what I could not do with a camera was capture the scene itself; I could not provide a sense of what it was like to be there. This was the first time I realized that the visual system was both informative and engaging. For the record, I finally apologized to Mr. Bateman when I wrote to him to ask his permission to include “Blown Grasses” in this chapter.

To further illustrate my point about the duality of vision I have included another of Mr. Bateman’s paintings. Figure 1.20 is titled “Along Walker’s Line” and whereas I am not sure where “Blown Grasses” was painted, I am sure of where Figure 1.20 was.

Figure 1.19: “Blown Grasses” by Robert Bateman, 1968, used with permission. This painting was hung in the main hall of Nelson High School in Burlington Ontario during the time the author was student there.

Figure 1.20: “Along Walker’s Line”, by Robert Bateman, used with permission. The author used to live on Walker’s line in Burlington Ontario and used to pass this barn regularly. When I first saw the painting in a gallery years ago I immediately recognized the scene even though I did not know the title or the artist.

Walker's Line is a road in Burlington Ontario and it runs from the shore of Lake Ontario, north for about 20 miles, terminating at a conservation area called Rattlesnake Point. In High School, I used to live on a side road off Walker's Line and I regularly passed this barn on my way to walking the family dog, a Golden Retriever called Lindy, at Rattlesnake Point. I first saw a print of the painting a number of years ago hanging in a gallery. I recognized the scene immediately even though I at first did not know either the title or the artist. The fact that I recognized the barn so quickly, even though the image was hung in a gallery thousands of miles removed from its origins, is proof that the image is decidedly informative. However, to me, having lived in the area, it does far more than just show me a picture of barn that sadly was replaced by residential developments years ago.

What struck me most about this painting was how well it captured the entirety of the scene. Anyone who has lived in Southern Ontario will be familiar with the cold chill day that it depicts. The chill that goes down the back, the cold wet feet from the heavy wet snow, the warning that winter still has a long way to go. When I look at this painting all of this comes back to me; it reawakens in me long forgotten feelings and memories; it takes me back to my youth and places me back in an old Morris Minor taking a then very young dog for a very long cold walk.

If I had stood at that spot on the day Mr. Bateman painted the picture, I could have taken a thousand pictures and not one would capture the scene as well. Each photograph would be a better-resolved image, the barn and the trees would be clearer, the tracks in the snow more distinct but the essence of the day would not be there. To update my oft regretted comment of so many years ago, "Any idiot with a camera can show you how it looked but it takes an artist to make you sense how it felt!"

The reason that I have included this section is that it clarifies the difficulties that I experienced with my early SeisScape display. I started this section with a quote on honoring both art and science and I have shown that the first SeisScape displays did exactly that, they honored both art and science. However, here is the point; the purposes of art and science are fundamentally different. There are many different definitions for both but for the purposes of this thesis and in the context of visualization I adopted the following:

- The primary purpose of art is to engage.
- The primarypurpose of science is to inform.

Here is the lesson that I learnt from "Ancient Evenings", visualization is not art, it is science. The purpose of visualization is to inform and not to engage.

1.3.3 The Early use of Color for Seismic Images

SeisScape and "Ancient Evenings" were not the first time in my career that I ran into the duality of vision. I was very much involved with the introduction of colored seismic displays to the industry and I ran into the same problem then. The routine use of color in seismic displays did not begin until the late 1970's. A number of authors had discussed the uses of color in geophysics in the late 60's and early 70's (Smith et al., 1972; Grossling, 1969; Balch, 1971) but because of (a) a lack of perceived need and (b) a lack of appropriate technology, color was rarely used to display seismic data.

This changed in the late 70's due to a number of reasons. The first reason addressed the need for color. In June of 1979, Taner, Koehler and Sheriff published their groundbreaking paper "Complex Seismic Trace Analysis in Geophysics" which introduced the field of seismic attributes. Prior to their introduction, the wiggle trace display served the needs of seismic very well. After their introduction, it did not because most of seismic attributes are not readily interpretable as wiggle trace displays. Attributes needed color!

The second reason for the introduction of color was the development of the Applicon color plotter. This was drum-based plotter that was capable of producing large-scale images in a reasonable time. The color resolution was very poor by today's standards and it produced most colors using a process called dithering. Nevertheless, when they worked (which was rarely) Applicon plotters produced very good color plots. This gave us both the need for color and a tool that was capable of producing it for us.

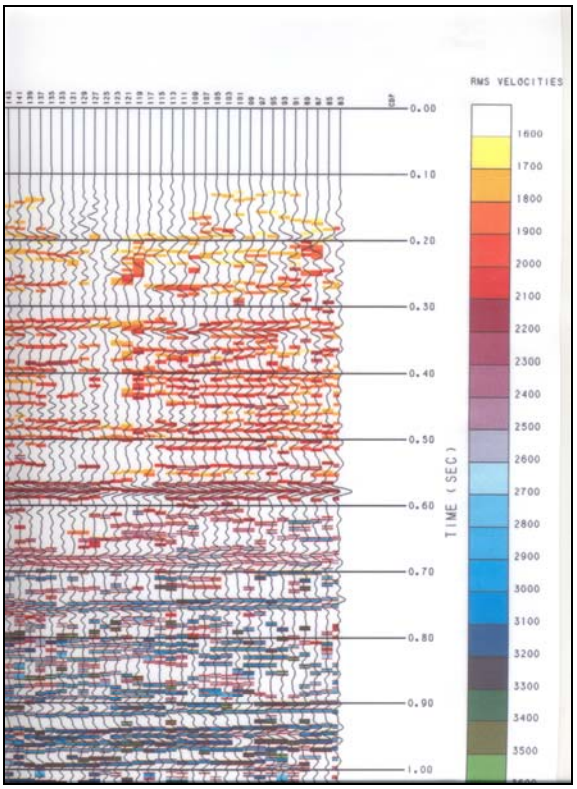


Figure 1.21: An early use of color in seismic, circa late 1979. The colors boxes shown behind the wiggle traces represent the maximum coherency at the given time as determined by Gulf's Continuous Coherency Velocity Analysis (CCVA) program. Color is coded to RMS velocity with the color table being shown on the right.

Figure 1.22: A further section of the display at the left. By tracking the color along an event, the user was able to observe the stacking velocity of the event in detail and determine where multiples were interfering with the primary events.

At the time of their introduction, I was working in the special projects group for Gulf Canada Resources in Calgary and in the fall of 1979, we received only the second Applicon plotter sold in Calgary. I made use of this plotter almost immediately by developing a technique whereby we superimposed wiggle traces from a stacked section over a color display of the RMS stacking velocities for the section as determined from

Gulf's Continuous Coherency Velocity Analysis program (CCVA). To the best of my knowledge, Figure 1.21 and Figure 1.22 are the only remaining plots of this display.

The display was useful in that it provided a way that the interpreter could see what lay behind the stacked data. It was excellent for identifying where prominent multiples, which were presumably stacked out of the section but present in the prestack gathers, were possible interfering with the final stack. It gave the interpreter a further degree of confidence that a potential stratigraphic anomaly observed on the section was related to geology and was not merely a seismic artifact. It also helped to identify areas where there could be problems with statics and the line geometry. The display very quickly became popular within Gulf Canada and I was asked to present a talk on it at Gulf's Houston office in late 1979. The point here is that this display was very informative but not very pretty. The displays were hard to look at and often visually confusing. Even so, they provided information that was otherwise not available and despite their lack of artistic merit, they became very popular. So popular that in the early 80's Western Geophysical copied and subsequently patented the display.

My manager at the time was Carl Nyberg who was in his own way very forward thinking. He saw the benefit of these displays and was very quick to give me credit for them. This was important to me because I was only just out of university and it was important for my confidence to produce something useful. This praise, however, was tempered by Carl's reaction to my next color displays. Having succeeded with the velocity displays I then decided to create color displays of several complex seismic trace attributes including instantaneous amplitude and instantaneous frequency.

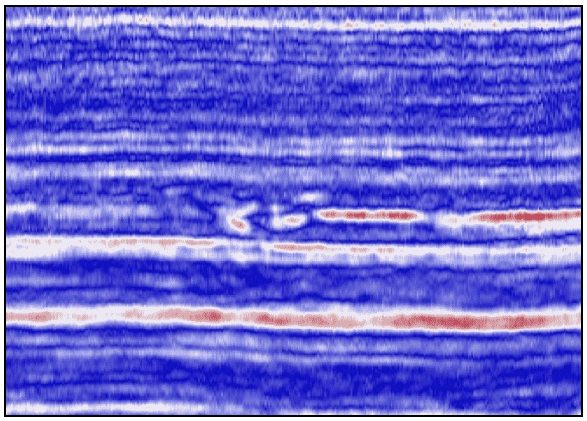


Figure 1.23: Instantaneous amplitude display of the data used in Figure 1.5. The color palette used grades from blue to white to red. Although I created this display using modern technology, it is similar in quality to one I created in 1979 using an Applicon plotter.

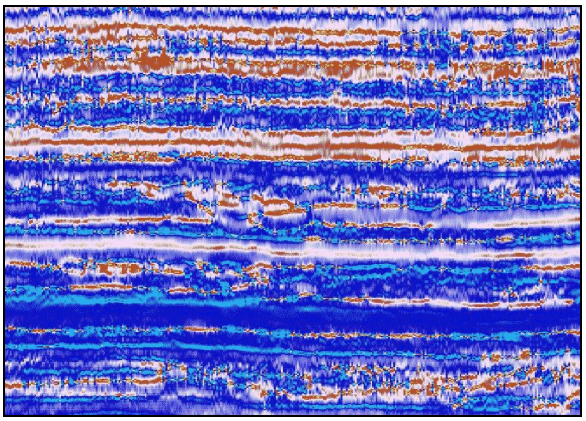


Figure 1.24: Instantaneous frequency display of the same section of data shown in Figure 1.23. The color palette uses magenta for low amplitudes and yellow for high. This display is similar to an instantaneous frequency display that I created in 1979. The blockiness and the lack of color depth in the image simulate the quality of the Applicon plotter displays.

Unfortunately, none of those displays remains in existence and consequently I have had to recreate what they looked like using modern technology. Figure 1.23 and Figure 1.24 represent what to the best of my recollection they looked like. Both displays are blocky and have a limited color depth just as the displays from the Applicon plotter had. Carl's reaction to these displays was very different to how he reacted to my velocity displays. The displays were, he said, excessively pretty. He believed that the interpreters would spend more of their time gazing at the images than interpreting them. I am not critical of Carl here, because his point was well taken. According to my wife, who was working as a seismic processor at the time, his comments mirrored the industry's reaction in general as she remembers hearing much the same comments from her own clients.

I consider that scientific visualization started with the ability to use color in our work because color adds a new dimension that can be both informative and engaging. In my velocity display, color was purely functional and the display was accepted. However, the color in the complex attribute displays accidentally served both purposes. It was used to communicate the information in the attribute but it was also "way too pretty". My early attribute displays, which accidentally mixed the functional, and the engaging were

primitive by today's standards. Even so, they were heavily criticized and did not enter into general use.

1.4 The Science of Visualization

Everything that I have written so far is to convince the reader of the one underlying fact of visualization, that whenever you put an image in front of a viewer you are both informing them and engaging them. It is naive to suggest that science is strictly the former and art strictly the latter because in practice both art and science inform and engage. In my opinion, in the context of visualization, the **primary** purpose of art should be to engage and the **primary** purpose of science should be to inform. However, these are never the only purposes and in the field of visualization, the difference in science and art is one of degree. Science is more focused on informing and art is more focused on engaging but it can never be an exclusive relationship.

In creating the first SeisScape display I did with computer graphics in 1999 what I had done with the Applicon plotter 20 years earlier, I simply threw an emerging technology at seismic data to see what would happen. In neither case did I follow any established principles. Therefore, it should not surprise you that my results were hit and miss. Some of my Applicon displays worked and some did not, some of my SeisScape displays worked but looked at critically, most did not. The problem as I came to learn was that since I did not know the principles of visualization I did not know how to shift the balance in the displays between art and science and all too often the displays ended up on the wrong side.

1.4.1 Early Conclusions

As I said in the very first line of this chapter, this thesis is ostensibly about SeisScape displays, what they are and how to use them. It is important to note, however, that by the time I started work on my thesis in January of 2003 I had already used them for a number of years and had reached several conclusions. It is those conclusions, which are in many ways the most important of this entire work. Before I talk about what they are, however, I want the reader to compare Figure 1.23 and Figure 1.24, which are my first ever color

attribute images dating back to 1979, with Figure 1.25 and Figure 1.26, which are the corresponding SeisScape images of the same attributes and data.

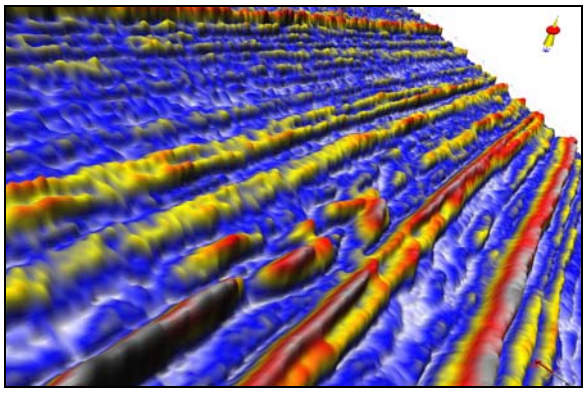


Figure 1.25: A SeisScape image of an instantaneous amplitude display using the same data as shown in Figure 1.23. The display is rotated to the right and tilted at a 45-degree angle to show more of the data and to show amplitude changes along events.

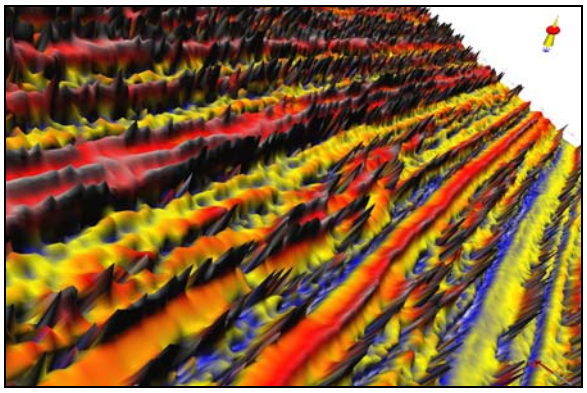


Figure 1.26: A SeisScape image of an instantaneous frequency display using the same data as shown in Figure 1.24. The display is rotated to the right and tilted at a 45-degree angle to show more of the data and to show amplitude changes along events.

In doing so keep in mind, that at the time I produced Figure 1.23 and Figure 1.24, I was heavily criticized because the displays, while showing useful information, were also too “artistic” to be useful. Given that criticism, if Figure 1.23 and Figure 1.24 are too artistic, how useful are Figure 1.25 and Figure 1.26?

Obviously, Figure 1.25 and Figure 1.26 are far more visual than the early (simulated) drum plotter images and most viewers would say that they were, therefore, “better” images. However, what does that term “better” really mean? Typically when we say something looks better than something else we mean it is more visually pleasing, we just like it better. Nevertheless, there is a second definition, that being that the image enhances the viewer’s ability to perceive the pertinent information. The first definition is art whereas the second is science.

These are the conclusions that I reached at the very beginning of this thesis:

Conclusion #1

One of the primary purposes of art is to engage the viewer. Art elicits emotion, which is important if you are putting on an exhibition at the National Art Gallery, but it is inconvenient if you are looking for porosity in a Devonian reef.

Conclusion #2

The primary purpose of visualization is to inform the viewer dispassionately and intellectually and the concept of what constitutes a "better" visualization display must be defined by strictly scientific principles.

In theory, visualization is the process of communicating information via the visual processing system; in practice, it is the art of extracting the informative from the engaging.

1.4.2 The Primate Visual Processing System

This thesis is based upon two premises:

1. To improve our ability at visually communicating complex information we must rely on applying strictly scientific principles.

2. The principles that we need to apply are buried deep within the primate visual processing system.

The primate visual system is unique among mammals. We share with other primates a surprisingly acute and highly refined vision. Our vision far exceeds that of all non-primate mammals and in many respects rivals that of other vertebrates such as large birds of prey. Nevertheless, it is not a general-purpose tool. It is highly evolved but it has evolved to see certain things in certain ways against certain backgrounds. The primate visual processing system does not see all things equally, it does not perceive all things equally, and most unfortunately, it is completely beyond our control. Humans are the

ultimate toolmakers. We use tools for everything and if the tool is not doing its job – we change it. However, in this one case we cannot do that. Our visual processing system may be the tool but it is fixed and we cannot alter it to meet our demands.

To improve visual communication then, the onus is on us to understand, at a very detailed level, how the visual system functions. If we have no direct control over what happens to the visual information once it enters our visual processing system, we have full control over what we put into it. It is up to us to learn how to create displays that are in tune with how our visual system functions. Given this, the underlying sciences of visualization are the sciences of the primate visual processing system. It is those sciences that we must become familiar with before we can fully understand how visualization can be used as a tool.

CHAPTER TWO: TRIVARIANT COLOR VISION

"Color vision is known best by man's perception of it. It creates a unique dimension to sight that is impossible to appreciate by any non-visual means. It depends on wavelength more than on the energy of light but it is an illusion of reality resulting from a comparison of the responses of nerve cells in our brain. Color and all vision are in a sense illusory depending only on messages that pass between millions of neurons that reside within the darkness of our skull. These visual messages allow us to project ourselves into a universe that would be unknown to us without vision."

Dr. Peter Gouras

2.1 Introduction

It is generally accepted that Aristotle was the first to enumerate the five "classical" human senses of sight, sound, hearing, touch and taste. Today we recognize that we have many more sensory systems including a system for a kinesthetic sense and a system for sense of balance. Not all of our sensory systems give rise to a direct appreciation of a sensation because many of our sensory systems function subconsciously. Of those that do give rise to sensation, the process by which the sensory stimulation is translated into an experience is called perception and the experience itself is called a percept[1].

Of all of the systems that communicate to us via perception, the visual system is by far the most underappreciated because we rarely notice its percepts. We notice auditory percepts, we notice olfactory percepts but we rarely notice visual percepts. That is because unlike the other senses, the visual system is never quiescent. We notice other senses as part of the world around us but the visual system **is** the world around us. It is our primary sensory modality, it dominates our neocortex and it functions even when the eyes are closed and the brain is asleep.

The percepts of vision are the objects that surround us and interact with us. It is the purpose of the visual system to identify these objects, determine their spatial position and relationships, evaluate their movement and discover their properties. It must do all this with all of the myriad objects that surround us; and do it in real time. The complexity of

[1] An impression of an object obtained by use of the senses.

this task dwarfs all other sensory tasks. We do not notice most of the percepts because the visual system is so efficient that we simply do not have time to become consciously aware of everything that it produces.

The products of the visual processing system are percepts and since we intend to use it to communicate scientific information the product of visualization must also be percepts; we must perceive that which we are trying to show. To that end, this chapter introduces the concepts of visual percepts and trivariant color vision. The former is what we want to produce; the latter is the system by which the visual system goes about its task of producing them. It introduces them in the context of the seismic variable density display and goes on to prove that whereas we perceive this display as a seismic display the display itself does not perceive seismic at all.

2.2 The Concept of Seismic Perception

"Perceptions are internal representations of the external world"
R.L. Gregory, Eye and Brain, Fourth Edition, 1997

Figure 2.1 and Figure 2.2 are both images of a complex scene. The former is a picture of the author's wife and dog in an alpine meadow; the latter is a variable density display of a seismic line from the Trujillo area of Peru.

In both instances, the brain creates a model of the scene in the mind. Creating this model is a two-stage process. In the first stage, the visual system segregates the scene into discreet objects, in the second it interprets these objects as percepts. These percepts are provisional in nature because as we acquire new knowledge our percept changes. Take, for example, the percept of the dog in Figure 2.1. Almost everyone with normal vision would recognize it as some form of animal; beyond that, most adults would recognize it as a dog; beyond that, most people familiar with dogs would recognize it as a Springer spaniel; beyond that, most people familiar with my family would recognize it as my dog etc.

Figure 2.1: The author's wife and dog in an alpine meadow.

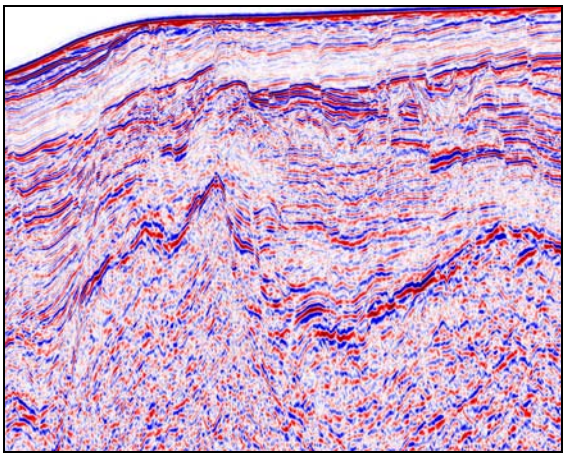


Figure 2.2: Variable density seismic display of a faulted data set from the Trujillo area of Peru, data courtesy PeruPetro. Color palette represents –ve amplitudes in blue, zero amplitude in white and +ve amplitudes in red.

Perception is also a multi-stage process because we assemble the whole from the parts. We do not just perceive the dog, we first perceive its components, its eyes, its nose, its tongue, its smile, and its tail etc. and we interpret each of those in turn as its own provisional percept. We develop our perception of the dog as a whole from a consideration of its parts but only after we assemble our perceptions of its parts from our perceptions of its parts parts, and we assemble …. Ad infinitum.

Figure 2.2 is a seismic image of a complex geological cross-section and it contains as many constituent parts as does the real world scene shown in Figure 2.1. It contains numerous reflection events each of which is broken into smaller sections by the myriad collection of major and minor faults. Each of these smaller sections is further subdivided by even smaller faults and each of these even smaller sections has its own characteristic amplitudes, dominant frequency etc.

However, unlike Figure 2.1 there is no concept of a percept here. There is no perception of the major faults, for example. We cannot see them directly and therefore we have to imply their locations. The situation is worse for the minor faults, which we can barely detect at all. We also do not perceive the reflection events as objects and although we know that each reflection event has amplitude changes along it, we do not perceive

those either. In fact, there is no sensation of perception at all in this image; there is nothing in it that grabs our attention and no part of it that is visually more distinct than any other is. When we look at the real world scene, we develop sensations of perception but when we look at the seismic scene we develop none at all, our visual system has failed us.

This is a strange concept that the visual system can fail even though we can see something clearly. At first, it sounds nonsensical to imply that perception can fail even when we can see. Nevertheless, the products of the visual system are percepts, sensations of perception, and if the visual system cannot produce them for us, then it has by definition, failed. The ultimate goal of visualization must be to produce a display that is as equally interpretable to the visual system as is a real world scene. It is doubtful, given the physical differences between the objects in the two images, that we will ever achieve that goal. However, there is a fundamental reason why Figure 2.2 is so poor, why it produces absolutely no sensation of perception. The reason is human trivariant color vision, which I will discuss in the remainder of this chapter.

2.3 Simple Visual Experiment

In the spring of 2006, I conducted a simple psychophysical[2] experiment involving over 100 participants from the University of Calgary and Divestco Inc. I showed two images (shown full size in Figure 2.3 and Figure 2.4) to the survey participants and asked them the same two questions for each:

1. What is it?

2. Did you recognize it automatically or did you have to think about what it was?

---

[2] Study of the quantitative relations between psychological events and physical events or, more specifically, between sensations and the stimuli that produce them.

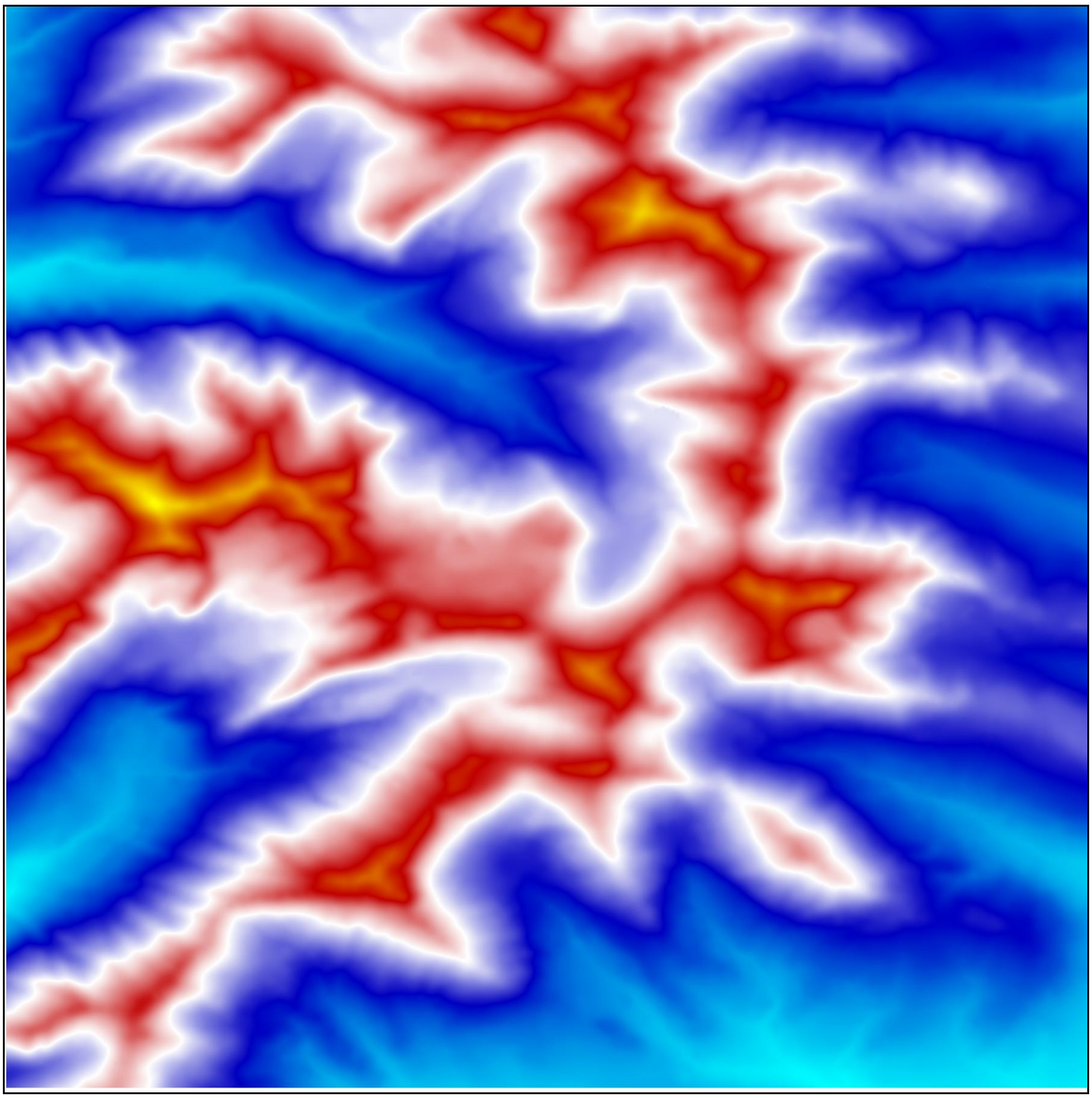


Figure 2.3: An image of an object created using a typical seismic color palette: cyan-blue-white-red-yellow. I sent this image to the participants second; even so 60% of the survey participants could not identify what the underlying object was. The majority of the 40% of the respondents who did identify it reported that they did not develop their identification directly but had to use secondary information, in other words, they had to think about what the object was before arriving at a conclusion.

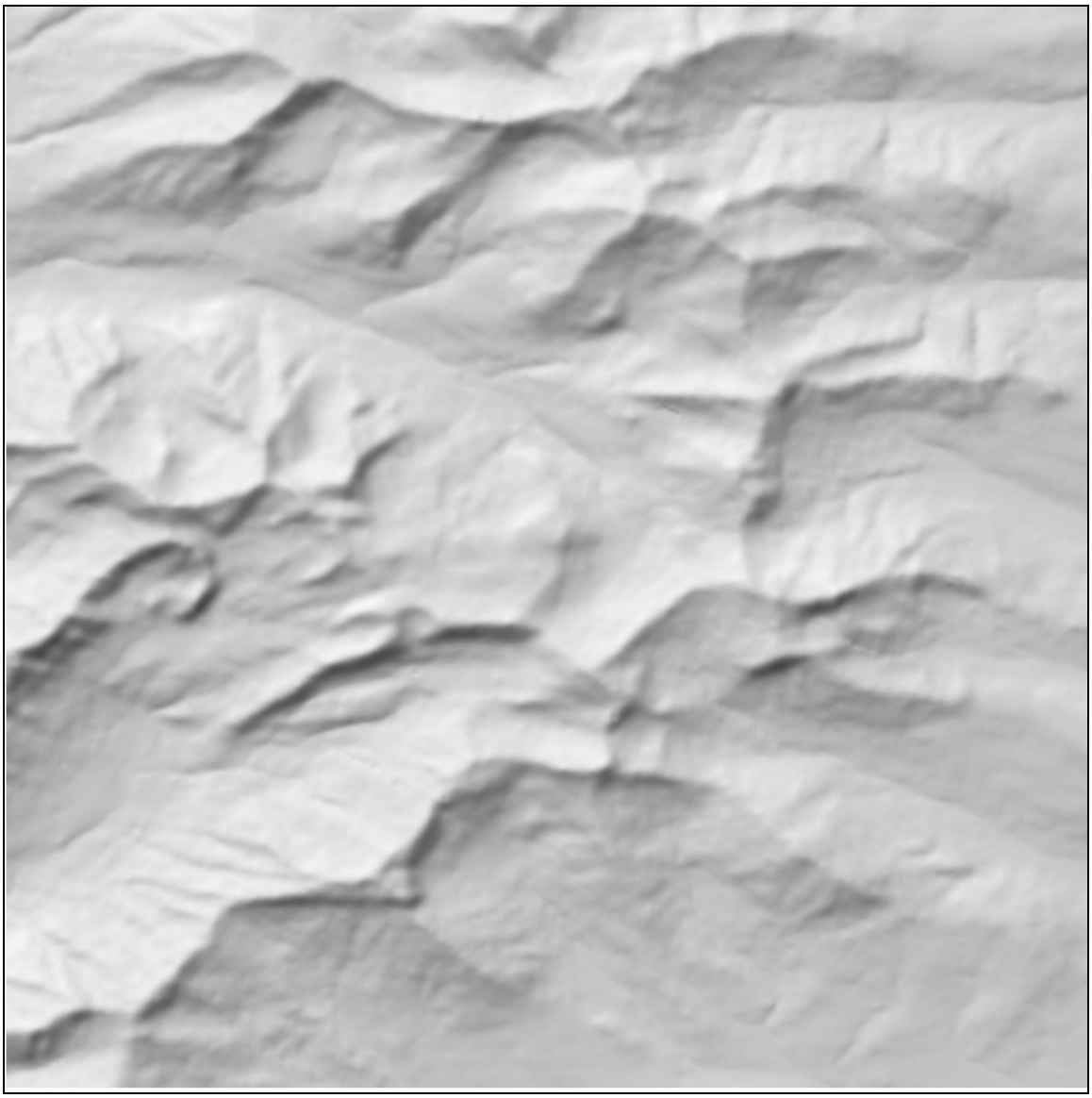


Figure 2.4: A shaded relief image of the same object shown in Figure 2.3. I sent this image to the participants first and 85% of the respondents reported that they identified the underlying object automatically. The identifications varied with majority perceiving it as a mountain range and others perceiving it as being either crumpled paper or a blanket.

Of the two questions, the second was the more important because the answer is indicative of whether or not the visual processing system succeeded. The term “think about it” is crucial because the visual processing system does not include conscious thought. If it functions as we would like, then we sense the object and if it doesn’t then we pass the image onto the higher brain functions for analysis.

The object that I show in both images is a mountain range, specifically, the Crowsnest Pass region of southeastern British Columbia. The color image of Figure 2.3 is essentially a variable density elevation display because elevations are mapped to color in the same way that a variable density seismic display maps seismic amplitude to color. For this example, I used a typical seismic palette (cyan-blue-white-red-yellow). Cyan represents the lowest elevations and yellow represents the highest elevations.

By contrast, Figure 2.4 is a shaded relief image of the same elevation data. Shaded relief (Batson, 1975), is a picture of the light that would reflect off a given surface for a given direction of illumination. For this image, the light source points from the upper left of the image towards the center. I sent this image to the survey participants first.

2.3.1 Survey Results

Eighty five percent of the people who responded to the survey reported that they were able to identify an object in Figure 2.4 very quickly. Most people correctly identified it as a mountain range but others saw it as crumpled paper and others as a blanket. Interestingly, there was a high degree of correlation between the respondents experience and what they reported seeing the image as. People with experience with aerial photographs identified the image almost exclusively as a mountain range whereas people in Divestco’s accounting department saw it as crumpled paper. As much as this is amusing, the correlation between occupation and identification is significant because it is indicative of the provisional nature of percepts. People perceive things according to their own personal experiences and knowledge. In this experiment, however, it was not important what they even eventually sensed the object to be. What was important was

that the vast majority of respondents had the sensation that what they were seeing in the shaded relief image was a real object.

This perception of an underlying object is in contrast with the respondents experience with Figure 2.3. I sent this image to the participants several days after the shaded relief image and even though they had already identified the underlying object, most people did not sense anything in the color. Sixty percent of the respondents reported that they could not determine what the underlying object was. Of the remaining 40%, the majority reported that the recognition was not automatic and that they had to think about the image before arriving at an answer.

This was not a rigorous survey and I caution the reader not to take too much from it. I include it here because it serves to introduce what is the most important and most fundamental fact of the visual processing system, one that psychologists have known for over a hundred years. This simple experiment shows that the visual processing system is a multi-channel system and that each channel contributes to perception but in a different way. In this example, a channel that produces the sensation of perception processes the shaded relief image. By contrast, a channel that fills in details but gives very little sensation of the object that the details apply to processes the purely chromatic image.

2.4 Primate Trivariant Color Vision

Our modern understanding of color begins with a series of experiment conducted by Sir Isaac Newton in the late 1660's. Before his experiments, people believed that color was a mixture of light and darkness. Hooke, Newton's antagonist, was a proponent of this theory and proposed a scale from brilliant red, which he believed was pure white light with no darkness added, to dull blue, the last step before black. He believed that darkness was a physical property and that black was the complete extinction of light by the hypothetical dark. Newton overturned this theory by use of the prism. In a revolutionary experiment, he first split the light into its spectrum and then refracted it back together. By reforming the original light, he proved that light itself was responsible for the color.

Newton's experiments led to an understanding of the nature of color but said nothing about how we see it. The Trichromatic Theory of Color Vision, first proposed by Thomas Young in 1802, was the first widely accepted theory of how we actually see colors. Young based his theory of color vision on the premise that there are three classes of cone receptors sub serving color vision. One of the more important empirical aspects of this theory is that it is possible to match all of the colors in the visible spectrum by appropriate mixing of three primary colors. Which primary colors are used is not important as long as mixing two of them do not produce the third.

Anyone who has looked at a television or a computer monitor will be familiar with Trichromatic color. Each pixel on a computer monitor consists of three smaller pixels, a red pixel, a blue pixel and a green pixel. By varying the intensity of the light emitted from each, the display can produce a complete spectrum of colors. We call this the additive mixing of colors and it is how we produce the colors of a computer monitor. As one might expect, however, nature is more complex and even though we have three separate color receptors in the retina, we do not combine them in the same way.

Whereas the Trichromatic theory of color explains many aspects of generating color it is seriously deficient when it comes to the human perception of color. We can use it to simulate colors but it does not explain why there are certain colors that we never see together. For example, we see yellowish-greens and bluish-reds but we never see bluish-yellows or reddish-greens. The trichromatic theory cannot explain this and we now generally accept that the theory of trichromacy only applies to the color receptors in the retina and not to our perception of color.

2.4.1 Hering Theory of Opponent Color Vision

The 19th century physiologist Edwald Hering proposed a different model for color vision. He proposed the Opponent Color Theory (Hering, 1964; Hurvich 1981) which we now accept as generally correct. Hering hypothesized that the trichromatic signals from the cones were subject to subsequent neural processing. He proposed two major opponent

classes of processing, a spectrally opponent process and a spectrally non-opponent process.

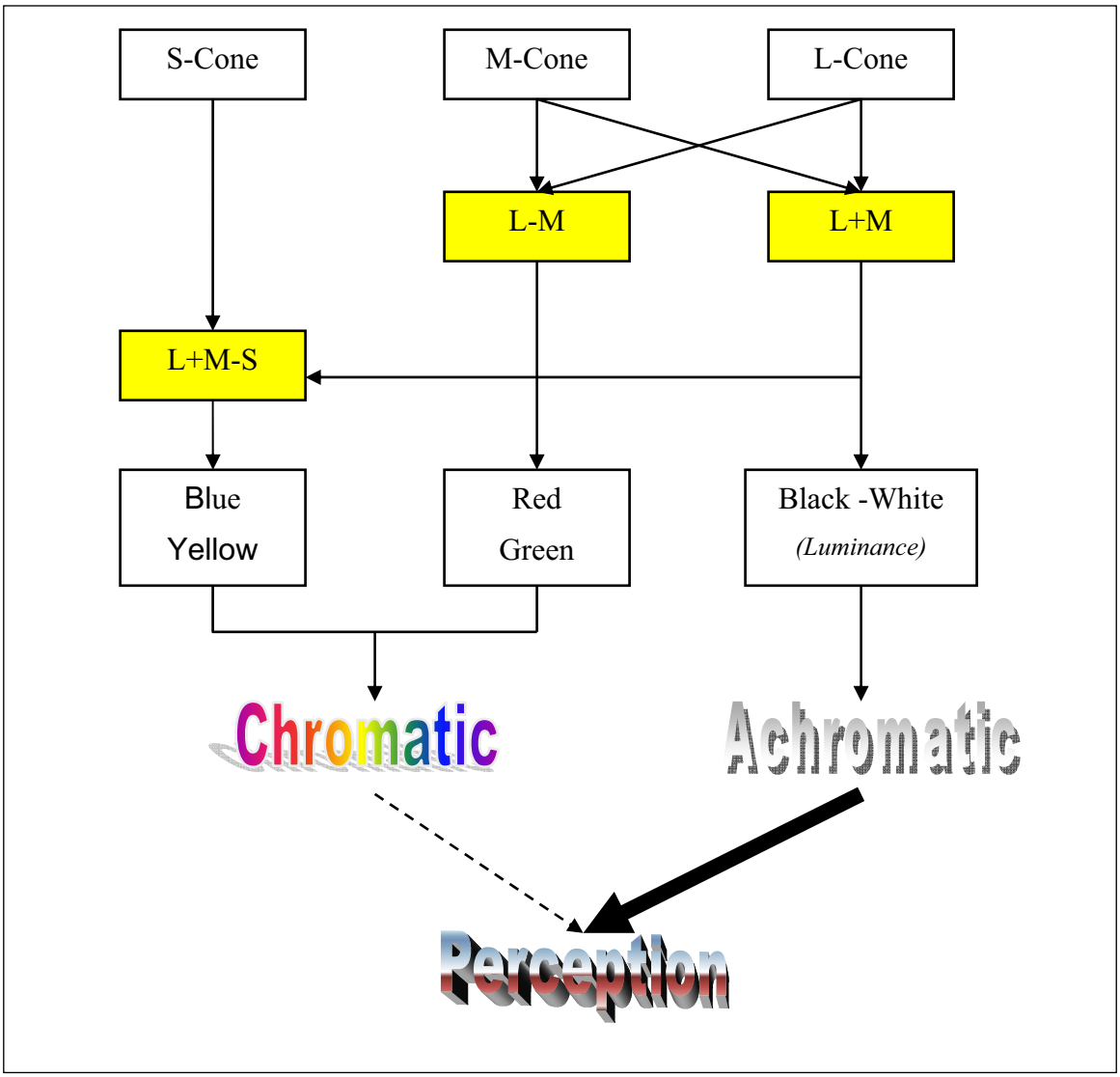


Figure 2.5: The Hering theory of Opponent Color Vision. Neural processing produces three channels of visual information each of which is processed by separate neural circuitry in the visual cortex. The first channel is the opponent black-white (achromatic or luminance) channel, it provides the bulk of our perception. The two chromatic channels, the opponent blue-yellow channel and the opponent red-green channel also contribute to perception but to a lesser degree.

In the opponent color theory, the spectrally opponent processes of red vs. green and blue vs. yellow provide our ability to separate hues. The spectrally non-opponent process produces our black and white vision. This opponent process model lay relatively dormant for many years until a pair of visual scientists working at Eastman Kodak at the time,

conceived of a method for quantitatively measuring the opponent processes responses. Leo Hurvich and Dorothea Jameson invented the hue cancellation method to evaluate psychophysically the opponent processing nature of color vision. Due in large part to their work we no longer question opponent processing. We call the modern model for how humans (and other primates) see colors "the Stage Theory" and it incorporates both the Trichromatic theory and the opponent color theory. The first stage, the Trichromatic stage, can be considered as the receptor stage, which consists of the three photo, pigments (blue, green and red cones). The second is the neural processing stage and this is where the color opponency occurs. It begins as early the first post-receptoral layer in the retina and continues through the visual system and on into the visual cortex itself.

Figure 2.5 shows, in general terms, Hering's spectrally non-opponent and opponent processes. We process the Trichromatic signals from the cones into three separate channels of visual information, an achromatic channel and two chromatic channels. Hering's non-opponent process occurs first. We then combine the signal from the L and M cones to produce the Black-White (luminance) channel. Once we produce this channel, we difference the same two inputs to produce a Red-Green channel and then we difference the luminance channel with the S cone signal to produce a third, the Blue-Yellow channel.

We subsequently process these three by two separate circuits in the visual cortex. The primary circuit, which we call the Achromatic Neural circuit, processes the intensity channel. The secondary circuit, which we call the Chromatic Neural Circuit, processes both the Red-Green and the Blue-Yellow channels. This processing of the Trichromatic cone signals into three channels of information is, in very general terms, how we see in daylight conditions. It is known as Trivariant color vision and among mammals; it is unique to Old World primates.

2.4.2 Trivariant Color Vision in Practice

In section 2.3 I discussed the results of a small visualization survey that I conducted in the spring of 2006. In this survey, I sent two images of the Crowsnest Pass (Figure 2.6)

to a group of over a hundred participants. I reproduce the images here at a smaller scale for comparison purposes.

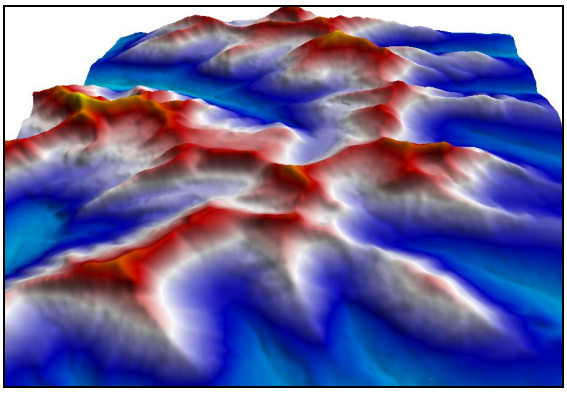


Figure 2.6: Three-dimensional view of the Crowsnest Pass using the same lighting and color used for the images in the visualization survey.

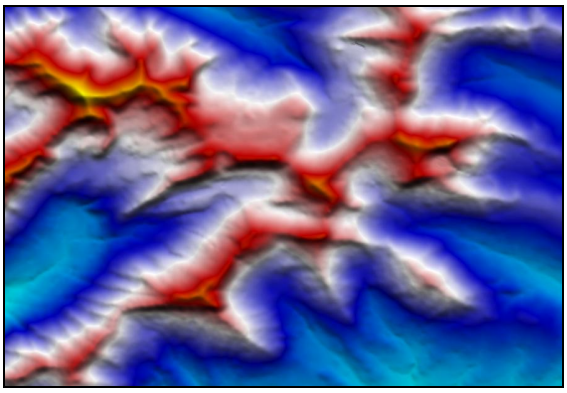


Figure 2.7: Bump mapped image formed by multiplying the color values of Figure 2.9 with the intensity values of Figure 2.8. This display is analogous to looking at Figure 2.6 from directly above (i.e. straight down).

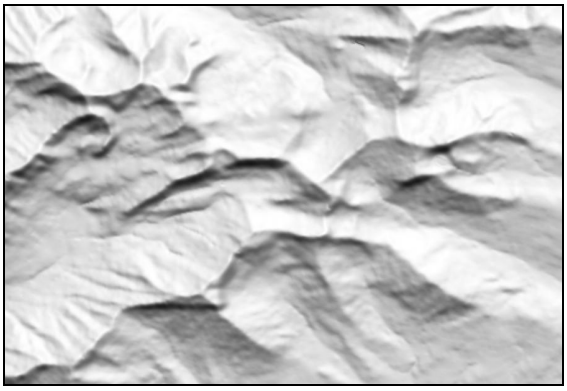


Figure 2.8: Small-scale version of Figure 2.4 shown for comparison. This image is purely achromatic and is processed by the achromatic neural circuitry in the visual cortex. Note that in comparison to Figure 2.7 the underlying perception of the mountain range doesn't change by removing the color.

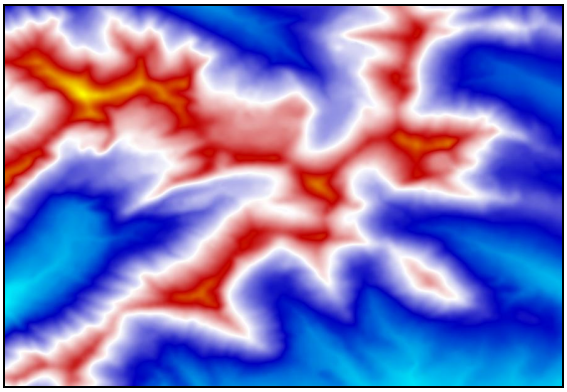


Figure 2.9: Small-scale version of Figure 2.3 shown for comparison. This image is purely chromatic and is processed by the chromatic neural circuitry in the visual cortex. Note that in comparison to Figure 2.7 the underlying perception of the mountain range is lost when we remove the lighting.

These four images are an illustration of the underlying processes of trivariant color vision. Figure 2.6 and Figure 2.7 represent "real world" views of the elevation data. The

former is a three-dimensional image and the latter a two-dimensional bump mapped[3] image. They both represent the single integrated image that we are conscious of whenever we view a scene in the real world.

By contrast, Figure 2.8 and Figure 2.9 simulate what happens to these integrated images once they enter the visual system. According to the now widely accepted trivariant theory, the visual processing system splits the integrated image[4] into three separate images. The first is an achromatic, intensity only image, which I simulate in Figure 2.8. The other two are purely chromatic images, which I simulate by the single combined image Figure 2.9. I have used a single chromatic image here instead of two because the important point is that there are two neural pathways from processing information, one for achromatic information and one for chromatic. When we look at Figure 2.8 what we perceive is the result of processing by the achromatic channel, when we look at Figure 2.9 what we perceive is the result of processing by the chromatic channel.

Comparing these images can provide an understanding of the fundamental nature of these channels. The only difference between Figure 2.7 and Figure 2.8 is the absence of the chromatic information. When we compare the two, it is clear that our underlying perception of the scene does not change because we get almost exactly the same sensation of perception in the two images. This is not to imply that nothing is lost, however, because clearly we lose details. For example, there is a small "island" like structure in the lower right corner of Figure 2.7. This percept of an island is not there in the achromatic image. Clearly, it is the colors used that make this appear as an isolated structure.

---

[3] Bump mapping (Blinn, 1978) is technique that produces the perception of three-dimensional wrinkles on a two-dimensional surface.

[4] The reader is cautioned that at no time does an "image" appear anywhere in the brain. I use the term here as a colloquialism to refer to streams of visual information.

By contrast, the only difference between Figure 2.7 and Figure 2.9 is the absence of the achromatic shaded relief information and consequently the comparison is more dramatic. This is because when we look at the purely chromatic image we lose the sensation of perception almost entirely. Interestingly, the “island” is even more apparent on the chromatic image than on the bump mapped image. Its visual appearance, though, is not a percept because it produces very little sensation of perception.

What this simple survey exposes is that the visual processing system is dependant upon both achromatic and chromatic information. In subsequent chapters, I provide a detailed description of the properties of these images as well as how they are formed and processed. For now, though, all that is important is to understand that the achromatic and the chromatic images are processed by separate but parallel neural circuits in the brain. Both channels contribute to our sensation of perception but it is the achromatic circuit dominates.

2.4.3 Trivariance and Seismic Data

The trivariant nature of vision has direct implications for our ability to communicate seismic information. There are two conventional seismic displays, the wiggle trace display and the variable density display. In this discussion, I am only interested in the variable density display because we create them using color palettes similar to the one I used throughout this chapter. I discuss the wiggle trace display, which is more achromatic in nature, in Chapter 4.

Figure 2.10 and Figure 2.11 are variable density displays, the former being of the elevation data from the Crowsnest pass and the latter being of a small section of a faulted seismic line from the Trujillo area of Peru. Both displays use the same gray-dark blue-white-dark red-yellow color palette. For the elevation data, gray represents the lowest elevation and yellow the highest; for the seismic data, gray represent the lowest negative amplitudes, white zero amplitude and yellow the highest positive amplitude. Both sets of data contain exactly the same number of vertical and horizontal samples, in this case 512 by 512.

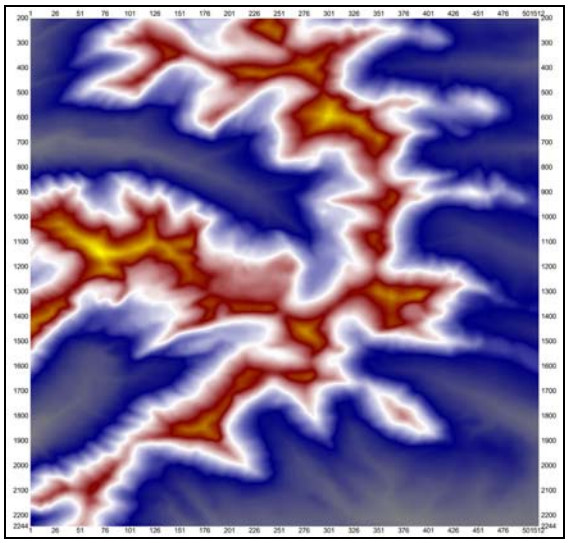


Figure 2.10: A variable density image of the Crowsnest pass elevation data using a gray-dark blue-white-dark red-yellow color palette. The data came from a 512 x 512 digital elevation model.

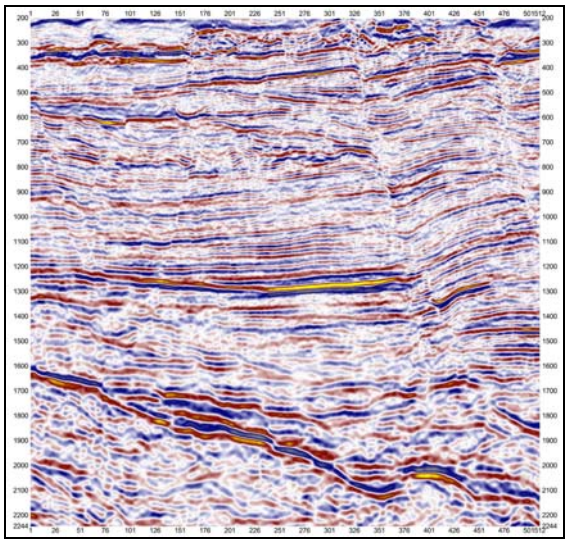


Figure 2.11: A variable density image of a small section of a seismic line from the Trujillo area of Peru (data courtesy PeruPetro). The data shown has the same number of vertical and horizontal samples as Figure 2.10 (512 traces by 512 samples) and uses the same color palette.

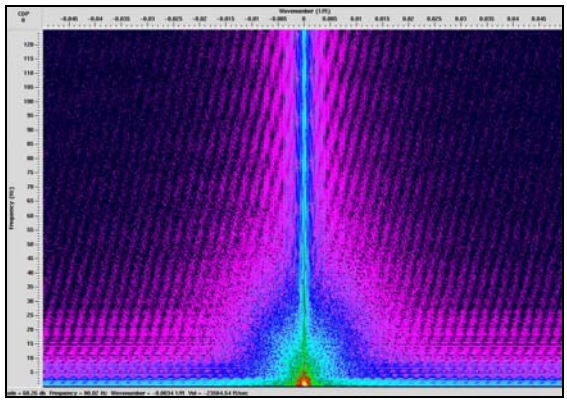


Figure 2.12: F-K Spectrum of the data shown in Figure 2.10. The elevation information is concentrated at very low spatial and temporal frequencies. We could decimate this data set several times both spatially and temporally before we lose significant information.

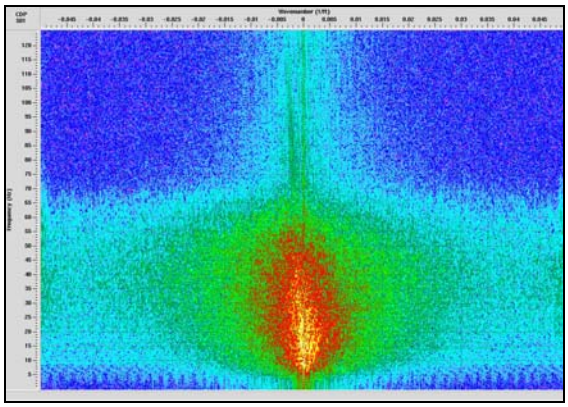


Figure 2.13: F-K Spectrum of the data shown in Figure 2.11. The seismic data is spread over a wider range of both temporal and spatial frequencies. Decimating this data by a power of two would result in a significant loss of both spatial and temporal information. This indicates that it is a more complex or information rich data source than is the digital elevation model.

Figure 2.12 and Figure 2.13 show the F-K spectrum of the two data sets out to their spatial and temporal Nyquists. They show, in F-K space, what the readers can judge by themselves by comparing the two variable density images; that in terms of information content, the seismic data is the richer data source. Visually, the seismic just looks busier;

there is a lot more going on in this section than in the image of the elevation data. In terms of the F-K spectra, it is clear that we could decimate the elevation data several times before we lost any significant information. This contrast with the seismic data because as is evidenced by the energy beyond the half-nyquists in Figure 2.13, decimating the seismic data even once would result in a significant loss of information.

All this leads to the point that a seismic section is an extraordinarily complex object. It contains a remarkable amount of information, all of which we must communicate visually. We assumed that variable density displays show us this information but in light of what we now know about trivariant vision, they cannot be; the assumption is false. If the reader doubts this then ask yourself this question:

We know what a mountain ranges look like, given that Figure 2.10 is not a mountain range then how can Figure 2.11 be a seismic section? The answer is that it is not a real seismic section at all, it is just a cartoon of a seismic section.

CHAPTER THREE: RETHINKING RESOLUTION

When all the world would keep a matter hid,
Since Truth is seldom friend to any crowd,
Men write in fable, as old Æsop did,
Jesting at that which none will name aloud.
And this they needs must do else it will fall
Unless they please they are not heard at all.
"The Vortex"

3.1 Introduction

I began my career in the fall of 1977 in the seismic processing division of Gulf Canada Resources. At the time, digital processing was still in its infancy and every week, the processing department held a show and tell in which one or more processors discussed sections they had been working on and the processing flow they had been following. The V.P. of Exploration, the Manager of Geophysics and most of the Divisional Managers, also attended these sessions.

After I had been with Gulf for a few weeks I was invited to attend the meetings and in my first meeting a section was presented, along with its processing flow which ran; demultiplex, notch filter, deconvolution, spherical divergence correction ... . I pointed out that this flow was out of order, that spherical divergence correction should come before deconvolution. Everyone else pointed out that decon before s.d. was the standard processing flow, a flow dictated, to Gulf worldwide, by the head office in Houston. Not deterred, I pressed my point and I remember quite distinctly standing at the white board and detailing just how applying s.d. after decon would, to use the exact expression, "degrade the resolution of the section". Even though I was very junior the managers were so concerned by my proof that I was asked to rework the section in question using s.d. before decon and to show the results at the next meeting.

I did as they asked and as one would expect, the section was dramatically improved over the first. In fact, the difference was so great that it looked as if decon had been applied to it for the first time. The results were incontrovertible and with that one simple

test, I proved to all present that data processed using Gulf's then standard processing flow, was, as I put it, "a very poor substitute for seismic data".

The room went very quiet! Nobody spoke for a few minutes until finally the Exploration V.P. turned to the Processing Manager and asked how long we had been using the old flow. About 18 months he replied. "You mean", said the V.P., "that for the past 18 months all of our data has looked like that" (pointing to the old section) "when it should have looked like that" (pointing to my section)). "That would be correct," said the manager. "Ok", said the V.P. with respectable self-control, "from now on we will use Steve's new processing flow on all our data". He then turned to me and said casually, "Thanks Steve, that was good work – you can go now". I bent down to pick up my sections but when I did, he placed his hand upon them and looked me straight in the eye. "Steve", he said, "leave the sections!"

The point of the previous story is to set the stage for the following discussion on the relevance of visualization to seismic resolution. The early part of my work on SeisScape displays involved two struggles. The first (detailed in the first chapter) was my struggle to extract the science behind the displays from their engaging nature. The second, which I talk about in this chapter, was my struggle to establish the relevance of the science once I had extracted it. I will confess that when I started work on this thesis I had yet to prove to myself that SeisScape displays were anything more than an engaging curiosity.

The story that I begin with is about technology or more precisely, about how the mis-application of technology can have dramatic and detrimental effects upon seismic resolution. It illustrates a fundamental law of geophysics that resolution is hard to find but very easy to lose. In exploration seismology, we live and breathe resolution. All of our research is dedicated to improving it, all of our processing efforts are dedicated to producing it and all of our interpretation efforts depend upon it. Moreover, we think that we understand it and to a very large part, we do.

In 1977, however, we did not understand it all that well. Applying s.d. after decon to surface seismic data is so obvious a mistake that nobody today would make it.

Nevertheless, back then, digital processing was less than a decade old and most of the managers and interpreters had started their careers in the analog recording era. The very basics of digital processing, statics, velocity analysis, deconvolution and cdp stacking, made such an improvement to their ability to resolve the subsurface that they didn't recognize the simple mistake they were making. Everyone was so pleased with the improvement; they did not realize they should have been getting a whole lot more!

I remember that week quite well. Almost all of the divisional managers came down each day to see how far I had progressed but they did not pressure me to come up with one answer or another. There was a certain amount of ambiguity in what they wanted me to prove. In the short term, it would have been better if I were wrong because the consequences of my being right were ominous. In the long term, however, it was better for me to be right because it meant a dramatic improvement in the data. To be fair most of them had accepted I was right because I had proved my point with the math. I was fresh out of graduate level courses in time series analysis and was up to date with the theory. This recent knowledge gave me the confidence to speak up in a very intimidating situation, I knew I was right because the math said so – and I knew the math.

This chapter begins my consideration of the relevance of visualization and I have begun it this way because I need to draw a parallel between where I was then and where I am now. Now as then I will make a statement about resolution that will challenge established concepts and at first be very hard to accept.

Conventionally, we consider that there are two principal forms of resolution; temporal which is the ability of the seismic wavelet to resolve reflections (in time) from thin beds and spatial which is the ability of the wavelet to resolve closely spaced geological details. It is a principal theme of this dissertation that there is a third form of resolution, namely visual resolution, that if ignored and not understood can have a significant negative impact upon seismic resolution. I will prove that visualization plays a critical role in establishing seismic resolution. I will also prove that because we have all but ignored it to this point in time, the resolution of every seismic record, section or time

slice that you have ever seen is significantly lower than it should have been. By ignoring visualization, we have filtered out an entire level-of-detail of relevant information.

With this statement, I say essentially the same thing now that I said in 1977. The difference is that this time I do not have the mathematics to back me up and because of that, I cannot prove the point mathematically. The theoretical aspects of visualization, which I expose in later chapters, involve mostly biological sciences and computer sciences and there is very little in the way of familiar equations. Because of that, unlike deconvolution or migration, I cannot point to any single equation and predict what it will do.

My initial difficulties in establishing the relevance of visualization were a consequence of my mindset. I was fixated on developing the theory first and I could not establish any theoretical link between visualization and resolution. In trying to follow this established path, I missed the point entirely and I should not have because I made it back in 1977. Back then, I initially proved my point theoretically but even so, nobody would accept it until I had proven it empirically. I had to show a before and after example; the theory was important but proving its effect upon the data was critical.

Ultimately, I gave up trying to establish the theory first and went backwards. I started with empirical results and extracted the theory from them; and that is what I do now. I start with the empirical proof of the importance of visualization to seismic resolution and then I develop the theoretical aspects that underlie it. Before I do anything else, I will show you the empirical results that finally convinced me just how relevant visualization was.

3.1.1 Note on Display Scaling

Throughout the rest of this thesis, I will often make comparisons between SeisScape displays and conventional seismic displays with particular emphasis on comparing SeisScape displays with wiggle trace displays. This presents a problem in that to be directly comparable the two images must show the same data. However, the optimal scales for both types of display can be very different. For example, Figure 3-1 to Figure

3-4 compare a wiggle trace display of small channel with a SeisScape display of the same data, first using a relative display scale of 12 tpi and 7.5 ips and then at an expanded scale of 12 tpi and 30 ips.

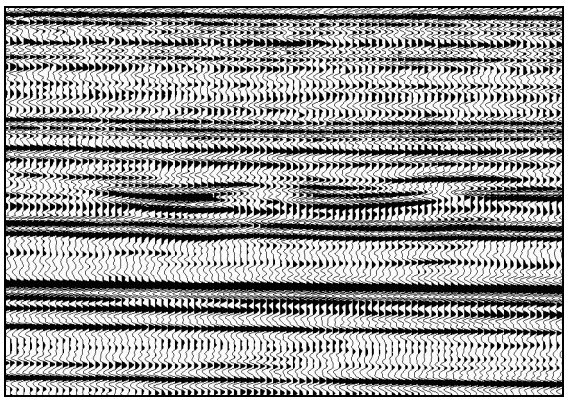


Figure 3-1: Wiggle trace display of a small section of data over a channel. Relative display scale is 7.5 ips and 12 tpi. This scale or one very close to it is often used when interpreting stratigraphic plays.

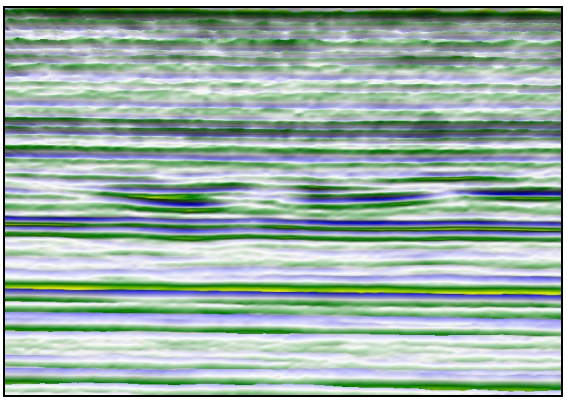


Figure 3-2: The SeisScape display equivalent to Figure 3-1. This compressed time scale is not "natural" for the display and degrades its appearance.

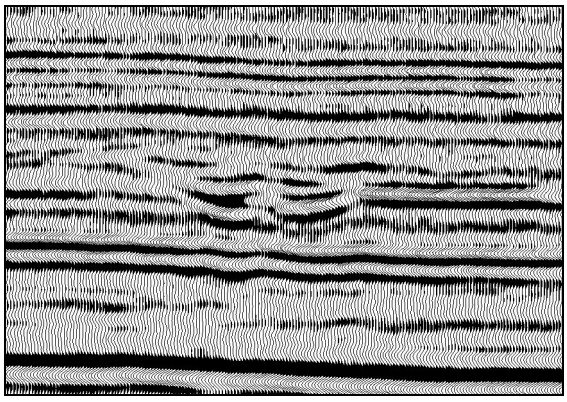


Figure 3-3: Compressed x-scale display of the channel. Relative display scale is 30 ips and 12 tpi. Scales of this type are used on structural data but they tend to degrade the appearance of smaller, more subtle stratigraphic displays.

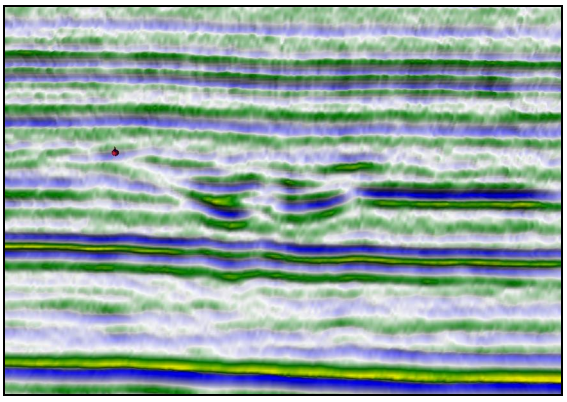


Figure 3-4: The SeisScape display equivalent to Figure 3-3. SeisScape displays have a natural scale, which is greatly expanded vertically. At this expanded scale, the displays look better and reveal more of the subtle features of the data.

Although each interpreter has their own preference for display scaling, in general, stratigraphic plays use a compressed time scale such as the one used in Figure 3-1 and Figure 3-2. This compressed scale, however, does not translate well to SeisScape displays, which are far more effective at the expanded scale shown in Figure 3-4. SeisScape displays, being three-dimensional, have a correct or natural display scale. Unlike wiggle trace displays whose scaling tends to be personal and to depend on the

type of play, the optimal SeisScape display scale tends to be constant and not to depend upon the type of play under examination.

Unless otherwise noted I will by convention use the display scale that is most appropriate for the SeisScape display. If this results in a significant degradation of the wiggle trace display, I will display it again at a more appropriate scale.

3.1.2 Notes on the Color Palette

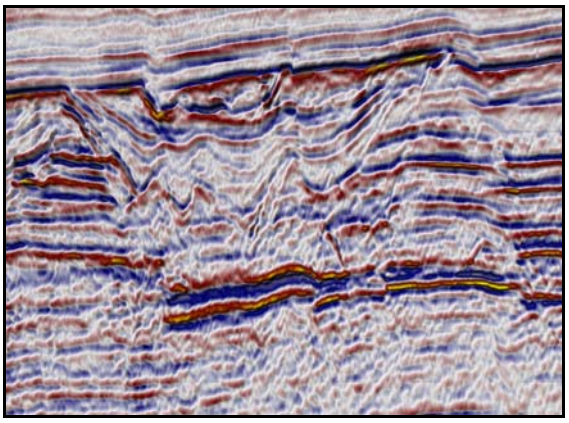


Figure 3-5: SeisScape display using an industry standard color palette that relies heavily upon blue, white, red and yellow.

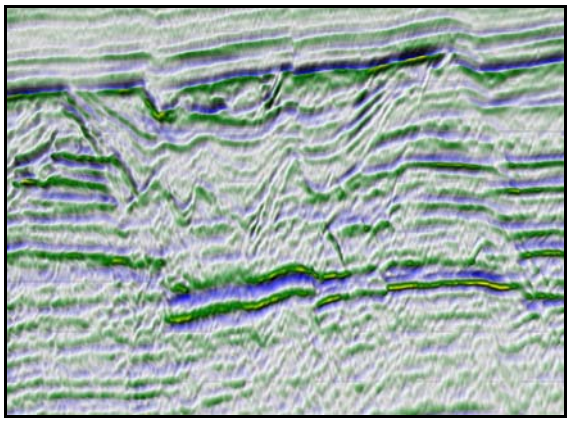


Figure 3-6 : The same display as shown in Figure 3-5 but colored using a palette based upon research into both how and why primates have developed trichromacy.

Throughout the remainder of this thesis, I rely heavily upon a seismic color palette (Figure 3-6) and others like it that I developed as part of my research into primate color vision. The palette, known after this as palette *HA-1*, is very similar to ones in standard use throughout the industry (Figure 3-5) in that it uses blue for negative amplitudes, white for zero amplitude and yellow for the highest positive amplitudes. Where it differs is that it uses a shade of green for low and intermediate positive amplitudes.

HA-1, which I discuss in detail in Chapter 11, is based upon an understanding of both how and why Primates developed trichromacy. The green background for positive amplitudes is not by accident! Although this palette is less visually appealing than the standard palettes, it does play a very significant role in enhancing our ability to form visual percepts from seismic events. As I discussed in Chapter 1, visualization is about

the communication of information and in that context, *HA-1* is one of the most important components.

As a final note, the shade of green used is important and was established on a computer monitor. The shade, in print, is somewhat darker than it appears on the screen.

3.1.3 Level-of-Detail

One of the terms that I use frequently is the term "level-of-detail" which I have borrowed from computer graphics. In computer graphics, it refers to the level of geometrical detail of objects in a scene. As an object moves away from the viewer in a scene, it covers fewer and fewer pixels. Eventually it reaches a point where the geometrical details of an object start to overlap a single pixel. Once this happens, the object's appearance starts to degrade. To get around this problem different levels of object's geometry are created each containing a subset of the previous level's geometry. As the object moves away from the viewer, the software performs a test to see how many pixels the object covers and then renders an appropriate level-of-detail of the objects geometry.

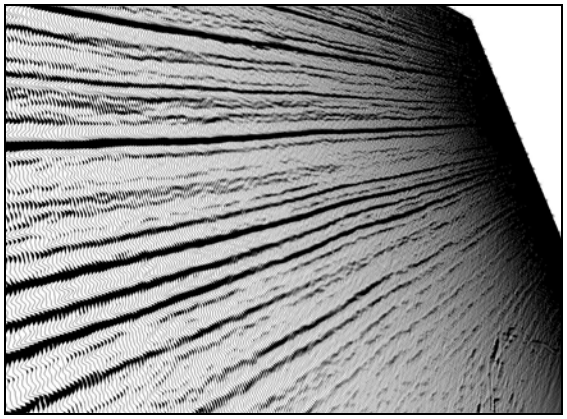


Figure 3-7: Wiggle trace display showing every trace. Notice that as the traces recede from the viewer that they overlap the same pixels and obscure the data.

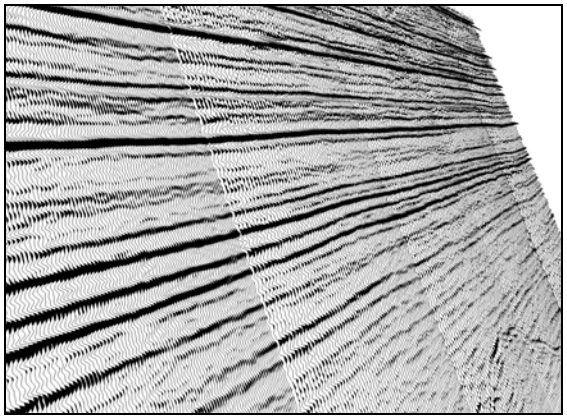


Figure 3-8: Wiggle trace display showing discreet levels of trace decimation. By decimating the traces by discreet amounts, the data in the section remains visible even as the traces recede from the viewer.

I graphically illustrate this concept for seismic data in Figure 3-7 which shows every trace of a rotated wiggle trace display and Figure 3-8 which shows the same data but with discrete levels of trace decimation. Each panel in Figure 3-8 would, in computer graphics,

be referred to as a level-of-detail. They are analogous to the geometric levels-of-detail in that they essentially reduce the complexity of the display in order to make it viewable at a distance.

The term level-of-detail implies that complexity changes by discrete amounts and that changing levels gains or loses an entire set of features. I use the term here to indicate the effect that various technologies have upon seismic resolution. Technologies such as deconvolution and migration produce level-of-detail increases in seismic resolution. This concept is illustrated in Figure 3-9 and Figure 3-10 which are the structure stack and the time migrated stack respectively of a sub-basement structure (data courtesy Divestco Inc.)

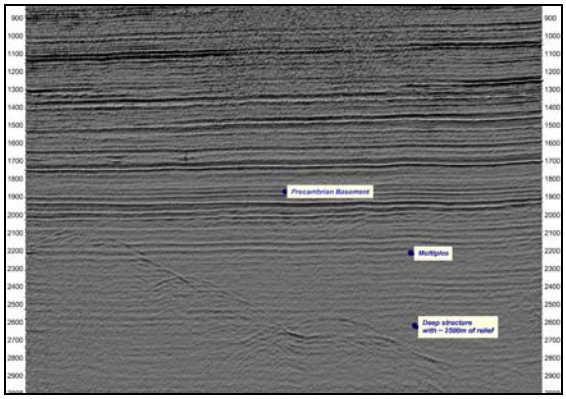


Figure 3-9: SeisScape display of a sub-basement structure shown using a gray-scale color palette.

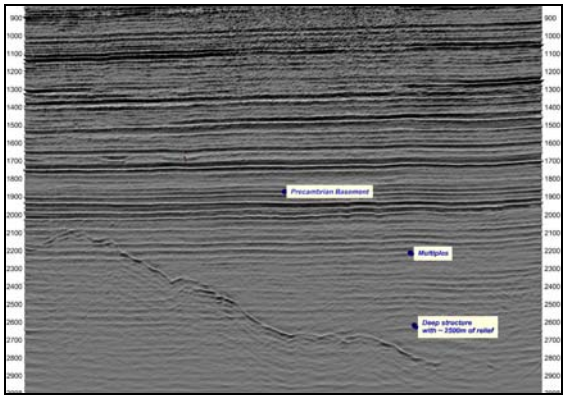


Figure 3-10: The same data shown in Figure 3-9 but after time migration.

The structure in question in this data is below the Precambrian. Consequently, the processor did not consider it when determining the migration velocities. As such, it is probable that the structure is poorly migrated. However, even migrating with incorrect velocities has increased the level of resolution by an entire level-of-detail. The same is true of technologies such as deconvolution and coherency, which also provide discrete jumps in resolution. When referring to objects in a scene, each level-of-detail provides a discreet change in geometrical complexity of an object. When referring to seismic data, each level-of-detail provides a discreet change in complexity of seismic resolution.

3.2 The Trujillo Dataset

Ultimately, I proved the link between visualization and resolution empirically but I had to work with SeisScape displays for a number of years before that happened. It took so long because in those early years the engaging nature of SeisScape displays dominated my perceptions. When it came to the science of them I had no idea what I was looking at and no idea what I was looking for. To use an analogy, I wandered about like a neophyte prospector, lost in the woods with no idea where he was, where he was going or what he was even trying to find. After 25 years of looking at seismic data through conventional displays my mindset did not allow that they were filtering out significant levels of resolution. If we think of the display as a filter, then it had been applied to every section that I had ever seen and because of that, I had no concept that there was anything else to see.

I spent considerable time looking for an empirical proof of whether or not SeisScape displays were better than conventional displays but because I did not know what conventional displays filtered out, I did not know what to look for. What I needed was a dataset that, to go back to my theme, hit me in the head hard enough that I finally saw the obvious. That data set was the Trujillo data set from PeruPetro. Although I show other data sets throughout, Trujillo is by far the most important because it was the first one that graphically illustrated just how much pertinent detail conventional displays filtered out. By itself, it changed my mindset and set me on the path to finally understanding the relationship between visualization and resolution and it proved empirically just how important the relationship was.

PeruPetro released Trujillo for research purposes a number of years ago. It is always easy to get seismic data to look at and to work with but getting data that one can show to others is difficult. In my position, I was able to obtain data from a wide range of areas but most of it was private and consequently I could not show it. This removed my incentive to study it in depth and most of the data that I initially looked at I only gave a cursory inspection to. However, Trujillo was different, I could show it and so I spent more time studying it in depth. When I did, everything became clear. I can only say that I am

particularly appreciative of PeruPetro for making such an extensive and excellent data set available for research. Without it, I may never have realized just how important the display is to establishing resolution.

The Trujillo Basin itself is geologically fascinating. I took the following description of the basin from an internal PeruPetro document, which, among other things, has the distinction of using an early SeisScape display (circa 2001) for its cover image. “Trujillo is a wrench-type basin that developed as a series of ‘en echelon’ narrow transtentional gashes attributed to left-lateral displacement along a subduction-parallel NNW-trending slip fault during the Tertiary. These strike-slip basins are characteristically narrow and relatively deep, and are connected via basement horsts or extensional faults, depending on whether the strike-slip faults are left stepping or right stepping. The evolution of the Trujillo Basin can be described by a series of compressive and extensional events generated by wrench tectonics”. In simple terms, the Trujillo basin is a microcosm of geological features. It contains extensive stratigraphic traps, horst and graben structures, complex folding and very subtle antithetic faulting.

The seismic lines in the project are large by our standards, some ranging up to several hundred kilometers in length. They are also quite deep with structures as deep as ten kms or six to seven seconds (two way time) in some places. Most importantly, however, the Trujillo data is some of the finest seismic data that I have come across. The data quality throughout the survey is excellent and the processing appears to have done a respectable job of bringing out many of the more subtle geological features. This combination of geological complexity and exceptional data quality means that throughout the entire dataset there are relevant geological features visible on SeisScape displays that are not visible on any other type display. To go back to my prospector analogy, Trujillo is the motherlode of visualization. It taught me what to look for and it played the crucial role in establishing that visualization, far from being a curiosity, was the third element of resolution.

Figure 3-11 and Figure 3-12 are images of Trujillo data and without going into detail, they indicate exactly why this data set was so important. The first image is a

conventional wiggle trace display; the second is a SeisScape display. I display both images at a relative scale of 12 traces per inch and 35 inches per second, which is optimal for the SeisScape display. For the reasons discussed in 3.1.1 the wiggle trace display shown above would be better displayed at 15 ips but is rendered at the same scale as the SeisScape display to facilitate comparison.

I have also decimated the wiggle trace display to show only every second trace. I did this because displaying every trace degrades the image on paper. Although this trace decimation reduces the effectiveness of the wiggle trace display, the comparison is still valid. Effectively displaying this amount of data on a computer monitor requires a trace decimation factor of four to five. Figure 3-11 shows less detail than a large-scale paper plot but far more detail than you would see on a computer monitor. Whereas I have decimated the wiggle trace image, the SeisScape image of Figure 3-12 shows every trace, decimation is not necessary. At the time of writing, it represents about the best SeisScape image that I can produce.

There is nothing particular in this image that I will ask you to look for. There are in other areas but I have deliberately chosen this particular area because it is free of any obvious features that might dominate the comparison. All I would like you to do is to spend a few minutes studying each image. Then, once you are familiar with both ask yourself the following. Consider that you are as familiar with the SeisScape image as you are with the wiggle trace image and consider that you are as unfamiliar with the wiggle trace image as you are with the SeisScape image. Under those circumstances, if someone showed you the wiggle trace display what would you think it had done to the data? Would you conclude that it had filtered out an entire level-of-detail?

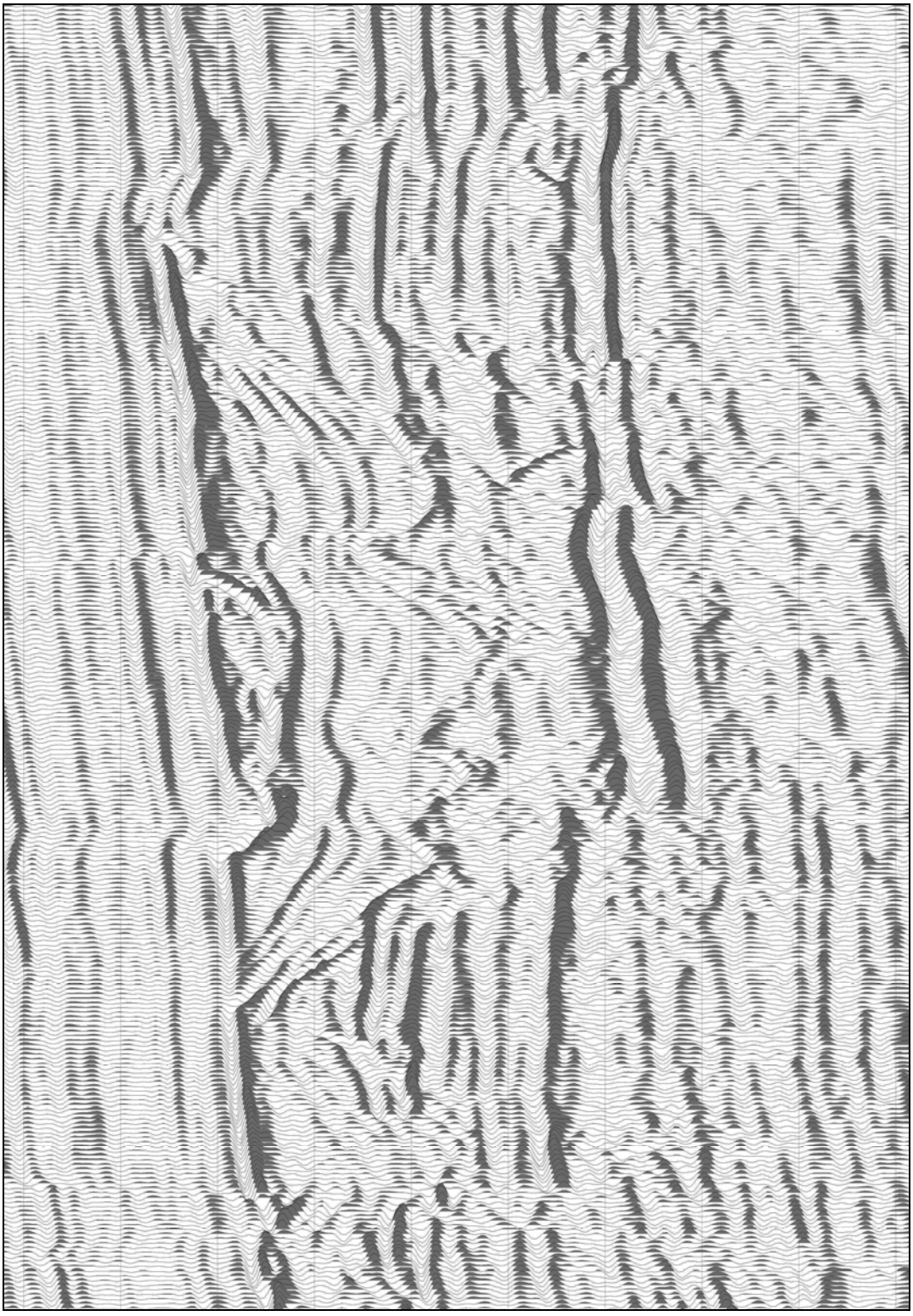


Figure 3-11: Wiggle trace display of a faulted data set, data courtesy PeruPetro. The display is shown at a relative scale of 12 tpi and 35 ips. Timing lines are 100ms apart and the trace spacing is 12.5 m.

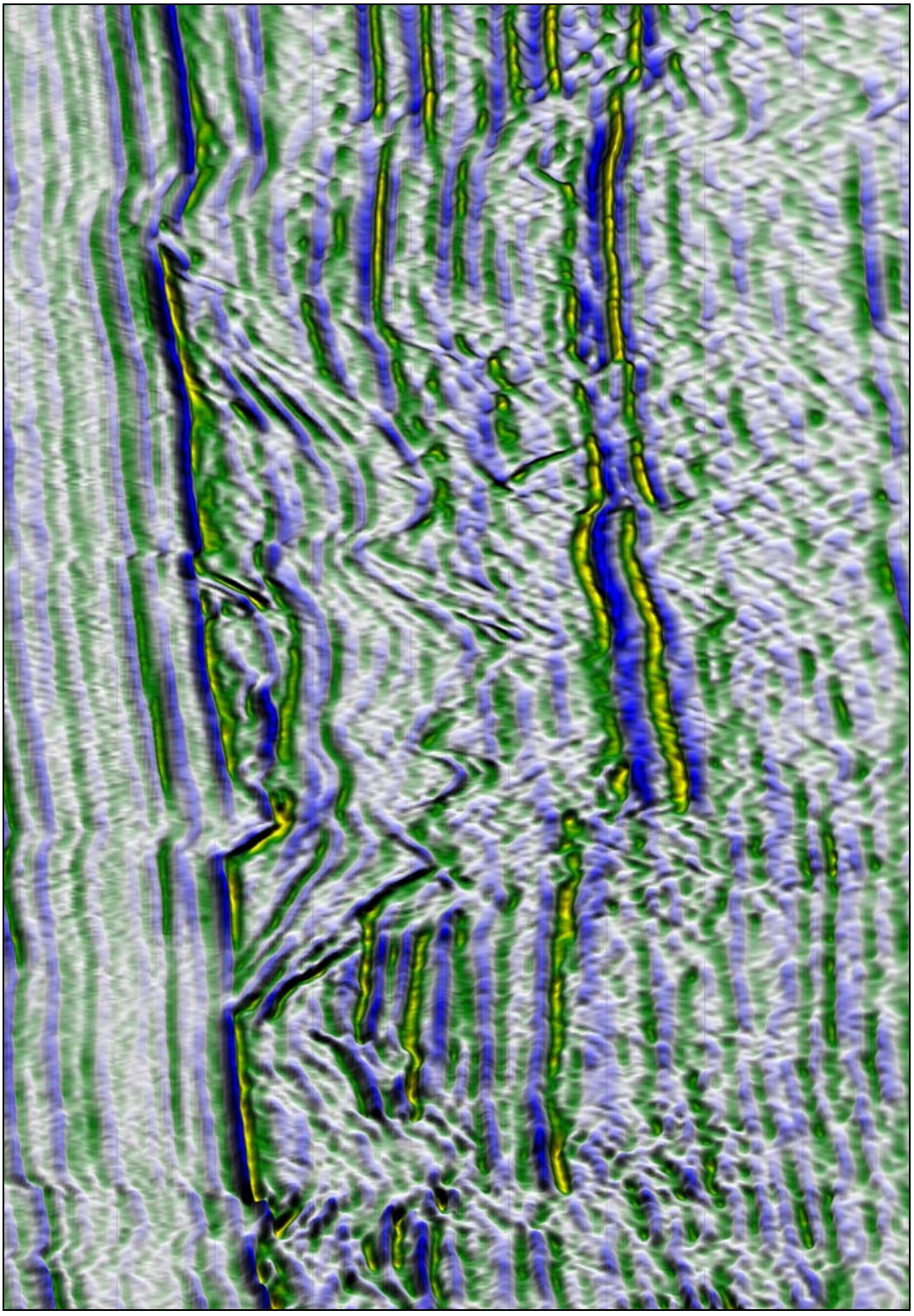


Figure 3-12: SeisScape display of the same data shown in Figure 3-11.

3.3 The Seismic Display Filter

The purpose of this chapter is to establish the link between visualization and resolution and to that end Figure 3-11 and Figure 3-12 by themselves go a long way towards establishing it. On one level, the SeisScape display shown in Figure 3-12 represents an end because it is the best display that I can produce now. On the other hand, it also represents a beginning because despite its obvious improvement over the wiggle trace display it is far from perfect. In order to make improvements we need to place the seismic display into its correct context with respect to resolution.

For the purposes of this thesis, I have chosen to treat the display as a filter because in terms of its effect upon our perception of resolution, that is exactly what it is. It is important to understand at the outset that the display filter does not directly affect seismic resolution, which is a property of the data. It only affects the perception of seismic resolution and in general, it degrades it. This is the concept that I was establishing with Figure 3-11 and Figure 3-12. The level of seismic resolution perceived in the two displays is dramatically different and yet the input to both displays is identical, with the exception of the trace decimation which is necessary on the wiggle trace display, they both show exactly the same samples.

To put things into the correct frame of reference I now define two terms that I use throughout the text:

Absolute Resolution

The collection of geologically relevant coherent events that leave a discernible imprint upon the seismic section.

Apparent Resolution

The subset of the geologically relevant discernible events that produce visual perceptions.

That the display is a filter upon seismic resolution may be a difficult concept to grasp. Consider Figure 3-13 and Figure 3-14 which are wiggle trace displays of the same small channel. The only difference between them is that Figure 3-14 is low-pass filtered. Even without knowing the filter applied, we would still have no difficulty determining that someone had applied one. This is because we are using the same type of display for each and consequently we can focus on changes in the data itself.

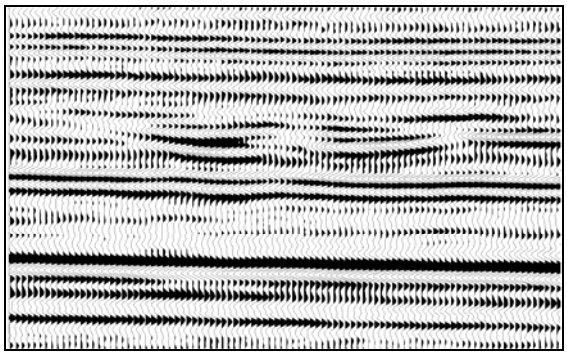


Figure 3-13: Wiggle trace display of a small channel, data courtesy unnamed source.

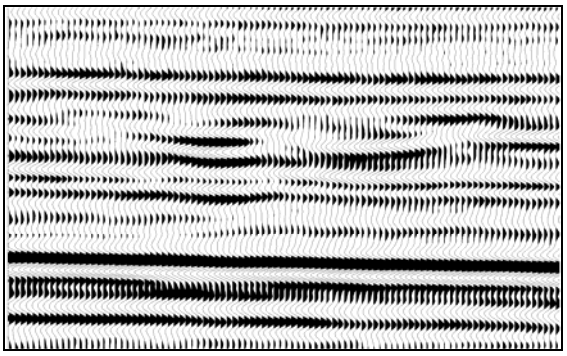


Figure 3-14: The same data shown in Figure 3-13 but low-pass filtered.

By contrast, consider now the five images Figure 3-15 to Figure 3-19. Whereas the previous images showed different sets of data using the same type of display, these images all show the same set of data, a small section of a Trujillo line, but using different displays (all of the displays use a relative scale of 12 traces per inch and 30 inches per second). In sequence, the displays are; (1) a SeisScape display using color palette HA1 and with the lighting oriented from the upper right side of the section, (2) a variable density display using a gray-scale color palette, (3) a variable density display using a blue-white-red color palette, (4) a wiggle trace display showing every trace (to match the resolution of a paper section), and (5) a wiggle trace display decimated to show every third trace (to match the resolution of a computer monitor).

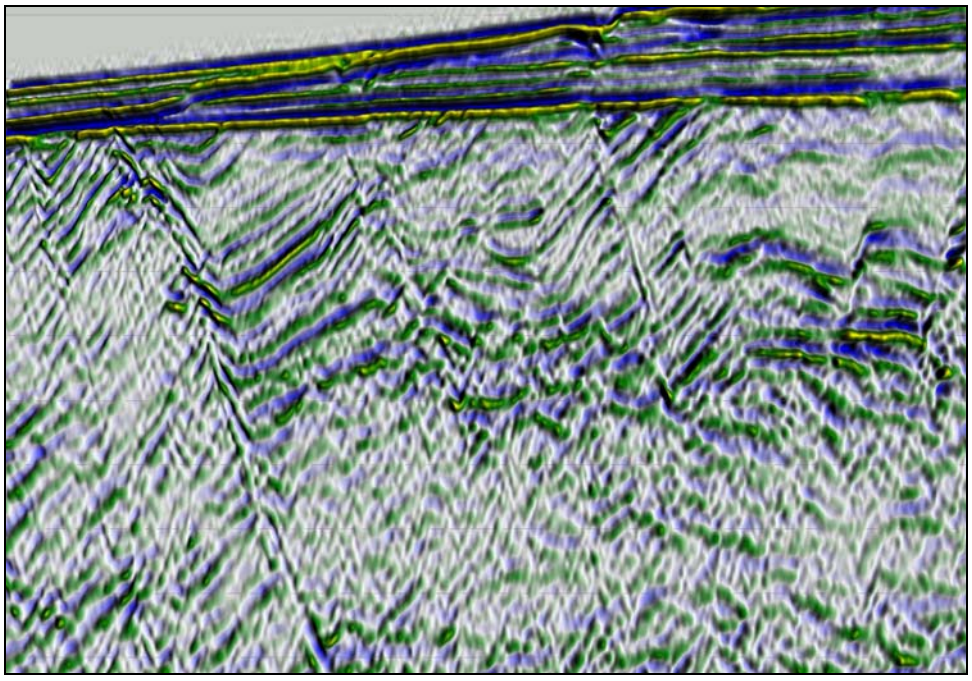


Figure 3-15: A SeisScape display of the same data shown in Figure 3-19. The display is colored using color palette HA1 and with the lighting oriented from the upper right side of the display.

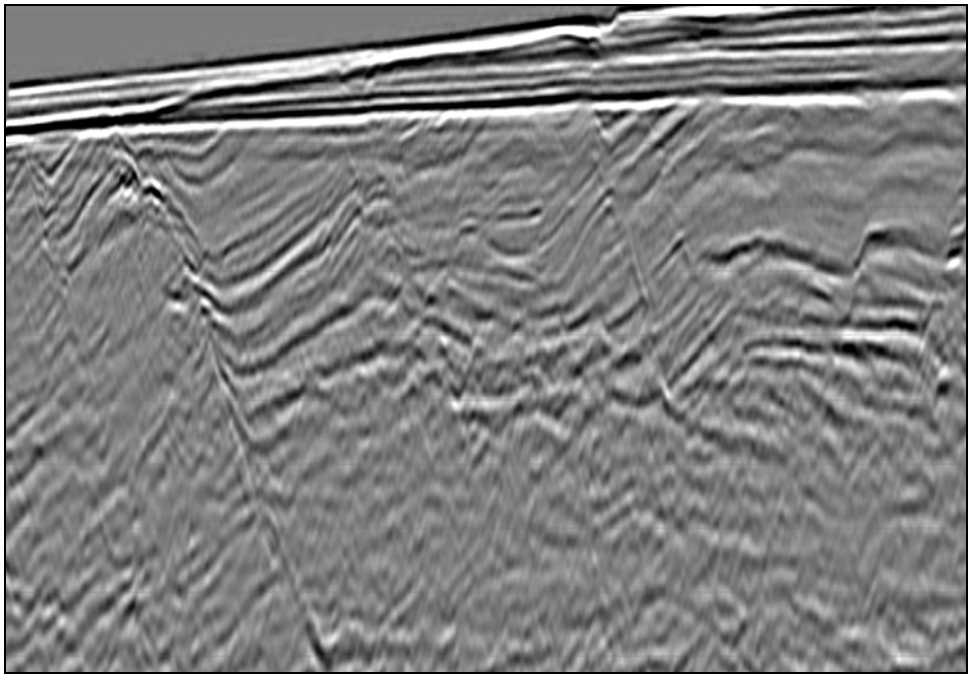


Figure 3-16: A variable density display of the same data shown in Figure 3-19. The display uses a gray scale color palette with black representing extreme negative amplitudes and white representing extreme positive amplitudes.

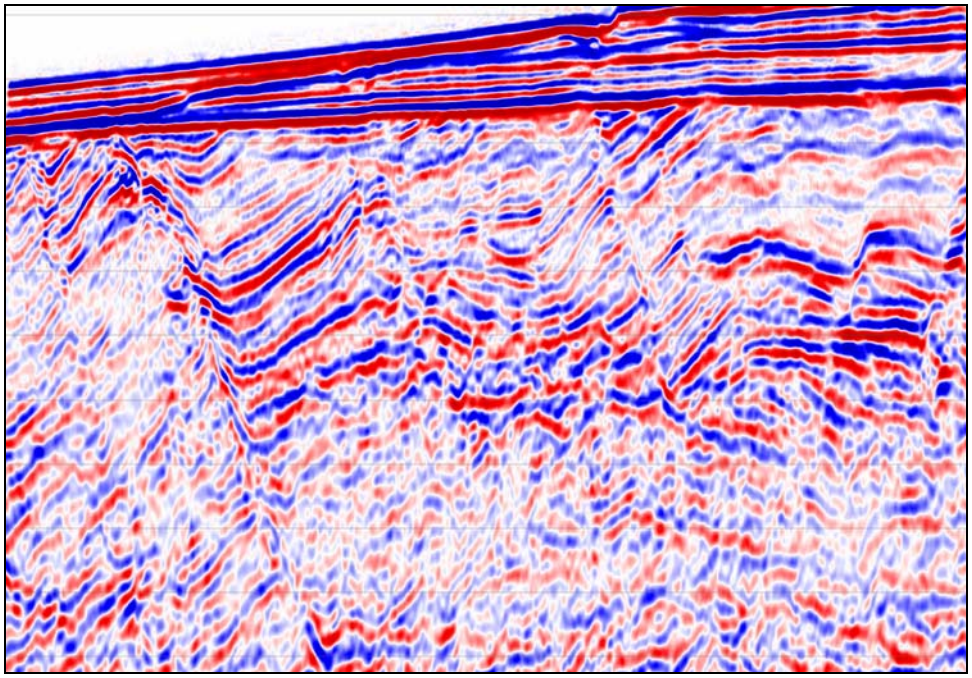


Figure 3-17: A variable density display of the same data shown in Figure 3-19. The display uses an industry standard blue-white-red color palette with blue representing extreme negative amplitudes, white representing zero amplitude and red representing extreme positive amplitudes.

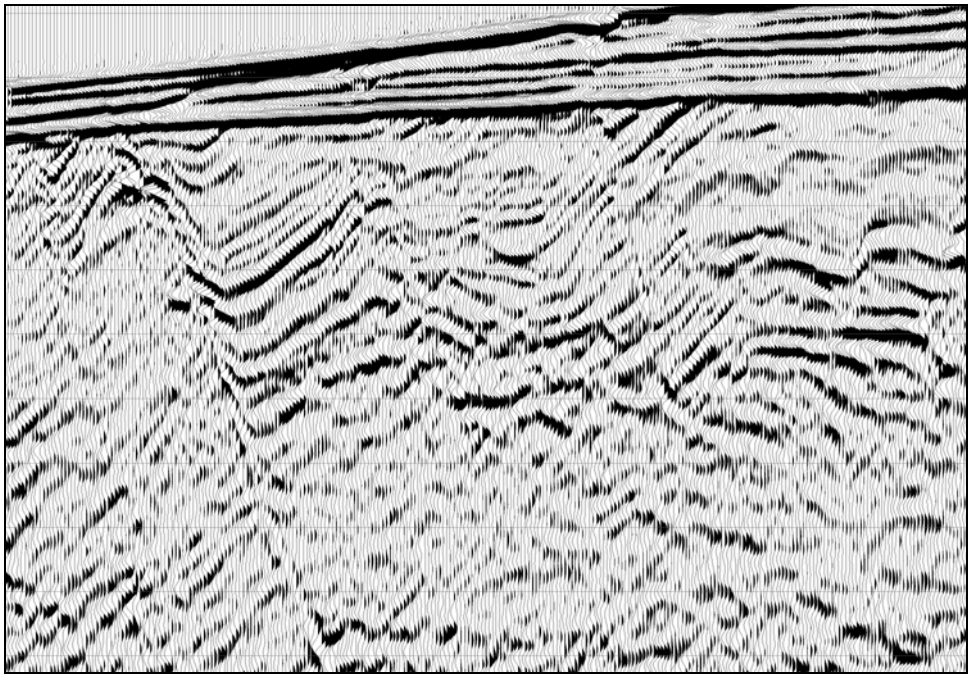


Figure 3-18: A wiggle trace display of the same data shown in Figure 3-19 but showing every trace in the section. This mirrors how the data would look when displayed on a printer. Typically, printers are higher resolution than computer monitors and can display more traces per inch without degrading the image.

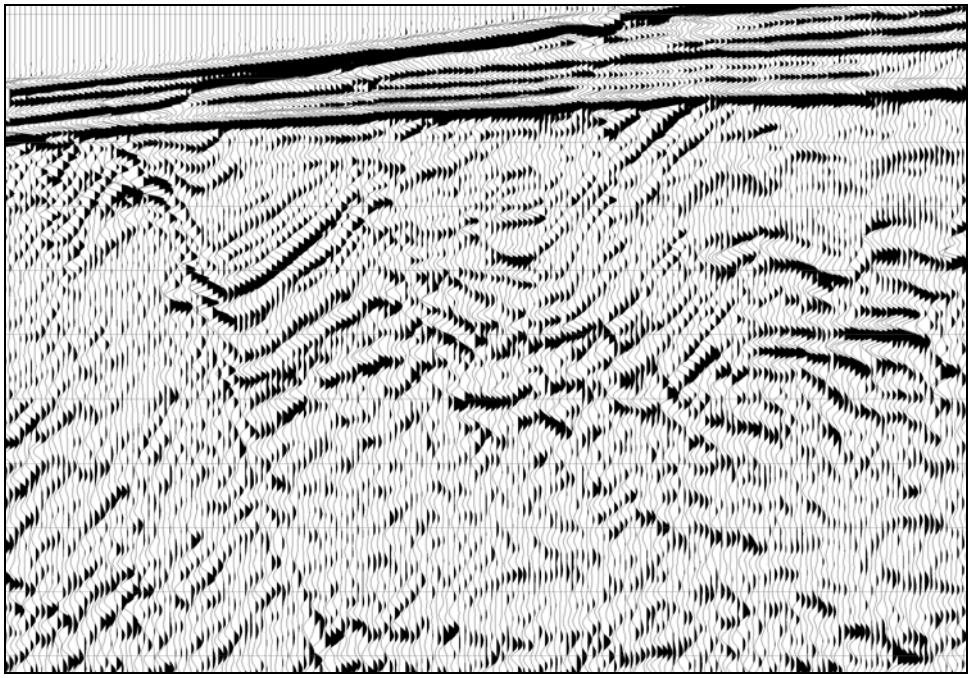


Figure 3-19: A wiggle trace display of a section of highly faulted data, data courtesy PeruPetro, showing every third trace. The trace decimation mirrors that needed to display the data effectively on a computer monitor, which typically has a resolution of 72 pixels per inch.

As you progress through these images, it is difficult to perceive that each is a filtered copy of the same data. This is because each display is significantly different from its predecessor and their purely visual differences seize our attention. The fact that each display is so different makes it hard for us to perceive changes in the underlying data, or rather our perception of the underlying data. With that in mind, please go back to the SeisScape display and then proceed through the displays again, one at a time. This time, however, ignore the form of the display and concentrate entirely on what you can see in them.

By doing that you will see that, as you progress through the images, what you see of the data in each image generally decreases. If, focusing on the data itself and not the form of the display, you compare the SeisScape image with the final decimated wiggle trace image it is clear that, in your perception, the data itself changes just as much as it did between Figure 3-13 and Figure 3-14.

In many respects, the display filter is similar to all of our other filters in that it operates upon the data to affect resolution. Conceptually, however, it is very different. The primary difference between the display filter and conventional filters is that the target of conventional filters is always the data. Conventional filters act upon the data and produce, hopefully, better-resolved data. The target of the display filter, however, is perception and not the data itself. The filter may act upon the data but ultimately the effect is upon our perception of resolution. The rule for conventional filters is; data in – resolution out (hopefully enhanced). The rule for display filters, however, is; resolution in – perceived resolution out (hopefully less degraded).

The last point, that our perception is hopefully less degraded, implies that the effect of display filters is always to degrade resolution. As a rule that is correct. This leads to a second conceptual difference between conventional filters and display filters. Conventional filters begin with low-resolution data and apply the filter to increase the resolution. Display filters start with the highest resolved data and apply the filter to produce the highest resolved perceived resolution. Unlike conventional filters, which we design to have maximum effect, we must design display filters to have a minimum effect.

There is one final point to make here about display filters and that pertains to coherent noise. SeisScape displays enhance the ability to see steeply dipping events. This includes such things as fault traces but it also includes coherent noise trains such as migration artifacts. For example, migration artifacts are particularly evident throughout the SeisScape display shown in Figure 3-15 and they visually contaminate our perception of the data. Much of what you can see on Figure 3-15 is geologically relevant but much of it is not, which raises the question if this is a weakness of the display filter.

For the purposes of this thesis, I take the approach that enhancing the ability to see coherent noise trains is not a weakness. I consider that the display filter should not differentially filter coherent signals; that is the job of conventional filters. The display filter bridges the gap between data and perception. Irrespective of whether a signal arises from geology or is an artifact of processing or noise, provided the signal is coherent the display should render it perceptible.

3.4 The Three Forms of Seismic Resolution

In the previous section, I showed that the display acts as a filter upon our perception of resolution. In this section, I lay the foundation for how to consider visualization in relation to the conventional sciences of seismic resolution.

Typically, when we consider seismic resolution we consider it in one of two contexts. The first context is temporal resolution, which is the ability of the seismic wavelet to resolve thin beds. The second context is spatial resolution, which is the ability to resolve closely spaced geological details. What is of particular interest to explorationists is what are the limits of those resolutions; just how thin a thin bed can we resolve and just how small a geological detail can we see?.

Temporal resolution is the ability of the seismic wavelet to resolve thin beds. A seismic signal is never a spike, we do our best to make it resemble one but even in theory, there are limits as to how close we can come. The best we can achieve is a zero-phase band limited representation of a spike and this necessarily imposes limits as to how small a geological anomaly we can resolve. There are various criteria for how we define this mathematically but all depend upon the wavelength of the wavelet. One of the most widely accepted definitions for resolution was developed by Widess in 1973. He considered the case of a constant velocity layer encased in a second constant velocity layer. Ignoring transmission losses his example produced two, closely separated reflection coefficients of equal magnitude but opposite sign.

Under these ideal conditions he suggested that the lower limit of resolution would be $\lambda/8$. Any bed thinner than that he claimed would not be resolvable by the seismic method. Other researchers have suggested that this $\lambda/8$ limit is overly optimistic. Kallweit and Wood (1982) also addressed the issue of temporal resolution but from the perspective of optics. They suggested that we use Raleigh's optical limit for the diffraction pattern of an optical instrument, which is $\lambda/4$.

Spatial or lateral resolution is the ability of the seismic wavelet to resolve closely spaced points. The limit of spatial resolution is determined by the Fresnel zone of the wavelet. The Fresnel zone was defined by Lindsey (1989) as "The area of constructive reflection accumulation surrounding the ray theory reflection point".

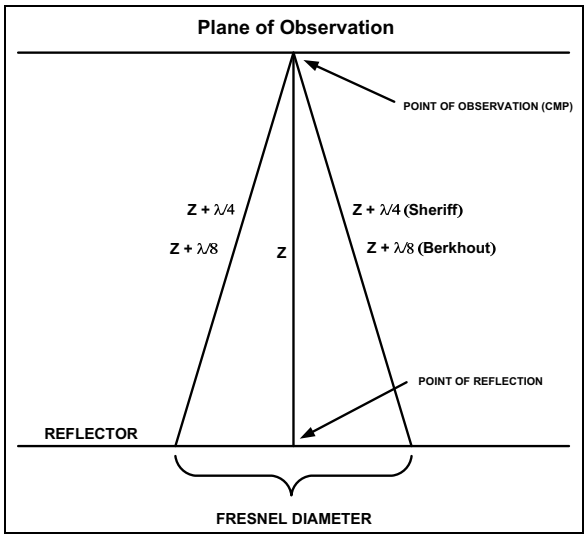


Figure 3-20: Schematic of spatial resolution.

Without going into detail (for a complete derivation see Lindsey 1989) the Fresnel diameter for an unmigrated seismic section is given by $Fd = \sqrt{\lambda Z}$ where Fd is the Fresnel diameter, λ is the dominant wavelength of the wavelet and Z is the depth of burial. The Fresnel diameter for a migrated section is λ/4.

It is important to note that despite the fact that both the temporal and lateral resolutions have mathematical derivations; neither of the limits is exact in practice. Rather they should both be considered more as guidelines; the theoretical limit of temporal resolution is somewhere around λ/4 to λ/8 and for spatial resolution it is somewhere around λ/2 to λ/4 but whether or not we can achieve these limits is very much dependant upon other factors. Both of these terms predict the limit at which any geological event will leave an observable impression upon the seismic section. However, here is the key to resolution, the term observable.

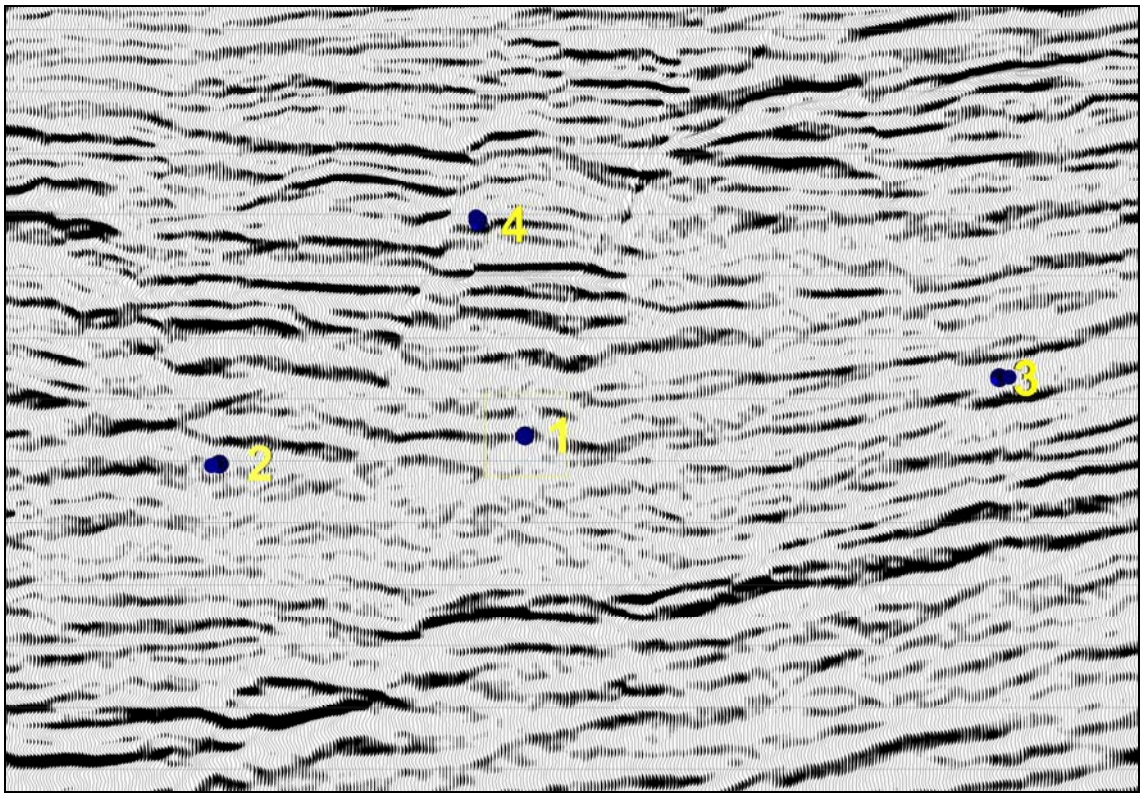


Figure 3-21: Wiggle trace display of a faulted data set, data courtesy PeruPetro. I show the display at a relative scale of 12 traces per inch and 15 inches per second, which I concluded, was optimal for this data. Note the prominent fault trace at # 1 and its offset higher in the section. There are other fault traces in the data, however, most notably at locations 2, 3, & 4. The traces are not visible on this display although they are clear on the SeisScape display.

We predict mathematically the theoretical limits of both temporal and spatial resolution but because we must eventually observe all seismic data, we establish the real limits of resolution empirically. To illustrate this, consider Figure 3-21 and Figure 3-22. These images show a comparison of a section of Trujillo data but unlike my first comparison (see Figure 3-11 & Figure 3-12) where the display scale and orientation was the same for both, I show these two at what I consider the best scale and orientation for each. There is a second difference in this example; whereas in the first comparison I chose an area of data that was void of any definitive events that might dominate the comparison, in this case I chose an area that has several. Specifically I have marked four locations each in an area where there are significant faults.

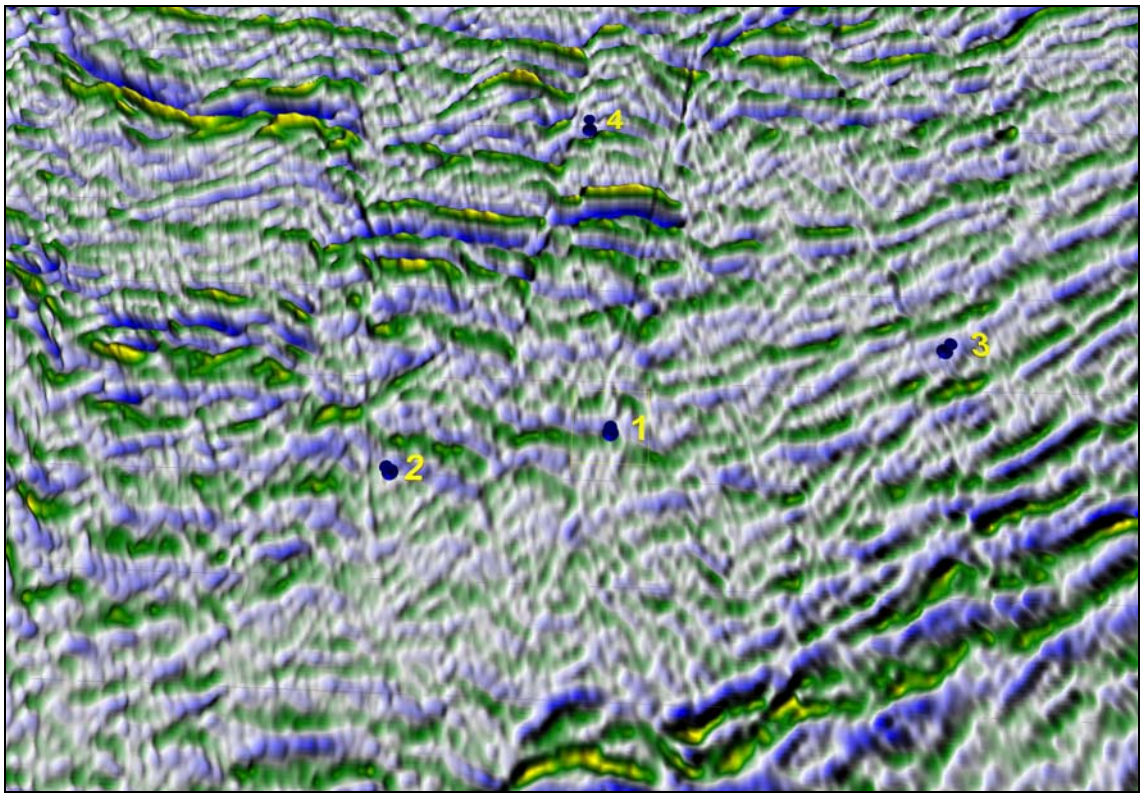


Figure 3-22: SeisScape display of the same data shown in Figure 3-21. The relative display relative scale is 12 traces per inch and 35 inches per second, which is optimal for this display.

Figure 3-21 is the best display that I could produce of the data using wiggle trace technology whereas Figure 3-22 is the best display that I could produce using SeisScape technology. To obtain the data that is shown in both the processor applied all of the conventional processes of deconvolution, multiple removal, velocity analysis, statics, filtering and migration and did the best with each. Let us assume that with these processes the section has the maximum amount of absolute resolution. The question is, considering just the marked faults, are the limits of resolution the same on both displays?

The answer is clearly that the limits are not the same. Fault # 1 is clear on both images but it is clearer and more detailed on the SeisScape image. The other faults are implied from the wiggle trace display but there is nothing definitive in them and they lack detail. On the SeisScape image, however, #2 is sharp and well-defined, #3 is clear but not as clear as it would be if the light were oriented differently and #4 is again implied as it

was on the wiggle trace display but it is still clearer. In terms of fault definition, the SeisScape display is a level-of-detail improvement over the wiggle trace display.

In both these images, the limits of resolution are not determined by the mathematical limits of either temporal or spatial resolution. The mathematically derived limits are the same for both displays but the empirically derived limits are vastly different. These empirical limits restrict us when we interpret a display, they define the practical limits of resolution. Conventionally we consider that there are two forms of seismic resolution, those being temporal and spatial resolution. We must now consider that there is a third form of resolution, visual resolution, which acts upon the previous two but is more difficult to quantify. Visualization is the science that defines and develops visual resolution.

Visualization is the third form of seismic resolution.

3.5 Summary

In this chapter, I defined the terms absolute resolution and apparent resolution. The absolute resolution is a product of the conventional processes of enhancing spatial and temporal resolution and is a quality of the data itself. Apparent resolution is a product of the visual system and the display, and is the subset of the absolute resolution perceived on any given display. In the context of these two definitions, the display serves as a filter upon resolution. The data fed into the filter is the absolute resolution; the output is the apparent resolution.

Unlike a conventional filter whose output is a modified set of data, the output of the display filter is a set of perceptions, all of which occur in the mind. Moreover, in direct contrast with a conventional filter, the objective of the display filter is to do as little to the data as possible. In designing a display filter, we want the output, i.e. the apparent resolution, to match as closely as possible the input, i.e. the absolute resolution. It is the primary purpose of this thesis to begin the process of designing such a filter.

Conceptually, the display filter has two stages. The first stage is the physical display, which serves to produce the input to the second stage. The second stage is the primate visual system, which produces as the final output, a set of perceptions. For the purposes of this thesis, I consider the second part of the filter, the primate visual system, to be immutable. The only place where we can change the filter is in the first stage, the display. To modify this filter such that it has minimal impact, however, we need to understand how the second stage goes about its job. We need to understand the nuances of the visual system and to that effect, Part 2 of this thesis covers the primate visual system in detail. Once we understand how the second stage of the filter operates, then and only then can we begin the process of designing the first stage filter. That process takes up Part 3 of this thesis.

3.6 Conclusions

The theme of this work is that "if you hit a man in the head often enough, hard enough and for long enough, he will eventually come to see even those things that are blatantly obvious". That is nowhere as true as it is in this chapter. Almost everything that I have said in this chapter is blatantly obvious. Obviously, the display acts as a filter upon resolution and obviously different forms of displays filter differently. Nevertheless, whereas in hindsight these conclusion are clear, I had to hit my head on the screen many times before they finally sank in.

This chapter was not so much about proving anything revolutionary as it was about putting things into context. It was about establishing my third conclusion from which I can now build.

Conclusion #3

Visualization is the third science of seismic resolution. Whereas deconvolution and migration establish the limits of absolute resolution, visualization establishes the limits of apparent resolution.

CHAPTER FOUR: CONVENTIONAL DISPLAY FILTERS

For undemocratic reasons and for motives not of State,
They arrive at their conclusions - largely inarticulate.
Being void of self-expression they confide their views to none;
But sometimes in a smoking room, one learns why things were done.
The Puzzler
Kipling

4.1 Introduction

In the previous chapter, I introduced the concept of a two-stage display filter. The first stage is the physical display, which serves to produce the input to the second stage. The second stage is the primate visual system, which produces as the final output, a set of perceptions. These perceptions constitute the apparent resolution of the display.

Conventional display filters take three basic forms; (1) wiggle trace displays, (2) achromatic (gray-scale) variable density displays, and (3) chromatic variable density displays. It is important to understand that, considering each display as a filter, both stages of the filters are different. Not only are the first stages (the display) different from each other but the second stages (how they are processed by the visual system) are also very different. That the visual system processes each display differently is by accident because we developed each display without apriori understanding of how the brain interprets them. Why the visual system processes them differently will become clear as the reader studies Part 2, the theory of the primate visual system.

In this chapter, I briefly study the first two conventional display filters with the primary objective of determining what the limit of their apparent resolution is. Geophysicists established long ago that none of the conventional displays was appropriate for all uses and that they each have their place in the visualization of seismic data. With this in mind, a secondary objective of this examination is to determine if wiggle trace displays and gray-scale images have a place in the future. The third filter, the chromatic variable density display, is an integral part of the SeisScape display and I considered it in detail in Part 3.

4.1.1 Test Data

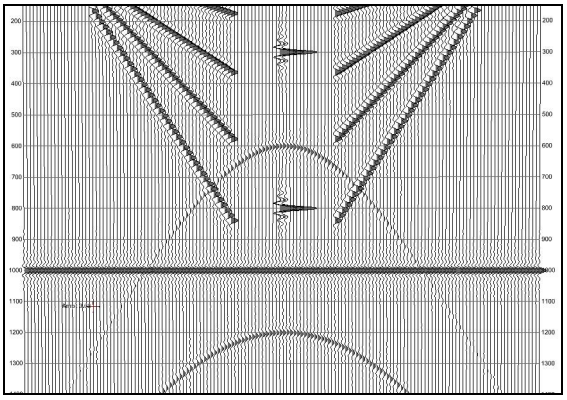


Figure 4-1: Wiggle trace display of the model data used as input to the FK migration that produced images.

In examining conventional displays I needed a dataset that had continuous seismic events with peak amplitudes that varied by several orders of magnitude. The best dataset that I could find for this was a synthetic dataset produced from an FK migration of the model data shown in Figure 4-1 (model data courtesy of Dr John Bancroft). This model data is used to test migration algorithms. It contains a series of dipping events, two diffractions that should, in theory, collapse to band limited spikes and two band-limited spikes that should ultimately show the signature of the migration algorithm used.

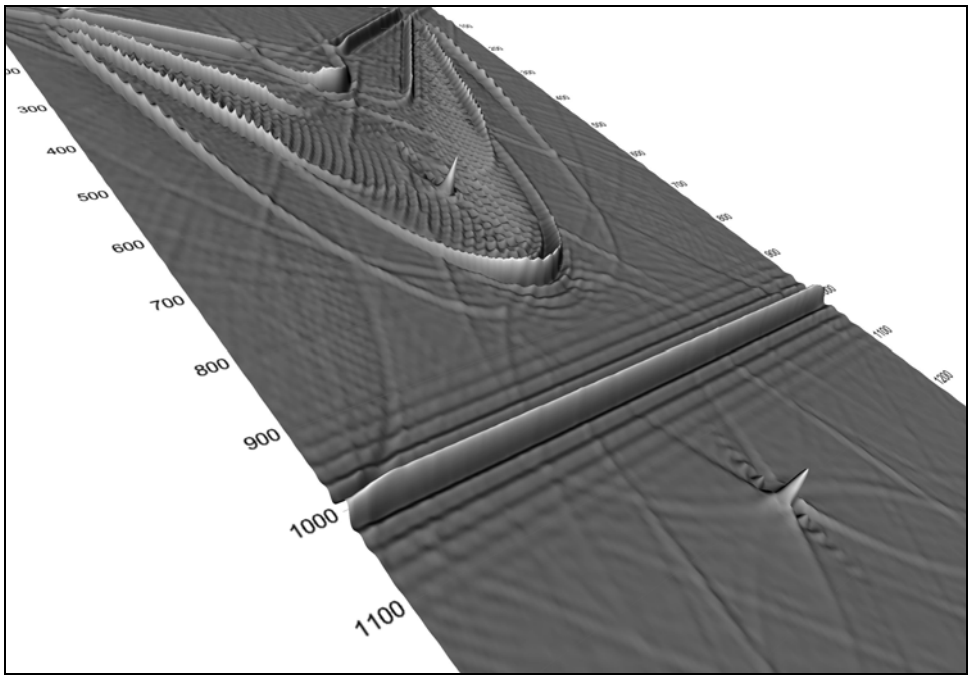


Figure 4-2: SeisScape display (using a gray-scale color palette) of the test data set used in determining the visual dynamic range tests. The section is the result of applying an FK migration to the data shown in Figure 4-1. The migration produced a series of artifacts, which differ from the primary event amplitudes by up to three orders of magnitude.

To produce the synthetic data I performed an FK migration on the model data (see Figure 4-2). The details of the migration itself are not important, what is important is that the migration produces artifacts that are varying orders of magnitude weaker than the events being migrated. After migration, the primary events have peak amplitudes in the range of 20,000 to 40,000 whereas the artifact events have peak amplitudes that are up to three orders of magnitude less.

4.1.2 Quantitative Display Qualities

Apparent resolution is a product of two variable factors. The first is the absolute resolution of the section, which is itself the product of both spatial and temporal resolution. The second are the qualities of the display, which include the visual dynamic range, and the concept of "just noticeable difference".

4.1.2.1 *Visual Dynamic Range*

Of those two qualities, dynamic range is the term that is most likely familiar to the reader. Other fields use it frequently to refer to the ratio between the smallest and largest values of a changeable quantity.

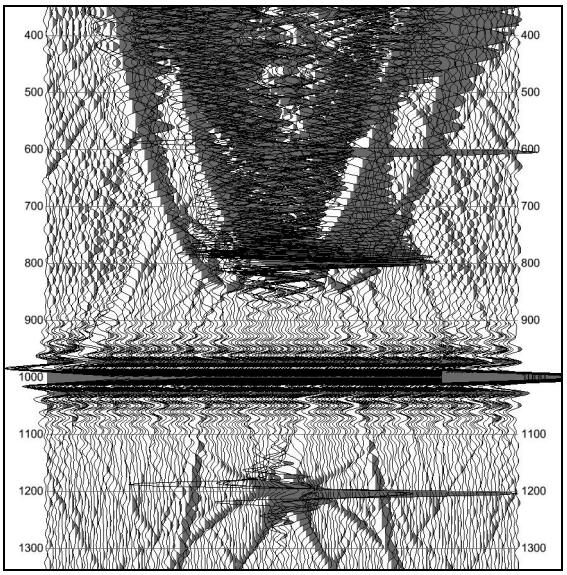


Figure 4-3: Wiggle trace display of synthetic data shown using a trace overlap of 60. Whereas a wide range of events is visible on this display, most of them are not interpretable because the overlap obliterates their details.

In the case of seismic visualization, this changeable quantity is the amplitude of the largest and smallest events distinguishable on the same display. In a seismic context, I define visual dynamic range as:

Visual Dynamic Range

The ratio (in decibels) between the highest and lowest amplitude interpretable events that can be clearly distinguished on a single display.

The important part of this definition is the term “interpretable event”. It is always possible to scale a display such that the weakest events become visible. The question, however, is what effect this scaling has on the more prominent events. To expand upon this, consider the wiggle trace display shown in Figure 4-3. This display has a wiggle trace excursion of 60. At this excursion, the weak amplitude events are clearly interpretable because the traces do not largely overlap. However the strong events overlap to the degree that they obliterate each other and whereas it is still possible to identify they exist, it would be impossible to interpret them at this scale. The scaling here has not altered the visual dynamic range of the display; it has merely altered the level of events that fit within it.

4.1.2.2 *The Weber-Fechner Law and Just Noticeable Difference*

Wiggle trace displays differ from all other seismic displays in that they require the user to identify physical deviations in the seismic trace. All other seismic displays, be they chromatic variable density displays, achromatic gray-scale displays or SeisScape displays, rely upon the visual system to detect changes in brightness. Brightness is a sensation or perception and as such, it is generally accepted that there is no objective way to measure it. Determining brightness falls into the sphere of psychophysics, which is the science that relates stimulus to sensation.

Psychophysics dates back to the work of the German scientist and philosopher Gustav Theodor Fechner who coined the word, developed the fundamental methods, conducted elaborate research and essentially initiated the field of experimental psychology. He based his work upon the earlier findings of the German physiologist Ernst Heinrich Weber who had discovered the concept of the “just noticeable difference”. Fechner ultimately developed a mathematical formula for relating stimulus to sensation, which he named Weber’s law although it is often known as the Weber-Fechner law.

Weber initially worked with weights and he discovered that the amount of change in magnitude of a given stimulus necessary to produce a just-noticeable change in sensation always bore an approximately constant ratio to the total stimulus magnitude. In terms of

weight what this means is that: if two weights differ by a just-noticeable amount when separated by a given increment, then, when the weights are increased the increment must be proportionally increased for the difference to remain noticeable. Fechner applied Weber's law to the measurement of sensation in relation to a stimulus. The formula that he developed was a simple relation that the magnitude of a stimulus must be increased geometrically if the magnitude of the sensation is to increase arithmetically.

I took the above from:

"**Psychophysics**." Encyclopædia Britannica. 2007. Encyclopædia Britannica Online. 9 July 2007 www.britannica.com/eb/article-9061735.

Weber's law is particularly important for the study of visualization. As I show in Chapter 8, the visual system depends upon contrast. We never perceive color or brightness in absolute terms; we only see them in contrast to their surrounds. Weber's law states that the contrast must change by a discreet amount before we can perceive that change. For brightness, the ratio of $\Delta L/L$ of the just-noticeable difference ΔL and the luminance L is constant and equal to approximately 0.02 for a wide range of luminance. Today there are better definitions for the just-noticeable difference and it is clear that it is not a constant but that it depends upon adaptation level of the visual system. For example, Ward (94) established the relationship between adaptation luminance, L_a and just noticeable difference in luminance $\Delta L(L_a)$ *as:*

$$\Delta L(L_a) = 0.0594 \times (1.219 + L_a^{0.4})^{0.25} \tag{4.1}$$

Even so, the j.n.d. of 0.02 is generally correct and I will use it as a rough guide in what follows.

In the case of weight, Weber described his relationship as follows:

$$dp = k\frac{dS}{S} \tag{4.2}$$

Where dp is the differential change in perception, dS is the differential increase in stimulus, S is the stimulus at that instant and k is a constant to be determined experimentally.

Integrating (4.2) gives:

$$p = k\ln(S)+C \tag{4.3}$$

Steven's power law, which is another proposed relationship between the magnitude of a physical stimulus and its perceived intensity, generally supersedes Weber's law. Regardless, it is not important in the context of this thesis to be precise about the relationship between stimulus and sensation. What is more important is what the Weber-Fechner law says about our perception of brightness. In particular, it indicates:

1. The relationship between luminosity (stimulus) and brightness (perception) is logarithmic, every time the luminosity doubles the brightness increases by an arithmetic factor.

2. Because of the logarithmic relationship, the visual system can better detect small changes in luminosity at low light levels than it can at bright levels.

3. Our perception of brightness depends upon contrast and there is a lower limit to the detectable contrast of roughly 2%. Contrasts in brightness less than this are not detectable by the visual system. This has direct implications for our ability to track weak signals that are superimposed upon stronger events.

4.2 Conventional Wiggle Trace Displays

Wiggle trace displays date back to the very origins of exploration seismology and despite the advances made in technology, they remain today one of the primary means of displaying seismic data. The questions at hand are, (1) what is their apparent resolution and (2) should we continue to use them in the future. These are important questions and the answer to the second question in particular is not as obvious as one might think.

Wiggle trace displays are different from all other seismic displays in that each seismic trace forms an independent visual object. Consequently, the visual system cannot process it as a whole but rather it forms a "sum of parts". This "sum of parts" produces a display with low apparent resolution and so it would be easy to discard the display for the future. However, "sum of parts" is another word for pattern recognition – and we discard it at our peril.

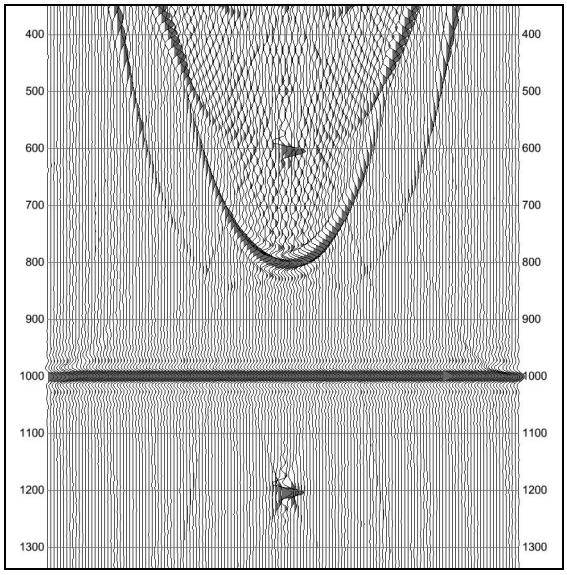


Figure 4-4: Wiggle trace display of the results of applying an FK migration to the model data shown in Figure 4-1. The trace excursion is set so that the horizontal event at 1000 ms has a trace overlap of 3.

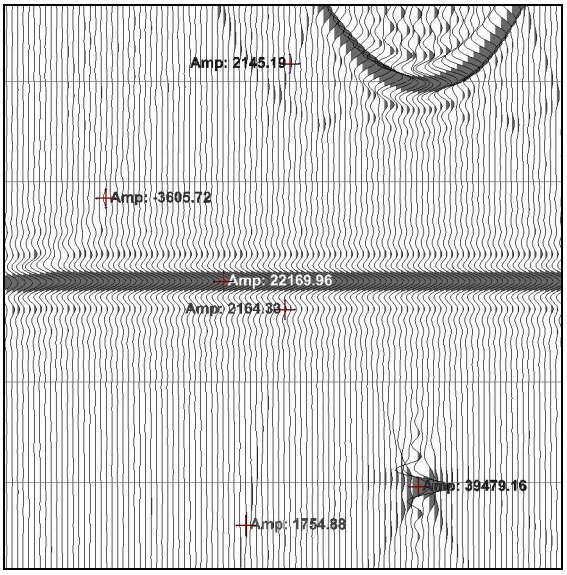


Figure 4-5: Expanded view of Figure 4-4 showing the amplitude of the various interpretable events. The trace overlap is set so that the horizontal event of amplitude ~22,000 is still interpretable. Given that, the lowest amplitude interpretable events are approximately one order of magnitude less.

4.2.1 Visual Dynamic Range

Figure 4-4 is a wiggle trace display of the results of the FK migration shown in Figure 4-2. The trace overlap is set so that the traces of the horizontal event at 1000 ms overlap by a factor of three. Any greater overlap tends to make an event uninterpretable and consequently I use this event as the highest amplitude interpretable event on the section. Given that 22,000 represents the highest interpretable amplitude on the display, the question is "what is the lowest interpretable amplitude".

Looking at Figure 4-5, which shows the amplitude values for some of the events that are harder to distinguish; it appears that the lowest interpretable amplitudes from this section appear to be about one order of magnitude less than the main event. For example, the side lobe to the main event has amplitude of around 2,100 whereas the main event amplitude is ten times large at 22,000. The other discernable events, such as the migration smile from the edges of the main event, all have amplitudes in this range.

Ultimately, the dynamic range of a wiggle trace display is tied to both the resolution of the display device and the zoom factor of the display. Detecting a subtle feature requires that there be some perceptible deviation of the trace. This perceptible deviation is akin to the previously mention "just noticeable difference". The lower the display resolution the greater the amplitude of the subtlety must be in order to cause an observable deviation. This works in reverse for the zoom factor of the display. The closer one is zoomed into the display, the smaller the amplitude change must be in order to cause a deviation in the trace.

In general, considering the events that are distinguishable in Figure 4-5, wiggle trace displays have at best a 20-30 db visual dynamic range. This range, however, is not absolute for any given display because, as I mentioned previously, how closely one zooms into the display affects the ability to detect weak amplitudes. Consequently, the apparent resolution of any given display can change significantly as one changes its zoom factor.

This simple analysis demonstrates that wiggle trace displays have low dynamic range and consequently low apparent resolution. However, this is only part of the story. The visual dynamic range suggests the range of amplitudes detectable by the visual system but it does not say anything about whether or not the visual system can form a coherent event out of what it sees.

4.2.2 Forming "Ghost" Events

What is surprising about wiggle trace displays is how much detail you can actually see in them. This is a testament to the veracity of the primate visual processing system. Wiggle trace displays, more than any other form of display, depend heavily on the pattern recognition capabilities of the achromatic channel of the visual processing system. For low amplitude events in particular, i.e. events where the wiggle fill does not overlap and form a solid object, the visual system can still form the impression of an event although it interprets each trace separately.

Seismologists have relied upon wiggle trace displays for almost a century but without any understanding of how the brain interprets them. Much of what we perceive on a wiggle trace displays results from the ability of the visual system to form an event (or object) from a series of isolated visual clues (seismic traces). Physiologically, this is due to the high proportion of orientation specific, achromatic cells in the hypercolumns of the visual cortex (for details see Chapter 8). These cells are tuned to detect lines and edges and form patterns from them. Given that a wiggle trace display is all lines and edges it should not be surprising that the visual system can form the impression of an event even when the hint of it is extremely faint.

Although the properties of the visual processing system make it surprisingly effective at detecting events, it is far from perfect. When we view a display such as Figure 4-4 the visual system provides us with two types of events. The first are the clear-cut events, the ones whose traces overlap to form a continuous visual object. These are the main events and when we look at the display, we see them regardless of the display scale. The other types of events, however, are ghost-like events. The traces of low amplitude events do not overlap and therefore do not provide the visual system with a clearly defined object. In these cases, the best that the visual system can do is to form an impression that they are there. There is nothing visually definitive for it to interpret and so it can only form a tenuous impression of an event.

It is these faint or subtle events, which are the focus of visualization, and it is with these events that wiggle trace displays can break down. Figure 4-4 contains numerous examples of ghost events but one in particular is interesting. The event in question starts at 350ms on the right edge of the display and traces all the way back to the migration signature located at 800ms in the center of the display. This event is not distinct enough to classify as interpretable but it is definitely there. The impression of the event, however, is nebulous and can be very easily lost as in Figure 4-6.

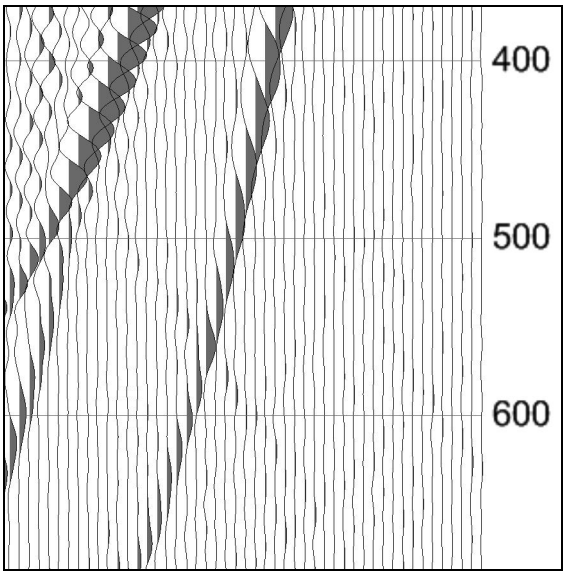


Figure 4-6: Expanded view of the upper right section of Figure 4-4. Note that the faint event terminating at 350ms on the right edge of Figure 4-4 is hardly discernable at this scale.

Here, I enlarge the display to show just the upper right area and when viewed at this scale the impression of the event is all but gone. I show Figure 4-4 and Figure 4-6 at the same relative scale and with the same trace excursion; the only difference between them is the zoom factor of the display. The act of zooming into the display, which one would normally do to reveal more detail about an object, has actually erased the event almost entirely. This is why I call these faint events "ghost" events. They are in many ways a figment, if not of the imagination then certainly of the visual processing system, and the closer one looks at them the less one can see.

The disappearance of these faint events as one zooms into them is one of the primary limitations of wiggle trace displays and it has an explanation in physiology. Anthropoid primates' possess a fovea, which is an area of the retina that contains a high density of cones. This fovea (which I discuss in detail in Chapter 6) gives us, within the limits of optics, almost perfect visual acuity, and it is not surprising that the majority of the cells in the visual cortex are dedicated to processing the signals that emerge from it. However,

here is the catch, the human fovea centralis is only about 1500 μm in diameter (Polyak, 1941, Ahnelt and Kolb, unpublished data) and covers roughly 6 degrees of visual arc. For the typical human sitting at a computer monitor, the area of the screen seen by the fovea is roughly a circle two to three inches in diameter. The visual processing system interprets anything within this circle in a different manner and for different purposes than anything outside it. This is why events disappear as you zoom in – more of the traces project outside of the high-resolution fovea and onto the very low-resolution area of retina. Put this way, using the analogy "sum of parts" to describe these events, the visual system is very adept at forming a whole from its parts, but only if the parts lie within the fovea.

4.2.3 Uses of Wiggle Trace Displays

The primary conclusion that must emerge from the previous analysis is that the apparent resolution of a wiggle trace display is very low. At best, the visual dynamic range is only 20-30 db. Moreover, even if we can visually detect low amplitudes it is questionable if the visual system can form a coherent event out of it. The ability of the visual system to form coherent events is both scale and zoom dependant, which places restrictions on the amount of data a person can effectively view on the limited dimensions of a computer monitor. Altogether, these limitations initially suggest that we phase wiggle trace displays out of use. However, a closer analysis suggests otherwise.

Geophysicists initially developed wiggle trace displays because they were the only way to view seismic data given the available technology. They gave no thought to how the visual system interpreted them and so it is by accident that wiggle trace displays are still useful. One of the questions that I originally asked about conventional displays was "if seismic were invented today, would we develop them?" This is an important question and to illustrate it consider Figure 3-11 and Figure 3-12. If the display that you were most familiar with were the SeisScape display, would you then choose to develop the wiggle trace display?

The answer is most likely no and the reason why is that the wiggle trace display obviously has a lower apparent resolution. However, not developing wiggle trace displays would be a mistake because just considering apparent resolution ignores one significant advantage of wiggle trace displays. Wiggle trace displays are all lines and edges and as I show in Chapter 8, it is lines and edges that feed our primary pattern recognition engine and pattern recognition is crucial for stratigraphic interpretation.

To conclude, wiggle trace displays have very low apparent resolution and consequently the industry should not use them as the primary seismic display. However, in interpretation scenarios that require pattern recognition, they are still extremely useful.

4.3 An Analysis of Gray Scale Displays

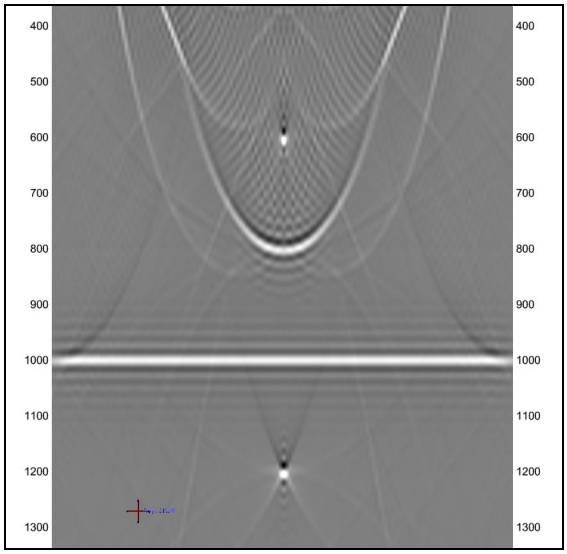


Figure 4-7: Gray scale display of the same data shown in Figure 4-4. The grayscale covers an amplitude range of ± 23000 which corresponds to the amplitude range of the event at 1000 ms.

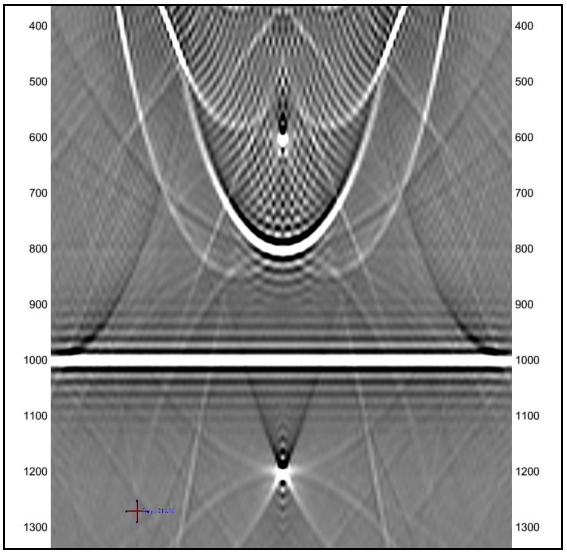


Figure 4-8: Gray scale display of the same data shown in Figure 4-4. The grayscale covers an amplitude range of ± 5000. Although you can see more detail at low amplitudes, the main event at 1000ms is saturated.

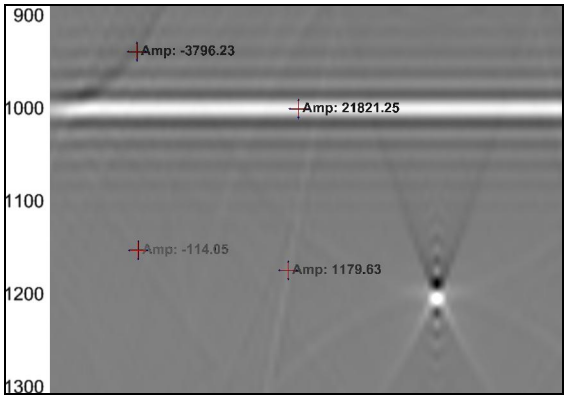


Figure 4-9: The lower left quadrant of Figure 4-7 showing amplitudes of several noticeable events. The lowest amplitude noticeable event has peak amplitude of approximately 100-150.

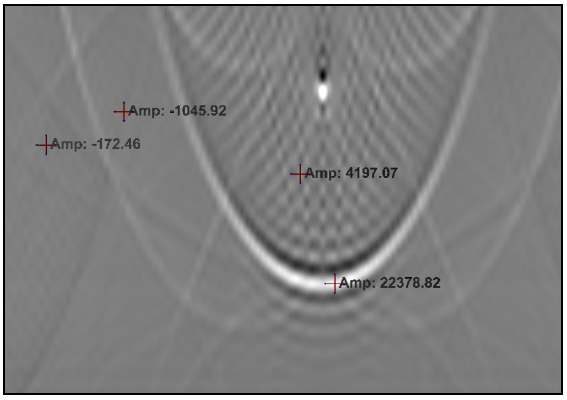


Figure 4-10: The center of Figure 4-7 showing amplitudes of several noticeable events. The lowest amplitude noticeable event has peak amplitude of approximately 150-200.

The pixel resolution of the display device largely determines the dynamic range of a wiggle trace display. A gray-scale seismic display, however, converts seismic amplitude to brightness and therefore can take better advantage of the dynamic range of the visual system. That a gray-scale display has better dynamic range than a wiggle trace display is

apparent in Figure 4-7 which is the gray scale equivalent of the wiggle trace display shown in Figure 4-4.

4.3.1 Visual Dynamic Range

The gray-scale covers the amplitude range of ± 23,000, which correspond to the amplitude range around the main horizontal event at 1000 ms.. At this scale, many of the artifacts that one would normally associate with a migration are now clearly discernable. The edge effects from the horizontal reflector, which were in the range of the lowest discernable amplitudes on the wiggle trace displays, are now evident as are other events, which were at the limit of visibility on the wiggle trace displays. Figure 4-9 and Figure 4-10 are blow-ups of the lower left and the center of the display respectively. They show the amplitudes of some of the artifacts, including what I consider the lowest discernable amplitudes on the displays. In the case of Figure 4-9 this is around 100-150 and in the case of Figure 4-10 it is between 150 and 200.

The events with peak amplitudes in the range of ± 150 are very faint but still clearly discernable. Given the peak amplitude for the gray-scale is 23,000, this results in a visual dynamic range of approximately 40-45 db, which is considerably better than the 20-30 db of a wiggle trace display.

Although the visual dynamic range of a gray-scale display is an improvement over the wiggle trace display, it still does not match the absolute resolution of the data. Even though weaker events are now discernible, a comparison of Figure 4-7 and Figure 4-8 indicates that there are events in this data, which the viewer cannot see. The gray-scale in Figure 4-8 covers the amplitude range ±5000 and so highlights very weak events at the expense of higher amplitudes. All throughout this display, especially at the edges, there are visible events, which are too faint to distinguish on Figure 4-7. In addition, it is possible to follow the edge effects from the horizontal reflector to much higher in the section before it fades into the background. Clearly, even though a gray scale image has better dynamic resolution than a wiggle trace display, in the case of this data set it does not approach the dynamic range of the events in the data itself.

4.3.2 Properties of Gray-Scale displays

I earlier described wiggle trace displays as low-resolution displays. Gray-scale displays are also low-resolution but for a different reason. Most modern display systems use 24-bit color, which provides 8 bits each for the red, green and blue components. For a color image, this means that the display can show 256^3 or approximately 16 million colors. For a gray-scale display, however, the red, green and blue components are identical. Consequently, maximum number of shades of gray is limited to 255. The properties of the visual system itself further reduce this 255-shade limit. As I show in Chapter 8, the visual system depends upon contrast. We cannot determine absolute luminance we can only determine relative luminance via contrast with its surrounds. Continuing from section 4.1.2.2 where I introduced the concept of "just noticeable difference", this leads to the question of just how low a contrast can we actually see.

Recalling section 4.1.2.2, in the case of humans this is approximately 2%. This means that for two adjacent shades of gray to be individually identifiable they must differ in contrast by around 2%. If the contrast is less then the visual system has a hard time separating them. I graphically illustrate this problem in Figure 4-11 to Figure 4-16 inclusive. In the vertical center of each of these images is a wiggle trace display of a single event. The wiggles and fill are set to a gray-level of 128 or midway between black and white. The backgrounds, however, are set at 125, 126, 127, 129, 130 & 131 respectively. Although the wiggles are hard to see on any of the displays, they are definitely perceptible when the background differs from the trace color by three (2.3%), they are barely perceptible when the difference is two (1.5%) but are almost invisible when the difference is one (.78%).

Figure 4-11: Wiggle trace display with a background gray level of 125 and wiggle trace and fill level of 128.

Figure 4-12: Wiggle trace display with a background gray level of 126 and wiggle trace and fill level of 128.

Figure 4-13: Wiggle trace display with a background gray level of 127 and wiggle trace and fill level of 128.

Figure 4-14: Wiggle trace display with a background gray level of 129 and wiggle trace and fill level of 128.

Figure 4-15: Wiggle trace display with a background gray level of 127 and wiggle trace and fill level of 130.

Figure 4-16: Wiggle trace display with a background gray level of 127 and wiggle trace and fill level of 131.

Further complicating matters is the fact that our ability to detect contrast is a ratio. When attempting to observe a weak signal superimposed upon another event, we see it better against weak events than against strong ones. I illustrate this effect on Figure 4-9 where there are numerous high angle, low amplitude events crossing the main horizontal reflector. These events are clearly noticeable when viewed against the low amplitude regions of the display but they disappear across the main reflector. The absolute magnitude of the contrast remains roughly the same regardless of where one observes the event but against the bright main reflector, the event is not perceptible because the ratio of the contrast is too small. This has direct consequences for our ability to track signals such as faults or diffractions.

Gray-scale seismic displays are in common use throughout the industry. There are a number of reasons for this not the least of which is that they convert seismic amplitudes to pixel brightness, which means that they require far less screen space to display. Recalling the discussion on the fovea from the previous section, in physiological terms, we can fit more of the section into the area covered by the high-resolution fovea than we can with a wiggle trace display, which is a distinct advantage for perception. A second reason, as outlined above, is that gray-scale displays provide an almost order of magnitude improvement in dynamic range over wiggle trace displays.

There is a third reason, however, found buried deep within the human visual processing system. One of the primary reasons that gray-scale displays are popular is that they appear to be pseudo three-dimensional and this makes them feel almost familiar. There is a reason for this, gray-scale displays are purely achromatic and are processed by the achromatic neural circuitry in the visual cortex (see Chapter 2). This circuit is responsible for creating most of our perceptions of the world around us and so it should come as no surprise that gray-scale displays look almost three-dimensional.

So far, I have established that gray-scale displays have a higher apparent resolution than do wiggle trace displays and that the brain processes them differently. The question remains as to whether or not they have a place in the future of seismic visualization. To

answer that question I will briefly discuss the lighting component of the SeisScape display, because it is also a form of gray-scale display.

4.4 Lighting Component of SeisScape Displays

A SeisScape display is composed of three primary elements; (1) a tessellated mesh that composes the topographical seismic surface, (2) a lighting (shaded-relief) component that illuminates the surface and (3) a variable density display draped over the surface. As such, it has elements that have analogies within the three conventional types of seismic displays. The tessellated mesh, is loosely analogous to the deviations of a wiggle trace display, the lighting is similar, in principal, to a gray-scale variable density display whereas the coloring is exactly a chromatic variable density display.

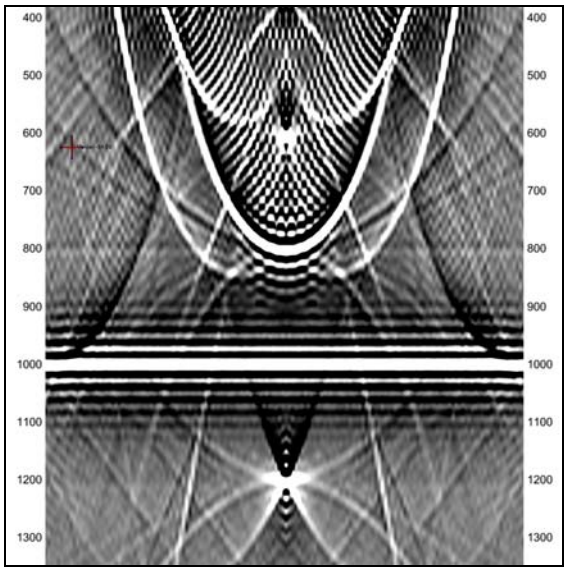


Figure 4-17: Gray-scale display of the FK Migration data. The gray-scale covers the amplitude range -1000 to 1000 highlighting low amplitude artifacts.

Because SeisScape displays are composed of three elements, they are by nature more complex than the other displays and a full analysis of their properties is beyond the scope of this section. Each of the three elements, the tessellated mesh, the lighting and the coloring, make significant contributions to apparent resolution and I consider them in detail in Chapters 9, 10 and 11 respectively. Lighting, however, is the critical component and provides the bulk of perception. It is the principal reason why SeisScape displays have a higher apparent resolution than do conventional displays and I will illustrate that here.

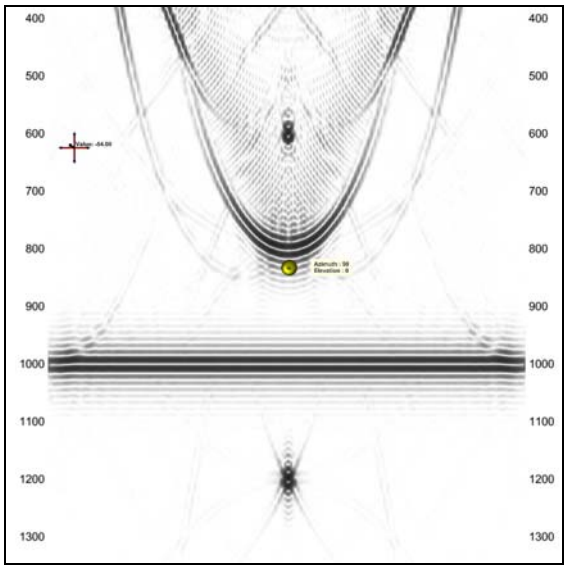


Figure 4-18: Same data as shown in Figure 4-17. The sphere and arrow in the middle of the image indicates the direction of lighting. Light elevation is 0 degrees to the vertical, which renders the azimuth irrelevant.

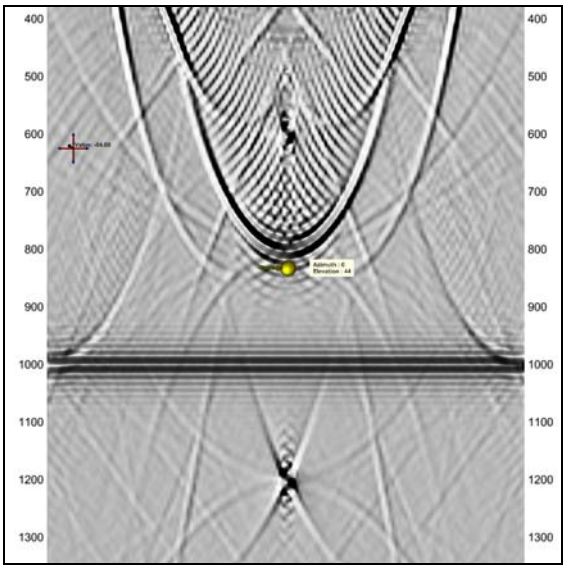


Figure 4-19: Same data as shown in Figure 4-17. Light azimuth is 0 degrees; elevation is 45 degrees to the vertical.

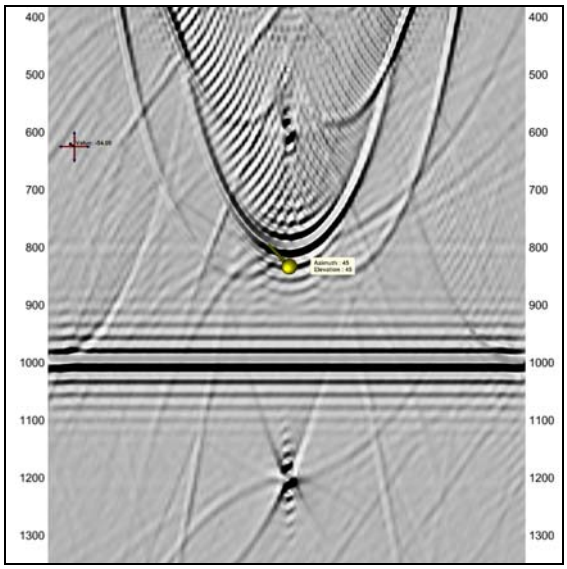


Figure 4-20: Same data as shown in Figure 4-17. Light azimuth is 45 degrees, elevation is 45 degrees to the vertical

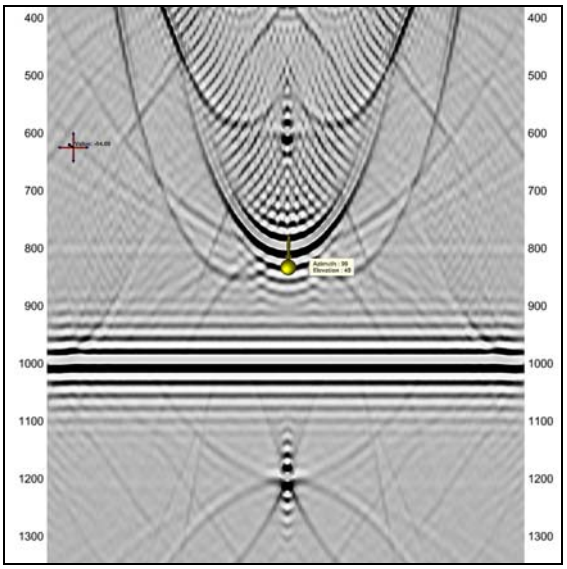


Figure 4-21: Same data as shown in Figure 4-17. Light azimuth is 90 degrees; elevation is 45 degrees to the vertical.

Figure 4-17 is a gray-scale display of the same area of data as shown in Figure 4-4, Figure 4-7 and Figure 4-8. I clipped the gray-scale in this image to cover the amplitude range of -1000 to +1000, which makes it possible to see the full suite of low-amplitude events. The prominent events are, of course, fully saturated and would be uninterpretable

on a normal seismic display. Figure 4-17 is included here for comparison against Figure 4-18 to Figure 4-21 which are four examples of SeisScape displays of the same data.

The only difference between the four SeisScape images is the direction of the light source. In the first image, I direct the light from directly overhead whereas in the other three I direct it at a 45-degree angle to the horizon. In the second image, the light azimuth is 0 degrees, which means that it points directly from the left edge of the display. In the third image, the azimuth is 45 degrees and in the final image, it is at 90 degrees. In all of the examples, I set the apparent height of the topography to zero so that the only observable effects come strictly from the lighting calculations. The orientation of the lightening on a SeisScape display is dynamic and alterable in real time; the four possible images shown here are just snapshots of this dynamic process but even so, they establish a critical point. With the correct elevation and azimuth, the viewer can highlight any of the low-amplitude events observed on Figure 4-17 without saturating the main events.

The reader can prove this to themselves by first selecting low-amplitude features on Figure 4-17 and then trying to find them on one or more of the SeisScape images. Keep in mind that; (a) these are still gray-scale images and consequently have the same low-resolution nature as conventional gray-scale displays, and (b) that they are just snapshots of a very dynamic process. Even so, it is still possible to identify all of the low-amplitude events and artifacts visible on Figure 4-17 on at least one of the images without loosing the ability to interpret high amplitude events.

Although it is not possible to see all of the low-amplitude events on any one of the SeisScape images, keep in mind that you should consider all four images together. Unlike wiggle trace or variable density displays, SeisScape displays are dynamic and real-time manipulation of the lighting is an inherent component of the displays. Comparing the four SeisScape images with their wiggle trace equivalent (Figure 4-4) or gray-scale equivalent (Figure 4-7), it is clear that for this synthetic example the apparent resolution of the SeisScape display is higher than the other two and approaches the absolute resolution of the data itself.

4.4.1 Lighting vs. Conventional Gray-Scale Displays

Both gray-scale displays and the lighting component of a SeisScape display are purely achromatic and the visual system processes them identically. Considered in the context of the display filter, this means that the second stage of the filter is identical for both. The difference between them lies purely in how the gray-scale is calculated. Conventional gray-scale images map seismic amplitudes directly to brightness. Lighting, on the other hand, maps the dot product of the light direction vector and the surface normal to brightness.

Lighting is the principal component of SeisScape displays. I discuss its calculation, and implications for seismic visualization in detail in Chapter 10. To date, its importance to visualization is not generally recognized. Besides the author's own works, the only other reference to the use of lighting with respect to seismic data was from Barnes in a short note to Geophysics. His work primarily centered on the use of shaded-relief for time slices and he concluded that using shaded-relief on vertical seismic provided little benefit.

A comparison of the conventional gray-scale images and the SeisScape shaded-relief images indicates that Barnes's assessment is incorrect. The apparent resolution of the shaded-relief images is clearly higher than the apparent resolution of the conventional gray-scale images. In addition, because of how lighting displays map amplitudes to brightness, the visual system is better able to interpret the images. The reasons for this I will leave until Chapter 10 but for now, to answer the question as to whether or not amplitude mapped gray-scale images have a place in the future. The answer is that they do not.

4.5 Apparent Resolution in Practice

In the previous sections, I used examples of artifacts, produced by migration, to illustrate the properties of wiggle trace displays, amplitude mapped gray-scale displays and shaded-relief displays. I showed that the apparent resolution of the first two conventional displays was significantly lower than both the shaded-relief component of a

SeisScape display and the absolute resolution of the test data. In this section, I answer the question of what this means in practice.

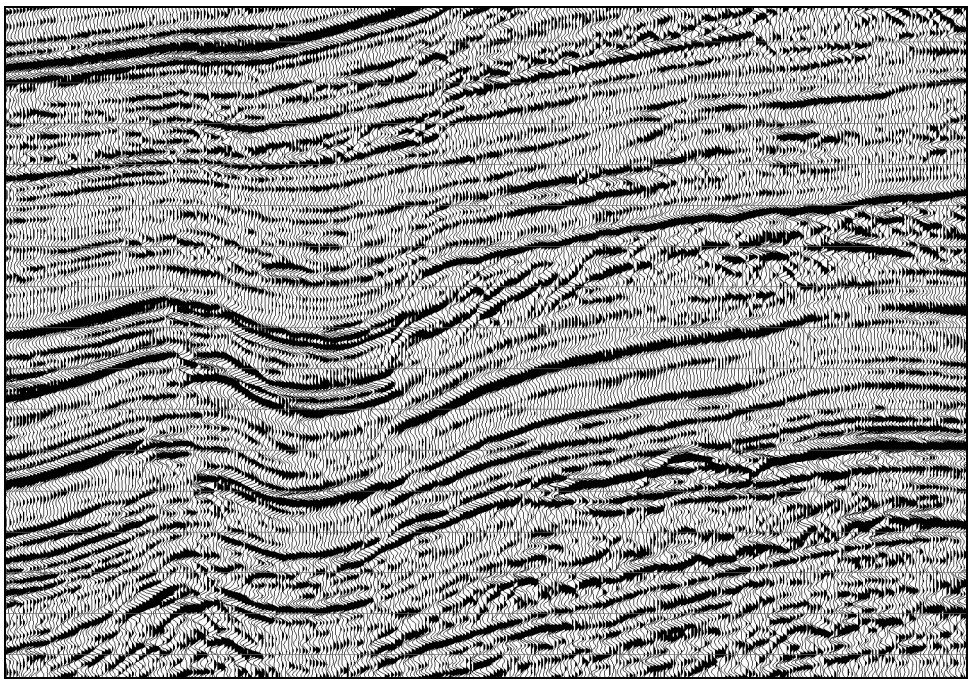


Figure 4-22: Wiggle trace display of a section of data from the Gulf of Mexico (data courtesy unnamed source). The section shown is unmigrated and so, given that the section is highly faulted, there should be diffractions evident. In fact, diffractions are visible throughout the section but they are indistinct.

One of the great mysteries of seismic data concerns diffractions. In theory, you should think of an unmigrated seismic section as being produced by a series of point source diffractors. As a result, a highly faulted section such as the one shown in Figure 4-22 should be literally cut to pieces with diffractions. Every time there is an edge, termination or fault in the geology, there should be a diffraction tail coming off it. Yet, you rarely see diffractions on unmigrated data and when you can see them, they tend to be very short. This effect is evident on the wiggle trace display of the unmigrated Gulf of Mexico data shown above. The section contains numerous faults and there are obvious diffractions coming off the event termination associated with them. However, the diffractions do not appear to extend more than 50-100 ms below their origins (timing lines are 100 ms apart) and many terminations do not appear to cast diffractions at all.

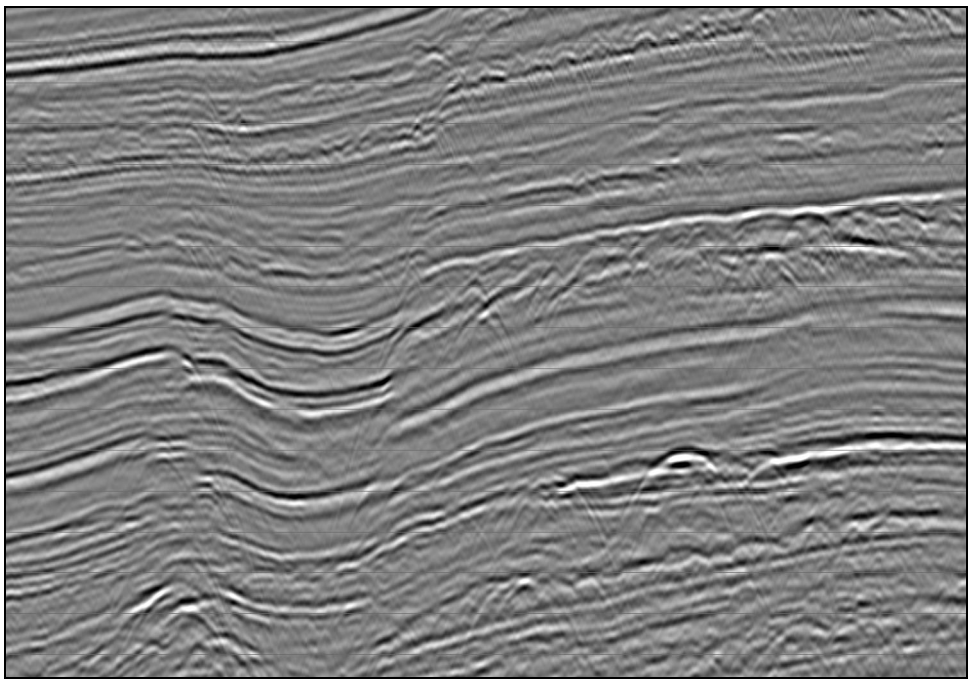


Figure 4-23: Variable density display of the same section of data shown in Figure 4-22 using a gray scale color palette. The diffraction are much more evident here than on the wiggle trace display just as the low-amplitude migration artifacts were clearer on Figure 4-7 than on Figure 4-4.

The visual effect of the diffractions on Figure 4-22 is very similar to the visual effect of the migration artifacts on Figure 4-4. We can see diffractions but only where the amplitude of the diffraction is similar to that of the main events. However, as the amplitude drops our ability to perceive the diffraction becomes nebulous and eventually they become “ghost” events that disappear as we zoom into them.

Figure 4-23 is an amplitude mapped gray-scale image of the same section and as we could with the gray-scale image of the migration artifacts, we can now see the diffractions more clearly. However, because of the low-resolution nature of gray-scale images and the fact that we are mapping the entire amplitude range of the data onto 255 levels of gray, we perceive that the diffractions are present but they lack visual significance.

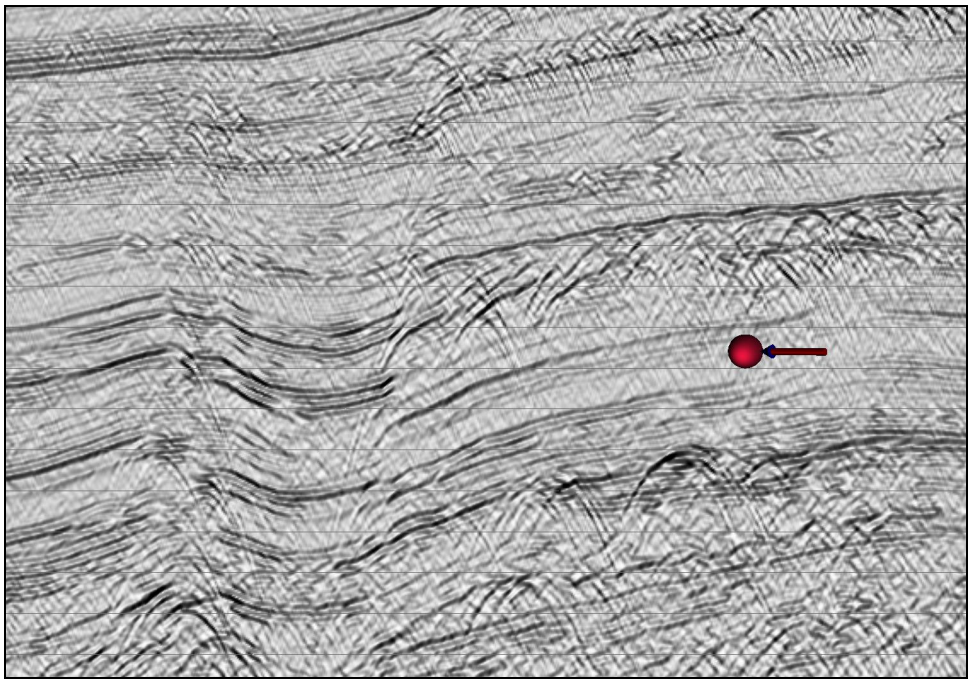


Figure 4-24: A SeisScape display of the same section of data shown in Figure 4-22. Light is oriented from the right side of the image.

As much as the gray scale display reveals more diffractions than the wiggle trace display the SeisScape display reveals even more. Figure 4-24 answers the question "what happened to the diffractions" - the answer is that they are lost in the display. With the migration artifact example it was possible to conclude that the apparent resolution of the SeisScape display matched the absolute resolution of the data. You cannot say the same thing here. Eventually the amplitude of a diffraction should decrease until it leaves no distinguishable imprint upon the absolute resolution of the section. It is impossible to say if we are following the diffractions that far but we can say that given what we can see on the SeisScape display that the apparent resolution of the other displays is very low.

4.6 Conclusions

SeisScape displays are a fundamentally new type of seismic display. This does not mean, however, that they are ideal for all circumstances. There are three conventional seismic displays, (1) wiggle trace displays, (2) achromatic (gray-scale) variable density displays, and (3) chromatic variable density display. Geophysicists generally recognize that none of these displays is ideal for all circumstances. It was the purpose of this chapter to review the properties of the first two displays (I cover the third in a later chapter) and determine if, given the nature of SeisScape displays, they should continue to be used in the future.

Conclusion #4

Wiggle trace displays have low apparent resolution and we should not use them in the future as the primary seismic display. However, since we construct the display purely from lines and edges, they feed our primary pattern recognition engine better than any other display and therefore they remain an essential component of seismic visualization.

Conclusion #5

In terms of apparent resolution, amplitude mapped gray-scale displays are higher resolution than wiggle trace displays but lower resolution than shaded-relief displays. Given that the visual system processes amplitude mapped gray-scale displays and shaded-relief displays identically, we should deprecate the former in favor of the latter.

Part 2: Theory

The Primate Visual Processing System

Standing aloof in giant ignorance,
Of thee I hear and of the Cyclades,
As one who sits ashore and longs perchance
To visit dolphin-coral in deep seas.
So though wast blind! – but then the veil was rent,
For Jove uncurtained Heaven to let thee live,
And Neptune made for the a spumy tent,
And Pan made sing for thee his forest-hive;
Aye on the shores of darkness there is light,
And precipices show untrodden green,
There is a budding morrow in midnight,
There is a triple sight in blindness keen;
Such seeing hadst thou, as it once befell
To Dian, Queen of Earth, and Heaven, and Hell.

To Homer
John Keats
1795-1821

The next four chapters represent the theory section of this thesis and as one might expect, they contain the majority of the references. What one might not expect, however, is just what those references are. One in particular, "Cones in the Retina of the Mongolian Gerbil", catches the attention. It brings to mind two questions. The first is how do you tell your parents that you are dedicating your life to studying the Mongolian Gerbil? The second is what is this paper doing in a geophysical thesis?

When I began work on this thesis, I had very little interest in the visual system. My undergraduate degree was in biophysics and so I knew more about it than the average geophysicist. Still, when I began in 2003, the primate visual system was the farthest thing from my mind. One of my stated goals, however, was to uncover the sciences behind visualization. To that end, in late 2005, I began to consider just how we went about communicating visual information. My early work led me to the concept of trivariant color vision, which I discussed in Chapter 2, and it made me think about just how much we take our vision for granted. Vision is our most dominant sense, so much so that we do not notice the sensations that it produces. We take note of sensations from all of our other senses but not from the ones from our visual sense. Vision produces our perceptions of the world around us and we make the subconscious mistake of thinking that it is perfect.

As I studied trivariant color vision, however, I began to discover that our vision is far from perfect and that it is tuned to see certain things in certain ways. It is far from the general-purpose tool that we take it to be. I began to discover that all vision is not the same and that primates, especially, have a very unique and very specific way of seeing. I realized that understanding the nuances of how we see, had very real and very practical implications for visualization in general. I eventually concluded that the principals of how we see are the first principals of visualization. This is the first reason why in the next four chapters I go into such depth. I wanted to produce a concise reference manual of the visual system.

I had a second reason for going into detail; one that it is has more to do with the future than the present. As I began to study the visual system, I quickly realized that it has geophysical parallels. Exploration seismology is grouped into three broad fields, acquisition, processing and interpretation. The visual system is grouped into the same three fields but, as one would expect, the biological analogues are far more sophisticated and far more elegant. It was interesting to learn how Mother Nature goes about the same tasks as we do and I began to wonder if the day would come when we could make use of her approaches.

Not that we can today. Our biological systems are far too complex and our computational capabilities are still far too limited. Nevertheless, consider this. When I programmed the first SeisScape display, the state of the art graphic processing unit had a rated speed of five hundred megaflops. Seven years later, the current state of the art gpu is rated at almost one teraflop. If we continue this level of progress then, by around 2015, we should start to see petaflop (10^{15}) gpu's and a few years after that, exaflop (10^{18}) gpu's should become the norm. What this means is that probably, within my career, we will see desktop computers whose speeds rival those of biological systems. And if they become as fast as biological systems, maybe we can program them as biological systems. This was my second reason for going into so much detail. Conceivably, sometime in the near future, someone will read this, see the parallels and discover an entirely new approach to our old familiar problems.

My third reason is that studying the visual system took me on a journey through myself. I was raised in the Church of England and as a young boy; I went to Sunday school just like everybody else. There, we were taught all of the old Bible stories. We were taught very carefully because the C. of E. is a progressive religion and they were careful to present the stories as allegorical. Nevertheless, even though I rejected religion when I was 14, as I progressed through my study of the visual system I ran head on into two old ideas that I never consciously suspected were still very much a part of my makeup.

The first idea is that humans are somehow different, that we are special and above all other creatures. Consciously, I would never admit that I believed this or ever had. However, as I progressed with my study, many of the things that I discovered were disconcerting and uncomfortable. I could only explain my feelings by considering that they challenged a subconscious primal belief in the supremacy of humans above all other species. This belief dates back to my earliest religious teaching; what follows in the next four chapters inadvertently but directly challenges it.

The second idea is that humans are engineered and as such are perfectly suited to their environment. Again, I was never consciously aware that I thought this but studying the visual system proved that I did. It proved it by shocking me with the truth that the vision I thought was so well adapted to my life, was in fact, well adapted to a life that humans have never lived. The truth is that our visual system is ill suited to the lives we lead today but perfectly suited to the lives our ancestors lived 35 million and 55 million years ago.

One of my principal motivations for my going into so much detail about the visual system was to search for the answer to a question that has become fundamental to my philosophy on life. The question is when you look in a mirror, who looks back? To me, in my rationalistic world, this is the most important question that humans must answer to survive into the future. You will not find the answer in the next four chapters but perhaps you will find, as I did, that the answer begins 22mm behind your reflection.

All of these reasons explain why I studied the visual system but none explain why I went as far into it as I did. To understand that, you have to understand my epitaph, which will read, "Think it through, stupid!" My life is cluttered with the detritus of hasty decisions that commit me to long term but ill-conceived projects. Studying the visual system is just one of them. I started with the simple idea that I wanted to learn how humans established their perceptions of form and color. At the time that seemed a simple enough question but a smarter and more cautious person would first have asked if anybody actually knew the answer.

The further I went into the visual system, the farther away the answer seemed to get and what started out as a short literature search ultimately became a quest. I reached the point where I had invested so much time and effort that stopping short of the goal was just not an option. Eventually, I went just far enough to satisfy myself with a partial answer, and then I stopped. I am not sorry that I took the journey because I learnt so much along the way. But looking back on it now, it was a good job that I did not think it through a little better. If I had known, when I started, that I would have to go to the very limits of human knowledge before I would get even a fleeting glimpse of the answer, perhaps I would have stayed at home.

CHAPTER FIVE: ANTHROPOID ORIGINS

We are so arrogant, we forget that
we are not the reason for evolution,
we are part of evolution.
Unfortunately, we believe that
we've been created to dominate
the planet, to dominate nature.
Ain't true!
Ted Danson

5.1 Introduction

For the purposes of this thesis, I defined visualization as "the science of communicating information through the visual processing system." With that in mind, this chapter, and the three that follow, attempts to define that science, first by placing it into its evolutionary context and then second by describing the physiological processes that convert the photons impacting the retina into our perceptions of the world around us.

This chapter covers the first part, the evolution of primates, and in particular the evolution of the catarrhine[1] primates, the clade[2] to which we belong. It may be surprising to the reader to learn that I do not attempt to discuss the evolution of humans or any of the other hominids that share the planet with us. This is because the primary motivation for this chapter is to place our visual system into its evolutionary context and I can accomplish that by stopping well short of the diversification of the hominids.

---

[1] Catarrhini is a parvorder of Primates. It contains the Old World monkeys, the gibbons or lesser apes and the hominids, which includes humans, chimpanzees, gorillas, bonobos and orangutans.

[2] In cladistics (a philosophy of classification that arranges organisms by their order of branching in an evolutionary tree), a clade is a group of organisms that consists of a single common ancestor and all of the descendants of that ancestor. It is a scientific hypothesis of the evolutionary relationships among the organisms in an analysis.

The goal of the visual processing system is perception, perceptions being loosely defined as "*the internal representations of the external world*" (R.L. Gregory, Eye and Brain, Fourth Edition). Perception is ultimately a physiological process but we are a long way from fully understanding those processes. In the following three chapters, I will do my best to explain what we do know about them with the intention of laying the foundation for a more scientific approach to visualization. In the end though, understanding the physiology of vision is not enough. Without evolutionary context, we may learn how we see but we will not know what we are meant to see, and that proves to be just as important. The catarrhine visual system is far from being a general-purpose tool. It evolved as we transitioned to new ecological niches and in those niches, we needed to detect certain things against certain backgrounds. We do not live in those niches anymore, but surprisingly our visual system still does. We may be orders of magnitude more massive and intelligent than our ancestors but we still see the same way.

The purpose of this chapter is to find the places in our evolution that were significant in the development of our visual system. There are two that are of major importance. The first is when we transitioned from being nocturnal to diurnal. The second is when we converted from being primarily insectivorous[3] to being primarily frugivorous[4]. That I have gone into so much detail here you should take as an indication of the importance that I place in understanding our relationship to the natural world. Before I started this chapter, I already knew that humans were primates; I just did not realize how much of a primate we really are.

---

[3] Insect eaters

[4] Fruit eaters

5.2 The Order Primates

Primates, form an order[5] within the class Mammalia and subclass Eutheria. The order includes lemurs, lorises, tarsiers, monkeys, apes and humans. The primate order contains over 300 species making it the third most diverse order of mammals after rodents and bats.

5.2.1 Primate Subgroups

There are two major lineages of primates well established in the literature, those being the anthropoids (comprising monkeys, apes and humans) and the strepsirrhines (Strepsirrhini, comprising *Lemuriformes* [lemurs] and *Lorisiforms* [lorises and bushbabies]). Tarsiers are thought of as a third primate lineage whose relationship to the other two is highly controversial (e.g. Groves, 1989; Kay *et al*., 1997). The two main hypothesizes are their placement of tarsiers as a sister-group to anthropoids (forming the clade Haplorrhini), or as a sister group to Strepsirrhini (forming the clade Prosimii). Recent molecular studies have produced conflicting results with Schmitz *et al*. (2001) lending support to the Haplorrhini hypothesis and Murphy *et al*. (2001a) supporting Prosimii. For the purposes of this thesis, I will adopt the more widely accepted (by Paleontologists) Haplorrhini hypothesis even though it may eventually prove to be incorrect.

The Strepsirrhini are regard as the more primitive group and subdivide into the infraorders *Lemuriformes* and *Lorisiforms*. They are characterized by having moist noses and bare muzzles, similar to a dog. Their noses are also covered with downy hair and they have a reflective layer, the tapetum lucidum, behind the retina. The tapetum increase the amount of light available to the retina and permits better functioning under scoptic[6]

---

[5] In the scientific classification used in biology, the order (Latin: *ordo*, plural *ordines*) is a rank between class and family. The eight "ranks" are as follows: Domain, Kingdom, Phylum, Class, Order, Family, Genus, and Species.

[6] Low-light condition such as nighttime.

conditions. While it makes the eye more sensitive to faint light there is a trade off because it also cuts down on visual acuity. Most, but not all, Strepsirrhini are nocturnal and they lack well-developed color vision.

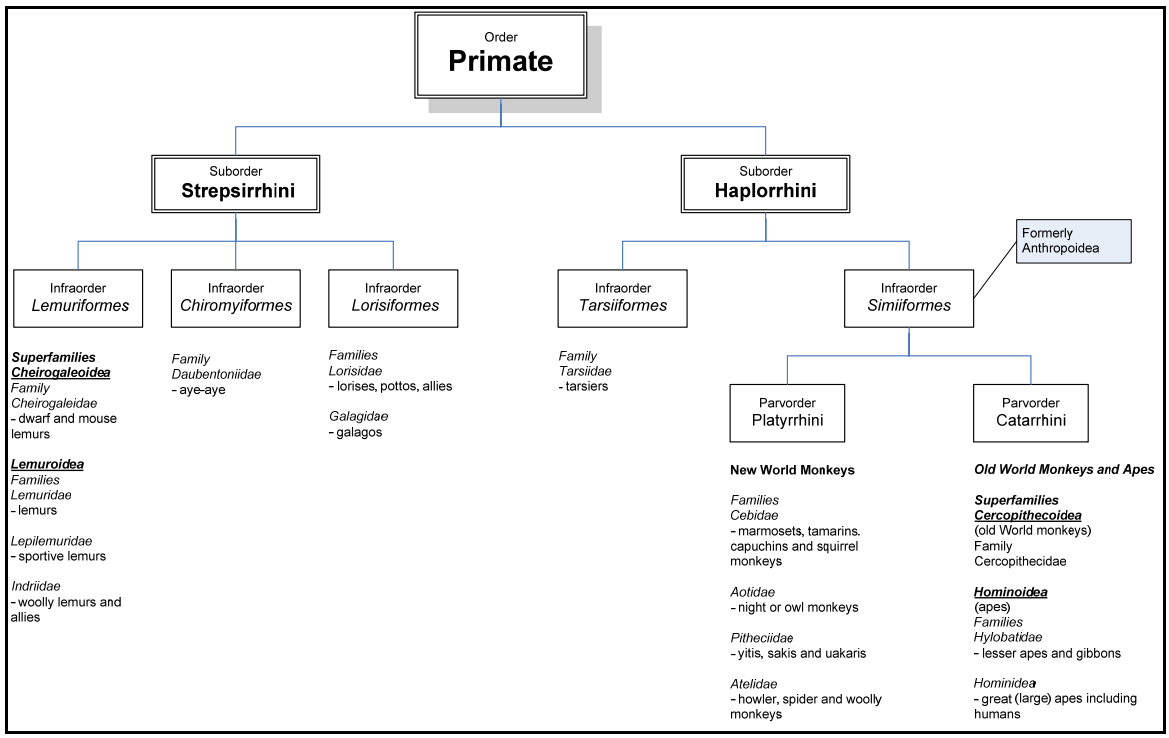


Figure 5.1: Primate family tree using the Haplorrhini hypothesis that groups tarsiers with anthropoids. Humans are found at the extreme lower right and are classified as order *Primates*, suborder *Haplorrhini*, infraorder *Simiiformes*, parvorder *Catarrhini*, superfamily *Hominoidea*, family *Hominidea*.

Haplorrhines, which include the tarsiers, the New World monkeys and the Old World monkeys and apes (including man) are considered the more highly evolved of the two subgroups. The Haplorrhini suborder subdivides into two infraorders, the Tarsiiformes (tarsiers) and the Simiiformes[7]. The Simiiformes in turn divide further into the Platyrrhini (New World monkeys) and the Catarrhini (Old World monkeys and apes and hominini).

[7] Whereas simiiforme is now the generally accepted term, throughout this thesis I use the older but more familiar term anthropoidae (anthropoid).

Haplorrhines have dry noses and have no tapetum. They also posses an area of enhanced vision in the retina, the fovea, which is of particular interest to this thesis. In contrast to the Strepsirrhini, most Haplorrhini are diurnal. In general, Haplorrhines are the more intelligent of the two subgroups and they are among the most intelligent of all mammals being on the same level with dolphins and elephants.

5.2.2 Physical Description and General Structure

Primates are a highly successful order, this success being based upon their relatively unspecialized structure. This has permitted primates to exploit a wide variety of evolutionary opportunities. Some of the lower primates, the aye-aye's, the tarsiers, the porto's and the lorises have become highly specialized but in general, the higher primates, the anthropoids, have avoided the dangers of evolutionary over adaptation.

Primates are essentially an arboreal species and they are well adapted for a life of climbing, leaping and running in trees. Even extant[8] ground dwelling species, including man, retain much of their original arboreal physical structure and all primates are comfortable in the trees. The general primate body type has long arms and legs and grasping hands and feet that aid in movement among the branches and provide stability on slender branches high above ground. With three exceptions, the spider monkeys, woolly spider monkeys and coloubus monkeys, all primates have five fingers and toes. All other mammals have either claws or hooves but only primates have flat nails. Some species of primates do have claws but even these have flat nails on the big toe (the hallux).

In all primates except humans, the hallux diverges from the other toes forming a pincher with the other toes that permits the grasping of objects such as branches. It is presumed that the loss of this divergence in the hominini reflects an adaptation to

---

[8] There are two classifications for species, extinct, which means that the species no longer exists and extant which means that the species is not extinct.

bipedalism. All catarrhini (old world) primates and some of the lemurs and lorises have an opposable thumb. Other arboreal mammals, for example the squirrel or opossum, also have an opposable thumb suggesting that the common ancestor of all primates was also arboreal. Primates also posses specialized nerve endings called Meissner's corpuscles in the hands and feet that provides increased tactile sensitivity. There is no evidence to suggest that any other placental mammal has them.

Primates also have forward facing eyes with a significant overlap between visual fields. This is a common feature of predators and it implies that the primate common ancestor was therefore a predator and possible insectivorous. This forward facing stereoscopic vision also provides a superior ability to judge distances than our ground dwelling mammalian brethren and it also provide far greater visual precision. Whereas the optic fibers of all mammals cross over to the other side of the brain, in certain primate species, including humans, up to 40% of the optic fibers do not cross over.

Primates do not have a highly developed sense of smell, this being relatively unimportant for an arboreal species that finds its food visually. Consequently, many primates have a much-reduced olfactory mechanism, the number of bones in the nose being reduced in both number and complexity when compared to nonprimate mammals. Primate teeth also differ from those of other placental mammals. Primates have low, rounded molar and premolar cusps whereas all other placental mammals have high pointed cusps. This makes the identification of primate teeth easy to recognize in the fossil record.

The principal evolutionary adaptation of primates has been the elaboration of the brain. In relation to body weight, the brain of a primate is larger than that of other terrestrial mammals and it has a unique fissure, the Calcarine sulcus, which separates the first and second visual areas on each side of the brain. The principal evolutionary trend in the primate brain has been towards elaboration.

Primates also differ from other mammals in the organization of the neocortex[9]. If an object is offered to a nonprimate mammal its immediate reaction will be to first smell it and then second to taste it. This is because the olfactory sense is the primary sensory modality in all nonprimate mammals and this is reflected in the dominance of the olfactory sense in their neocortex. The same object offered to a primate, however, invokes a very different response. When offered an object, a primate's primary response will be to examine the object by both touch and sight. In most primates, visual acuity and manual dexterity have replaced the sensitive and inquiring nose. The primate olfactory system has been severely reduced and replaced with a dominating tactile and visually dominant sensory system. This emphasis is reflected in the neocortex, which in primates is dominated by the areas dedicated to vision and touch.

The brain of the higher primates, the monkeys and the apes, is larger both relatively and absolutely than those of the lower primates. This quantitative increase is attributed partly to the elaboration of the tactile and visual sensitivity regions of the neocortex and partly to the elaboration of the intrinsic pathways connecting one part of the brain with another. The large, complex brain of humans is attributed to an increase in the size of and not number of nerve cells and to the greater complexity of connections linking one cell with another.

5.2.3 Primate Diversity

Primates are one of the most diverse orders of mammals showing a wide range of both size and adaptive diversity. The smallest primate is Madame Berthe's mouse lemur of Madagascar, which weighs in at 35 grams and the largest is the gorilla with a weight of 140-180 kg almost 4,000 times heavier.

---

[9] The neocortex, a part of the cerebral cortex, is characteristic of higher vertebrates such as mammals, which operate under the control of multiple sources of sensory input. It is the top layer of the cerebral hemispheres and it is made up of six layers labeled I to VI, It consists of grey matter surrounding the deeper white matter.

Primates occupy two major habitats: tropical forests and woodland-grassland vegetational complexes, the savanna. There is a greater diversity among the forest dwelling primates than among those occupying the grassland. This is most likely due to there being more ways of locomotion in the trees than there are on the ground. Tree dwelling primates can move about by leaping from branch to branch, a form of motion dictated by the hind limbs; by arm swinging, a function of the forearm or by quadrapedalism, which is a function of both forelimbs and hind limbs. In the savanna, the methods of locomotion are more limited being restricted to forms of quadrapedalism, hominines excluded.

Nonhuman primates also enjoy a wide distribution in terms of both geographic location and vertical range. They are widely distributed throughout tropical latitudes of Africa, India, Southeast Asia and South America. In Ethiopia, the gelda lives at elevations up to 5,000 meters and the Gorillas of the Virunga Mountains have been found at elevations of up to 4,200 meters.

The habitat of most nonhuman primates is predominantly tropical with only a few species extending their range to subtropical and temporal latitudes. The Barbary ape lives in the temperate forest of the Atlas Mountains and some populations of rhesus monkeys live in northern latitudes in China. The most remarkable primate habitat is that of the Japanese macaque that lives in the mountains of northern Honshu, an area that is snow covered for eight months of the year. Here they show the remarkable primate ability for adaptation (and hedonism) as they spend most of their days lounging in hot springs that form pools in volcanic areas. In general, however, most primates including humans prefer moist warm climates.

Primates also show a wide diversity of diet. The principal food of nonhuman primates is divided into vegetable (fruits, flowers, leaves, nuts, barks, pith, seeds, grasses, stems, roots and tubers) and animal (birds, eggs, lizards, small rodents, bats, insects, frogs and crustaceans). Nonhuman primates rarely eat the flesh of larger mammals, the exception being the chimpanzee, our closest extant relative.

In general, primates are omnivorous and in keeping with their generalized structure, there are relatively few primates who exhibit dietary specificity. Even species with definite dietary preferences are not exclusive. The leaf eating monkeys, the coloubus monkeys and the langurs, prefer young leaves but only in season. In other seasons, they will dine on fruit, flowers and seeds. Broadly, however, most primate species do exhibit dietary preferences. The apes (other than the mountain gorilla) are essentially frugivors. Many of the smaller primates, the galagos, dwarf lemurs, loris and the aye-aye are substantially insectivores. There are no known insectivores above approximately 500 grams and there are no known folivors[10] below. This is probably due to the difficulty that a large animal would have in catching small and nimble insects and the difficulty that a small animal would have in digesting the cellulose and hemicellulose components of leaves.

The tarsier is the only know primate to be exclusively carnivorous. Humans and by association extinct species of hominini are omnivorous and very general in their diet. It is difficult to discern from the fossil record what the original dietary preference of hominini was but it is almost certain that it was not heavily reliant on the flesh of large mammals. Although the fossil record of the tool making hominini does show examples of butchery, it is most likely that this represents scavenging activities.

---

[10] Leaf eaters

5.3 Anthropoid Evolution

5.3.1 Mammalian Evolution

Anthropoids descend from Primates which themselves descend from placental[11] mammals. Mammals are an ancient lineage that derived from members of the reptilian order Therapsida in the Triassic, between 248 and 206 million years ago. The therapids were a subclass of Synapsida, one of the earliest known reptile groups. The synapsids first emerged during the Carboniferous Period (354 to 290 million years ago) and were the dominant reptiles of the Permian (290 to 248 million years ago).

Despite the excellent fossil record of mammals, there is no direct line linking mammals and therapsids. The features that separate modern reptiles and mammals appear to have evolved at different rates and in response to a variety of interrelated conditions. At any point in the transition from reptiles to mammals, there were forms that combined the various characteristics of both groups. Such a pattern of evolution is termed a mosaic and it is common during the evolutionary transition of a major new adaptive type. Because of this mosaic evolution, it was originally difficult to classify fossils as either reptilian of mammalian. To simplify the classification many authors now classify mammals based upon a single characteristic, the articulation of the jaw between the dentary and the squamosal bones and the attendant movement of the accessory jawbone to the middle ear as auditory ossicles.

Mammals diversified into four main groups during the Mesozoic era: multituberculates, monotremes, marsupials and placentals. The multituberculates went extinct during the Oligocene about 30 million years ago but the other three main groups remain in existence. Evolutionary biology has long suggested that early mammals were small shrew-like nocturnal insectivores. Recently, however, this idea has been challenged

---

[11] Any member of the mammalian group (cohort Placentalia) characterized by the presence of a placenta, which facilitates exchange of nutrients and wastes between the blood of the mother and that of the fetus

by the discovery of several large mammals from the Early Cretaceous. Li *et al.* (2000) reported the discovery of a large Cretaceous mammal called *Repenomamus Robustus* while Hu *et al.* (2005) reported the discovery of *Repenomamus Giganticus***.** *R. Giganticus* was more than 3 feet long whereas *R. Robustus* was approximately 20 in long. Both species were carnivorous and one specimen of *R. Robustus* was found with a small dinosaur preserved in its stomach.

The two are the only known members of the family *Repenomamidae*. The discovery of the two is important because it was previously believed that the niches of animals larger than a few feet were fully occupied by the dinosaurs and were not filled by mammals until after the extinction event at the end of the Cretaceous. The discovery of *Repenomamus* proves that at least some species of mammals were able to compete against the dominant dinosaurs. All other known mammalian species from this period, however, are very small and it remains to be seen if other larger and more competitive species existed. It is still generally accepted that most mammals of this early period were small and rodent like and that they only exploded into medium and large sized mammals during the Paleocene period (64-58 mya).

5.3.2 Monotremes

Of the modern mammalian orders, the monotremes are the most ancient order and there are now only two known extant families, the amphibious platypus and the terrestrial echidnas. As well as being egg-layers, they share primitive skeletal features such as a shoulder girdle that has been lost in the other mammalian orders. Their exact evolutionary relationship is hard to determine because they not only share many evolutionary features with extinct early mammals but genetic information places them close to the more advanced marsupial order. Although there are no universally accepted theories for monotreme evolution, there are two known extinct orders of monotremes from the early cretaceous that prove their ancient lineage.

5.3.3 Marsupials

The second order of mammals is the marsupials[12]. Marsupials are characterized by premature birth and continued development of the newborn in the marsupium[13] of the mother. The majority of marsupials are found in Australia, New Guinea and the surrounding islands (over 200 species). Approximately 70 species live in the Americas, almost all in South and Central America, and one, the Virginia opossum lives throughout North America.

The structure and behavior or marsupials in many cases parallels that of placental mammals indicating that they filled evolutionary niches at the end of the cretaceous in much the same way that placental mammals did. Marsupials are substantially less intelligent than placental mammals, the ratio of the brain size to body weight being much less. They are also notably less social. Beyond short-lived pair bonds during mating, there is little social organization and even species such as kangaroos, which move about in groups, display little, if any, socialization. Fossil evidence indicates that marsupials developed in the New World, the oldest know marsupial fossils dating to the Late Cretaceous.

5.3.4 Placentals

Like the marsupials, Placental mammals (infraclass Eutheria) evolved during the late Cretaceous period. During this period of early evolution, they were usually smaller than present-day rabbits. Placental mammals existed well before the KT[14] boundary but exploded in its aftermath. The true placenta allowed for a longer development period

---

[12] Any of the 250 species belonging to the mammalian infraclass Metatheria (also know as Marsupialai)

[13] A flap of skin covering the nipples of marsupials.

[14] The K-T boundary refers to the Cretaceous-Teriary extinction that occurred approximately 65.5 million years ago.

within the protection of the womb, a major factor that made them ideal for seizing the terrestrial environments vacated by the dinosaurs at the end of the cretaceous. Extant placental mammal diversity is now generally accepted to be the result of post KT rapid evolutionary radiation and has culminated in 4800 living species placed into 18 modern orders.

The phylogenetic[15] pattern and temporal framework of the diversification of placental mammals has been an area of intense controversy (Benton, 1999; Brohman *et al*., 1999). Until recently, only two superordinal clades were recognized from both morphological and molecular data sets: Paenungulata (*Proboscidea*, *Hyracoidea*, *Sirenia*) (Lavergne *et al*., 1996; Simpson, 1945) and Cetartiodactyla (a group comprising *Cetacea* and *Artidactyla*) (Gatesy *et al*., 1999; Shoshani and McKenna, 1998).

Recently, two studies using nuclear genes and large molecular data sets (Madsen *et al*., 2001; Murphy *et al*., 2001a) established the existence of four primary placental mammal clades: (a) Afrotheria (*Proboscidea, Hyracoidea, Sirenia*); (b) Xenarthra (sloth's, anteaters and armadillos); (c) Euarchontoglires (a taxon[16] combining Euarchonta [*Primates, Dermoptera, Scandentia*], and Glires [*Rodentia* and *Lagomorpha*]); and (d) Laurasiatheria (*Eulipotyphla* [core insectivores], *Chiroptera*, *Cetartiodactyla*, *Perissodactyla*, *Pholidota* and *Carnivora*).

Molecular dating analyses based upon these data sets strongly suggest that the superordinal divergences among placental mammals preceded the Cretaceous – Tertiary (KT) boundary, ranging from 64 to 118 mya (Eizirik, *et al*., 2001; Madsen *et al*., 2001; Murphy *et al*., 2001a). A more recent study using an expanded data set confirmed support

---

[15] Phylogenetics (from the Greek: *phylon* = tribe, and *genetikos* = relative to birth) is the study of the evolutionary relatedness among groups of organisms. Phylogenetics treats species as a group of lineage-connected individuals over time.

[16] A taxon (plural taxa), or taxonomic unit, is a grouping of organisms. Once named, a taxon will usually have rank such as class, order, family etc.

for the four major clades and identified the position of the placental root to be between Afrotheria and the other major groups (Murphy *et al*., 2001b). These results suggest a southern origin for the extant lineages of placental mammals, with the earliest separation (ca. 103MYA) coinciding with the final breakup of Africa from South America in the late Cretaceous.

Although many of the relationships and the temporal framework of the sister-groups within the clades remains in question, it is clear that placental mammals were already a widely established and divergent mammalian order prior to the K-T boundary. The survival of the four clades (and the other mammalian orders) thorough the K-T event was potentially made possible by their being (a) homoeothermic; (b) nocturnal; and (c) insectivores. What is clear is that within a few million years of the K-T event, placental mammals had experienced rapid evolutionary radiation and were well on their way to occupying most of the vacant evolutionary niches.

5.4 Primate Origins

Primates are members of Euarchonta, a super order of mammals that contains four orders: (1) The *Dermoptera* or colugos; (2) the *Scandentia* or tree shrews; (3) the extinct *Plesiadapiformes* (closely related or a precursor to primates) and (4) *Primates*.

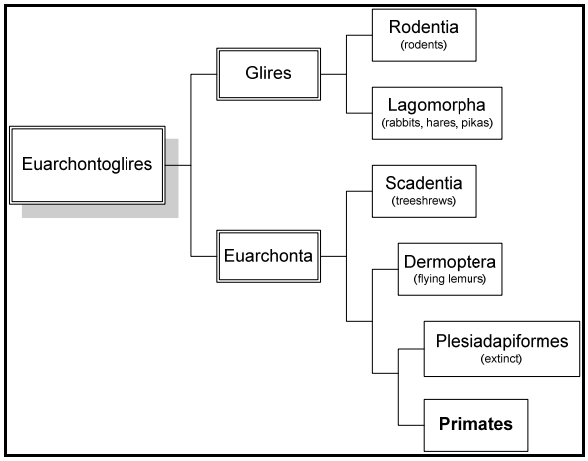


Figure 5.2 : The primate family tree. Primates descend from the clade Euarchontoglires, one of the four primary groups of placental mammals.

Determining exactly when these lines split is an active area of research. The phylogenetic analysis of primate lineages and dating is made difficult by the incompleteness of the fossil record. Even so, paleontologists and neontologists generally accept the following “facts” about our heritage:

1. Tarsius is the closest living relative to the anthropoids and that strepsirrhines, lemurs and lorises, are more distantly related.

2. The oldest Tarsius relatives occur in the Asian model Eocene,

3. The oldest undisputed fossil record of anthropoids is from the late Eocene localities in Afro-Arabia.

4. Platyrrhines[17] first appear in the late Oligocene in South America.

This is not much to go on and beyond these four, very little is clear. A recent study using molecular dating and Bayesian analyses (Eizirik *et al*. 2004), however, has clarified the dates of early primate divergence and the divergence of primates from the other Euarchonta. For this study, DNA sequences from 15 nuclear genes were analyzed for 14 species (*Home sapiens, Pan Troglodytes, Hylobates concolor, Macaca mulatta, Atles fuscipeps, Callimico goeldii, Lemur catta, Otelemur garnetti, Tarsius bancanus, Tarsius syrichta, Tupaia minor, Urogale everettii, Cynocephalus variegatus,* and *Cynocephalus volans*). Figure 5.3 summarizes the results of this work.

Interestingly, the molecular dating technique indicates that most of the major divergences within the primate order and its divergence from the other Euarchonta occurred prior to the KT boundary. It indicates that the major divergence of primates from the other Euarchontas, Scandentia and Dermoptera, occurred approximately 83 million years ago, long before the end of the cretaceous. Further analysis reveals that Anthropoids separated from the other primates approximately 77 million years ago and that Tarsiiformes split from strepsirrhines at approximately 71 MYA. This dating indicates that the Tarsiers are probably most closely related to the strepsirrhines and that

---

[17] Platyrrhines or South American monkeys comprise one of the two infraorders of anthropoid primates. Humans belong to the second infraorder, the Catarrhine primates (also know as Old World monkeys and apes).

Prosimii hypothesis, as opposed to the Haplorrhini hypothesis used in this thesis, is most likely correct.

Although this analysis opens the door to several questions, including clouding the haplorrhini/strepsirrhini debate, it definitely indicates that the base primate lineage underwent radiation into its three main modern lineages, anthropoids, strepsirrhines and tarsiers very rapidly. It also implies that this radiation occurred so rapidly that was little time for the three to develop common evolutionary traits.

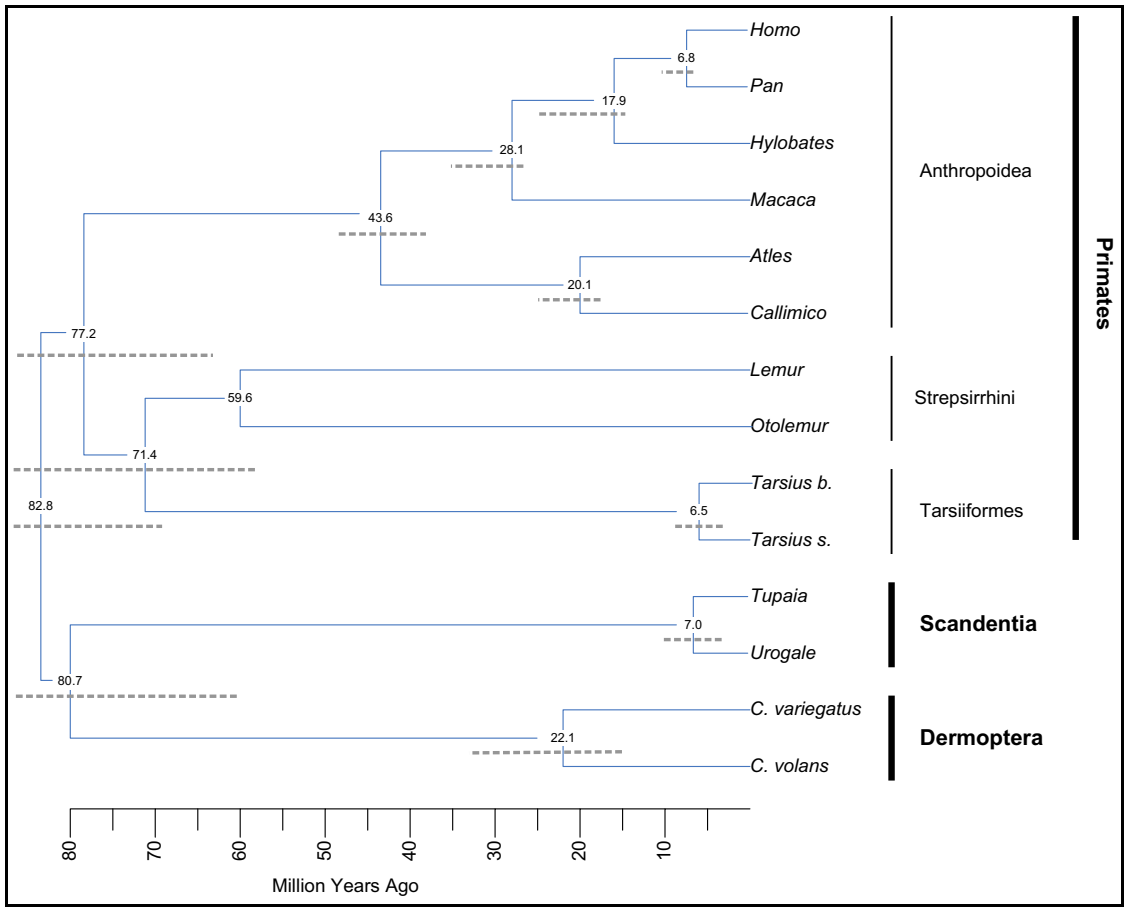


Figure 5.3: Phylogenetic tree depicting the timing of diversification of major lineages of Primates and their closest relatives. Numbers indicate the point estimate for the time of divergence (in MYA) at the adjacent node. The shaded line under each number represents the 95% credibility interval for the estimates. *From Anthropoid Origins New Visions, pp 58, Kluwer Academic/Plenum Publishers, New York.*

The credibility intervals (95% credibility) for the study were at times quite broad (broken lines in Figure 5.3). The results of this study, however, are consistent with other

studies using different techniques (Springer at all., 2003; Shoshoni and McKenna, 1998) and an extensive error analysis indicated that the dates given in Figure 5.3 are most likely too young, thus indicating that the major divergences occurred even earlier. This analysis along with others now strongly indicates that the anthropoid lineage, the line that would eventually evolve humans, is very ancient and that it had already diverged from its sisters long before the KT boundary.

5.5 Primate Fossil Evidence

As detailed in Chapter 5.4 on Primate Origins, DNA analysis has indicated that primates are an ancient lineage and that the anthropoid line split from the other primates in the Late Cretaceous. Unfortunately, this is not confirmed by any hard evidence in the fossil record. Recent fossil finds in North Africa (De Bonis *et al*., 1988; Godinot and Mahboubi, 1992; Simons, 1992) have pushed the fossil record of Anthropoidea from the Oligocene[18] to the Eocene[19] but to date no intermediate fossils linking anthropoids and either tarsiers or strepsirrhines has been found. The oldest well-known uncontested anthropoids (*Parapithecidae* and *Oligopithecidae*) come from the late Eocene of Egypt (~37MYA) (Simons, 1992). They are more primitive morphologically than any living anthropoid but their morphological organization was substantially identical to living platyrrhines.

Early fossil tarsiers and strepsirrhines are so poorly known that they contribute virtually nothing to narrowing the gap between anthropoids and other primates. There are virtually no Lemuriforme fossils prior to the Miocene; just a few doubtful records exist in

---

[18] Oligocene (from the Greek *oligos* (few) and *ceno* (new): A geologic period that extends from about 34 million to 23 million years before the present.

[19] Eocene (from the Greek *eos* (dawn) and *ceno* (new): A geologic period that extends from 55.8 million to 33.9 million years before present. The name refers to the "dawn" of modern mammals that appeared during the epoch. The end of the Eocene is marked by a major extinction event that may be related to impact of several large extraterrestrial bodies.

the African Paleogene[20] (Simons, 1995; Simons *et al* 1994). Recently discovered tarsier fossils have pushed the earliest occurrence of tarsiers to the late middle Eocene of Asia (~45MYA), establishing a much greater minimum age for the differentiation of the haplorrhine suborder from other primates. However, this group is only known from dental remains so the record is incomplete.

This lack of direct fossil evidence for early primates should not be taken as an indication that early primates were rare. The fossil record in general is not indicative of an entire ecosystem and the frequency that a species occurs within the record does not directly indicate how common they were in their environment. The frequency of occurrence in the fossil record is more an indication of the type environment that a species lived in. For an animal to be preserved as a fossil requires that it be quickly buried after death and escape savaging of the remains. It is generally accepted that early primates were small (< 500 gm) and arboreal. The most frequent cause of mortality in similar extant species is predation from either snakes or birds. It is highly unlikely then that a small, arboreal creature would live long enough to die of natural causes and even then, it would likely not fall into any area suitable for fossilized preservation.

Two Paleogene groups have been identified as possible sister or parent taxa for Anthropoidea. The Eocene-Oligocene Adapidae, from North America, Asia, Europe and Africa is thought by some researchers to be related to anthropoids (Franzen, 1994; Gingerich, 1980; Simons *et al*., 1989). It is now generally accepted however, that adapids are the group from which Lemuriformes emerged and so have no bearing on anthropoid origins.

---

[20] Paleogene is a unit of geologic time that began 65.5 million years and ended 23.03 million years before the present. It began with the mass extinction that ended the cretaceous period and is most notable for the rapid radiation of mammals. It consists of three epochs, the Paleocene, the Eocene and the Oligocene.

A second group, the Paleogene Omomyidae of North America Asia, Europe and possible Africa is generally considered to have given rise to tarsiers and anthropoids either via separate omomyid stocks (Szalay, 1967) or via a common stem lineage. Both groups are now well represented in the fossil record by skull and limb bones and thus a good phylogenetic assessment is now possible. A third possibility is that the omomyids are sister to anthropoids and that the anthropoid lineage predates them and is much more ancient (Hofstetter, 1980).

Recently, the remains of a new family of Eocene primates, the Eosimiidae, were discovered in Asia (Beard *et al*., 1994; Beard *et al*., 1996). This has prompted the question of whether or not they too are sisters to anthropoids.

Kay *et al*. (1997) evaluated the conflicting hypotheses about anthropoid relationships. They undertook a cladistic analysis of 256 dental, cranial and postcranial characters for 50 taxa, including platyrrhines, Lemuriformes, the best-known ancient fossil catarrhine, Aegyptopithecus, Tarsius, fossil adapids and omomyids, several Eocene-Oligocene African anthropoids and Eosimias (dental remains only).

Their results supported several major conclusions pertinent to anthropoid origins: (1) the primary dichotomy in living primates is between haplorrhines and strepsirrhines (as opposed to anthropoidae and prosimii). (2) Adapidae is the sister group of Lemuriformes and is therefore not ancestral to anthropoids. (3) Omomyidae is part of the anthropoid lineage. (4) Eosimiidae is a sister group to Anthropoidae and not a direct ancestor. (5) Tarsius is a sister group to Anthropoidea and may be nested within Omomyidae (this is uncertain). Figure 5.4 summarizes this work.

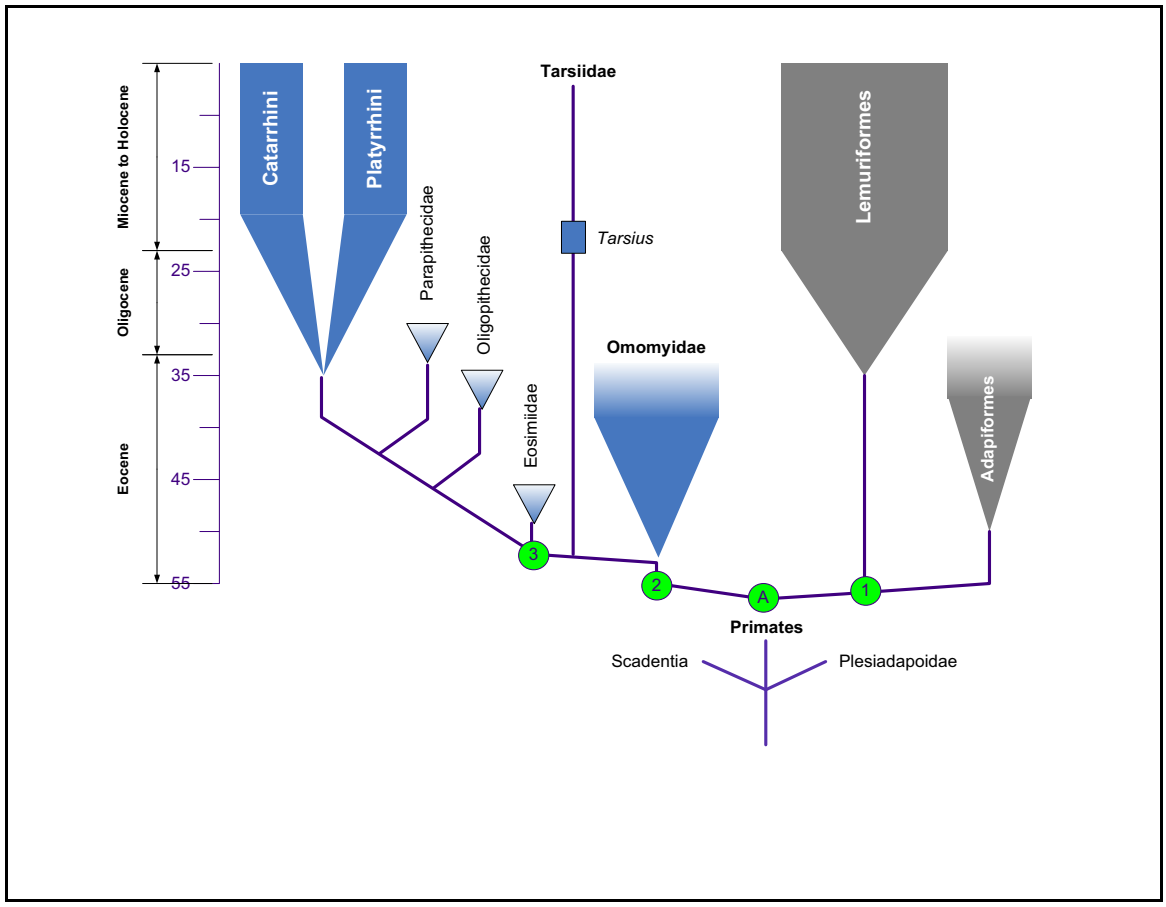


Figure 5.4: Phylogenetic relationships of the better-known taxa of Eocene and Oligocene primates as well as extant haplorrhines and strepsirrhines. Symbols at the nodes are as follows: A, Order Primates; 1, Strepsirrhini; 2, Haplorrhini; 3, Anthropoidea. *From Fig 3, Kay et al, 1997*

5.6 The Early Anthropoids

It is still not clear whether Omomyidae are a direct anthropoid ancestor or are a sister taxon. However, what is clear is that even if they are not part of the direct anthropoid lineage, they are definitely haplorrhines and they are very close to the direct lineage. This means that we can use them as a model for what the early anthropoids we like. The most important thing to realize about the early anthropoids is that they did not look at all like us. The average homo sapien weighs in at ~70kg or 70,000 g. This is approximately 250 times heavier than the typical omomyid. They were very small compared to us, so small in fact, that it is difficult to see any remnant of them in ourselves.

When we look at the modern catarrhine primates, it is clear that their body size, general morphology and high intelligence indicate that they have undergone extensive evolution since the time of the early anthropoids. It may seem then that understanding the nature, morphology, habits and diet of our early ancestors may not be relevant to us today. However, as I will show in more detail in later chapters, the visual system of the stem anthropoids was very similar to extant species. Primates in general and catarrhine primates in particular have, for mammals, a highly acute and well-developed sense of color vision. This acute color vision is the underlying science of visualization and unlike our body size or intelligence; it developed early in our history and has remained intact.

Catarrhine primates and humans in particular have unique lives. It is important to understand that these lives evolved because of our color vision and not the other way around. An understanding of our color vision then must start with knowledge of the lives of these early anthropoids. The evolutionary pressures on them led to development of our acute, trichromatic vision. Moreover, it is that vision that has led to evolution of our modern lifestyles.

Whenever you look in a mirror, you see in the reflection, a highly intelligent catarrhine primate. That is not who looks back. Twenty-two mm behind your reflection in the mirror lies a retina and a visual system that belongs to someone else. Half of it, the acute part, belongs to the earliest stem anthropoid and the other half, the trichromatic color vision, to the earliest catarrhine.

5.6.1 Diet and body size

Fortunately, the enamel coating of primate teeth makes them ideal for fossilization and in consequence much of our knowledge of early primates including omomyids, comes from their teeth and their dental morphology. This, in combination with their body size, can give us a direct indication of their diet. Body size is important because there is great similarity between the dental morphologies of insectivores and folivors. Because insects have hard exoskeletons and leaves have large amounts of fiber, both insectivorous and folivorous primates have relatively long cheek-tooth shearing crests. Because of this,

it is difficult to distinguish whether a species primarily ate insects or whether it primarily ate leaves based on dental morphology alone. Body size, however, can play a direct part in determining whether a given primate species is insectivorous or not. Purely insectivorous primates' weigh less than 500 g. This is because only small-bodied animals are able to acquire enough insects to fuel their high metabolisms. Purely folivorous primates, however, never weigh less than 700 g because digestion of high-fiber leaves is time consuming and not compatible with high metabolisms. Thus, small primates with well-developed shearing crests are most likely insectivorous whereas larger primates with the same dentition are most likely to be folivorous. In contrast, purely frugivorous or gummivorous[21] primates occur at all body sizes and have poorly developed cheek-tooth shearing crests because their food lacks substantial fiber.

The fossil record indicates that early anthropoids were small (< 1000 g) with most omomyids ranging between 50 and 400g. This suggests that the divergence of anthropoids from the non-anthropoid ancestors may not have been associated with a significant change in body size or diet. Early anthropoids were clearly too small to be folivorous and must have been insectivorous or frugivorous or possible both (Williams *et al*., 1994). It is believed that leaf eating evolved late in anthropoid evolution, evolving separately in catarrhines and platyrrhines. By the time of the Late Eocene - Early Oligocene two anthropoids (*Parapithecids* and *Propliopithecids*) had developed low-crowned molars, which suggest that they had made the change from insect eating to more low-fiber herbivorous diets.

5.6.2 Diurnal or Nocturnal?

It is well established that the early mammals and primates were small and nocturnal. Small, visually oriented extant nocturnal mammals have larger eyes and orbits than do diurnal ones. This difference is not reflected in larger extant species. Once the body size exceeds ~1300 g, differences in activity are not reflected in relative orbit size.

[21] Animals that eat primarily gums, saps and other tree exudates.

Consequently, it is assumed that diurnality/nocturnality cannot be determined for larger fossil primates as well. However, most omomyids (e.g. *Necrolemur*, *Pseudoloris*) are relatively small and their orbit size clearly indicates nocturnal habits (Beard *et al*., 1991, Kay *et al*., 1977).

Several late Eocene anthropoids (*Apidium, Simonsius* and *Catopithecus*) (Simons *et al*., 1994), however, are inferred to be diurnal, as are most extant anthropoids. This indicates that the switch from the ancestral nocturnality to diurnality occurred early in anthropoid evolution. This is further indicated by a study in orbital convergence (Ross, 1995) which indicates that anthropoids adopted diurnality at body sizes less than ~1300g.

A significant development in the history of anthropoids was the development of the fovea (which will be of significance later). Haplorrhines (tarsiers and anthropoids) are the only mammals with a fovea. The fovea is an area of the retina with a high density of visual receptors and a low ratio of photoreceptors to ganglion cells. The fovea improves visual acuity in a small area of the visual field and among vertebrates is found in fishes, reptiles and birds that are diurnal visual predators (Walls, 1942). This suggests that haplorrhines evolved a fovea to increase acuity for diurnal visual predation (Ross, 1996).

Further evidence for this can be found among extant anthropoids. Many small-bodied species, (e.g. squirrel monkeys) spend much of their time searching for insects and small invertebrates (Garber, 1992). This is further indication that haplorrhines evolved fovea to assist in diurnal visual predation. Given that it is generally accepted that early primates were nocturnal and insectivorous, the development of the fovea must have occurred early in anthropoid evolution while they were still predominantly insectivorous, probably as they were converting from nocturnal to diurnal habits.

5.6.3 Who were the Early Anthropoids

With all this in mind, it is possible to assess who the early anthropoids were and who the earliest catarrhine primate was. Omomyidae[22] is generally accepted to be the parent Haplorrhini and thus to be closely related to the stem Anthropoidea (the line that eventually led to Platyrrhini and Catarrhini). Omomyids though are believed to have been nocturnal and are not considered as the earliest anthropoid, which was diurnal.

Figure 5.4 indicates that the earliest anthropoids evolved around 46 million to 47 million years before the present. Although we know when it evolved, the fossil record is not yet clear enough to indicate whether stem Anthropoidea arose in Africa or Asia. *Eosimiids*, the earliest well-known Asian anthropoid and *Algeripithicus*, the earliest well-known African anthropoid, are of the same age, having emerged sometime in the early to middle Eocene. Both of these species, and other less well-known species, resembled Omomyidae in body size and general morphology. This leads us to believe that the Eocene species that gave rise to anthropoids must have closely resembled omomyids. Even though we are not clear on exactly where they evolved we can at least say that it appears that the earliest anthropoids evolved in the Old World and subsequently colonized the New World.

Unfortunately, the brain development of these early anthropoids is not particularly clear. The neocortex of all extant nonprimate mammals is dominated by their olfactory sense whereas in extant primates it is dominated by a visual/tactile sensory system. This indicates that the olfactory sense dominated the original mammals and that the unique primate neocortex evolved after their divergence. At present, it is unknown exactly when this transition took place and there is considerable uncertainty if the early stem anthropoids had already undergone this evolution. For example, the stem anthropoids,

---

[22] The reader should note that the Haplorrhini theory is still very much in question and tarsiers may well be sister to strepsirrhines forming Prosimii instead. For the purposes of this thesis, however, this is irrelevant since in either case, the omomyids are considered to be parent to anthropoids.

Apidium and *Catopithecus* and the early catarrhine *Aegyptopithecus* had relatively smaller brains and larger olfactory lobes than existing anthropoids (Fleagle *et al*., 1987; Kay *et al*. 1977). This indicates that the reduction in the olfactory sense may have occurred later in our development and possible arose separately in catarrhines and platyrrhines.

Extant anthropoids have highly acute vision. This is due to both our evolution of a fovea centralis and the loss of our tapetum lucidum. As I showed in Chapter 5.2.1, both adaptations are reflective of diurnality. Since both tarsiers and anthropoids share these adaptations, it is most likely that the stem anthropoid was also diurnal and probably also possessed a fovea. The fovea here is of great interest because it shows how an adaptation originally developed for one activity can be usurped for another. The primary driving force behind the evolution of the fovea is thought to be the diurnal predation of insects. However, we do not use it for that purpose very often anymore; now use it for reading and for interpreting seismic data. Without the enhanced visual acuity that it provides, neither activity would be possible.

Of particular interest to this thesis is the trichromatic nature of our color vision. All catarrhine primates are trichromatic, having three different cones each maximally sensitive to a different part of the spectrum. Tarsiers and most platyrrhines, however, are dichromatic, as are all nonprimate mammals. This indicates that the stem anthropoid was also dichromatic and that the final step in our visual development, the evolution of trichromacy, would occur much later in our history, and probably not until after the divergence of platyrrhines and catarrhines in the Oligocene. I will talk about why this occurred later in the thesis.

Given then what we know about omomyids and other early anthropoids we can make a reasonable statement about what our earliest anthropoid ancestor might have been like. Kay *et al*. (1997) hypothesized that it was probably a small-bodied (<1000g), partially insectivorous, primarily leaping and active, arboreal quadruped, and, for reasons not covered here, it was most probably solitary.

This then was our earliest anthropoid ancestor. From this small active creature would eventually emerge all of the monkeys, apes, hominines and finally humans that now dominate the arboreal forests of the tropical world. In many ways we are very different from this creature. Our body size, great intelligence and general morphology, diet and habits have evolved to be almost unrecognizable. Nevertheless, this earliest ancestor is still there in our visual system. They still, to a very large extent, dominate who we are.

CHAPTER SIX: THE VISUAL ACQUISITION SYSTEM

"To any vision must be brought an eye adapted to what is to be seen, and bearing some likeness to it."
(Plotinus, Section 9, 6th tracate 1st Ennaid; 3rd Century C.E.)

6.1 Introduction

6.1.1 Geophysical Analogs

As one might expect, you do not have to go too far into the visual system before realizing that it is magnificently complex, wonderfully majestic and staggeringly beyond any attempt at simplistic explanation. Nevertheless, being a geophysicist, I have found that it is both possible and useful to draw analogies between the visual processing system and the world of seismic with which I am more familiar. A reader familiar with seismic will recognize that it can be split into three broad areas, each with links to the others; these are:

1. Acquisition
2. Signal Processing
3. Interpretation

The visual system too can be split into these three same areas, the difference being that the feedback systems between the three are infinitely more complex. Although the reader is cautioned not to take the analogies too far, it is possible to consider the visual system in terms of these same three areas. The visual acquisition system captures the photons that enter the eye and converts them into streams of neural impulses, essentially projecting the three-dimensional physical world onto a two-dimensional plane. The visual signal processing system then modifies the neural impulses leaving the photoreceptors and converts them into parallel streams of information that travel the optic nerve. Finally, the visual cortex interprets these streams of information and attempts to reconstruct, in the mind, the original three-dimensional world.

All three together form the visual processing system. This chapter deals with the first part of system, the visual acquisition system. It deals with how light is received, how the retina is organized, the significance of the fovea and the development of the various cones that feed the entire system.

6.1.2 Vision and Evolution

It is interesting and somewhat paradoxical that vision, which is so much a product of evolution, would initially be such a problem for it. Darwin viewed human optics as being perfect. Considering that natural selection could not take place in a perfect world this posed a serious problem to the theory of evolution and it was quickly and violently grabbed onto by his detractors. They argued that the natural world was chaotic and imperfect and that it was irrational to believe that anything as perfect as the human eye could arise from it by random chance. The eye must, therefore, have been designed by a supreme intelligence and it must have been a product of creation. The debate was finally resolved, to Darwin's great relief, by Helmholtz who proved conclusively that the human eye suffers from chromatic aberration (Cronin, 1993) and thus was far from perfect.

It is impossible to study vision without also considering its evolutionary context. That is because the eye and the retina in particular are one of the most obvious places that we can look to see evolution in action. The photoreceptor layer of the retina serves as the visual acquisition system sampling the visual field both spatially and chromatically. In this, it resembles a 3D seismic survey with the various rods and cones acting as the biophysical equivalents of geophones. The resemblance, however, is purely superficial. In geophysics, uniformity is the order of the day but in the real world, nature abhors uniformity and nowhere is this maxim better expressed than in the retina.

Geophysical surveys sample the acoustic wavefield both spatially and chromatically but they do so in a consistent manner. For example, when designing a survey the goal is usually to provide evenly sampled pre-stack and post-stack wavefields. Because of this, we generally lay out shots and receivers on regular grids and we put a lot of effort into the design of the grid patterns. Similarly, the geophones that sample the wavefield

chromatically, ideally have a flat response over the seismic bandwidth. A geophone that has a different response to various frequencies is undesirable and manufacturers go to great lengths to eliminate any such effects.

However, this is not the case with the retina. It also samples the visual field spatially and chromatically but the sampling is anything but uniform. Humans, for example, have highly acute spatial vision and to the limits of optics, human vision is about as good as it gets. However, this "super vision" is restricted to a very narrow cone approximately six degrees wide. Outside this cone, our visual resolution drops by a factor of 20. We are never consciously aware of this lack of resolution because much of our conscious self is tied to this six-degree cone, which is a direct function of our evolutionary lineage. Likewise, when it comes to chromatic sampling, humans with normal color vision have a better ability to discriminate colors than does any other non-primate mammal. This color discrimination, however, is also highly non-uniform and there are colors that we can barely discriminate at all. Once again, as with our spatial resolution, we are not consciously aware of it.

This lack of uniformity in both spatial and chromatic sampling is a common trait of all vertebrates. What is not common is the form of the non-uniformity itself. Each different species has a different type of non-uniformity each of which is directly indicative of the evolutionary pressures that directed the development of the species visual system. What is particularly fascinating in humans is that our patterns of spatial and chromatic sampling developed separately but once established have remained virtually intact over tens of millions of years. Our spatial sampling abilities, for example, developed when our ancestor weighed less than 500 gm's and at a time we were converting from the nocturnal to the diurnal predation of insects. Most of us do not do that anymore but we still could if we wanted too, we still have that vision.

I started this chapter with a quote from the Greek philosopher Plotinus, the father of Neo-Platonism. I wonder when he wrote it, if he had any concept of evolution or of how many centuries would pass before we would prove he was correct. Nevertheless, he is

correct because stamped upon the retina of every vertebrate is a detailed description of an ecological niche and in our case it is one that has long since disappeared.

6.1.3 Note on Sources

This chapter and the two that follow introduce the physiology of vision. In this chapter, I discuss how the visual system goes about its job of capturing light; in Chapter 7, I discuss how it converts the raw chemical inputs from this capture into streams of usable information; and finally in Chapter 8, I discuss how we convert these streams of information into our perceptions of the world around us. I will say at the outset that our knowledge of these processes is far from complete. The ratio of what we know to what is too be known starts at close to one at the photoreceptor layer and declines to almost zero by the time we reach the higher levels of the visual cortex. Even so, it is possible to trace the effect of a photon from the time it enters the eye through to the visual cortex and make some reasonable comments as to how it contributes to perception.

What follows in each chapter are only brief summaries of very complex subjects. I will try to keep my descriptions concise and at a level that is pertinent to the overall theme of scientific visualization. For a more comprehensive description of the entire visual processing system and how it evolved, the reader is directed to the following sources; (1) The Primate Visual System: a comparative approach, Jan Kremers, Wiley Publishers. (2) Anthropoid Origins: New Visions, Ross & Kay, Kluwer Academic/Plenum Publishers. (3) Webvision: http://webvision.med.utah.edu, a teaching resource from the University of Utah. (4) The Perception of Color, ed. Peter Gouras, in Vision and Visual Dysfunction 6, CRC Press Inc (5) EYE and BRAIN: The psychology of seeing, Fourth Edition, R.L. Gregory.

Figure 6.1, Figure 6.3, Figure 6.4, Figure 6.5 and Figure 6.6 are all taken from Webvision: http://webvision.med.utah.edu and are used with permission.

6.2 Early Development of Vision

There is a tremendous diversity in eye types and structures found within the animal kingdom. Phylogenetic studies have shown that almost all animal phyla have evolved some form of light-sensitive organ (Salvini-Plawen, 1977, Land, 1992). These range from simple eyespots with only a few photoreceptor cells to eyes with highly developed optics. There are roughly 10 known distinct forms of optical system ranging from the pinhole eye of the nautilus, eyes that resemble reflecting telescopes, two kinds of camera lens eyes and several kinds of compound eyes.

The three largest animal phyla, vertebrates, mollusks and arthropods, have each evolved a different solution to the problem of obtaining an optical image. Both vertebrate and mollusks use a single lens optical system (although vastly different) whereas the arthropods have multifaceted compound eyes. These morphologies have long been considered to have evolved separately. Compound and single-lens eyes are so different that it is difficult to conceive of a common ancestor to both. Similarly, the single-lens optics of vertebrates and mollusks are fundamentally different which implies that the systems must have evolved separately. The fundamental differences in morphology, development and photoreceptor infrastructure of the various eyes found within the animal kingdom have led to the suggestion that eyes must have evolved separately as many as 40 times (Salvini-Plawen, 1977).

Recently, however, studies of conserved regulatory and structural genes (Halder 1995) have pointed to a single monophyletic origin of photoreceptor cells in evolution. All of the visual system studied shared homologous proteins call opsins that serves as the primary photoreceptor. In addition, the Pax-6 gene, which plays a master role in eye development for both vertebrates and invertebrates, has been shown to be highly conserved between studied phyla. Pax-6 regulates the development of the embryonic eye of vertebrates and in the early eye development of Drosophila (fruit fly). It has also been found in squid and has a similar function. Therefore, it appears that the same regulatory gene is present in vertebrates, mollusks and arthropods. This gene has also been found in other phyla although its role has not yet been studied.

It now seems probable that rather than evolving independently, all visual apparatus share a common monophyletic origin that dates back to the early Cambrian. This is far from certain, however, and the subject remains an active area of research (Fernald, 2001).

For the purposes of this thesis, I am only concerned with the development and functionality of the vertebrate eye. Figure 6.1 shows a simplified model for the evolution of the vertebrate eye. In this model, the earliest form of eye is called an "eyespot" and it was simply a slightly indented patch of photoreceptors. This early version of the eye models the modern sensory apparatus for both taste and smell. Eyes of this nature could distinguish light from dark but not direction (Land 1992).

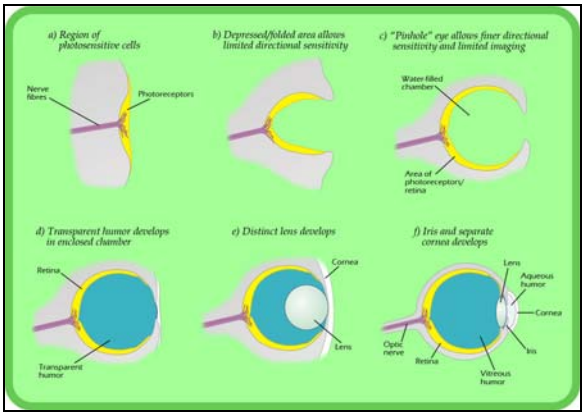


Figure 6.1: Diagram of major stages in the evolution of the eye.

The next stage in evolution was the deepening of the indent to form a shallow cup shape. This allowed for the discrimination of directional brightness but little else. Over time, this cup deepened and the opening diminished in size thus producing what is essentially a pinhole camera. This was the first eye capable of detecting directional brightness and distinguishing shapes and is still used today by the Nautilus.

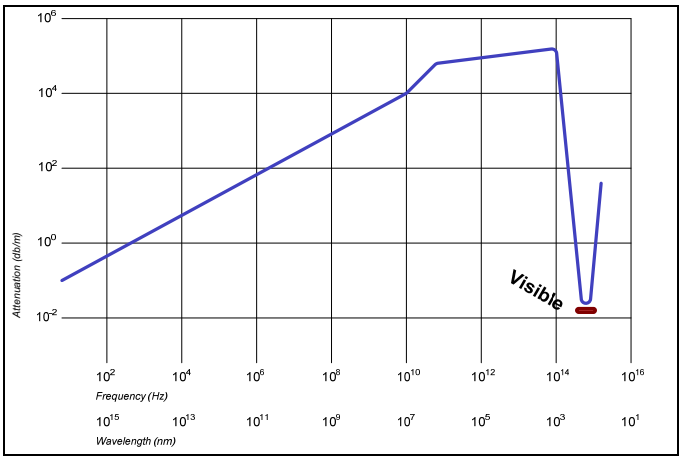


Figure 6.2: Attenuation (dB/m) of electromagnetic radiation in seawater plotted as a function of frequency and wavelength. The narrow band that corresponds to visible light is shown as a red band on the right.

The next stage was the development of a thin overgrowth of transparent cells over the aperture by a transparent crystalline protein (Fernald, 2001). This served to protect the eyespot but it also separated it from its early aquatic environment. That vision evolved in an aquatic environment is evidenced in Figure 6.2, which plots attenuation, in seawater, of electromagnetic radiation. Seawater rapidly attenuates all electromagnetic radiation, with the exception of the visible spectrum, which suggests that the early development of the eye must have occurred in a marine environment.

The eye cavity filled with a transparent humor, which preserved the refractive index of seawater (which it still does today). This permitted the development of the eye in non-aqueous environments. The transparent protective cells eventually split into two layers, with circulatory fluid in between that allowed wider viewing angles and greater imaging resolution. This was the first rudimentary lens and eventually the thickness of the transparent layer gradually increased naturally forming a biconvex shape of a modern lens. Independently, a transparent layer and a nontransparent layer split forward from the lens, these became respectively, the cornea and iris. Further separation of the forward layer formed the aqueous humor that increased the refractive power of the eye and eased circulatory problems (Fernald, 2001).

6.3 The Anthropoid Retina

Simplistically, the retina can be thought of as a sampling device because its most obvious function is to sample the visual field both spatially and chromatically. The retina, however, is a multi-layered tissue and the sampling occurs only in its innermost layer, the photoreceptor layer. The other layers of the retina belong to the next layer of the visual system, the signal processing system. They will be discussed in the next chapter. In this section, I consider the retina as a spatial sampling device, the chromatic sampling being dealt with in the next section.

6.3.1 General Morphology

The retina is the part of the eye that receives the light from the external three-dimensional world and converts it into chemical energy. This chemical energy activates nerves that transmit the light energy out of the retina and onto the higher functions of the visual processing system and the brain. In adult humans, the retina is approximately 0.5 mm thick and comprises about 72% of a sphere approximately 22mm in diameter. In adult humans, when measured from a central fixation, the monocular visual field is approximately 160 deg (width) x 175 deg (height); the binocular field is approximately 200 deg (width) x 135 deg (height) and the region of binocular overlap is 120 deg (width) x 135 deg (height) (Wandell, 1995). Given then the size of the retina and the visual field it covers, one degree of visual angle projects onto approximately 288 μm on the retina (Draso and Fowler, 1974).

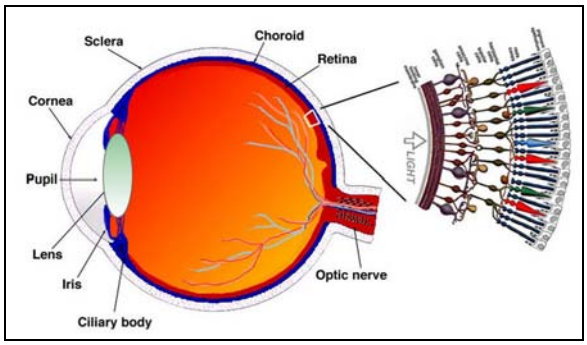


Figure 6.3: A schematic diagram of the components of the Anthropoid eye. The cutout illustrates the basic layered structure of the retina. Contrary to intuition, the photoreceptor layer lies at the back of the retina in close proximity to the pigment epithelium that nourishes it.

There are ten layers of cells in the retina that can be seen with a microscope (Figure 6.4). These are arranged into four primary layers; (1) the innermost layer is the pigment epithelium; (2) beneath this layer is the photodetector layer containing the rods and cones; (3) the next layer is a layer of nerve cells (neurons) called the bipolar cells. These cells receive the energy imparted by light to the photodetectors; (4) the final layer in the retina is a layer of neurons called ganglion cells; these cells connect to the bipolar cells and subsequently transmit the visual information out of the eye along their axons, which make up the optic nerve fibers.

The photosensitive cells are the rods and cones, the rods being much thinner than the cones but morphologically similar. The photosensitive pigment is located in the outer segment of the rods and cones, resting against the epithelium. The absorption of a photon

by the visual pigments is translated first into a biochemical message and then into an electrical message that stimulates the succeeding neurons of the retina. The raw biochemical energy is preprocessed by the bipolar and ganglion cells and subsequently transmitted to the brain via the spiking electrical discharge pattern of the ganglion cells.

Interestingly, the retinal layers are reversed from what one would expect. Intuitively, it appears that the rods and cones should face towards the outside of the eye, i.e. the lens, but this is not the case. The epithelial layer is at the very back of the eye and the ganglion layer is at the front. This is because the pigment bearing membranes of the photoreceptors have to be in contact with the eye's pigment epithelial layer, which provides a steady stream of the vital molecule retinal (vitamin A). Retinal binds with the photoreceptor's opsin protein and changes its conformation in response to photons. Once a retinal molecule is exposed to light, it undergoes a conformational change and is recycled back into the pigment epithelium. The retina is one of the most metabolically active of all tissues and this continual charging and discharging of the photoreceptors takes a considerable amount of an organism's energy budget.

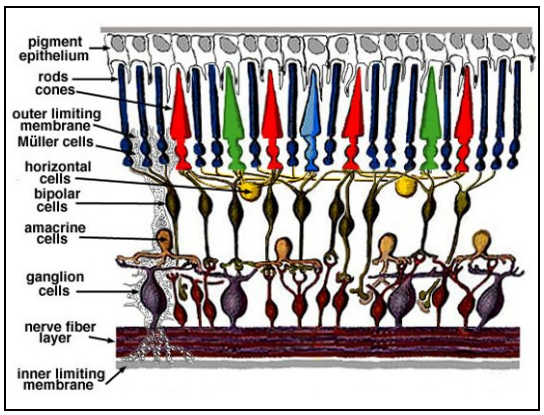


Figure 6.4: The ten layers of cells in the anthropoid retina. Light has to traverse the lower eight of these layers before impacting the photoreceptor cells.

What this means for vision is that for light to sensitize a photoreceptor it first has to travel through all of the non-photosensitive layers of the retina. The tissue behind the retina is usually very dark because its cells are full of melanin granules. The pigment granules absorb stray photons, preventing their reflection back into the photoreceptors, which would cause images to blur. They also protect the cells from too much exposure to light radiation.

In anthropoids, the retina contains between 120,000,000 and 150,000,000 photoreceptors. Of these, the vast majority are rods, the ratio of rods to cones being

approximately 20-1. In the human retina there are approximately 6,400,000 cones and between 110,000,000 and 150,000,000 rods (Osterburg 1935). One of the many surprising facts about the retina is that although there are a vast number of individual photoreceptors, there are only approximately 1.2 million axons or neural fibers in the optic nerve (Polyak, 1941, Quigly at al., 1982). This indicates that the raw visual signal recorded by the photoreceptors undergoes a substantial amount of preprocessing before being passed onto the brain. Obviously, the retina is as much a processing system as it is an acquisition system.

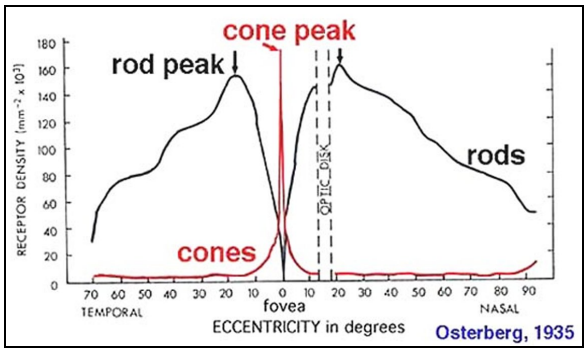


Figure 6.5: Density of rods and cones in the human retina as a function of eccentricity. When the gaze if fixed upon an object the lens and the fovea are in direct alignment.

Rods and cones are not equally spaced throughout the retina. The central area of the retina is dominated by cones whereas the peripheral area is dominated by rods. Cone density peaks in the area of the retina called the fovea centralis and rapidly falls off to an even density (~5,000 cones/mm^2) outside of it. Rods, on the other hand, are entirely absent in the fovea and reach a peak density (~160,000 rods/mm^2) in a ring approximately 5 mm or 18 degrees from the foveal pit. There are, of course, no rods or cones covering the optic nerve (the blind spot).

This spatial organization of the rods and cones is important for understanding how we use vision. Cones, which are responsible for our daylight vision, are less sensitive to light than are rods but they respond much faster. Rods, on the other hand, are responsible for our night vision and whereas they are sensitive enough to respond to an individual photon, the signals from them are thought to arrive up to 1/10th of a second slower (McLeod, 1972). This is one of the primary reasons that sporting events such as baseball become progressively more difficult as daylight fails.

6.3.2 The Fovea Centralis

Obviously, if a photon has to pass through all of the non-light sensitive layers before reaching the photodetectors then its chance of being accidentally absorbed increase greatly. To overcome this, primates have developed what is known as the fovea centralis, an area of the retina where the inner layers are almost absent. Primates are the only mammals to develop a fovea although one may occur in the retina of species from all other vertebrate classes (with the exception of amphibians). The structure of the primate fovea, however, is different that of all other non-primate species indicating that our fovea evolved very late in our evolutionary history.

Understanding the fovea (meaning pit) is critical to any study of vision because even though it is small, it provides catarrhine primates with most of their visual information. The primate fovea is only about 1500 μm in diameter (Polyak, 1941, Ahnelt and Kolb, unpublished data) and covers roughly six degrees of visual arc. It is seen as a depression close to the optical axis of the eye that, in size, is slightly larger than a pinhead but substantially smaller than a bb pellet. However, when the gaze is fixed on an object the lens and the fovea are in direct alignment and so, although it only occupies approximately 1% of the total area of the retina, it provides us with most of our critical visual information.

To put this into perspective, the fovea covers about 6° of visual arc, the moon occupies about 0.5° and a typical thumbnail (~0.75” x 0.75”) at arms length covers about 1.5°. Given that the typical viewer sits at about arms length from a computer, the visual arc covered by the fovea projected onto the computer monitor is roughly a circle, 3” in diameter.

Through this narrow portal comes most of our visual information because the cones in the fovea are packed more densely than anywhere else on the retina. In adult humans, there are about 250,000 cones in the fovea packed at density ranges from 100,000 to 300,000 cones/mm^2 (Osterburg, 1935; Curcio et al., 1987). Outside of the fovea typical cone density is around 5,000 cones per mm^2, which indicates that even though most

cones lie outside the fovea our vision has evolved to provide very close scrutiny of very small objects.

Table 6-1: Peak retinal cone densities in various mammalian taxa. From Chapter 20, Table 3 in Anthropoid Origins: New Visions, Kirk and Kay ed., Kluwer Academic/Plenum Publishers, New York

Scientific Name	Common Name	Activity Pattern	Maximum cone density (cones / mm^2)	Source
Alouatta caraya	Black howler monkey	Diurnal[1]	359,000-430,000	Franco et al., 2000
Homo sapiens	Human	Diurnal	100,000-324,000	Curico et al., 1990
Macaca mulatta	Rhesus macaque	Diurnal	134,000-160,000	Perry and Cowey, 1985
Saimiri sciurus	Squirrel monkey	Diurnal	90,000-140,000	Franco et al., 2000
Meriones unguiculatus	Mongolian gerbil	Diurnal	45,000-50,000	Govardovskii et al., 1992
Spermophilus beecheyi	California ground squirrel	Diurnal	41,800-50,000	Long and Fisher, 1983
Felis catus	Domestic cat	Cathemeral[2]	~27,000	Müller and Peichl, 1989
Oryctolagus cuniculus	Old World rabbit	Cathemeral	~14,000	Juliusson et al., 1994
Mus musculis	House mouse	Cathemeral	11,000-18000	Szél et al., 1992
Didelphis marsupialis aurita	South American opossum	Nocturnal[3]	2,400-2,600	Ahnelt et al., 1995

[1] Diurnal animals are primarily active during the day.

[2] Cathemeral animals are active throughout the twenty-four hour period on a sporadic basis.

[3] Nocturnal animals are primarily active at night.

The fovea is essentially an area of super vision. In contrast, non-primate mammals, all of whom lack a fovea centralis, have maximum cone densities that range from a low of 2,400 cones/mm^2 for the South American opossum to a high of around 50,000 cones/mm^2 for the Mongolian gerbil (of all things). The typical house cat has a density of around 27,000 cones/mm^2 with the dog being somewhat lower (although they are generally loath to admit it).

Anthropoid visual acuity is further enhanced by the development of the foveal pit. The foveal pit is a small depression, totally devoid of rods, approximately 250 microns in diameter located in the center of the fovea (Yamada, 1969). Within the foveal pit, the cone photoreceptors are concentrated as closely as possible and arranged in a hexagonal mosaic (Figure 6.6).

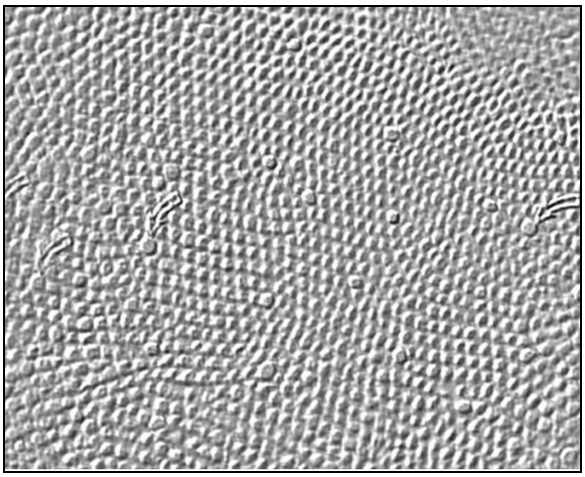


Figure 6.6: Tangential section through the human fovea. The smaller cones are either M-cones or L-cones (which are morphologically identical), the larger cones are S-cones.

Inside the foveal pit, the overlying layers of the retina are concentrically displaced leaving only the cone cells and some of their cell bodies. Outside of the foveal pit, the complete layering of the retina appears gradually along what is known as the foveal slope until the edge of the fovea is reached. Here, the displaced neurons of the cones of the central pit are located, forming a rim around the fovea. This area, known as the foveal rim or parafovea, is the thickest part of the entire retina. The foveal pit contains about 17,500 cones and because of the lack of any overlying interference, it provides us with our clearest and sharpest vision. It covers approximately one degree of arc, which on a computer monitor at arms length it covers a circle roughly 0.5” in diameter, or about the width of a four-letter word in 12-point font.

As with all things in nature, the high visual acuity provided by the fovea comes at a price. Because of the displacement of the overlying retinal layers, the fovea has poor blood supply and consequently must obtain its oxygen from the choroid that is across the

retinal pigment layer. Under bright light, this blood supply cannot satisfy the metabolic needs of the fovea and so the cones in the fovea, and especially in the foveal pit, exist quite often in a state of hypoxia (oxygen starvation).

6.4 Visual Acuity in Anthropoids

"*Visual acuity or resolving power refers to the ability to distinguish between closely-spaced visual stimuli. Where the optical properties of the eye do not limit acuity (e.g. through refractive errors or large pupillory diameter), visual acuity refers to the ability of the brain to distinguish fine spatial details of the image projected by the cornea and lens onto the retina*" (Kirk and Kay, 2004).

Enhanced spatial resolution is one of the main driving forces behind the evolution of anthropoid vision and humans are no exception. Early behavioral studies expressed visual acuity in terms of the retinal angular separation (in degrees, minutes and seconds) of the closest two stimuli (usually parallel lines) that a test subject could perceive as separate entities. Modern studies use a grating pattern, which is usually just an oscillating pattern of light and dark bands on a computer. The patterns may be simple black and white bands or they may be gradational according to a specified mathematical formula. Grating acuity is measured in cycles per degree (c/deg) which is the maximum number of full cycles that can be distinguished from a uniform gray background of equal total luminosity.

Behavioral test on diurnal anthropoids indicate that our visual acuity consistently ranges between 40-80 c/deg (Table 6-2). This is significantly better than behaviorally determined acuity of all other mammals, which ranges from 0.05 c/deg for the little brown bat (Suthers, 1996) to 23 c/deg for the horse (Timney and Keil, 1992). The visual acuity of diurnal anthropoids even exceeds those of many diurnal birds such as pigeons (16 c/deg) and blue jays (15-27 c/deg) and ostriches (17-23 c/deg) (Boire et al., 2001; Fite and Rosenfield-Wessels, 1975). The only known vertebrates that surpass the visual acuity of diurnal anthropoids are the large diurnal raptors such as the wedge-tailed eagle (140 c/deg) which has much larger eyes than diurnal anthropoids and a much higher cone density (Fite and Rosenfield-Wessels, 1975).

Table 6-2: Behavioral tests of visual acuity in various mammalian taxa. From Chapter 20, Table 1 in Anthropoid Origins: New Visions, Kirk and Kay ed., Kluwer Academic/Plenum Publishers, New York

Scientific Name	Common Name	Activity Pattern	Visual Acuity (c/deg)	Source
Homo sapiens	Human	Diurnal	50.0-77.0	Cavonius and Robbins, 1973
Macaca mulatta	Rhesus macaque	Diurnal	46.2-53.6	Cowey and Ellis, 1967
Saimiri sciurus	Squirrel monkey	Diurnal	40.5	Cowey and Ellis, 1967
Equas caballus	Domestic horse	Cathemeral	10.9-23.3	Timney and Keil, 1992
Felis catus	Domestic cat	Cathemeral	5.0-8.9	Bisti and Maffei, 1974
Zalophus californicus	California sea lion	Cathemeral	5.4-5.7	Schusterman and Ballait, 1970
Tursiops truncatus	Bottlenose dolphin	Cathemeral	1.6-2.5 (air) 2.5-3.8 (water)	Herman et al., 1975
Oryctolagus cuniculus	Old World rabbit	Cathemeral	1.5-3.0	Van Hof, 1967
Mus musculis	House mouse	Cathemeral	0.5	Sinex et al., 1979
Myotis lucifugas	Little brown bat	Nocturnal	0.05-0.1	Suthers, 1996

That human's possess visual acuity that exceeds all non-primate mammals and is comparable to birds of prey may come as a surprise to many readers. This visual acuity is due, in large part, to the evolution of the fovea centralis previously mentioned. The fovea is only part of the story, however, because the diurnal anthropoid eye contains other significant adaptations that affect acuity. These include; (1) the size of the eye; (2) filters and cone distributions that minimize chromatic aberration; and (3) the absence of a tapetum lucidum.

6.4.1 Size of the Anthropoid eye

All primates have larger eyes in comparison to other mammals and anthropoids are no exception (Hughes, 1977; Ross, 2000; Walls, 1942). Larger eyes can hold a larger retina and thus a larger retinal array. This increases visual acuity by spreading the image over a greater number of photodetectors. This is one of the few areas where a direct analogy between physiology and technology is possible because a larger eye size directly mimics using a higher resolution computer monitor. The more pixels in the monitor, the more highly resolved the projected image becomes. In a direct analogy, the anthropoid eye is larger and thus has more photodetectors and thus the resultant image is more highly resolved.

While this increased size enhances resolution it also reduces brightness. Ross (2000) reports that under comparable lighting conditions, the brightness of the retinal image is five times less in diurnal anthropoids than in other mammals. Further, the small size of the anthropoid cornea limits the light-gathering capabilities of the eye by restricting the maximum size of the pupil. All this suggests that the eye size of diurnal anthropoids evolved to increase visual acuity in a bright light setting where the brightness of the surrounds could compensate for the proportional dimness of the retinal image.

6.4.2 Chromatic Aberration

The degree to which light is refracted by the eye's cornea and lens is an inverse function of its wavelength. Short-wavelength light in the blue and ultraviolet part of the spectrum is refracted more than red and green light and thus comes into sharp focus in front of (rather than on) the retina (Walls 1942). Blue and violet light thus appears to be slightly blurred or out of focus relative to red and yellow images. This blurring of the vision is problematic for species such as the early anthropoids that rely on high visual acuity for predation. As I show in Chapter 11, it is also a problem for geophysicists who rely on a red-blue color spectrum to communicate seismic information. The problem of "chromatic aberration" is minimized in diurnal anthropoids by the presence of two types of optical filters.

The first filter is found in the anthropoid lens, which contains a series of tryptophan-derived compounds that screen out most light below approximately 400 nm (Bova et al., 1999). For example, anthropoid lenses filter out all but 0.0001% of incident 360nm UV light (Cooper and Robson, 1969). In contrast, the lens of all other mammals with the exception of diurnal squirrels are virtually transparent to light above 320nm (Gorgels and van Norren 1992) and in some case they even contain photopigments that absorb maximally in the UV range (Calderon and Jacobs, 1999). It appears that anthropoids have sacrificed the ability to perceive a large part of the spectrum visible to other mammals for the sake of improving visual acuity.

The second filter is found in the retina itself, which plays an important role in filtering out blue and violet (visible) light. The central retinal area that encompasses the fovea and its surrounds support the highest visual acuity and has the best ability to discriminate colors. This area contains high concentration of the carotenoids lutein and xeaxanthin (Bone et al., 1985). These compounds absorb light maximally at about 460 nm (blue light) but transmit nearly all light above 550 nm (Nussbaum et al., 1981). On visual inspection, this area shows itself as a yellow spot on the retina and is known as the macula lutea. The macula lutea is also present in extant Tarsiers (Castenholz, 1984) which suggest that it was present in the last common ancestor of living haplorrhines (both tarsiers and anthropoids). It has been suggested that the macula lutea has a dual purpose, the first being the reduction in chromatic aberration and the second to protect the foveal area from oxidative damage caused by short-wavelength light (Thomson et al, 1992).

6.4.3 Cone Distribution in the Fovea

The various cone types are not evenly represented in the fovea. The anthropoid retina contains three types of cones, S-cones, M-cones and L-cones (see 6.5). The M-cones and L-cones are actually slightly mutated forms of the same gene and so they are almost impossible to distinguish. Because of this, it is difficult to determine the ratio of M and L cones in the human retina. Recent studies, however, (Roorda and Williams, 1999, Hofer et al. 2005), have shown that there is considerable variation amongst individuals with some people having an almost 1:1 ratio of M and L cones and others up to a 16:1 ratio.

The S-cones, however, are a different gene and are morphologically different from the other two (see Figure 6.6). This makes their identification in the fovea possible. Ahnelt (1987) showed that the large diameter S-cones distorts the regular mosaic of M and L cones in many places and break them into irregular subunits. He showed that S-cones have their lowest density in the foveal pit, comprising only 3-5% of the total cone population. They reach a maximum density of 15% on the foveal slope and then form an even 8% of the total population elsewhere in the retina.

The absence of S-cones in the fovea is best explained by recalling Helmholtz's result on chromatic aberration in the human eye (see 6.1.2 Vision and Evolution). Given the size of the human eye, chromatic aberration makes it impossible to sharply focus both short and long wavelength light. However, by dramatically reducing, in the fovea, the number of cones maximally sensitive to short wavelength light this loss of sharpness is avoided. Therefore, even if a short wavelength photon passes through the cornea and subsequent filters, there is probably nothing there to detect it when it hits the photoreceptor layer.

6.4.4 The Tapeta Lucida

The tapetum lucida (literally "bright carpet") is a layer of reflective tissue that lies behind the photoreceptive cells of the retina. In a tapeta possessing eye, light that passes through the retina without being absorbed is reflected back towards it by the tapeta. This greatly increases the chances that a given photon will be absorbed by a photodetector and produces the glowing eyes that many creatures seem to posses at night. The tapeta greatly increases a retina's sensitivity to low light conditions but by scattering back the incoming light it also necessarily reduces visual acuity (Buttery et al., 1990).

All haplorrhine primates lack this tapetum lucidum although by contrast tapeta are found in all other mammalian orders and appear to have arisen in them convergently (Nicol, 1981). It was originally thought that diurnality was the root cause for the absence of tapeta in haplorrhines (Martin, 1990). However, there does not appear to be any direct cause and effect relationship between nocturnality/diurnality and the presence or absence

of the tapeta. Many nocturnal animals lack the tapeta and many diurnal ones posses it. Around the house, both dogs and cats posses a tapeta but they are both active in low light and bright light conditions. In tapeta possessing species that are active in daylight it now appears that the tapetum places an upper limit on visual acuity because the resolving power of the photoreceptor mosaic cannot be increased beyond the threshold set by the tapetal scattering (Pirie, 1966). The absence of a tapetum lucidum in extant anthropoids therefore appears to be a further adaptation directed towards increasing visual acuity.

6.4.5 Spatial Sampling

In manufactured optics, even the best grade lens is unable to focus a point source to a point. The same is true of a biological lens. Physiological lenses are typically high quality and aside from chromatic aberration, they can be perfect – limited only by diffraction (Land and Nilsson 2002). Because of diffraction, a point is focused into a blurred spot called an Airy disc whose intensity function is known as the point spread function (Land and Nilsson, 2002). The half width of the disc is given by: $w = \lambda/D$ rad where λ is the wavelength of light and D the diameter of the aperture (in humans, the aperture is the pupil).

The Airy disc is analogous to the seismic Fresnel zone as both define the upper limit of spatial resolution. Using the Widess criterion for seismic resolution of $\lambda/8$, seismic resolution in the case of unmigrated data is given by $F_d = \sqrt{\lambda Z}$, F_d being the Fresnel diameter, λ the wavelength and Z the depth of burial. For perfectly migrated data this reduces to $F_d = \lambda/4$, or for the half width, $\lambda/8$ (Lines and Newrick, 2004).

The formulas for the Airy disc and the Fresnel zone are very similar and both place constraints on the economics of acquisition. The Fresnel zone places a theoretical limit on seismic resolution. Acquiring a 3D seismic survey is a financially expensive proposition and so designing a survey with a subsurface sampling less than the target Fresnel zone is discouraged by economics. The same is true of physiological acquisition. Phototransduction, the conversion of light energy to electrical signals is metabolically expensive and is a significant part of the energy budget of any animal. The retina is

amongst the most metabolically active of all tissues and so it is imperative that it be designed to optimize the spatial sampling. Sampling below the Airy disc limit increases the physiological cost of acquisition but adds no new information because intensities registered by neighboring receptors become indistinguishable.

Conventional principals of optical design suggest that the smallest useful receptor spacing is approximately the half width (radius) of the point spread function (Figure 4.2, Synder and Miller, 1977). In optics, this is known as matched or Nyquist sampling because it allows for the recovery of the finest details transmitted by the optics. Sampling at a coarser spacing will alias the spatial signal in much the same way that under sampling a seismic signal will alias the frequency information.

With this in mind, what can we say about human vision? For a human with a 2mm pupil and light with a wavelength of 550nm (yellow light) the Airy disc has a half width of .000275 rads, which is approximately $.016^0$ or roughly 1' arc. For a human eye with diameter of 22mm, this corresponds to a disc of radius 0.006mm and area of .0001185 mm^2. In the central retinal area, cone densities exceed 100,000 cones/mm^2, which corresponds to a minimum cone spacing of approximately .0032mm. Given that the fovea is dominated by M-cones & L-cones and assuming that the distribution of the two cone types is roughly equal then the cone spacing for each cone type is roughly .006mm or approximately the half width of the previously mentioned Airy disc for yellow light.

This is in marked contrast to the area outside of the fovea. In the rest of the retina, cone density drops to around 5,000 cones/mm^2. This corresponds to a receptor spacing of approximately 0.015 mm or roughly five times the foveal spacing.

This analysis indicates that Anthropoid vision has evolved to provide the most finely resolved detail possible in a very small area of vision. Within the fovea, which accounts for only about 6^0 of the visual field, anthropoid visual resolution matches the theoretical optical resolution. Matched sampling itself is very rare in the animal kingdom. Most vertebrate eyes have sampling densities that are a factor of two lower than that needed for

matched sampling (Synder et al., 1977). The fovea of anthropoids and birds of prey are the only known examples of true nyquist sampling (Synder et al, 1986).

6.4.6 Visual Acuity and its Relevance to Seismic

All of the preceding sections on visual acuity have been to convince the reader of one inescapable fact, specifically: the smaller an object is, the better we see it. This is another of the paradoxes of the visual system. Intuitively one would think that we would perceive larger objects better but this is not the case. Our high visual acuity evolved when we (a) weighed less than 500 gm's and (b) were migrating from being nocturnal insectivores to being diurnal insectivores (see Chapter 5.6.2 – Diurnal or Nocturnal). Our primary evolutionary pressure then was to locate and identify insects, which are usually very small and often very fast.

The primate visual system is tuned to "seek and destroy" very small objects, which unfortunately make it less than ideal for viewing seismic data. Take, for example, Figure 6.7, which is a variable density display of a seismic section over a Devonian reef. The image is three inches wide, which means that when you look at it at arms lengths you are viewing it primarily with the fovea. This is the absolute limit of our high-resolution vision; this is about the size of what we can see with it.

This is not though the size of our "perceptive area" and you will notice that when you look at Figure 6.7 you do not perceive everything in the display. This is because within the fovea there is an even smaller area of enhanced vision – the foveal pit, which covers an area roughly the size of the reef in Figure 6.7. Moreover, if you look directly at the reef you will notice that your awareness is tied very much to this small, one-degree cone. When we look at this section, we are never aware of the whole, we are only truly aware of a very small part of it.

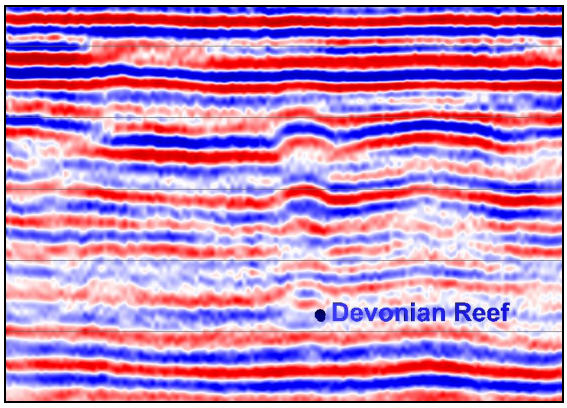


Figure 6.7: Variable density display of seismic over a small Devonian reef. The seismic events are primarily horizontal and resemble lines of text. When we view a horizon we use the same visual processes that we do when reading, moving the fovea centralis along the event in the same we do across a line of text.

Although we are not aware of it, the process of interpreting a seismic section is analogous to reading. When we read a document, we primarily use the cones of the foveal pit, which is roughly the size of a four-letter word on this page. Reading is the process of moving the eye so that the fovea centralis views adjacent words. To prove this to yourself look directly at a word on this page and then try to read another word without letting your visual focus move to it. This is very difficult to do, in most case when you think you have succeeded all that has happened is that the eye has very quickly flicked to the word in question, and then flicked back. It is all but impossible to read a word that you are not directly looking at.

The same is true of viewing seismic data although the situation is more complex. If we concentrate just on the horizons in Figure 6.7 one obvious detail is that they are primarily horizontal. Not all seismic events are horizontal of course but one feature that they do have in common is that they generally have a broad linear extent. In this, they resemble lines of text and if we were to scan along a horizon for amplitude changes or other features the visual process used would be identical to reading; that is, we would be moving the area viewed by the foveal pit along the event.

This tight central focus of our visual system makes it difficult for us to take in the whole of a horizon at once. It also makes it difficult for use to take in any other feature that has a large aerial extent. As an interesting aside, if primates are not specifically geared to viewing seismic, other species are. All eyes are not the same because they are all designed to match their particular ecological niche and retinal topography, in particular, can tell us a lot about the habitat and lifestyle (Hughes, 1977) of this niche. As I have shown, haplorrhine primates have a fovea, a central area of the retina with a very

high density of cones. Most other animals also possess areas of high visual acuity or "acute zones" but very few are circular like the fovea. Animals that inhabit open country often have a thin streak of high-resolution vision in the retina that views the horizon. This is presumably because this is where their visual activity occurs. Other mammals such as rabbits and squirrels have a long horizontal strip of specialized cells called a visual streak, which can detect the fast movement of predators (Kolb, 2003). It is not strictly relevant but interesting to speculate just what a species with a horizontal high-resolution streak would make of a seismic section.

The primate visual system is essentially a target acquisition system. It is geared towards identifying objects of interest in the periphery of our vision and then rapidly moving the eye so that the object is directly in line with the foveal pit. In addition, although this process can be under conscious control it can also, under certain circumstances, be automatic. It would benefit us if the foveal pit, to which our awareness is tied, covered a larger area of the visual field. The problem, however, is that the visual system is our most metabolically active system. As is, even with such a small area of high visual acuity, the visual system still takes up a disproportionate amount of our neocortex. Expanding the size of this "area of interest" would be prohibitive in terms of both energy expenditure and the size and complexity of the neocortex required to process the extra information.

The fovea centralis is then a tradeoff and, if it only covers a small area of the visual field, we have compensated by developing the ability to rapidly move it around to center on target locations. When we use it to view seismic, we are in essence looking for the proverbial needle in a haystack but using an electron microscope. The fovea centralis most likely covers less of a seismic section than we need and it probably provides better visual resolution than we need but it is the system that we have and we cannot change it. It is not ideal for viewing seismic but by understanding its properties, strengths and limitations we can, perhaps, learn how to direct it correctly.

6.5 Anthropoid Photoreceptors

In the previous section, I discussed how the anthropoid eye serves as a spatial sampling device. In this section, I will discuss how it also serves as a chromatic sampling device. I do not intend to go into as much detail here as I did in the previous section. This is because much of the discussion on why anthropoids, and in particular catarrhine anthropoids, evolved their peculiar set of photoreceptors is best left to a later chapter where its relevance to the visualization of seismic data is more apparent.

The vertebrate lineage emerged approximately 540 million years ago. Surprisingly, evidence now suggests that these early life forms already possessed four different visual pigments (Collin et al., 2003). Over time, this has been reduced to two in all Eutherian mammals with the exception of all catarrhine and some platyrrhine primates who posses three. However, we do not posses three of the original four; we possess only two with one of them being duplicated. Even so, the ability of catarrhine primates to separate hues is superior to that of all other mammals. Nevertheless, before we feel too proud, we may see colors better than the family dog but we do not see them as well as most birds and reptiles and probably we do not see them as well as our early Devonian ancestors.

6.5.1 Structure of visual pigments

All visual pigments are based upon the protein opsin that is attached to a chromophore[4]. In mammals, this chromophore is invariably 11-*cis*-retinal which is derived from Vitamin A which itself is produced in the retina (and not absorbed from diet). Since the chromophore part of the visual pigment is constant, the peak spectral absorption of the pigment (λ_{max}) is determined by the amino acid sequence of the opsin protein and not the chromophore itself.

---

[4] A chromophore is the part of a molecule that absorbs light and is responsible for its color.

In vertebrates, sensitivity to scoptic[5] conditions is generally achieved by a single class of photoreceptors in the retina called rods. Rod photoreceptors, so named because of the appearance, have a λ_{max} at approximately 500nm and are extremely sensitive to light, being activated by individual photons. This sensitivity, which is so important to scoptic vision, also means that rods become saturated very quickly and are therefore of no use under photopic conditions. Consequently, they are of no use in color vision and will not be considered further.

Photopic[6] vision in vertebrates is achieved by up to four classes of cones, each of which contains opsin with a different spectral sensitivity. Cones, like rods, are so named because of the appearance in cross-section. The four cone classes are distinguished based on the amino acid sequence of their respective opsin proteins. Their spectral sensitivities roughly correspond as follows: long wave sensitive (LWS) with a λ_{max} of 500-570nm, a middle wave sensitive class (MWS) with a λ_{max} of 480-520nm, and two short wave sensitive classes, SWS2 with a λ_{max} of 415-470 nm and SWS1 with a λ_{max} of 355-435nm.

In Eutherian mammals, this complement has been reduced to two classes, LWS and SWS1, probably because of the nocturnal lifestyle that dominated much of their evolutionary history. It has only been partially reversed in anthropoid primates.

6.5.2 Note on Absorption Spectra

Figure 6.8 shows the well-known graph of the absorption spectra of the three catarrhine visual pigments. This graph is not what it first appears to be and it is important to understand, in the context of future chapters, to what these absorption spectra refer.

Although the three chromatic cones, the S-cones, M-cones & L-cones, shown in Figure 6.8 are typically thought of as blue, green and red cones, they do not produce a

---

[5] Low light conditions, i.e. night time.

[6] Daylight conditions.

color-coded signal. Figure 6.8 implies that they do but the absorption spectra refer to the probability that a photon of a given wavelength will be absorbed by the photopigment. Once it is absorbed, however, its only effect on the cone is to either increase or decrease the hyperpolarization of the cone's membrane potential. For any given cone type, long wavelengths absorbed by the cone will have the same effect as short wavelengths if the amount of quanta absorbed by the cone is identical. This is important because in the end all the brain can infer from a single cone is whether the light absorbed by the cone has increased or decreased. It can never determine the spectra of the light absorbed.

6.5.3 Catarrhine visual pigments

Humans are catarrhine primates and we, along with all other catarrhine primates, have three cone receptors, (1) an SWS1 receptor and (2) two LWS receptors (the M-cones & L-cones). Over time, we have lost the genes for both the SWS2 receptor and the MWS receptor, which are present in the retina of non-mammalian vertebrates.

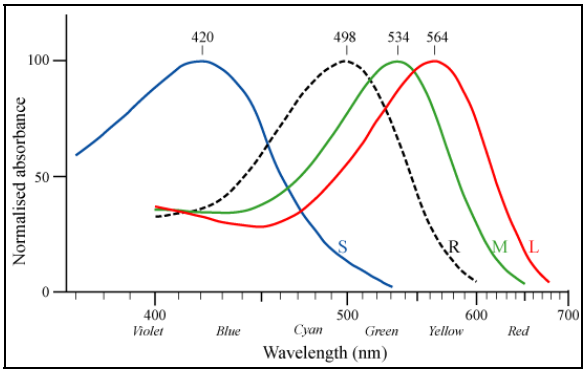


Figure 6.8: Absorption spectra of the three catarrhine cone pigments plus the rod spectrum (broken line). The L-cone and the M-cone are mutated forms of the same primordial LWS receptor and provide an enhanced ability to discriminate long wavelength hues.

The SWS1 receptor is present in all vertebrates but it has a varied spectral sensitivity ranging from around 355 nm in the ultraviolet region of the spectrum (UVS) to around 435 nm in the violet region of the spectrum (VS). The UVS pigment is present in lamprey, a member of the jawless vertebrates that separated from the main vertebrate lineage about 540mya (Collin et al., 2003). This is evidence that the ancestral SWS1 pigment was UVS (Hunt et al, 2001) and that it has shifted in many species to VS over time.

Primates in general posses a VS version of the SWS1 gene (on chromosome 7 in humans) with a λ_{max} of 415-420nm. This long-wave shift in sensitivity and the fact that

primate lenses strongly absorb UV light have all but eliminated our ability to see in the ultra-violet. This is thought to have occurred for two reasons; (1) To reduce chromatic aberration and thus improve visual acuity; (2) To avoid damage to the cornea, lens and retina, caused by the free radicals produced by UV light. This latter reason is thought to be especially important in long-lived species such as primates.

The LWS gene in mammals is found on the X[7] chromosome. Catarrhine primates' posses two copies of this gene (M & L) that have arisen from mutation of the ancestral gene (Nathans et al., 1986). These two genes are found in a head to tail arrangement on the X chromosome and are very similar to each other with a 97% nucleotide similarity. The M gene has a λ_{max} of around 532 nm, whereas the L gene has a λ_{max} of around 563 nm. This duplication is common to all catarrhine primates, which indicate that it must have evolved at the base of the catarrhine lineage approximately 30 million years ago. Some platyrrhine primates also posses this duplication but it is not common and thus it is most likely that the few platyrrhines who evolved it did it separately and after their divergence from the Old World lineage.

The reason why this gene duplication and mutation occurred in catarrhine primates is an interesting study in and of itself. I discussed in the previous section, how the primate visual system is essentially a target acquisition system that originally developed to locate and identify insects. The subsequent evolution of this original dichromatic vision into a trichromatic visual system occurred much later in our development as we were converting from being insectivorous to being frugivorous and folivorous.

Trichromacy provides us with a far greater ability to discriminate colors than all other mammals but only in a small area of the spectrum. In this, the development of trichromacy mirrors the development of the fovea. The fovea developed to provide us with high spatial acuity but only in a small part of the visual field. Trichromacy

---

[7] In humans, the X chromosome is also called the sex-linked chromosome.

developed to provide us with an enhanced ability to discriminate hues but only in a narrow part of the visible spectrum.

The reason why catarrhine primates evolved trichromacy is particularly relevant to this thesis because it gives us a clue as to what color means to us. In the context of seismic visualization, we previously used color as the primary means of communicating complex information. It will turn out that trichromacy is not designed to do this; it has a very different purpose. Because of its importance to visualization I will leave the discussion of why it evolved and what it means until Chapter 11, where its significance can be better illustrated.

CHAPTER SEVEN: THE VISUAL SIGNAL PROCESSING SYSTEM

The tale is as old as the Eden Tree – and new as the new-cut tooth –
For each man knows ere his lip-thatch grows he is master of Art and Truth;
And each man hears as the twilight nears, to the beat of his dying heart,
The Devil drum on the darkened pane: "You did it, but was it Art?"
"The Conundrum of the Workshops"

7.1 Introduction

This chapter is one of three that consider the physiological processes of vision. In my opinion, this one is by far the most important and the most challenging of the three. This is because it will require the user to change, at a very fundamental level; their understanding of what it is that we actually see. In keeping with my geophysical analogy, this chapter covers the signal processing stages of the visual system. However, the user should not take from this, any suggestion that the signal processing in the two are in any way similar. In geophysics, the vast majority of the signal processing occurs after we have the seismic signal, i.e. they occur after we have recorded the signals from the geophones. By contrast, the vast majority of processing in the visual system is dedicated to producing the visual signals. Once we have them, we interpret them almost as is.

Going back into the past for a moment, I spent the better part of my undergraduate fourth year (1974-1975) working part time for Dr G.H. Renninger in the physics department at the University of Guelph. Computing, at the time, was still in its infancy as was our understanding of the visual system and Dr. Renninger had me work on programming a model of the interactions between cones in the first layer of the retina. This turned into an experiment into; (a) figuring out how to draw Bessel functions using only a line printer; and (b) figuring out how many times a computer operator would let the same error message print out before cancelling the job. For the record, I eventually succeeded with the former and the latter maxed out at 130,000 lines (at 50 lines per page) give or take a few.

At the time, I held a very primitive view of the visual system. In my simple understanding, the retina recorded the visual image and passed it onto the brain where it was presumably interpreted. How it did that though was a mystery. I knew that we had

three types of cones, blue, green and red, and I presumed that their color signals were sent to the brain, which formed colors in some sort of RGB combination. That was the limit of my knowledge and considering that my focus very quickly afterwards switched from biophysics to geophysics, the reader should not be surprised to learn that when I started this thesis my opinion had not changed.

Fundamental to my simplistic understanding of vision was an assumption that the brain interpreted what I was conscious of seeing. In this, I mean that I assumed that the retina essentially took a picture of what I was seeing and passed the picture onto the brain. The analogy is that the retina would function as a movie camera with the brain being the screen that the movie was played on. This analogy is very nice and very simple but it does not work at any level at all.

The analogy starts to fail immediately because as I showed in Chapter 6, the photoreceptors do not emit a color-coded signal. Although we use the terms blue-cone, green-cone and red-cone, they are no such thing. The raw signals leaving each cone merely indicate if the amount of light being received by the cone has increased or decreased; it says nothing about the wavelength of the light being received. This raw information bears no relationship to what we are conscious of seeing. Our conscious vision is constructed out of multiple streams of information that are extracted from this raw chemical energy and all of these streams, in some way, depend upon contrast.

This is the most important point of this chapter, that there never exists, anywhere in the brain, a single image that bears any relationship to what we are conscious of seeing. It is also a critical point for the science of visualization. Each image that we produce is ultimately segregated into multiple, parallel streams of information, and each stream is interpreted separately and for different purposes. Knowing what those streams are is vital to understanding how we communicate visual information.

The visual signal processing system produces streams of information for detecting achromatic contrast, chromatic contrast(s), movement, position and orientation. There are also specific streams for positioning the eye and for rapidly moving it. Obviously, trying

to understand and describe all of these various streams is beyond the scope of this thesis. Rather than trying to describe them all, I will concentrate primarily on only those streams that I consider relevant to visualization. Those are the streams of opponent information that produce our perceptions of form and color. Even in this, I will restrict the bulk of the discussion to the "private line" streams of information that directly connect each cone in the fovea centralis to the brain. The processing of the raw visual signal into these "private line" streams primarily occurs in three locations; (1) the bipolar cell layer of the retina; (2) the ganglion cell layer of the retina; and (3) the lateral geniculate body. Each of these will be dealt with in turn.

7.2 The Signal Processing Layers in the Retina

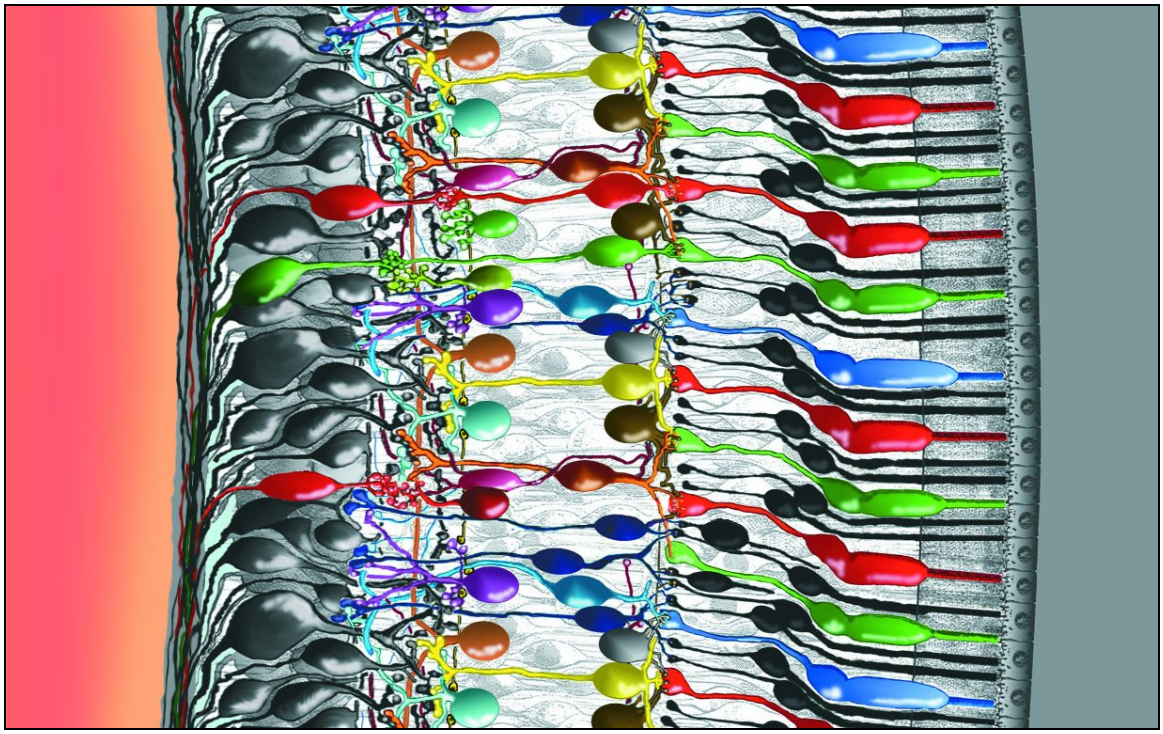


Figure 7.1: "Intricately wired neurons in the retina allow a good deal of image assembly to take place in the eye itself. In this rendering, light enters the eye from the left. The photons travel through the vitreous fluid of the eyeball and penetrate the entire retina, which is about half a millimeter thick, before reaching the photoreceptors—the cones and rods that respond to light (the colored and black cells attached to the epithelium at right). Signals then pass from the photoreceptors through a series of neural connections toward the surface of the retina, where the ganglion-cell nerve-fiber layer relays the processed information to the optic nerve and into the brain." (Drawing and caption by Kolb 2003, used with permission)

The anthropoid retina contains approximately 130,000,000 photoreceptors but the optic nerve contains only about 1,000,000 nerve fibers. Clearly, there is a significant amount of processing of the visual signal occurring in the retina itself.

The basic structure of the retina as shown in Figure 7.1 resembles a three-layer cake. In Figure 7.1, the light enters the eye from the left and has to traverse these layers before encountering the photoreceptors at the back of the eye. The nerve signals eventually leave the eye via the optic nerve, which is connected, to the ganglion cells located on the outer surface of the retina. Much of the processing and organization of the visual signals occurs here in the retina itself.

The retina contains three layers of nerve cells; (1) the photoreceptor layer containing the rods and cones; (2) the inner nuclear layer containing horizontal cells, bipolar cells and amacrine cells; and (3) the outer layer of the retina, which contains the ganglion cells. Between the three layers are two layers that contain the synapses linking the various nerve cells. The first layer, the outer plexiform layer, exists between the photoreceptor and inner nuclear layers. In this layer, the photoreceptors are linked with the dendrites[1] of the horizontal and bipolar cells. The second layer, the inner plexiform layer, exists between the inner

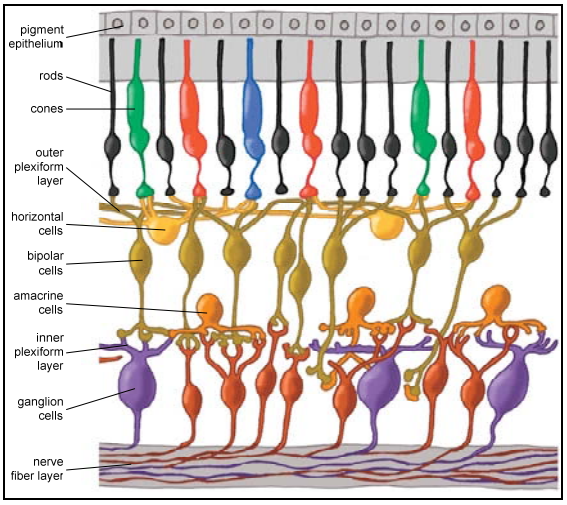


Figure 7.2: "Cells in the retina are arrayed in discrete layers. The photoreceptors are at the top of this rendering, close to the pigment epithelium. The bodies of horizontal cells and bipolar cells compose the inner nuclear layer. Amacrine cells lie close to ganglion cells near the surface of the retina. Axon-to-dendrite neural connections make up the plexiform layers separating rows of cell bodies." (Image and caption from Kolb 2003, used with permission)

[1] Dendrites are the branched projections of a neuron that act to conduct the electrical stimulation received from other neural cells to the cell body of the neuron from which the dendrites project.

nuclear layer and the ganglion cells and connects the bipolar and amacrine cells with the ganglion cells.

Seen from a perspective of Figure 7.2, there are two types of neural interactions in the retina, vertical and horizontal. The vertical wiring of the retina takes the visual signal from the photoreceptors through the bipolar cells into the ganglion cells and onto the brain. The horizontal wiring which occurs with the horizontal cells in the outer plexiform layer and with the amacrine cells of the inner plexiform layer provide feedback loops to the vertical wiring increasing spatial organization and modulating the intensity of its neural impulses. Whereas the vertical wiring is reasonably well understood, the horizontal wiring in primates, especially in the fovea, is only partially understood (Silveira et al, 2005).

7.3 The Bipolar Cells

The bipolar cells provide the link between the photoreceptors and the ganglion cells. Information is passed from the photoreceptors to the bipolar cells by way of the neurotransmitter[2] glutamate. Counter-intuitively, photoreceptors continually produce glutamate and only stop releasing it when stimulated by light. All bipolar cells have glutamate receptors on their surface but some have an inhibitory receptor and some an excitatory receptor (Boycott and Wässel, 1999). The cell with an inhibitory receptor is called an ON cell because it fires when it stops receiving glutamate (i.e. when it's connected photoreceptor fires). In contrast, the OFF bipolar cell has an excitatory glutamate receptor which means that it continually fires (releases glutamate) in the absence of light. It only stops firing when its connected photoreceptors fire.

[2] Neurotransmitters are chemicals that are used to relay, amplify and modulate electrical signals between a neuron and another cell.

This arrangement means that each photoreceptor is connected to two bipolar cells, one an ON channel and the other an OFF channel. This produces parallel sets of visual channels; the first is the ON channel, which detects light areas on a dark background; the second is the OFF channel, which detects dark areas on a light background. These dual channel properties of an image are fundamental to our visual processes.

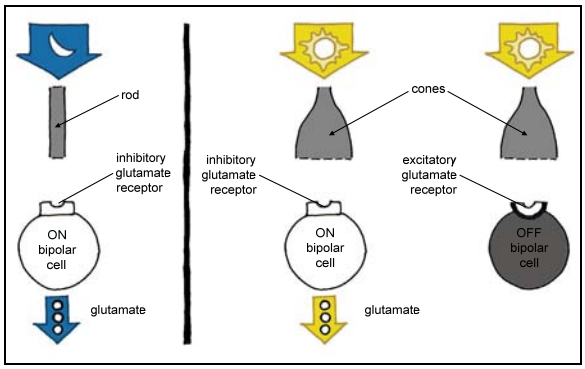


Figure 7.3: "Photoreceptors transmit information to bipolar cells using the molecule glutamate, but different bipolar cells respond differently to the presence of the molecule; some fire in response, whereas others cease firing, depending on the kind of glutamate receptor on their surface. Contrary to what one might expect, photoreceptors stop releasing glutamate when stimulated by light, in turn causing ON bipolar cells to release glutamate." (Image and caption from Kolb 2003, used with permission)

There is now generally recognized to be one class of bipolar cells that connect exclusively to rods and several classes of bipolar cells that connect exclusively to cones (Cajal, 1892). Cone bipolar cell classes differ with respect to the number of cones that they connect and the stratification level of their axons[3] within the inner plexiform layer (the layer that connects to the ganglion cells).

There are ten known types of cone bipolar cells present in human retina (Boycott and Wässel, 1991) but it is possible that more are likely to exist. Of those that we know, seven of them are concerned with converging information from many cones. They are known as diffuse cone bipolar types (DB). Three cone bipolar types are concerned only with single cone contacts in a one-to one relationship. These are known as midget bipolars and blue cone specific types (FMB, IMB and BB).

[3]An axon, also known as a nerve fiber, is a long, slender projection of a neuron that conducts electrical impulses away from the cell body.

7.3.1 Midget bipolar (MB) cells

For the purpose of visualization, the most important bipolar cells are the Midget bipolar cells because they provide us with our high spatial and chromatic acuity. These cells, which were first described by Polyak (1941), are the most numerous class of bipolar cell in the retina. MB cells have very small dendritic and axonal trees (input and output fields) and in the central and mid-peripheral retinal areas ($< 50^0$ of eccentricity) appear to connect a single cone with a single midget ganglion cell (ganglion cells will be discussed later in this chapter). In the far-peripheral region ($> 50^0$) they contact up to five cones (Wässel et al, 1994***)*** and several MB cells provide input to an individual parvocellular (PC) ganglion cell. Thus, in contrast to the central region of the retina, in the far-peripheral retina PC ganglion cells probably receive mixed cone input from L and M cones.

MB cells are further subdivided into flat (FMB) and invaginating (IMB) subclasses that correspond to OFF and ON cells respectively (Kolb, 1970). In the central retinal area, MB cells carry the chromatic signal from either an M-cone or an L-cone. Within this area, each cone type connects directly to one ON cell and one OFF cell, which suggests that the midget system has evolved to meet the demands of high spatial resolution and not for the processing of color signals (Lennie et al, 1990; Wässel and Boycott, 1991).

7.3.2 Blue-cone bipolar (BB) cells

The short wave sensitive cones (S-cones) do not connect with MB cells in the same manner as the M and L-cones. In the fovea, the blue-OFF signal is transmitted by the same FMB cells (Klug et al, 2003) as the M and L cones. The ON signal, however, is not transmitted by the IMB cells but by a special class of bipolar cell, the blue-cone (BB) cell. The S-cone or blue-cone bipolar cells are exclusively ON cells with wide dendritic

fields that connect to widely distributed S-cones (Mariani, 1984). They make synapses[4] with amacrine cells and ganglion cells in the IPL.

7.3.3 Diffuse bipolar (DB) cells

There are six subclasses of diffuse bipolar cells (DB1-DB6) three of which are OFF (DB1-DB3) cells and three of which are ON (DB4-DB6) cells plus one that has a very wide field (giant) in dendritic spread. Diffuse bipolar cells have wide dendritic and axonal trees and make non-selective contact with all cones. The smaller diffuse bipolar cells connect with 5-7 cones in the central retinal area and 12-14 cones in the peripheral retina. The giant diffuse bipolar cell connects with as many as 15-20 cones (Mariani, 1984) but its role in vision is not well understood.

7.3.4 Rod bipolar (RB) cells

Rod bipolar cells are exclusively ON cells. They have wide dendritic fields and make synapses with all rods within the field. RB cells make synapses with amacrine cells in the IPL.

7.4 Bipolar Receptive Fields and Simultaneous Contrast

All sensory neurons have what is known as a receptive field which is defined as the region of space in which the presence of a stimulus will alter the firing of that neuron. In the retina, the photoreceptors (rods and cones) have the narrowest receptive field possible because they respond only to photons that hit the receptor itself. In an analogy to technology, this narrow focus receptive field is virtually identical to the pixels in a digital camera, which respond only to photons that fall directly on them.

---

[4] Chemical synapses are specialized junctions through which cells of the nervous system signal to one another and to non-neuronal cells such as muscles.

The direct input from the cones does not, however, define the receptive field of the bipolar cells to which they connect. If it did and the retina were simply to transmit simple ON/OFF signals to the brain the resulting vision would most probably be blurred and grainy. Instead, the information streams sent to the brain are refined and honed in the retina, a process that starts with the horizontal cells of the outer plexiform layer (OPL). Bipolar cells function in concert with these horizontal cells and have what is known as a "center surround" receptive field.

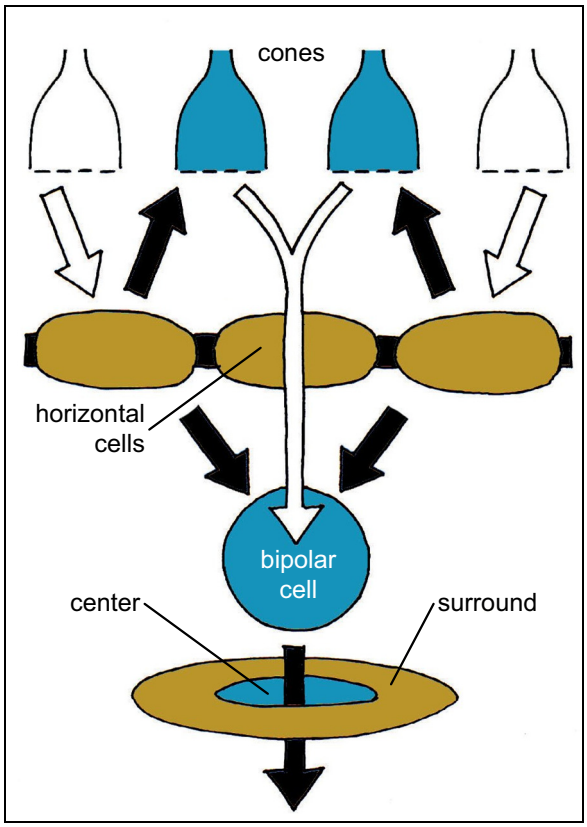


Figure 7.4: "Horizontal cells accumulate information from a wide field of cones and influence the signals bipolar cells transmit by adding an opponent surround signal to their receptive fields." (Image and caption from Kolb 2003, used with permission)

Using as very loose geophysical analogy, the photoreceptors would be the individual geophones and the bipolar cells would be a geophone array. The receptive field of a single geophone is very narrow and analogous to that of a photoreceptor in that it responds only to ground motion at its plant position. Geophone arrays, however, have a broader receptive field, because they sample the ground motion from an array of geophones that are distributed over an area.

It is now thought that horizontal cells of the OPL provide, through a mechanism of lateral inhibition, a surround arranged around the center of the receptive field of firstly the photoreceptor itself and then the bipolar cell contacting the photoreceptor (Baylor et al., 1971; Kaneko, 1970; Werblin and Dowling, 1969). The negative feedback synapse of the horizontal cell to the cone photoreceptor allows the larger receptive field of the horizontal cell network (horizontal cells are coupled across the retina by electrical synapses between neighboring cells) to provide a surround to the narrow central cone

response (Naka, 1976). This concentric organization is then transmitted to the bipolar cells making contact with the cone (Toyoda, 1972) and thence to the ganglion cell that the cone bipolar cell contacts.

This brings into question just what type of information is transmitted out of the retina. In Chapter 6, I showed that the only effect that a photon has on a photoreceptor is to increase its hyperpolarization. In consequence, the electrical signal leaving the photoreceptor cannot directly transmit chromatic information. I have previously made analogies between the biological processes of vision and both seismic acquisition and photography. It is now apparent that these analogies may be valid at the photoreceptor layer of the retina but not beyond. The signals, acoustic and chromatic, that are recorded by seismic acquisition systems and cameras are directly related to the intensity of the input signal. However, this is not the case for bipolar cells. Bipolar cells receive input from both the photoreceptors that signal ON/OFF and by feedback or lateral inhibition, input from a horizontal cell that signals the opposite, OFF/ON. This produces what is known as simultaneous contrast where a dark boundary inhibits a light area and vice versa.

Therefore, from the very earliest level of organization of the visual signal, we see that vision depends upon parallel channels of contrast. The brain never sees the raw output from the cones themselves. Instead, processing within the inner nuclear layer of the retina converts the raw inputs into parallel channels of contrast. The first channel, the ON channel, signals where a dark boundary inhibits a light area. The second channel, the OFF channel, signals the opposite. Both channels are connected to the ganglion cells that provide further processing.

7.5 The Ganglion Cells

Ganglion cells are the final output neurons of the vertebrate retina. They receive information about the visual world from the previously mentioned bipolar cells and from the amacrine cells (retinal interneuron's) of the inner plexiform layer. The ganglion cells are the "digitizing" engine of the visual processing system in that they convert the sensed

chemical energy into nerve spikes. Nerve spikes are a time-coded form of electrical signaling used to transmit nervous system information over long distances, in this case along the optic nerve and on into the brain visual centers.

Ganglion cells are also the most complex information processing system in the vertebrate retina. Different cells are selectively tuned to detect subtle features of the visual scene, including color, size and direction and speed of motion. It is important to realize that these so called “trigger features”, may not have a unique interpretation and that ultimately it is up to the brain to determine the most likely interpretation for them.

The preprocessing done by the ganglion cells is greatly aided by the amacrine cells of the inner plexiform layer. The amacrine cells play a large role in building the image that is ultimately passed onto the brain and in the IPL; there are more than 22 types of these cells that make connections with more than 18 different types of ganglion cells. Not surprisingly, it is estimated that only about half of the interactions in this layer are fully understood.

Understandably, a full description of the known interactions between the amacrine cells, the bipolar cells and the ganglion cells are beyond the scope of this thesis. Rather, I will concentrate on the basic functioning of the midget ganglion cells and on the streams of visual information that they eventually pass onto the brain for interpretation.

7.5.1 Midget Ganglion Cells

There are thought to be three ganglion cell types involved with spatial and color vision namely the midget ganglion cells, the blue/yellow ganglion cells and the parasol ganglion cells (Polyak, 1941; Kolb et al., 1992). In the context of this thesis, the most important of these are the midget ganglion cells, which make up approximately 80% of the ganglion cells and are thought to be high acuity cells that also carry a color specific signal. They project to the parvocellular (PC) layer of the lateral geniculate nucleus (LGN: simplistically the relay station between the retina and the visual cortex) of the brain (see 7.6) and are thus referred to as PC cells (or simply P cells).

It has now been proven that in the fovea a single PC cell connects to a single midget bipolar cell (Kolb and DeKorver, 1991; Kolb and Marshak, 2003). Considering that in this area a single MB cell connects to a single cone, the foveal PC pathway provides a "private line" of communication between a single cone and the brain. It is known (Dacey et al, 2000) that in primates, midget ganglion cells have the same ON center/OFF center organization as do the MB cells. It appears then, that in the foveal region and out to the borders of the central retina (about 4 mm from the fovea center) the midget pathways of the human fovea are organized in the following manner: one cone to two midget bipolar cells (ON- and OFF-center bipolar types) to two midget ganglion cells (ON- and OFF-center ganglion cell types).

Outside of the fovea in the near to mid-peripheral area, the PC cells connect with three MB cells each of which connects to an individual cone (Polyak, 1941; Kolb et al., 1992). Beyond this area, in the far-peripheral area of the retina, three MB cells (each of which connects to up to five cones) converge on an individual PC cell. Thus, outside of the central retina, each PC cell receives input from multiple cones, three in the case of the mid-peripheral retina and 9 or more in the far-peripheral (Kolb and Marshak, 2003).

7.5.1.1 *PC Cell Receptive Field*

The PC ganglion cell receptive field is center-surround in much the same way that the MB bipolar cell is. Both the center and surround have a Gaussian responsivity (Derrington and Lennie, 1984) which results in a "Mexican-hat" shaped receptive field shape (Figure 7.5) reflecting the integration of opposing information about centers and surrounds. This kind of processing helps sharpen the boundaries of images.

Human retinas have two types of ganglion cells: ON-center and OFF-center, which parallels the organization of the MB cells. ON-center ganglion cells are activated when a spot of light falls in the center of their receptive fields, whereas OFF-center ganglion cells fire in response to light falling on their fields' periphery leaving their center dark (Figure 7.5).

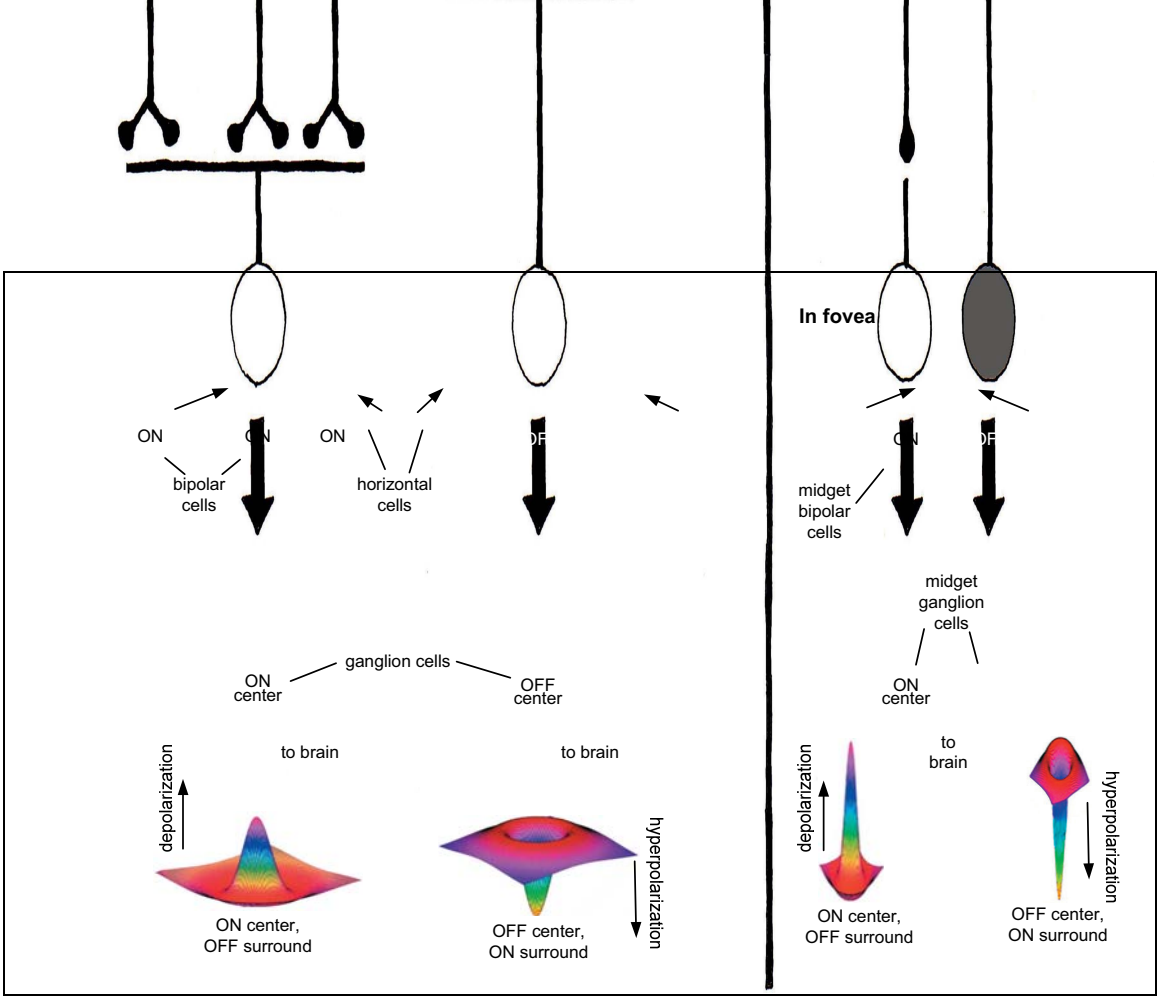


Figure 7.5: A comparison of the receptive field size for PC ganglion cells within the fovea (right) and in the mid to far peripheral areas of the retina (From Kolb 2003, used with permission)

The center of the receptive field comes directly from the MB cells whereas the surround comes from both the horizontal cells of the IPL and the amacrine cells of the OPL. The horizontal cells contribute to the surround by conveying antagonistic surround signals to bipolar cells and thence on to the to ganglion cells. Several classes of amacrine cells also contribute directly to the surround (Kolb et al., 2002), receiving input from MB cells and providing an antagonistic signal for the PC cells.

In the fovea (right Figure 7.5), PC ganglion cells have very narrow receptive fields each carrying information from a single cone. Outside of the fovea (left Figure 7.5) the PC receptive field becomes very broad and diffuse reflecting the fact that outside of the fovea both PC cells and MB cells receive inputs from multiple sources.

7.5.1.2 *Color Opponent PC Cells*

All catarrhine primates have trichromatic color vision in that they possess a mutated copy of the LWS gene. This gives them (and us) two different cones (M and L-cones) for detecting medium to long wavelength light. Because of this, PC ganglion cells and PC LGN cells always display some form of red-green color opponency due to M-cone and L-cone input to the center and surround of their receptive field.

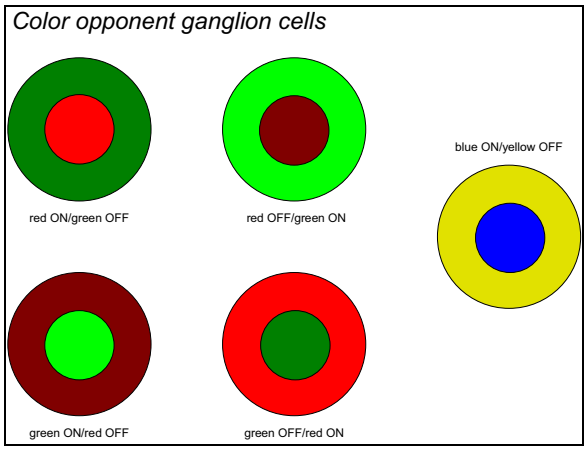


Figure 7.6: The receptive field of the four-color opponent PC ganglion cells and the small bi-stratified (blue/yellow) ganglion cell.

Figure 7.6 represents the generally accepted view of the color opponent PC ganglion cells. It is clearly established that within the central retina, the private-line PC cells receive their center input from a single M-cone or L-cone. What is not clear is if the receptive field surround of the cell receives its input exclusively from the cone that does not contribute to the center or from a mixture of M-cones and L-cones. Thus, it is possible, that the surrounds shown in Figure 7.6 are not purely from one cone as indicated. However, single cone-type surrounds would improve the PC pathway chromatic signal and there is some physiological evidence in its favor (Reid and Shapley, 1992; Lee et al., 1998). Further, there is evidence that in the peripheral retina, where the center and surround receive input from many more cones, that the PC cells are overtly red-green opponent (Martin et al, 2001). This again indicates that the surrounds are probably from a single cone type and not mixed. For the purposes of this thesis, therefore, I will assume that the private-line PC cell surrounds are from the opponent cone to the center.

The reader may be tempted to conclude that the PC color opponent ganglion cells correspond to the Hering red-green and blue-yellow opponent channels described in Chapter 2. For various reasons they are not, those channels are formed later in the

neocortex. The output from these cells is more complex and will be described in more detail next.

7.5.1.3 *Achromatic and Chromatic Single Opponent Signals*

The PC or parvocellular pathway is of primary importance to visualization and I have already shown that this system is based upon single opponent receptive fields (7.5.1.2). As shown in Figure 7.6, there are four types of midget ganglion cells; (1) red-ON/green-OFF; (2) red-OFF/green-ON; (3) green-ON/red-OFF; (4) green-OFF/red-ON. The beauty of this system is that these single opponent cells produce a signal that contains both achromatic and chromatic information multiplexed into one output.

The output signal is the sum of three filters; a line spread function, a spectral weighting function and a temporal weighting function. I will ignore here the temporal component and describe the output in terms of the first to two.

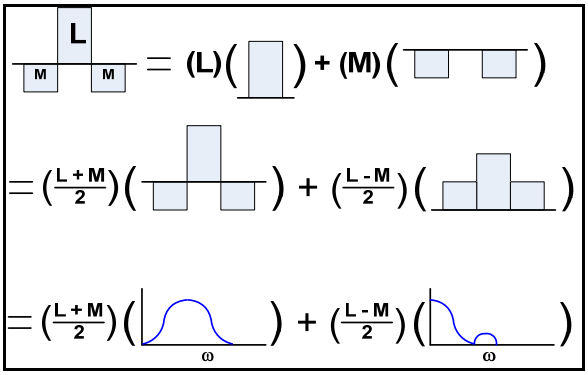


Figure 7.7: "The "double-duty" of the L/M single opponent unit. An L-cone on-center M-cone off center cell is shown to be separable into a cone additive (L+M) and a cone subtractive (L-M) component each with its own line-spread function. The amplitude spectrum of each component is also shown (ω = spatial frequency)." Image and caption adapted from Figure 12.3 in Vision and Visual Dysfunction 6: The Perception of Color, by Gouras, CRC Press Inc.

Figure 7.7 is a schematic of the line spread function and the spectral weighting, this function is shown on the left. The first two lines show how the receptive field can be described as the sum of two line spread functions each multiplied by the appropriate spectral weighting function (L & M). Line three shows the Fourier transform of the line spread functions (i.e. the response of the PC cell as a function of spatial frequency).

This analysis shows us that the final output of a PC ganglion cell can be described in terms of two components; (1) a cone additive function with bandpass spatial frequency characteristics; and (2) a cone subtractive function with low-pass spatial frequency

characteristics. The cone additive function represents an achromatic or luminance signal whereas the subtractive component represents a chromatic signal.

To summarize, the same single-opponent PC cell responds to both chromatic and achromatic changes in contrast. The chromatic system responds best to constant or slowly varying (spatially) chromatic contrasts whereas the achromatic system has a preferred spatial frequency for luminance variations and does not respond to uniform fields. Because we have ignored the temporal filter, the above is a simplification of the true output from a PC cell. Including the temporal filter does complicate the output but it does not change its nature of containing multiplexed achromatic and chromatic signals.

7.5.2 Small Bistratified Ganglion Cells

The previously mentioned midget ganglion cells, which dominate the retina, receive input exclusively from M and L-cones. There is some evidence that they receive some input from S-cones but if they do, the contribution of the S-cones is thought to be minimal. Short wavelength information is instead transmitted to the brain along a different pathway, the Blue ON/Yellow OFF pathway (right: Figure 7.6).

The short-wavelength system is very ancient and there are differences in the genetic structure and locus of the S-cone visual pigment compared with the L- and M-cone pigments (Nathans et al., 1986). The S-cones are common to all vertebrate retinas and always form a consistent 8-10% of the cone photoreceptor population (Marc, 1982). In primates, however, S-cones are infrequent in the fovea and totally absent in the foveal pit causing a so-called S-cone blind spot (Williams et al., 1981). The population of S-cones peak in number on the foveal slope at about 12% of the population.

Early electrophysiological investigation of monkey (and presumably human) retinal ganglion cells indicated that blue/yellow opponency was carried primarily by an S-cone ON center ganglion cell type (right: Figure 7.6) with a much larger receptive field center than is typical of the L- or M-cone midget ganglion cells (Mariani, 1984). This ganglion cell, called the small bistratified ganglion cell, receives its receptive field center input from the blue-cone bipolar cell (BB) and does not appear to have a spatially antagonistic

receptive field structure. Instead, the yellow opponency appears to be coextensive with the center of the receptive field. It is likely that the yellow opponency is already present in the BB cell via the horizontal cells of the OPL. Interestingly there are very few recordings of the opposite type of ganglion cell i.e. yellow-ON and blue-OFF

Not surprising given the domination of M and L-cones in the retina, the short wavelength system has a lower spatial and temporal resolution than the other two cone systems (Stockman et al., 1991; Humanski and Wilson, 1992) but because of its ancient lineage, it is probably the only system to truly carry color information through the retina (Rodieck, 1991).

7.5.3 Other Ganglion Cells

The midget ganglion cells and the small bistratified ganglion cells are the only cells that are known to exhibit color opponency and to contribute to color vision. They dominate conscious perception but they are not the only ganglion cells in the retina. There are numerous other types of cells that make significant contributions to other aspects of vision. For example, the parasol ganglion cells (MC), which receive mixed input from M and L-cones via the diffuse bipolar cells, make up about 10% of the ganglion cell population. MC cells respond in a phasic (transient) manner as opposed to the tonic (sustained) response of the PC cells and respond best to weak contrast signals that cover a large area (Kaplan and Shapley, 1986). The exact function of this parasol system is not yet conclusively understood but it is hypothesized that it functions in concert with the PC system to increase the brains ability to extract high-resolution information (Silveria, 1996).

In addition, there are ganglion cells exclusively dedicated to rods and numerous types of wide field cells that receive input from a much wider collection of cones than either the PC and MC cells. Neither of these groups, however, is known to contribute to color vision or our conscious perception and will, therefore, not be covered further.

7.6 The Lateral Geniculate Nucleus

In mammals, the most important visual pathway is the retino-geniculo-cortical (RGC) pathway. There are other pathways concerned with the processing of visual information but these deal with non-perceptual visual processes such as pupil size and eye movement. The RGC pathway is the pathway of visual perception and is therefore of primary interest in the context of this thesis.

To this point, I have dealt only with the first part of this pathway, the generation of the neural signals that travel along the optic nerves (plural in this case because there is one for each eye). Visual perception occurs in the visual cortex but this is not where the optic nerves terminate. Rather, they terminate in the lateral geniculate nucleus (LGN – the geniculo part of the pathway) which is a part of the thalamus. The thalamus is the first part of the brain where neural signals can be modified by input from other senses and this is especially true of the LGN. The LGN receives information not only from the retina, but it also receives extra retinal inputs from other senses and also via feedback loops from the visual cortex itself. As such, it is far more than a relay station. It heavily modulates the 'driving' input from the retinal ganglion cells and it projects the information directly to the primary visual cortex. Whereas the retina is silent when the eyes are closed, the LGN is not, continuing to feed information to the visual cortex even in the absence of direct visual stimuli.

The LGN is a critical part of the visual processing system. Unfortunately, it is the least studied part of the pathway and much of its function is still poorly understood. Even so, since it ultimately forms the streams of information that are used in visual perception, it must be considered here for completeness.

7.6.1 Input to the LGN

There are two functional categories of afferents[5] to the LGN, the 'drivers' and the 'modulators' (Sherman and Guillary, 1998, 2001). The drivers determine the qualitative characteristics of the receptive fields of the LGN cells whereas the modulators alter the quantitative aspects of the receptive fields without changing their basic properties. For the LGN, the driving inputs come directly from the axons of ganglion cells of the retina (which form the optic nerve). The modulating inputs come from other sensory organs, but mainly from feedback loops from the visual cortex itself.

7.6.2 Structure of the LGN

The LGN is a feature of the thalamus of all mammals and there are certain key features that are common to nearly all (Kaas et al., 1972). First, the LGN is subdivided into layers with each layer forming a retinotopic[6] representation of one part of the visual field. There are two LGN's, one for each hemisphere of the brain. Each LGN receives input from both eyes. These inputs are kept separate and so each layer within the LGN is activated by one eye or the other.

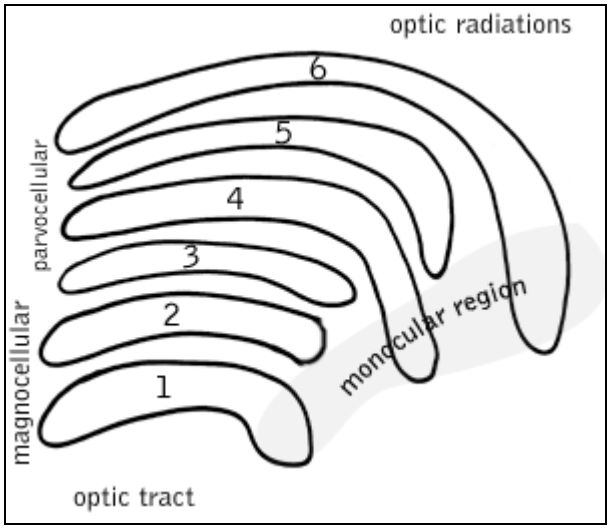


Figure 7.8: Schematic diagram of the primate lateral geniculate nucleus. The contralateral eye projects to layer 1, 4 & 6 and the ipsilateral to layers 2, 3 & 5.

---

[5] Afferent neurons, otherwise known as sensory or receptor neurons, carry nerve impulses from receptors or sense organs toward the central nervous system (in this case the brain).

[6] Retinotopy is the concept that certain areas of the visual cortex are organized in a way that adjacent points in the visual field (that fall on adjacent points on the retina) are processed by neurons in adjacent parts of the cortical area.

The visual field of each LGN is defined by the visual hemifield of the contralateral[7] eye. Because of this, the layers that receive input from the contralateral eye (Figure 7.8: layers 1, 4 & 6) are larger than the layers receiving input from the ipsilateral eye (Figure 7.8: layers 2, 3 & 5) because of the inclusion of the lateral extreme of the hemifield, i.e. the part of the visual field not seen by the other eye.

A second general feature of LGN organization is that the retinotopic maps on the various layers are precisely aligned. The LGN projects directly to the primary visual cortex, which contains a single fused map of the contralateral visual hemifield. Consequently, each location in the visual cortex receives input from a column or row of LGN neurons that extend across all LGN layers.

The main feature of LGN organization is the lamination. All primates have an LGN with four basic layers; however, these basic layers are elaborated upon depending upon the evolutionary line. Primates have layers with two types of cells, small cells and large cells. The large cell layers are called the magnocellular layers (MC) whereas the layers of small cells are called parvocellular layers (PC). Since each LGN receives input from both the contralateral and the ipsilateral eye, primates have four basic layers, two PC layers and two MC layers. All primates also have a third type of even smaller cells called koniocells (KC) between the main layers (Kaas et al, 1978). In anthropoid primates, these cells are too few to be classified as actual layers and instead the zones containing them are considered to be simply interlaminar zones. In prosimian primates, these zones are quite distinct and are called the koniocellular layers, which elevate the total number of prosimian layers to six.

The PC layers receive their primary input from the PC ganglion cells and thus are part of the midget pathway that carries high-resolution spatial and color information to the visual cortex. Likewise, the MC cells receive their primary input directly from the MC ganglion cells. The koniocellular layers appear to receive their direct input from the

[7] Contralateral refers to the opposite side of the body whereas ipsilateral refers to the same side.

small bistratified ganglion cells (Martin, 1998) and thus are part of the blue-ON/yellow-OFF pathway. This partially explains why the koniocellular interlaminar zone is so indistinct in anthropoids. That it is much less distinct than the other layers is expected considering that both PC and MC ganglion cells, which receive input from the M-cones and L-cones, are far more numerous in the retina.

At first glance Figure 7.8, which is a schematic drawing of the anthropoid LGN, seems to imply that anthropoids also have six layers in the LGN. Figure 7.8 implies that layers 3 & 5 (from the ipsilateral eye) and 4 & 6 (from the contralateral eye) are distinct layers but they are not, they are actually the same layers. The basic PC layers of all Haplorrhine primates are partially subdivided and interdigited and so layers 3 & 5 are the same layer as layers 4 & 6. Because of this subdividing, the LGN of anthropoids has been described as having six layers whereas in fact there are only four complete layers.

Regardless of the number and subdivision of the layers in the LGN, the layers are stacked in such a manner that a toothpick driven through all of the layers would hit the same point in visual space in each of the layers.

7.6.3 Receptive Field of LGN PC Cells

There is a high degree of correlation between the receptive field of the LGN cells and the retinal ganglion cells implying that an LGN cell receives dominant input from a single retinal ganglion cell (Sherman and Guillery, 2001). There is also considerable evidence that there is a tight retinotopic organization of the LGN, which again implies that there is an almost 1 to 1 relationship between retinal and LGN PC cells (Kaplan et al., 1987). This relationship between the properties of the receptive field of LGN and retinal cells is called the classical receptive field (CRF).

All this is not to imply that the LGN is simply a relay station between the retina and the visual cortex. In addition to the CRF, the LGN cells also receive input from surrounding regions of visual space, which is known as the extraclassical receptive field (ECRF). The ECRF inputs do not directly elicit a response from the cell but rather function as ‘modulators’ by modifying the response of the cell. This is done in a center-

surround relationship with the center coming from the CRF and the surround from the ECRF.

The ECRF inputs greatly complicate the understanding of the receptive field and function of the LGN. The primary visual cortex provides the major extraretinal inputs to the LGN of all species studied (Sherman and Guillery, 1996) but other inputs come from other sensory organs and also from within the LGN itself. Exactly how these ECRF inputs functions is the subject of much debate. What is not in debate, however, is that the CRF properties come directly from the retinal ganglion cells. Because of this the structure and organization of the visual image that is produced in the ganglion cell layer of the retina (including the “private line” information) is preserved and subsequently projected onto the visual cortex.

CHAPTER EIGHT: THE VISUAL INTERPRETATION SYSTEM

"The human visual system can detect and discriminate between an incredibly diverse assortment of stimuli that may be chromatic or achromatic, in motion or not, pattered or unpatterned, two-dimensional or three. Remarkably, the neural end-product of visual stimuli impacting upon the retina is, in one sense, always the same. After the complexities of phototransduction, lateral interactions provided by horizontal and amacrine cells, and integration of signals by ganglion cell dendrites only the constantly changing stream of action potentials propagating along ganglion cell axons is left to inform our visual perception. These seemingly identical signals must somehow be processed in the subcortex and cortex to create the full range of visual percepts we experience. How this is achieved is a puzzle that currently occupies the professional lives of thousands of researchers and the basic framework of a solution has only begun to unfold in the last several decades."

Dr. Mathew Schmolesky
The Primary Visual Cortex
Webvision

8.1 Introduction

And so we finally arrive at the "gray matter", that part of the central nervous system that is dedicated to converting the signals from the various sensory organs into our perceptions of sight, sound, touch, taste and smell. The purpose of this journey was to learn how humans established their perceptions of form and color. Everything to this point was simply the path but this is the destination. Unfortunately, having traveled down a very long and twisted path we have arrived to find that no one is at home! No one is at home because we simply do not know the physiological processes that convert the neural signals that project into the visual cortex, into perceptions.

There is a direct correlation between how much we know about any particular aspect of vision and how far it lies from the retina. On the very surface of the retina, the ratio between what we know to what there is to know, approaches unity. In the visual cortex, however, that ratio is just barely greater than zero. Consequently, the final answer to how we form perceptions lies buried deep within areas that are still outside our sphere of knowledge.

This is not to imply that studying the visual cortex is an esoteric exercise. The knowledge that we do have, extends mostly into the V1 area of the cortex, an area that is

the most actively and extensively studied part of the human brain. It is in this area that we first see how perceptions might be formed. In this region, we start to detect the form and the color of an object, the two things that are critical for visualization. We are a long way from knowing exactly how the perception of an object is eventually established in the mind. However, through our expanding knowledge of V1, we have started to learn at least the underlying mechanics for how we go about detecting it.

We will have to satisfy ourselves that getting just a fleeting glimpse of those processes justifies the trip. I must warn you, however, that you can only catch a glimpse of them by journeying deep into the caverns of the mind, much deeper than it is safe for a neophyte spelunker to venture. I only have a very sketchy map so you will need lots of rope if you ever hope to return!

8.2 The Cerebral Cortex

The cerebrum or cerebral cortex, originally functioning as part of the olfactory lobes, is involved with the more complex functions of the human brain. In humans and other advanced vertebrates, the cerebrum has grown over the rest of the brain, forming a convoluted (wrinkled) layer of gray matter. The degree of convolution is partly dependent on the size of the body. Small mammals (e.g., lesser anteater, marmoset) generally have smooth brains, and large mammals (e.g., whale, elephant, dolphin) generally have highly convoluted ones.

The cerebrum is the largest and uppermost portion of the brain. The cerebrum consists of the cerebral hemispheres and accounts for two-thirds of the total weight of the brain. One hemisphere, usually the left, is functionally dominant, controlling language and speech. The other hemisphere interprets visual and spatial information.

The cerebral hemispheres consist of an inner core of myelinated nerve fibers, the white matter, and an outer cortex of gray matter. The cerebral cortex is responsible for integrating sensory impulses, directing motor activity, and controlling higher intellectual functions. The human cortex is several centimeters thick and has a surface area of about 2,000 square cm (310 square inches), largely because of an elaborate series of

convolutions; the extensive development of this cortex in humans is believed to distinguish the human brain from those of other animals.

Nerve fibers in the white matter primarily connect functional areas of the cerebral cortex. The gray matter of the cerebral cortex usually is divided into four lobes, roughly defined by major surface folds. The frontal lobe contains control centers for motor activity and speech, the parietal for somatic senses (touch and position), the temporal for auditory reception and memory, and the occipital for visual reception. Sometimes the limbic lobe, involved with smell, taste, and emotions, is considered to be a fifth lobe.

Numerous deep grooves in the cerebral cortex, called cerebral fissures, originate in the extensive folding of the brain's surface. The main cerebral fissures are the lateral fissure, or fissure of Sylvius, between the frontal and temporal lobes; the central fissure, or fissure of Rolando, between the frontal and parietal lobes, which separates the chief motor and sensory regions of the brain; the calcarine fissure on the occipital lobe, which contains the visual cortex; the parieto-occipital fissure, which separates the parietal and occipital lobes; the transverse fissure, which divides the cerebrum from the cerebellum; and the longitudinal fissure, which divides the cerebrum into two hemispheres.

A thick band of white matter that connects the two hemispheres, called the corpus callosum, allows the integration of sensory input and functional responses from both sides of the body. Other cerebral structures include the hypothalamus, which controls metabolism and maintains homeostasis, and the thalamus, a principal sensory relay centre. These structures surround spaces (ventricles) filled with cerebrospinal fluid, which helps to supply the brain cells with nutrients and provides the brain with shock-absorbing mechanical support.

Note: The material in this section was excerpted from:

"Brain. (2007). In *Encyclopædia Britannica*. Retrieved January 10, 2007, from Encyclopædia Britannica Online: http://www.britannica.com/eb/article-9016178"

"Cerebrum. (2007). In *Encyclopædia Britannica.* Retrieved January 10, 2007, from Encyclopædia Britannica Online: http://www.britannica.com/eb/article-9022139"

8.3 The Visual Cortex

Unless otherwise noted, the material in section 8.3 and 8.4 are excerpted from:

"The Primary Visual Cortex" by Dr. Mathew Schmolesky, Webvision

"Eye, Brain and Vision", the online edition, by Nobel Laureate David H. Hubel.

"Color Vision" by Dr. Peter Gouras, Webvision

8.3.1 Basic Physiology

The visual or calcarine cortex is located in the occipital lobe, which is the visual processing center of the mammalian brain. The occipital lobes are the smallest of the four true lobes of the human brain. They lie at the back of the brain and are caudal[1] to the parieto-occipital sulcus[2]. This sulcus joins the calcarine sulcus in a Y-shaped formation. The area above the Y-shaped sulcus is called the cuneus gyrus[3] whereas the area below is called the lingual gyrus.

---

[1] Caudal – below and behind.

[2] In neuroanatomy, a sulcus (pl. sulci) is a depression in the surface of the brain. The large furrows (sulci) that divide the brain into lobes are often called fissures.

[3] A gyrus (pl. gyri) is a ridge on the cerebral cortex. It is generally surrounded by one or more sulci.

The cortex on both banks of the calcarine sulcus (calcarine fissure: left Figure 8.2) form the primary visual area, which receives input from the contralateral eye via the lateral geniculate nucleus and the optic radiations. The visual field is represented near the calcarine sulcus in a retinotopic fashion. The upper quadrants of the visual field are laid out along the lower bank of the sulcus (the lingual gyrus) and the lower quadrants along the upper bank (the cuneus gyrus).

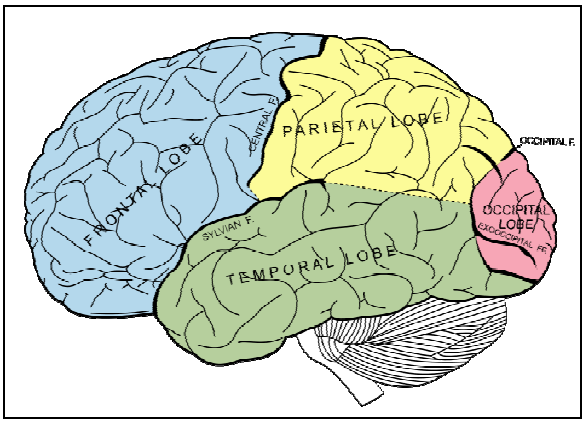


Figure 8.1: Lobes of the human brain. The occipital lobe is shown in red. From Gray's Anatomy 1918, public domain image.

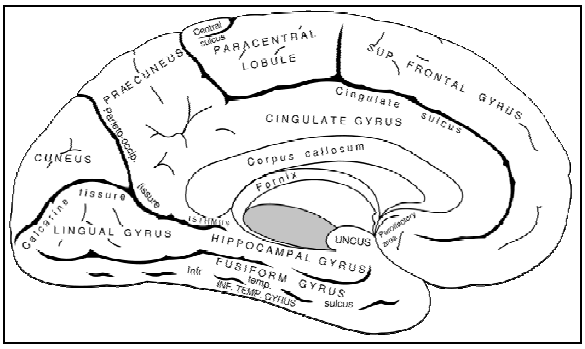


Figure 8.2: Medial surface of the left cerebral hemisphere. From Gray's Anatomy 1918, public domain image.

The term "visual cortex" actually refers to multiple regions of the cortex including the primary visual cortex (also know as the striate cortex or layer V1) and other extrastriate cortical areas known as V2, V3, V4, V5 (also known as MT). The primary visual cortex is the most heavily studied part of the central nervous system and a significant amount of its functionality has been revealed. However, the number and function of the extrastriate cortical areas is less well understood and what is known about them is most often conflicting and incomplete.

8.3.2 Dorsal and Ventral Streams

There are two primary pathways of information from the lateral geniculate nucleus into the primary visual cortex. The first is the parvocellular pathway, which arises in the midget (PC) ganglion cells of the retina. The second is the magnocellular pathway, which has its origins in the parasol (MC) ganglion cells. PC cells respond in a tonic or sustained manner whereas the MC cells respond in a phasic or transient manner. This difference in

response characteristics would tend to indicate that the information from the two types of cells is destined for different visual pathways. This is, in fact, what happens in the visual cortex. There, the two pathways of information, the parvocellular and the magnocellular, are processed in separate parallel streams, called the ventral stream and the dorsal stream respectively.

The ventral stream, which is fed by the parvocellular system, is sometimes called the "What Pathway" because it is associated with form recognition, object representation and the storage of long-term memory. The dorsal stream, which is fed by the magnocellular pathway, is sometimes called the "How Pathway" because it is associated with the perception of motion, the representation of object location and the control of the eyes and arms (Ungerleider and Mishkin, 1982). Simplistically, in this light, the ventral stream represents the pathway that we use when we study and object at rest. The dorsal stream, on the other hand, is responsible for our understanding of where an object is and for our hand-eye coordination. Although proposed in 1982, the ventral/dorsal stream concept is still highly contentious and most researchers now consider it an oversimplification of the true processes that occur in V1. Even so, given that the ventral stream processes the "private-line" information from the individual cones in the fovea it will be the only pathway studied here in detail.

8.4 Primary Visual Cortex

I should point out that most of what follows is not based directly upon the human visual cortex. Rather it is based upon studies performed on macaque monkeys because it is the best model that we have for how the human visual system and visual cortex in particular, functions.

The V1 is the part of the visual cortex that receives input directly from the retino-geniculate pathway. As such, the response of its neurons is more directly related to the retinal image than are the other areas, which are more complexly related. Because of this, the interactions and functions of its neurons are more easily understood and have been unraveled in some detail.

The subsequent layers of the visual cortex are just as important in the study of the perception of form, color and movement but their function and interactions are not well understood. Within these upper layers of the visual cortex, lies the final link between psychophysics and physiology but that link is elusive and so my study of the visual system is best terminated here within the V1. Even so, understanding the V1, even at the superficial level that I will cover here, is sufficient to give us insight into how the brain eventually analyses an image and begins the process of perception.

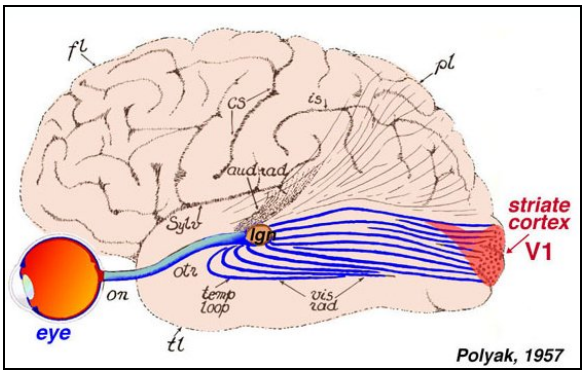


Figure 8.3: Visual input from the brain goes from the eye to the LGN and then to the primary visual cortex, or area V1, which is located in the posterior of the occipital lobe. From Polyak, 1957.

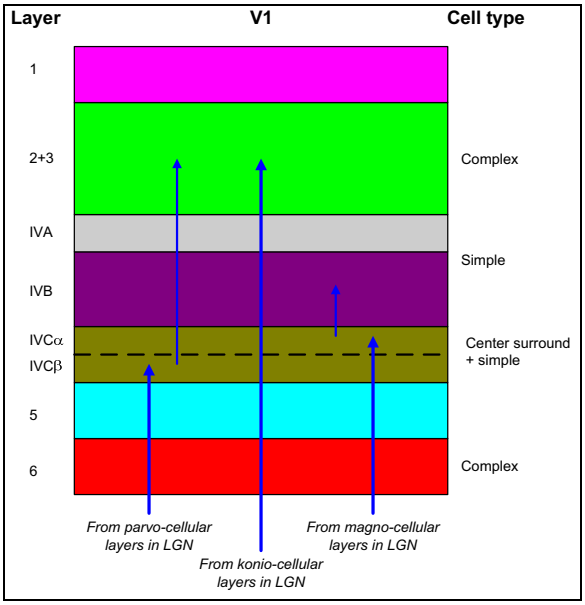


Figure 8.4: Layering of the striate cortex. The striate cortex receives input from the lateral geniculate nucleus. The parvocellular (ventral) stream projects into layer IVCβ whereas the magnocellular (dorsal) stream projects into sublayer IVCα. The konio-cellular layers of the LGN project directly into layers 2 and 3.

The primary or striate visual cortex (V1) is the best-studied visual area of the brain. It is located in the posterior pole of the occipital cortex and is highly specialized for processing information about static and moving objects. It is excellent at recognizing patterns and its anatomy has been worked out in detail. It is divided into six functionally distinct layers numbered 1 – 6.

V1 is a plate of cells approximately 2 mm thick with a surface area of a few square inches. Compared with the LGN it is an immense structure. The LGN contains roughly 1.5 million cells whereas the striate cortex contains around 200 million (Hubel, 1988).

The cells in V1 are topographically organized (retinotopic). This means that as you move along the retina from one point to another, the cells in both the LGN and the striate cortex trace a continuous path. For example, the spatial organization of the ganglion cells in one part of the retina is mirrored in both the LGN and the striate cortex. In essence if we think of the output from the retina as being an image, the same image is carried through the visual system and is projected upon the striate cortex in the same way that an image is projected upon a screen. The reader is cautioned not to take that analogy too far, the visual signal is highly processed along the way, but the basic principle holds.

V1 Layer 4 receives most of the input from the LGN and is subdivided into 4 sub layers, IVA, IVB, and IVCβ (roman numerals are used in the literature to avoid confusion with the primary layering). The chromatic parvocellular layers of the LGN project into sub layer IVCβ whereas the magnocellular layers project into sub layer IVCα. The blue-yellow konio-cellular intercalated layers of the LGN project directly into the complex cells of layers 2 & 3.

8.4.1 Cortical Magnification in V1

The eye, as we already know, is spherical which is an ideal shape considering that any other shape would have a hard time rotating in the limited dimensions of the skull. Likewise, the retina is also spherical which differentiates it from the film of a camera, which is flat. This spherical shape is important because it maintains a constant retinal magnification, roughly 3.5 degrees of visual field per mm. Consequently, the retinal image can be recorded without the peripheral distortion that is found on fish-eye camera lenses.

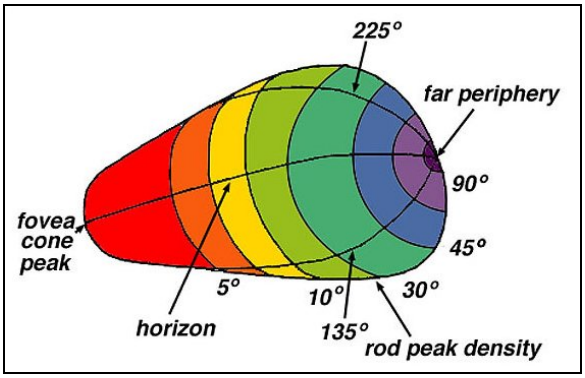


Figure 8.5: "The unfolded striate cortex has a shape like a pear. It would be a quarter sphere if the visual fields were equally represented everywhere but instead it is greatly distorted by the disproportionate representation of parts near the center of gaze (fovea), a feature termed "cortical magnification". In contrast the far periphery is greatly underrepresented." Image and caption from "The Primary Visual Cortex", M. Schmolesky, Webvision.

Although the image is recorded without distortion this does imply that it is recorded uniformly. I have already elaborated in Chapter 6 that there is a wide disparity in cone density between the fovea centralis and the peripheral regions of the retina with a ratio of approximately 20-1. Likewise, the ganglion cells, which transmit the visual information out of the eye, are also highly concentrated in the fovea. Given then that the retinotopic map is preserved through the LGN and the V1 it should not be surprising to learn that far more of the V1 is dedicated to processing foveal information than the peripheral regions.

Unlike the retina, the striate cortex is not restricted to being a sphere and we already know that it is a highly convoluted structure. However, when unfolded this structure does not resemble either a sphere or a plate, rather it approximates a pear like object as seen in Figure 8.5. Given the retinotopic organization of V1 and the concentration of visual information around the fovea, it is clear that the majority of the striate cortex is dedicated to processing foveal information. This serves as further evidence of the primate fixation on what is immediately before it.

8.4.2 Orientation, Ocularity and Hypercolumns in V1

Prior to the V1, the output from each eye is kept separate. For example, a single neuron in the LGN receives input from only one eye. This is undoubtedly due to the need to maintain two monocular streams to facilitate depth perception. Stereopsis[4] relies heavily on the shift in position of the two eyes when focusing on objects at distance. It would appear then to be important that the output from the two eyes be kept separate until the monocular information can be extracted by the brain.

The cells of layers IVCα and IVCβ, which receive their input directly from the LGN, are known to be strictly monocular. However, this starts to change in the subsequent V1 layers where neurons start to receive afferents from both eyes. In V1, cells maintain an

[4] The perception of depth.

ocular dominance but many of the cells receive input from both eyes but with one being dominant.

One of the most important discoveries in the understanding of how the visual cortex functions was the discovery that orientation plays a major role in stimulating V1 neurons (Hubel 1988). That orientation was important to vision had already been suggested by psychophysical experiments (Gibson, 1950) but the work of Hubel and Weisel was of such importance in the study of the brain that it earned them the Nobel Prize.

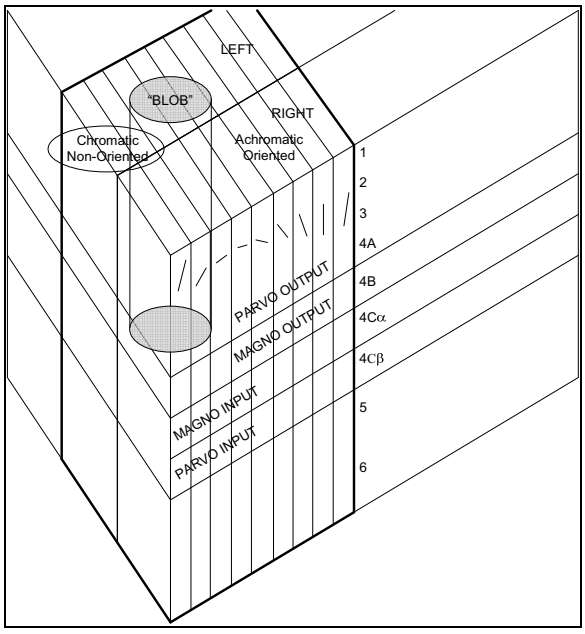


Figure 8.6: "Representation of a hypercolumn in visual area V1 of the primate brain. A hypercolumn is composed of two vertical ocular-dominance columns with cells representing one or mainly one eye juxtaposed to an adjacent slab containing cells favoring the other eye. Each ocularity column is composed of a stack of orientation columns each containing cells that favor a particular orientation; the change in orientation preference is continuous through the ocularity-columns. Within the center of each ocularity column is a cylinder of cells called a blob. Within the blob cells lack orientation selectivity and frequently react to a unique chromatic contrast." Figure and caption adapted from Figure 11.2 in *The Perception of Color* Vol. 6 *Vision and Visual Dysfunction*, (ed.) Gouras, 1991, Macmillan, London.

What Hubel and Weisel discovered was that there appeared to be clusters of cells that responded to the same orientation extending right the way through V1. These clusters formed what are known as "orientation-columns" and they are now thought to play an important role in the perception of form.

Superimposed upon the mosaic of the orientation-columns are ocular-dominance columns. An ocular dominance column comprises an entire suite of orientation-columns that all receive their dominant input from the same eye and which all appear to serve the same area of visual space. An ocular-dominance column is not just orientation specific, however, because within the middle of each ocularity-column is a distinct cylinder of cells called a blob that appears to be chromatically sensitive but not orientation sensitive. Two juxtaposed ocular dominance columns, each of which subserves the same area of

visual space, form what is called a hypercolumn, which I show in Figure 8.6. A single ocular-dominance column then forms a primarily monocular unit of visual space whereas a hypercolumn forms a binocular unit of visual space.

8.5 Determining Chromatic and Achromatic Contrast

Unless otherwise noted, the material in this section is excerpted from:

"*The Cortical Mechanism of Color Vision"* and *"Precortical Physiology of Color Vision"*, in Vision and Visual Dysfunction 6: The Perception of Color, (ed.) Gouras, 1991, Macmillan, London.

I described, in Chapter 7, how the single opponent signal from an individual cone contains two signals, a chromatic signal and an achromatic signal, multiplexed together. This multiplexed signal persists through the LGN and into the visual cortex where it is elegantly demultiplexed. In this section, I provide a brief summary of how this is thought to occur and I illustrate what the properties of the resultant achromatic and chromatic streams are.

On the surface, the mechanism behind how the brain demultiplexes its visual signals may seem to be irrelevant. However, this is not the case because how the brain extracts the information can tell us a considerable amount about what it is that the visual system is looking to find. One of the points that I made earlier is that the primate visual system is not a general-purpose tool; it has evolved to pick out certain things against certain backgrounds. In addition, it is here, at the very first stage of the striate cortex, that we can begin to see what those things are and how the brain goes about finding them. To that end, I detail here how this demultiplexing occurs for the signals that arise from the private-line midget ganglion cells of the fovea.

It is known that both retinal and LGN neurons respond to changes in either chromatic or achromatic contrast. Starting at the earliest layers in V1, however, this is known to change with most cells responding strongly to one or the other. Within each ocular-dominance column are groups of cells that strongly favor chromatic contrast and

cells that strongly favor achromatic contrast. The cells that favor chromatic contrast form groups known as “blobs” (Livingstone and Hubel, 1984a; Michael, 1987). These cells are easy to identify because they stain heavily for the enzyme cytochrhrome oxidase, which is one of the fundamental molecules of metabolism. All cells contain cytochrhrome oxidase and the fact that blob cells stain so heavily indicates that they are highly metabolically active. Some of the cells within the blobs do respond to achromatic contrast but the majority responds only to chromatic contrast. Most importantly, however, cells within the blobs do not show orientation selectivity.

On the other hand, cells that respond best to achromatic contrast are located outside of the blob regions, in the areas that are specifically tuned to orientation. Some chromatic cells in the border of the blob region respond to orientation but the majority of orientation specific cells appear to be achromatic. Each ocular dominance column then can be subdivided into two separate regions; (1) a primarily chromatic (blob) region; and (2) a primarily achromatic (interblob) region. There is some evidence that some of the cells in the interblob region also receive chromatic input (Gouras and Krüger, 1979) but their exact function is still unknown.

8.5.1 Determining Achromatic Contrast

Figure 8.7 illustrates a hypothetical cone mosaic representing a unit area of chromatic space in the primate fovea (a unit area of achromatic space is a single cone). In this mosaic, there are a large proportion of L & M cones. There are also a small number of S-cones but these cones do not have a separate channel and are only involved in chromatic contrast and low-resolution spatial vision.

Determining achromatic contrast turns out to be surprisingly simple because the achromatic system simply combines the signals from both the L-cones and the M-cones along predefined axis of orientation. The L-cone and M-cone ON channels are combined to produce slits, which respond best when light excites the cones along the axis of orientation. Conversely, the L-cone and M-cone OFF channels are combined to produce

bars, which respond best to a region of darkness that is likewise aligned along the axis of orientation.

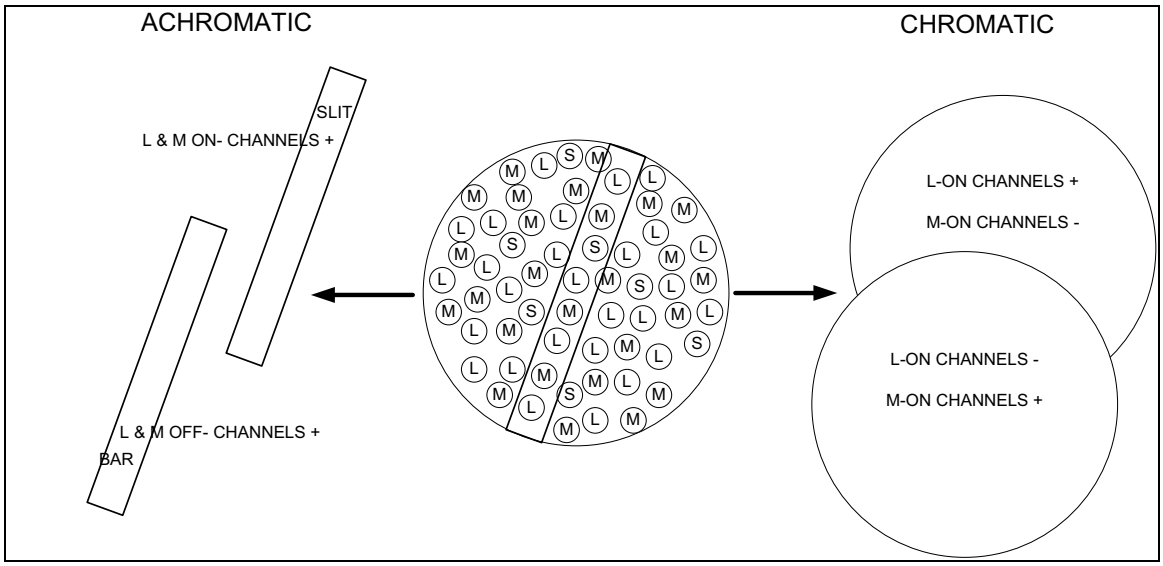


Figure 8.7: "A mosaic of foveal cones in the retina (center) representing a unit area of chromatic space. A unit of achromatic space is represented by a single cone (L or M); Single cones do not have single channels representing them in the cortex. (Left) The units of achromatic space can be organized along different angles of orientation, illustrated by insert. Each angle of orientation is represented by on-channels (slits) and corresponding off-channels (bars) reflecting on- and off-channels coming from each foveal cone. Here L and M on- (or off-) channels are synergistic. (Right) The same cone mosaic is used in parallel to build chromatic contrast detectors. In this case the L and M on- (or off-) channels are antagonistic." Image and caption adapted from Figure 11.3 in *The Perception of Color* Vol. 6 *Vision and Visual Dysfunction*, (ed.) Gouras, 1991, Macmillan, London.

Because this system combines the output from L-cones and M-cones it sacrifices chromatic contrast and so responds only to the total energy of the light. However, this sacrifice provides an offsetting benefit in that it can use the information from individual cones and thus maximize resolution. In fact, integrating an oriented array of single cones provides an even finer "hyperacuity" than by analyzing individual cones (Westheimer, 1981).

Evolutionary biology suggests that the achromatic system, which provides our ability to detect light from dark, is our original vision and now we can see how it actually works. Most importantly (and surprisingly), the achromatic system described here does not produce a single channel of information; it does not just sense luminosity. What this analysis tells us is that the achromatic system, from its earliest levels, is specifically tuned

to detect borders of energy contrast. This, in itself, is significant because as I showed in Chapter 2, our primary sensation of perception is delivered by the achromatic system; it is the system that we use to detect patterns and objects and it is, from the ground level up, primarily an edge detection system.

8.5.2 Determining Chromatic Contrast

In comparison to achromatic contrast, chromatic contrast is harder to determine and requires a larger mosaic of cones. Consequently, whereas a unit of achromatic space resolves down to a single cone, the chromatic color space is much larger as I show in Figure 8.7. To extract chromatic information, cone signals must be differenced as opposed to the combined approach that led to the achromatic contrast. In trichromats, which include all Old World monkeys, apes and man, this is done in two separate circuits. The first circuit compares the outputs of the L and M cone channels and the second compares the output of the S cone channel with the combined output of the L and M cone channels. Importantly, these circuits do not produce a direct sensation of color. That sensation comes about later. Rather, they produce streams of contrast information; the first channel detects contrast between red and green, the second between blue and yellow.

The first circuit (Figure 8.8) compares the outputs of the L and M cone channels among a group of cones thus providing a stream of information that senses red-green contrast. This circuit is interesting because as I showed previously, the signals from these cones contain both achromatic and chromatic information. To determine the contrast between the L-cones and the M-cones, this circuit is further subdivided into two; (1) a circuit that is maximally excited when the L-cones are maximally absorbing light and the M-cones minimally absorbing; and (2) a second circuit that is the reverse, i.e. it is maximally excited when the M-cones are maximally absorbing and the L-cones minimally absorbing.

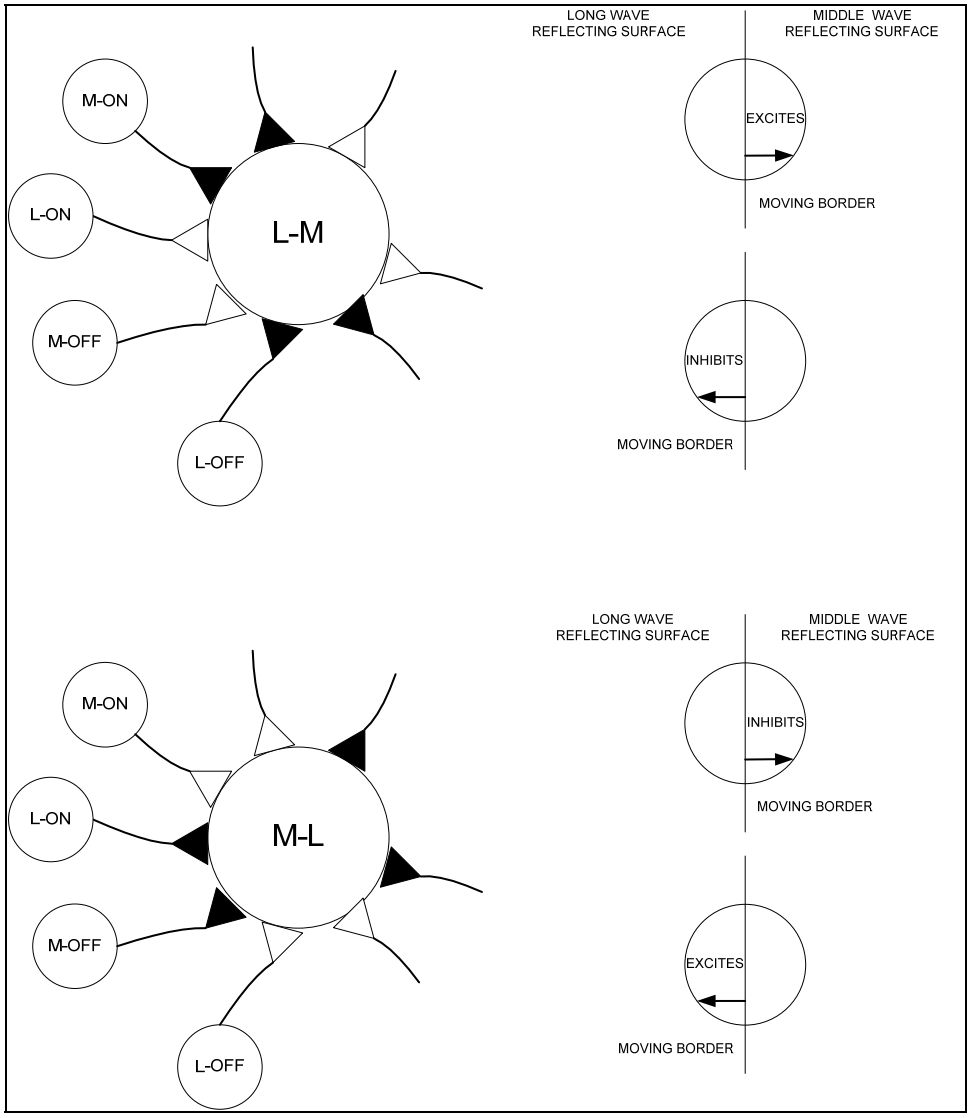


Figure 8.8: "Logical synaptic arrangements that lead to the construction of chromatic contrast detectors that sense red-green contrasts. The inputs come either directly or indirectly from cells in the parvocellular layer of the lateral geniculate nucleus. All of the inputs represent excitatory (open triangle) or inhibition (closed triangle) from a unit that receives excitatory input from one cone mechanism M or L, and an antagonistic input from the other cone mechanism (L or M). This cone channel may be either an ON- or OFF-channel. The unit (above left) is excited when a light stimulus entering its field is absorbed strongly by L cones and weakly by M cones, i.e. the difference, L-M, is maximal; it is also excited by the reverse relationship; when L-M is minimal, by a light stimulus leaving its field (an OFF response). It responds best (strongly excited) when middle wavelengths (green) are leaving its field as long wavelengths (red) are entering its field; the reverse movement inhibits it (upper right). The unit below left (M-L) shows the exact opposite behavior and has exactly the opposite synaptic input. It is excited strongly by middle wavelengths (green) entering as long wavelengths (red) are leaving its field; the reverse movement inhibits it (lower right)." Image and caption adapted from Figure 11.4 in *The Perception of Color* Vol. 6 *Vision and Visual Dysfunction*, (ed.) Gouras, 1991, Macmillan, London.

The first of these subchannels (upper right Figure 8.8) receives excitatory input from both the L-cone ON channels and the M-cone OFF channels. It also receives inhibitory inputs from the L-cone OFF channels and the M-cone ON channels. The second subchannel (lower right Figure 8.8) receives exactly the opposite inputs. It receives excitatory inputs from the M-cone ON channels and the L-cone OFF channels and inhibitory inputs from the M-cone OFF channels and the L-cone ON channels.

The determination of red-green chromatic contrast is then divided into two opposite push-pull cone opponent channels, which are similar to the push-pull ON and OFF channels that arise from the midget ganglion cells. This is not to say that they are analogous as they are different in two ways. Firstly, this system eliminates the effect of small achromatic spots upon the mosaic as a whole. Since the chromatic contrast detectors integrate over a large number of both L and M cones, a change in the response of an individual cone is all but eliminated. Secondly, in V1, the ON and OFF channels for both the L and M cones are brought together into a single cortical cell. This reduces the effect of achromatic changes upon the cell and enhances the effect of chromatic contrast. These cells respond uniquely to chromatic contrast and will respond the same whether the contrast is due to an increment or a decrement in achromatic (effective energy) contrast. In essence, this system functions as a demultiplexing system, extracting the chromatic information and discarding the achromatic.

The second stream (Figure 8.9) is common to all primates and compares the S cone response to the combined outputs of the L and M cones within the same cone mosaic as the red-green circuit; this stream senses blue-yellow contrast.

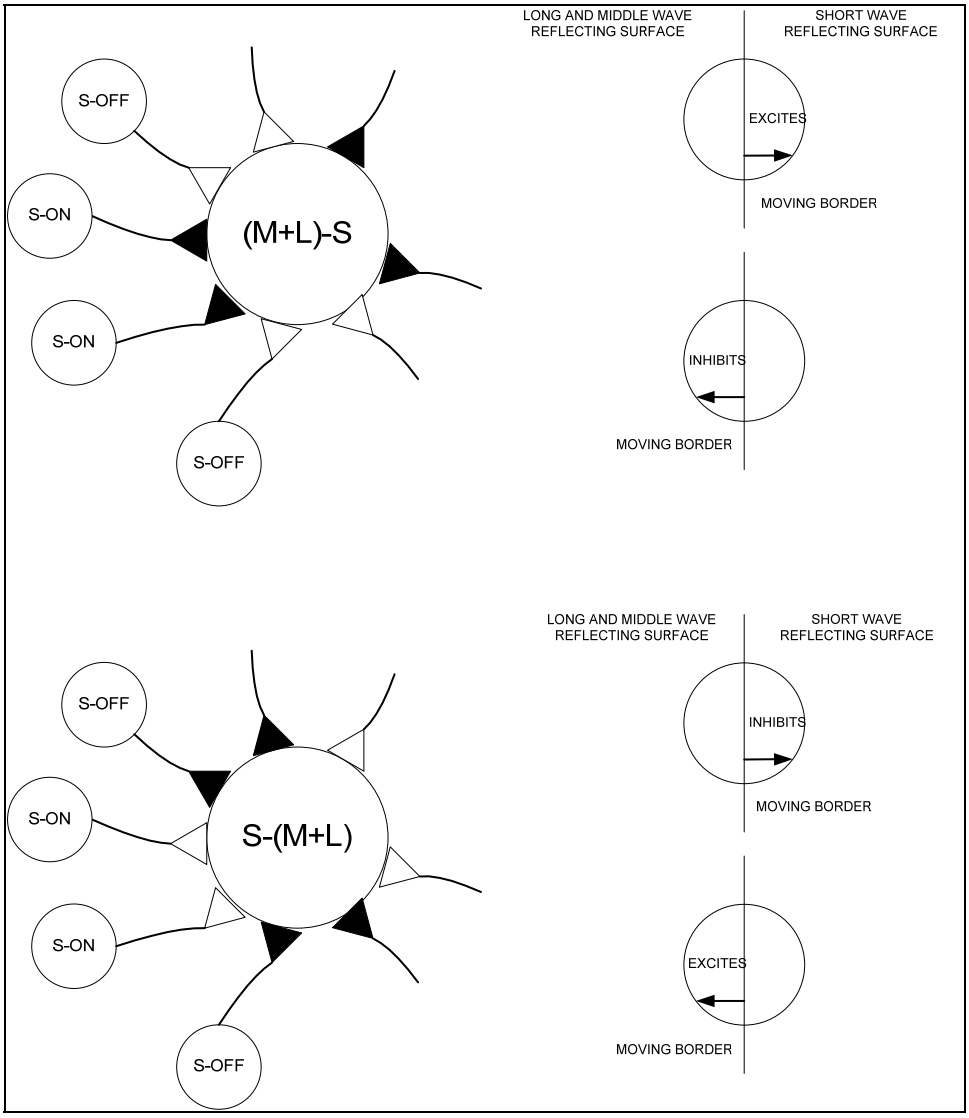


Figure 8.9: "Logical synaptic arrangements that lead to the construction of chromatic contrast detectors that sense blue-yellow contrasts. All of the inputs represent excitation (open triangle) or inhibition (closed triangles) from units that receive an excitatory input from an S-ON or an inhibitory one S-OFF from S cones and antagonistic inputs from L and M cones. The unit, above left, (M+L) – S is excited when a light stimulus entering its field is absorbed strongly by L and/or M cones and weakly by S cones, i.e. the difference (M+L) – S is maximal; it is also excited by the reverse relationship, when (M+L) – S is minimal, by a light stimulus leaving its field(off-response). It responds best (strongly excited) when long and middle wavelengths (yellow) are entering its field as short wavelengths (blue) are leaving the field; the reverse movement inhibits it (upper right). The unit below left S – (M+L) shows the exact opposite behavior and has exactly the opposite synaptic input. It is excited strongly by short wavelengths (blue) entering its field as long and middle wavelengths (yellow) are leaving its field; the reverse movement inhibits it (lower right)." Image and caption adapted from *Figure* 11.5 in *The Perception of Color* Vol. 6 *Vision and Visual Dysfunction*, (ed.) Gouras, 1991, Macmillan, London.

The blue-yellow contrast detector is also subdivided into two separate neural circuits. In both circuits the S-cone ON and OFF channels are compared to the combined signal from the L-cones and M-cones. One is excited maximally when the difference between the S – (L and M) cone absorptions is maximal and the other when the difference between the (L and M) – S is maximal.

This system of determining chromatic contrast has the disadvantage that because it integrates over a large area of visual space it looses the ability to detect fine scale changes. Because of this it posses far fewer orientation selective channels than does the achromatic system. On the other hand, it produces four chromatic contrast detectors for each unit of chromatic space. These detectors are highly sensitive to any change in the spectral characteristics of light falling upon the mosaic of cones but are relatively insensitive to a change in the total energy being received.

8.5.3 Significance to the Hering Theory of Color Vision

In Chapter 2, I introduced the Hering theory of color vision. This theory, now widely accepted, predicts that the brain forms three opponent channels of information, a black-white luminosity channel, a red-green opponent channel and a blue-yellow opponent channel. It is tempting at this point to relate the previously mentioned achromatic circuit and the four chromatic circuits directly to Hering's predicted channels. Looking further, however, it is clear that this is not the case. The opponent theory considers the *perception* of luminosity and the *perception* of color but perception does not exist at this level of neuronal organization. For example, the L-M detector (upper left Figure 8.8) responds strongly to appearance of red but it also responds just as strongly to the disappearance of green; it also responds well to the appearance of orange and the disappearance of bluish-green. The same is true of the other three chromatic circuits; they signal both the appearance of one element and the disappearance of the other.

Rather than corresponding to the Hering opponent channels the previously mentioned neural circuits correspond to the so-called cardinal directions in color space which were identified by psychophysical experiments: for example, Krauskopf et al,

(1982) showed that a sinusoidal modulation of an adaptation light reduces specifically the sensitivity along an [L-M], [S-(L+M)], and [L+M] axis. Obviously more neural processing must occur before the brain achieves its perception of color.

8.6 Simultaneous Contrast

In Chapter 2, I showed that the perception of form and color depends upon two parallel channels of information, an achromatic channel and a chromatic channel. In the preceding sections I have gone a long way towards showing how these channels are established in the brain and what some of their properties are. For example, in 8.5.1 I showed that the achromatic system is primarily an edge detection system because the first cells in the primary cortex that respond primarily to achromatic changes are the orientation specific cells of the ocular dominance columns. I also showed in 8.5.2 how the chromatic system originates with four separate channels of chromatic contrast, two red-green specific and two blue-yellow specific. What I have not discussed is the process behind how these four channels are used to form our perception of color. This process is called simultaneous contrast and I discuss it in this section.

It is well established that the perception of color depends upon contrast and that our perception of a given color changes depending upon its surround. For example, consider the sentence below:

The text color of the first two words is the same, - bright red. The backgrounds however are different, specifically yellow and violet. Although the text color for the two words is identical, a person with normal color vision will perceive it differently for each word. The text color appears brighter, more vibrant against the yellow background, and more orange against the violet background. This same phenomenon is further illustrated by the remaining two words whose text color is grayish-red and whose backgrounds are

blue and dark grey respectively. Again, the text color of the two words is perceived differently despite the fact that it is the same for both.

DaVinci was the first to be credited with the observation that adjacent colors influence each other but this observation was not quantified until the work of Chevreul in the 19th century. Chevreul was one of the most important chemists of the 19th century (among other things he was credited with discovering margarine). However, it was during his time as the director of the Gobelins Tapestry Works in Paris that he made his contribution to both art and science because it was during this period the he discovered the basic principles of simultaneous contrast.

Figure 8.10 : Chevreul's 72 part color circle.

Chevreul discovered that certain dyes, when placed next to each other in a carpet, failed to achieve the desired effect. In particular, he noticed that black yarn appeared differently when used next to blues. He had little interest in the artistic treatment of color; instead, he wanted to develop a systematic way to quantify the effect of one color upon another. To this end, he devised a 72-part color circle (Figure 8.10) whose radii, in addition to the three primaries of red, yellow and blue, depict three secondary colors of orange, green and violet. The resultant sectors were each subdivided into five zones and all radii were separated into 20 segments to accommodate the different brightness levels. Chevreul's work, though incomplete, led to his famous law of simultaneous contrast: "Two adjacent colors, when seen by the eye, will appear as dissimilar as possible". Although Chevreul eventually failed in mathematically quantifying the relationship between colors, his was the first attempt that scientifically confronted the active role that the brain has in the formation of colors.

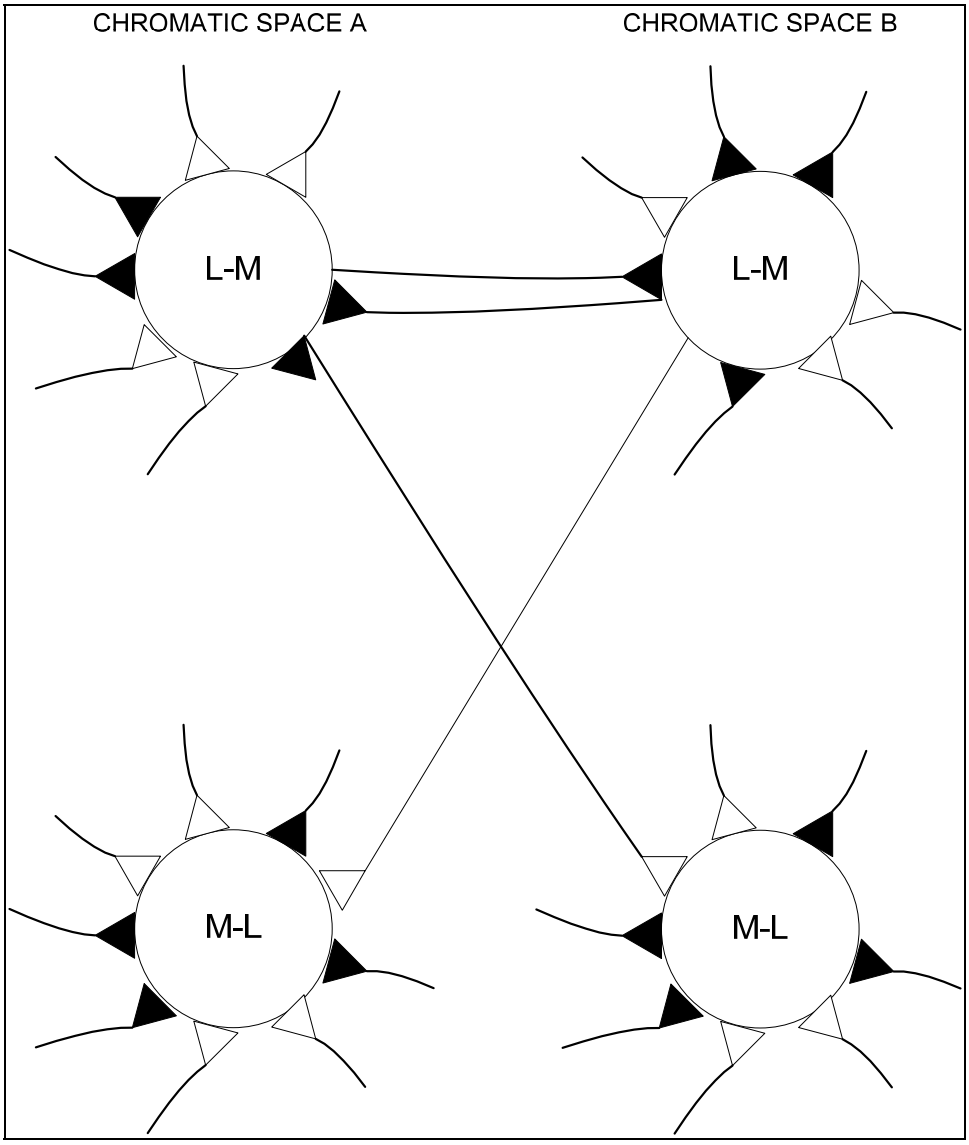


Figure 8.11: "Synaptic interactions between the chromatic contrast detectors of Figure 8.8 which lead to simultaneous color contrast. Two L-M chromatic contrast detectors in neighboring areas of chromatic space are shown to inhibit (closed triangles) each other and excite (open triangles) the opposite type (M-L) of chromatic contrast detectors in neighboring units of chromatic space. If the red side were on the left and the green on the right, the upper left L-M units and the M-L unit would be strongly excited; the reverse border would excite the opposite pair f chromatic contrast detectors." Image and caption adapted from Figure 11.6 in *The Perception of Color* Vol. 6 *Vision and Visual Dysfunction*, (ed.) Gouras, 1991, Macmillan, London.

During Chevreul's time very little was know about the physiology of vision. Today we know far more about it but even so, the exact mechanism for how simultaneous contrast is established in the brain is not completely understood. We know that the first step in producing our sensation of color is the development of the four channels of color

contrast but how exactly these individual signals are combined together is still not completely understood. Figure 8.11, however, does show one possible logical combination that would lead to simultaneous contrast.

Simultaneous color contrast cannot occur until after separate neural circuits in the visual cortex have extracted chromatic and achromatic contrast, this much is certain. Once this has been established then simultaneous contrast could be established by having a chromatic contrast detector subserving one area of chromatic space excite a chromatic detector of the opposite type and/or inhibit a chromatic detector of the same type in neighboring areas of chromatic space. In this manner, a border separating areas of long-wavelength (red) and middle-wavelength (green) would be enhanced. In this system, the perception of the redness of a red surface would be enhanced against a green background. This effect is experienced subjectively and is routinely reported by subjects with normal color vision.

The neurons, which lead to simultaneous contrast, have been called double-opponent neurons. They were initially reported in the retina of goldfish by Daw in 1968 but in primates, they have only been discovered in the visual cortex. This leads to the suggestion that in primates the formation of color only begins deep in the striate cortex.

Although the existence of the neuronal interactions shown in Figure 8.11 have not been proven conclusively, they are logical and they could lead to the development of the psychophysically established simultaneous contrast. Neurons exhibiting this behavior have been found in the visual cortex (Hubel and Weisel, 1968***;*** Michael, 1987a, b, c) and especially within the blob areas of V1 (Livingston and Hubel, 1984). These neurons are called double-opponent neurons and it is believed that once they have been established that orientation selective chromatic contrast detectors that are also double opponent and thus detect simultaneous contrast can also be constructed. Neurons of this type have been reported by (Hubel and Wiesel, 1968; Dow, 1974; and Gouras 1974).

8.7 The Determination of Form

In the previous sections, I discussed the physiological processes by which the visual cortex begins its task of reconstructing the world around us. In my discussions, I have concentrated on the private-line signals that traverse the optic pathways and carry high-resolution visual information directly from individual cones in the fovea straight into the primary visual cortex. These signals contain both a chromatic and an achromatic stream of information multiplexed together. The first task of V1 is to demultiplex this single stream into parallel streams of chromatic and achromatic contrast. In this final section on the visual system, I provide a simple psychophysical example of how these two streams of contrast are combined to determine our sensation of form.

As I showed in 8.5, most chromatic cells in V1 are non-orientation specific whereas most achromatic cells are orientation specific. Most visual scientists believe that orientation selectivity plays a large role in the perception of form (Marr, 1982). Consequently, in the V1 at least, most of form recognition must occur via the achromatic cells because they drive the majority of the orientation specific cells. However, this is not the complete story. There is also a small, but significant, number of cells in the blob regions of V1 that are both chromatic selective and orientation selective. This indicates there is a second method for determining form, one that is based upon purely chromatic information. This has been observed and reported upon psychophysically by Switkes et al, 1988, who showed the existence of channels in human vision that are selective for both color and orientation.

To illustrate these two mechanisms I use Figure 8.12, which shows a deliberately low-resolution (100 pixels per inch) image of a red ball. I produced this image by multiplying together a purely chromatic red circle (Figure 8.12 center) with a purely achromatic white ball (Figure 8.12 right). At first glance, these images seem to contradict theory. In theory, the achromatic form detection system is higher resolution than the chromatic form detection system but the edges of the red circle are smooth whereas the edges of the white ball are jagged. This implies that, contrary to theory, the chromatic system has the higher-resolution.

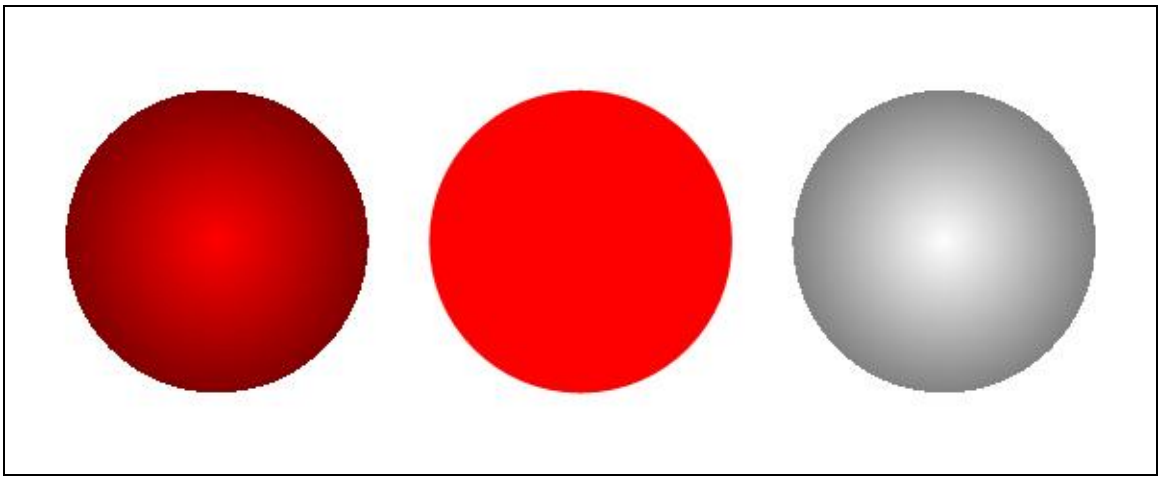


Figure 8.12: An image of a red ball (left) split into its chromatic (middle) and achromatic (right) components. The native resolution of this image is 100 pixels per inch. Because the image is low-resolution, the border of both the chromatic and achromatic balls should be pixilated. The border of the achromatic image is but the border of the chromatic image is smooth. This shows that the chromatic circuitry for detecting form is lower resolution because it cannot detect the pixilation. The red ball on the left also shows the pixilation indicating that the higher resolution form determined by the achromatic circuitry dominates and somehow overrides the chromatic.

However, this is not the case. These images all have the same resolution and the edges of all of the images are pixilated. The jagged appearance of the edges of the white ball is a true representation of the image but it can only be detected by the high-resolution achromatic form detection system. The low-resolution chromatic system cannot see the fine detail even though it is there and so the edges of the red circle appear smooth.

In the real world, when an object such as the red ball in Figure 8.12 (left) is viewed, the striate cortex produces two streams of visual information that are processed for form by separate neural circuits. The first is the achromatic image (right Figure 8.12) and the second the chromatic image (middle Figure 8.12). Of these two, the achromatic image is the sharper because a higher number of orientation-selective cells, each of which has a very small receptive field, detect it. The purely chromatic image appears fuzzier with less well-defined edges because only the low-resolution orientation-selective chromatic contrast detectors detect it.

In the case of purely achromatic contrast, a sharply defined but achromatic object is perceived; in the case of purely chromatic contrast, a blurry but colored object is perceived. What happens, however, when both types of contrast are present? In that case,

it is believed that they suppress each other; the chromatic detecting system suppresses the achromatic color and the sharply defined achromatic border suppresses the fuzzy chromatic border. There is significant physiological evidence that luminance can suppress chromatic contrast detectors (Gouras and Kruger, 1979); and there is significant psychophysical evidence that chromatic contrast can suppress luminance contrast (Switkes et al, 1988). This seems to be the way that the higher brain functions to combine both form and color into a single integrated perception.

8.8 Summary

This brings me to the end of my study on the primate visual system. At the beginning of this study, I asked myself a simple question, how do we go about determining the form and the color of an object? I had to ask that question because very early in my research, I realized that the success of visualization was very much hit and miss. There was a lack of established principles that indicated how we should communicate complex scientific information. The only principles of visualization that we understood were artistic in nature and they were inappropriate for what I was trying to do.

What I wanted to do was establish a simple set of principles that I could follow and to which I could refer back. Visualization is the science of communicating information through the visual system. I wanted to know the rules I had to follow to do it properly. In the end, I partially succeeded and I partially failed. Most of the failure comes from the simple fact that I underestimated the task. Earlier, I described the visual system as being majestic. Having gone through it from one end to the other I now have a long list of adjectives that describe it, majestic is still one of them but it is not high on the list!

The visual system is breathtaking and stupendous and I can only sit in awe of the minds that unraveled what little we know about it. Newton once said that he stood on the shoulders of giants, I had to use a rocket pack and even then, I only caught vague glimpses of what they saw. I am truly and honestly humbled by some of the things that I have studied.

Even though I did not find everything that I set out to search for, I consider that my study was a partial success and certainly worth the effort. This is because, at the very end, I started to see a hint of those principles that I set out to find. Consider this, despite the fact that wiggle trace displays have the lowest resolution of any seismic display; they are still the most heavily used conventional display. Now we know why. Humans are primates and all primates are comfortable in the trees. Ask yourself what a tree-dwelling animal needs to see the best; the answer is that it needs to see branches. It needs to see lines and edges; it needs to distinguish between what is safe and what is deadly, and it needs to do it very quickly and without conscious effort.

The very basis of our form detection system has evolved with this as its primary driving factor. From its very lowest levels, our ability to detect form is based on detecting lines and edges. Knowing this, we have the answer to wiggle trace displays – they are all lines and edges. I do not know of any research that supports this but it is interesting to speculate that the reason why so many of us feel so comfortable when we view them is that they invoke primordial feelings of home.

Understanding of the importance of achromatic lines and edges is the first principle of visualization and discovering its importance was, by itself, worth the study. In Chapter 2, I showed the importance of achromatic information using psychophysical techniques. Now, we see that there is a physiological reason behind it.

The second part to this study was to understand the role that color plays in perception. One of the principal conventional seismic displays is the chromatic variable density display and we now see that we use it for inappropriate purposes. If we use it as our principal seismic display then we are, in essence, using it to communicate seismic form. However, clearly, the communication of form was not the driving factor in the evolution of our color vision. We do not establish perception through color; therefore, it must serve other purposes.

If we do not use color to determine form then what do we use it for? Undoubtedly, color serves an important purpose for primates because we, alone among mammals, developed Trichromacy. It was obviously important to our ancestors to distinguish colors at the green-red end of the spectrum but why did they need to and more importantly, how can we use it. The quick answer is that we developed Trichromacy to pick red-yellow fruits out of the green forest canopy. The quick answer to how we can use it is that reds and yellows attract our attention whereas greens do not.

I will leave this particular subject here because I go into it in more detail in Chapter 11. I will end, however, with an observation that our perception of color is subjective on two levels. The first is the one we all know about, that everyone's particular taste in colors is different. The second is that because of simultaneous contrast, our perception of color is physiologically subjective but possibly universal.

Part 3: Practice

The Form and Color of SeisScape Displays

When Earth's last picture is painted and the tubes are twisted and dried,
When the oldest colours have faded, and the youngest critic has died,
We shall rest, and, faith we shall need it – lie down for an Æon or two,
Till the Master of All Good Workmen shall put us to work anew.

And those that were good shall be happy: they shall sit in a golden chair;
They shall splash at a ten-league canvas with brushes of comets' hair.
They shall find real saints to draw from – Magdalene, Peter, and Paul;
They shall work for an age at a sitting and never be tired at all.

And only the Master shall praise us, and only the Master shall blame;
And no one shall work for money, and no one shall work for fame,
But each for the joy of the working, and each, in his separate star,
Shall draw the Thing as he sees It for the God of Things as they are!

"When Earth's Last Picture is Painted"
Kipling

CHAPTER NINE: TESSELLATING SEISMIC DATA

They builded a tower to shiver the sky and wrench the stars apart,
Till the Devil grunted behind the bricks: "It's striking, but is it Art ?"
The stone was dropped at the quarry-side and the idle derrick swung,
While each man talked of the aims of Art, and each in an alien tongue.
Kipling
The Conundrum of the Workshops

9.1 The Structure of a Seismic Section

A seismic section is a complex mosaic of overlapping and often conflicting signals, some of which are geologically or seismically relevant and some of which are noise. Of the relevant signals, some have strong amplitudes and we can see them on all seismic displays. Some, however, have weak amplitudes and are superimposed on the stronger events. As I showed in Chapter 4, these weaker events are very hard to see on conventional displays. As a rule, however, the importance of a coherent event to an interpretation is not directly proportional to its amplitude. An interpretation often depends just as much on weak, hard-to-see events as it does on strong, visually dominating events.

For the purposes of the rest of this thesis, I consider that these two levels of events form different structures within the seismic as a whole. The strong amplitude, major events, forms the seismic macrostructure whereas the weak amplitude events form the seismic microstructure.

I established in Section 4.2.1 that the wiggle trace display has very low apparent resolution. I also showed, however, that because it is constructed purely of achromatic lines and edges, it is useful for pattern recognition and that, unlike gray-scale displays; it has a future in seismic visualization. Wiggle trace displays are prominent throughout the industry and geophysicists will continue to use them in the future, albeit for reduced purposes. Because they are so familiar and because they show primarily major events, I use them as the base for my definitions of macrostructure and microstructure.

9.1.1 Seismic Macrostructure

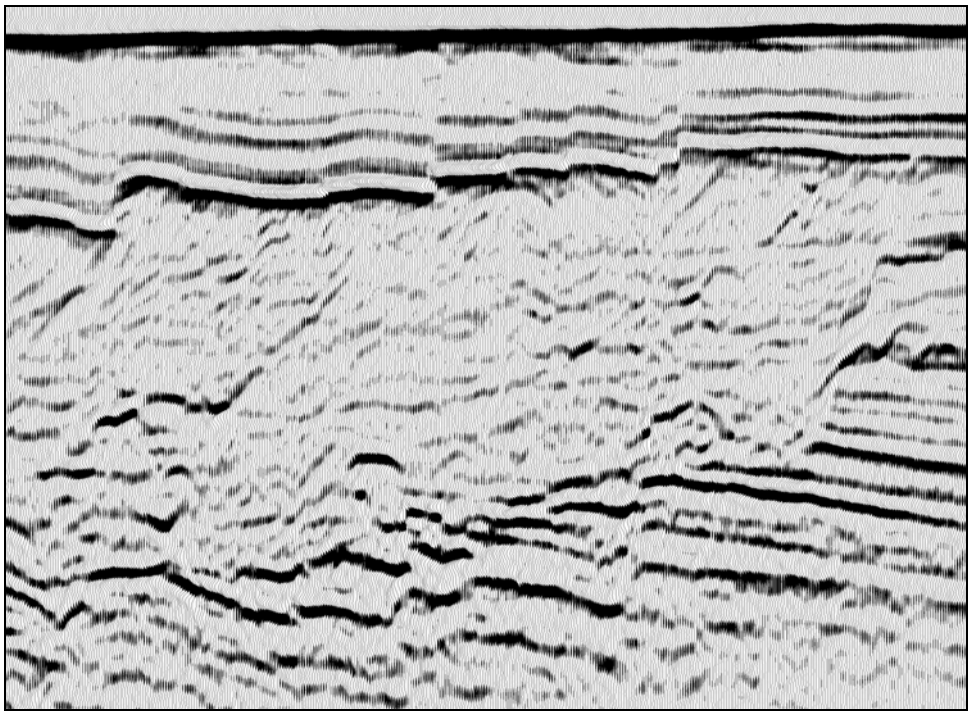


Figure 9.1: A wiggle trace display of Trujillo data (data courtesy PeruPetro). The section contains a series of prominent events that constitute the macrostructure.

Figure 9.1 shows a wiggle trace display of a portion of a Trujillo seismic line. When we look at this display, we see a series of strong amplitude, major events, which appear as almost solid black objects. Regardless of the display used, you expect to see these events, and you expect to see how they relate to one another. I consider these events constitute the seismic macrostructure, which I define as follows:

Seismic Macrostructure

For any seismic section, the seismic macrostructure is the collection of coherent signals observable on a wiggle trace display. In terms of absolute and apparent resolution, the seismic macrostructure equates to the apparent resolution of a wiggle trace display.

9.1.2 Seismic Microstructure

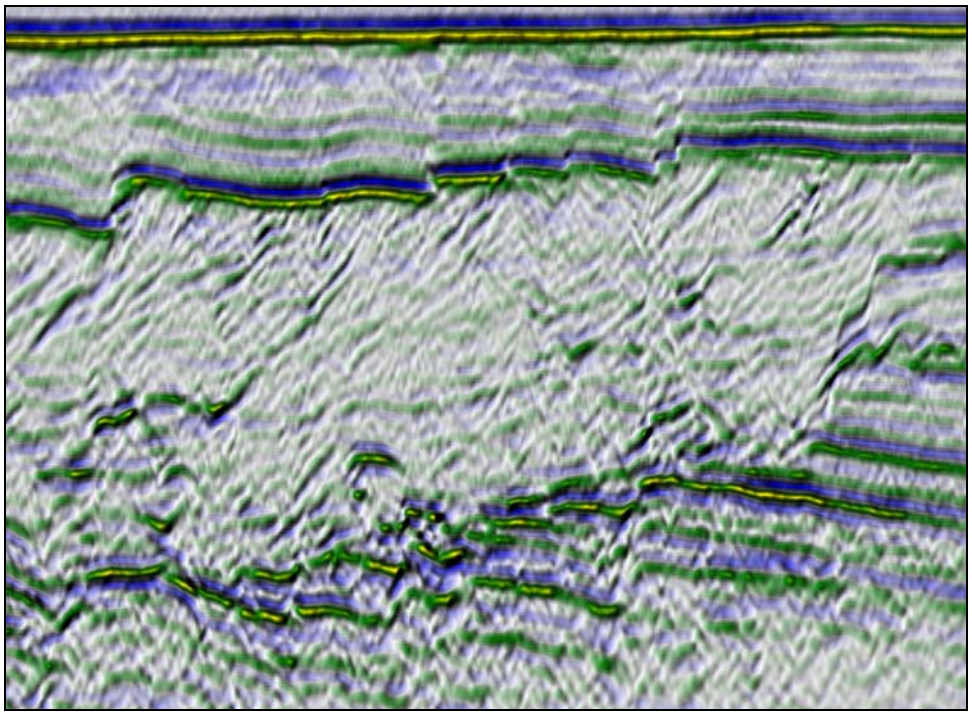


Figure 9.2: A SeisScape display of the same data shown in Figure 9.1. The apparent height of the seismic relief is set to zero. The central portion of this image contains a series of coherent signals that we do not perceive on the wiggle trace display. These subtle signals constitute the microstructure of the section.

By contrast, Figure 9.2 is a top down SeisScape display of the same data shown in Figure 9.1. I set the height of the seismic relief to zero so the display is flat like the wiggle trace display previously shown. All throughout the display and especially in the low amplitude area between the upper and lower faulted major events, you can see an entire level-of-detail of new information, the majority of which is low amplitude. The new information revealed (or not filtered out) by this display constitutes part of the seismic microstructure for this section which, for the purposes of this thesis, I define as:

Seismic Microstructure

For any seismic display, the microstructure is the difference between the absolute resolution of the data and the apparent resolution of a wiggle trace display.

With that definition in mind, Figure 9.2 reveals only part of the microstructure because, whereas it exposes more of the absolute resolution, it does not necessarily expose it all.

9.1.3 The Twin Objectives of Seismic Visualization

I chose to define the terms macrostructure and microstructure because the remainder of this thesis is dedicated to developing techniques to improve seismic visualization. These techniques generally fall into one of two classes depending upon which type of structure they are designed to enhance. Some techniques enhance seismic macrostructure whereas some enhance seismic microstructure.

In Chapter 3, I introduced the concept of considering the display as a resolution filter and showed that, in general, this filter degrades resolution. Recalling Figure 3-12 and Figure 4-24, the images all show how much lower the apparent resolution of conventional displays is in comparison to SeisScape displays. In terms of the previous definitions, these examples showed how conventional displays filter out seismic microstructure. This leads to the primary and most obvious purpose of visualization, that being:

Objective #1

The primary purpose of seismic visualization is to reveal seismic microstructure. In terms of absolute and apparent resolution, this equates to minimizing the difference between the two.

Beyond a consideration of apparent resolution, which mainly applies to microstructure, the display also affects our ability to perceive amplitude changes along macrostructure events. Consider Figure 9.3, which is another SeisScape view of the data shown in Figure 9.1. Unlike Figure 9.2, this display has a significant relief height and I rotated it counter-clockwise around the x-axis so that you can see amplitude changes along the macrostructure events.

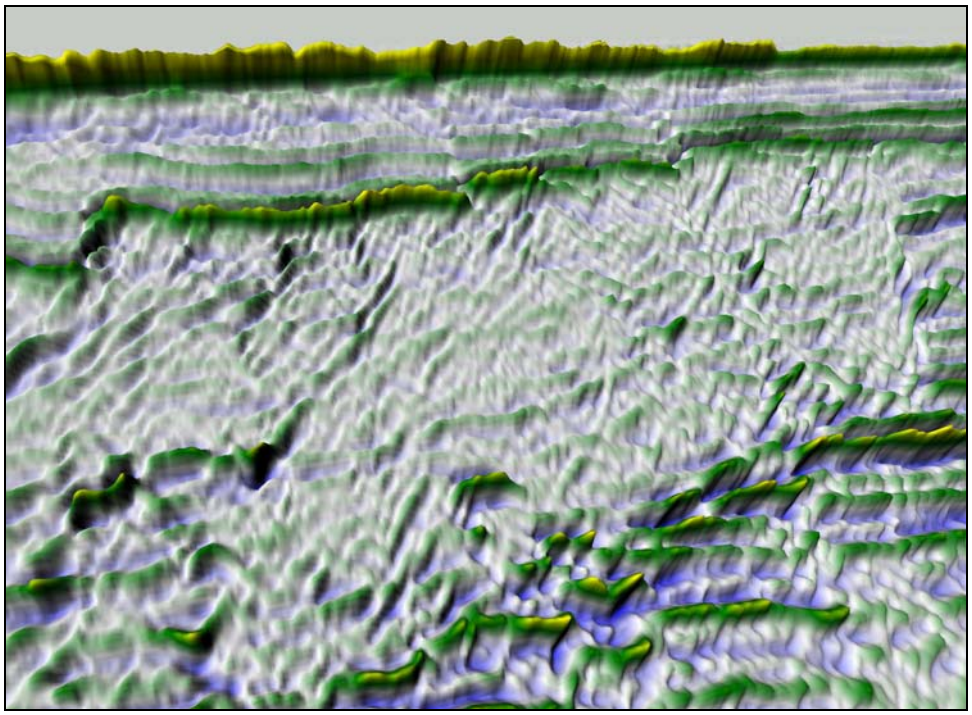


Figure 9.3: The same display as shown in Figure 9.2 but rotated counter clockwise around the x-axis and with a non-zero relief height. In this orientation, you can see amplitude changes along the macrostructure events.

The uppermost event in this section is a water bottom reflection. When you examine this event on either the wiggle trace display or the top-down SeisScape display, the amplitudes along it appear almost constant. Figure 9.3, however, reveals that there is significant variation in the amplitudes along the event, variations that you would not expect from looking at the first two images.

You see the same sort of short period and long period amplitude variations along the other macrostructure events as well. Consider the first set of faulted events below the water bottom. If all you had to go on were Figure 9.1 and Figure 9.2, you would not expect the trace-to-trace amplitude variation exposed by Figure 9.3. It is clear that neither the wiggle trace display nor the top-down SeisScape display adequately communicate the amplitude structure of the macrostructure events. This example illustrates that the display acts as a filter upon the macrostructure just as it does upon the microstructure and it introduces the second, less obvious, purpose of visualization.

Objective #2

The secondary purpose of visualization is to reveal the amplitude structure of macrostructure events.

Considering SeisScape displays are three-dimensional and show amplitude as topography, this secondary purpose appears satisfied by default. However, as with all things seismic, if it looks easy then you do not understand it! Revealing the amplitudes of macrostructure events is trivial, provided the events are flat. Once they start to dip, however, things become a little more complicated.

9.2 The Seismic Mesh

A SeisScape display is a three-dimensional representation of seismic data and is composed of three elements; (1) a tessellated[1] mesh of points that form the mosaic of the surface; (2) a lighting component that illuminates the surface; and (3) a variable density color display that is draped over the surface. Each of these components has analogies in the conventional displays. The tessellated mesh is loosely analogous to the wiggle trace display, the lighting is analogous to the amplitude mapped gray-scale display and the variable density color display is identical to chromatic variable density displays.

Each component of the SeisScape display plays its own part in establishing our perceptions of seismic data. I discuss the first of these components, the tessellated mesh, in this chapter. I discuss the lighting component and the variable density coloring in the next two chapters respectively. Tessellation, affects our ability to perceive both the seismic macrostructure and through its inter-relationship with the lighting calculations, our ability to perceive the seismic microstructure. In this chapter, I focus primarily on the first of these, the viewer's ability to perceive amplitude variations along macrostructure events.

---

[1] In computer graphics tessellation refers to the process of converting a complex polygonal surface into a series of non-overlapping triangles.

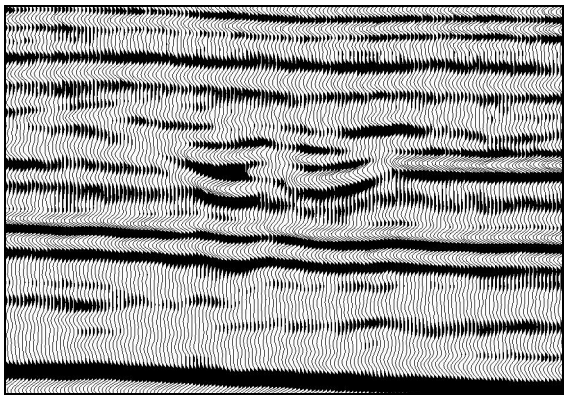


Figure 9.4: Wiggle trace display of a small channel. Data is shown at an expanded scale of 12 tpi and 30 ips to match the SeisScape display.

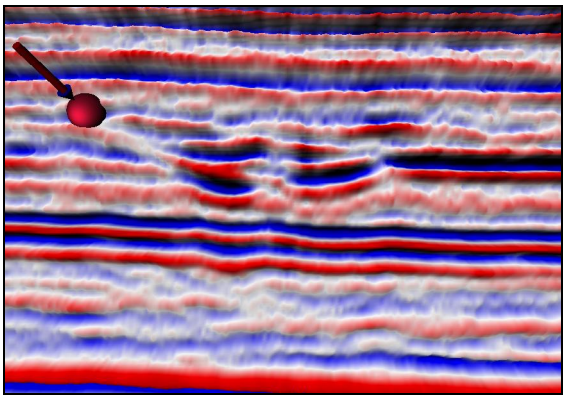


Figure 9.5: SeisScape display of the same channel shown in Figure 9.4. The direction of lighting is indicated by the arrow.

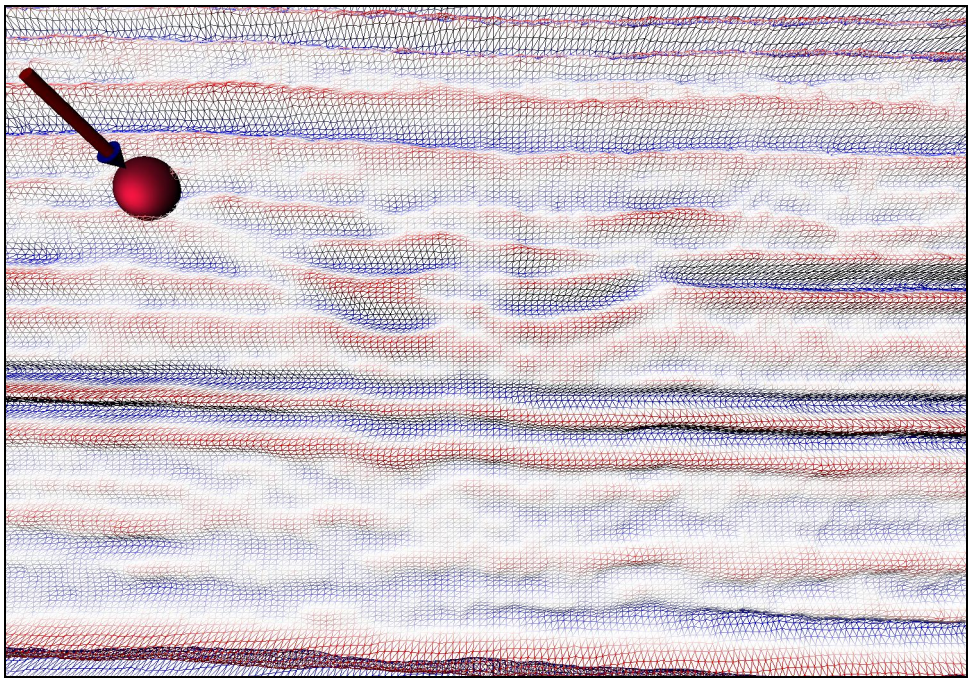


Figure 9.6: A wireframe image of the same data shown in Figure 9.4. The mesh is a series of triangles, each triangle connecting two samples on one trace to an adjacent sample on the next trace. Each quadrangle of four seismic samples produces two triangles.

SeisScape displays are three-dimensional meshes. Each vertex in the mesh has three coordinates; (1) the x coordinate that represents the spatial position of the trace; (2) the y coordinate that represents the time of the sample; and (3) the z coordinate or elevation that represents the seismic amplitude. In a spatial sense, a SeisScape display is generally planer. It is possible to take into account bends in the geometry of the line by supplying both an x and a y spatial coordinate.

However, in practice the three-dimensional nature of the display becomes confused when the real world bent-line coordinates of the line are used. Consequently, in what follows, all of the SeisScape displays shown use straight-line geometry. I illustrate the mesh structure of a SeisScape in wireframe mode in Figure 9.6. For reference, I show a wiggle trace display and a solid SeisScape display in Figure 9.4 and Figure 9.5 respectively.

9.2.1 Tessellation

As you see in Figure 9.6, the SeisScape mesh consists of triangles. Triangles are the basic unit of all 3D graphic objects and they are the building blocks or bricks of 3D graphics. Using an analogy from construction, you can build almost any structure from small bricks. However, when you look at the structure from a distance, you do not see the bricks themselves, you only see the structure. The same is true of triangles, regardless of the complexity of a surface, in 3D graphics; an object is always built out of triangles.

In theory, you could construct three-dimensional surfaces from higher order polygons such as rectangles which, given the regular geometry of a seismic line, look like a better alternative. Consider Figure 9.7, which shows three adjacent samples from two adjacent traces. From a casual perspective, it initially looks like the simplest way to build a surface is to form rectangles such as (1, 1), (1, 2), (2, 2), (2, 1).

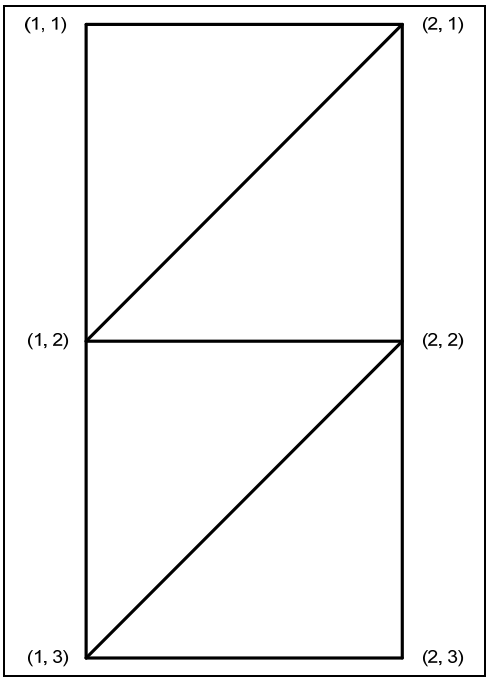


Figure 9.7: An illustration of how the SeisScape mesh may be tessellated.

In computer graphics, however, the software needs to know whether a given surface is facing towards the viewer or away from them. If it knows this, then it can simplify and speed up the rendering. This is the primary reason why we build all 3D surfaces with triangles;

triangles are planar and consequently the software can always determine which way they face. Conversely, a rectangle, especially one formed from seismic data, is rarely planer and therefore can face both towards the viewer and away from them at the same time. That is why we rarely use rectangles and other higher-order primitives when building three-dimensional objects and why we always tessellate seismic data.

9.2.1.1 *Tessellation Ambiguity*

Given seismic data occurs on a regular grid, it at first appears that constructing a tessellated seismic mesh is as simple as constructing either a wiggle trace or a variable density display. This is not the case because both wiggle trace and variable density displays are unambiguous whereas a SeisScape display is not. When you build either a variable density display or a wiggle trace display, you do not physically connect points on adjacent traces. Therefore, there is only one way to build them and the displays are unambiguous. However, when you construct a SeisScape display, you physically connect points on adjacent traces and there is always two ways to do the connections. Each connection produces a different surface and consequently, a SeisScape mesh is ambiguous at every sample. For each given sample, there are always two ways to connect it to its neighbors, each approach producing a locally different surface.

When you tessellate a seismic section, you form two triangles for each sample on every trace. You do this one of two ways, as shown by the left and right images on Figure 9.8. When I formed meshes in the early SeisScape displays, I tessellated the points using the schema shown on the left. Under this schema , if the sample position is (1,1), i.e. trace 1 sample 1, then the coordinates of the first triangle are (1,1), (2,1), (1,2) and the second (1,2), (2,1), (2,2). However, I could also have used the second schema shown on the right. Under this schema , the coordinates of the triangles become (1,1), (1,2), (2,2) and (1,1), (2,1), (2,2).

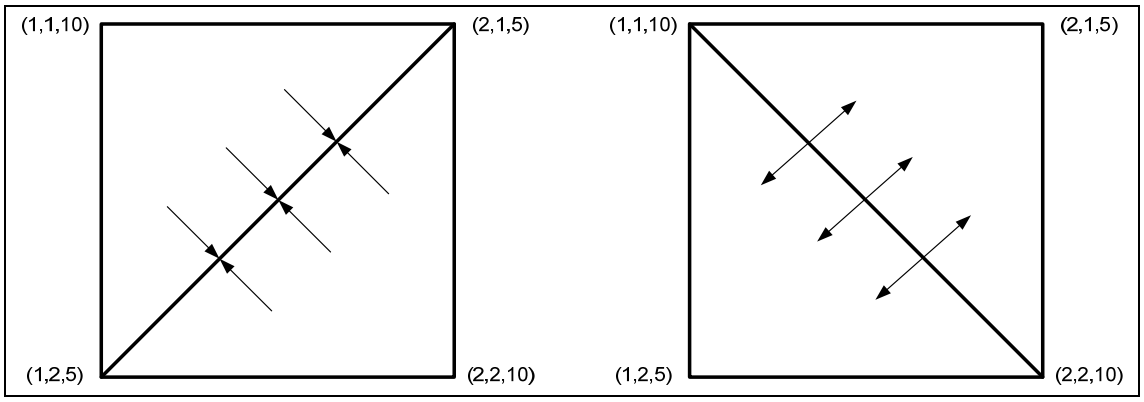


Figure 9.8: The tessellation of seismic data is always ambiguous. The tessellation in the left image above produces a trough whereas the tessellation in the right image produces a ridge.

Consider the amplitudes for the samples. If you tessellate the data using the above left schema, the result is a trough like feature whereas if you use the schema on the right, the result is a ridge like feature. Clearly, tessellation has a preferred dip alignment. If a seismic event has a positive dip then you should tessellate it using Figure 9.8 left (hereafter referred to as up-dip tessellation). If an event has a negative dip then you should tessellate it with Figure 9.8 right (hereafter referred to as down-dip tessellation).

I graphically illustrate the effect of this ambiguity with Figure 9.9 and Figure 9.10, which show the flanks of a salt dome, tessellated with the up-dip schema. The display is oriented left-to-right so Figure 9.9 shows the down-dip flank of the dome whereas Figure 9.10 shows the up-dip flank. Looking at the up-dip flank, the events are smooth and continuous and there is no visible evidence of the tessellation. However, because the tessellation cuts across the ridge of down-dip events, the events on the down-dip flank have a saw-tooth appearance. This saw-tooth appearance is an artifact of the tessellation and one that may occur whenever the amplitude difference of the four samples involved in the tessellation is high. For high amplitude dipping events, it is always necessary to tessellate along the apex of the events rather than along them. Otherwise, as is clearly indicated by Figure 9.9 and Figure 9.10, the tessellation will significantly degrade the ability to perceive the amplitudes of macrostructure events.

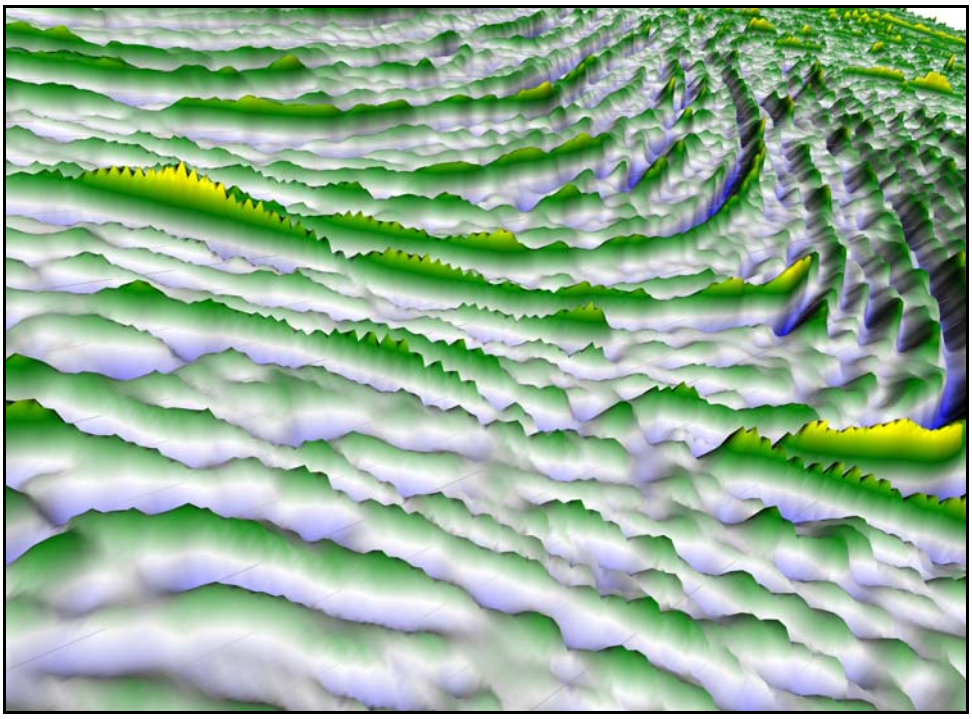


Figure 9.9: Down-dipping flank of a salt dome (data courtesy unnamed source) tessellated with the up-dip schema. The up-dip tessellation connects points across the apex of the events producing a saw-tooth effect.

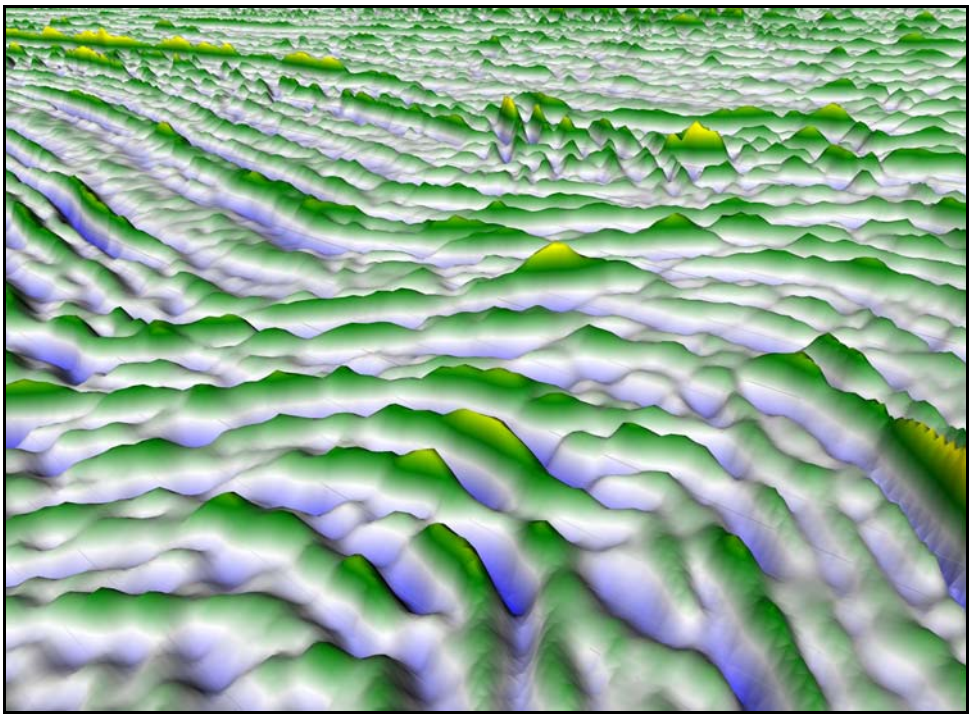


Figure 9.10: The up-dip flank of the same salt dome shown in Figure 9.9, again tessellated with the up-dip schema. The schema “prefers” up-dip events and consequently the events are more continuous.

On quick observation, the saw-tooth artifacts visible on Figure 9.9 are reminiscent of the spatial aliasing that is often observed on wiggle trace displays. On closer inspection, however, it is clear that saw-tooth artifacts are not related to spatial aliasing at all. Spatial aliasing is a significant problem for migration and will occur when the following condition is true:

$$\delta x > \frac{V}{4 f \sin\theta} \qquad 9.1$$

Where δx is the trace spacing, V is the seismic velocity of the medium, f is the frequency and θ is the angle that the event makes to the surface (Lines and Newrick, 2004). Saw-tooth artifacts visually imply that the data is not sufficiently sampled in the x-direction. The artifacts, however, are simply the result of tessellating across a dipping event rather than along it and do not directly depend upon the spatial sampling interval. Regardless of the dip of an event, if it is correctly tessellated then there will be no visible effect of any spatial aliasing.

The primary focus of this chapter is to study the effect that tessellation has upon the macrostructure of a section. Tessellation, however, defines the seismic surface. Consequently it also has a pronounced effect upon lighting, which is the primary tool used for the detection of microstructure. Lighting and the detection of microstructure is the subject of the next chapter. However, as further illustration of the importance of tessellation, in Figure 9.11 and Figure 9.12 I show an example of the effect of tessellation upon microstructure. These two images show the up-dipping edge of a small channel, I tessellate Figure 9.11 with the up-dipping schema whereas I tessellate Figure 9.12 with the down-dipping schema. In all other respects, including the direction of lighting, the two images are identical.

Immediately to the left of #1 in Figure 9.11 is an area of obvious seismic microstructure. Exactly what this microstructure represents and whether or not it is geologically significant, is not important. What is important is that you cannot see this microstructure on Figure 9.12. In the latter case, with the exception of the events at the

upper left of the image, using the incorrect schema has not significantly degraded the macrostructure events. However, this down-dipping schema all but erases the low amplitude, high angle microstructure events.

Saw-tooth or diamond pattern artifacts are indicative of a larger problem inherent in tessellating an arbitrary data set; before you can correctly tessellate a series of points, apriori knowledge of the surface is essential. Rendering a model in three-dimensions requires two sets of data; (1) a set of vertices that define the points in the model; and (2) a set of indices that define the vertices for each triangle.

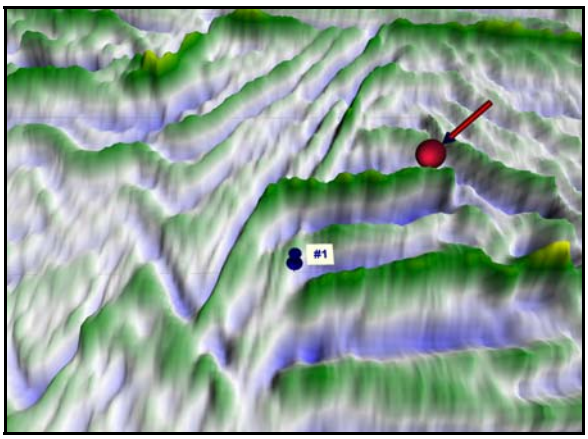


Figure 9.11: The up-dipping edge of a channel (data courtesy PeruPetro) tessellated using the up-dip schema. Note the presence of microstructure to the left of #1 and above.

Whereas the vertices define the general outline of the model, it is the indices that give it shape. Vertices are just points in space; indices form surfaces out of those points. Ultimately, tessellation is the process of determining what those indices should be.

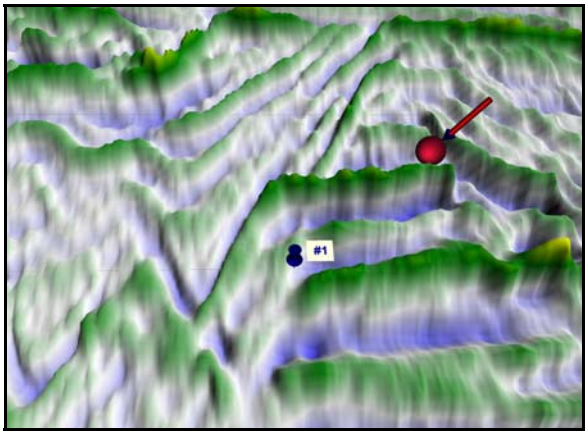


Figure 9.12: The same data as shown in Figure 9.11 but tessellated using the down-dipping schema. Note that the appearance of the up-dipping microstructure is considerably degraded.

Under controlled circumstances, such as generating a model of an object in a game, we know the underlying geometry of the object. Tessellating is simple under these circumstances because we know which points connect to which other points. Tessellating a seismic section is much harder, however, because we do not know the underlying geometry and that geometry may not even be unique. An unmigrated seismic section, for example, may contain conflicting, dipping events at the same point and therefore, any tessellation schema may

enhance one dipping event at the cost of another. This is the main reason why I said that whereas the apparent resolution of a SeisScape display is higher than that of a conventional display, it is still not equal to the absolute resolution of the data.

This is not to imply, however, that determining an appropriate local tessellation schema is impossible. It may well be, given the nature of seismic data, that the tessellation of any point is non-unique. However, there are approaches we can take, which will improve the overall definition of the seismic surface. In the next section, I introduce three possible techniques.

9.3 Tessellation Schemas

Tessellating a seismic section is a non-trivial task. It requires apriori knowledge of the dips and orientations of both the macrostructure and the microstructure events, which, in practice, is very difficult to obtain. Even in the case of a fully interpreted seismic line, the level of information provided by the interpreted events is insufficient for tessellation, which requires knowledge of the local structure of the data at every sample.

The tessellating software must also complete the tessellation quickly and without significant viewer interaction. For example, when you animate through a 3D seismic volume, you must tessellate each inline, crossline or timeslice before the viewer can interact with them. To maintain animation speed, therefore, you must tessellate each section in less than roughly $1/10^{th}$ of a second. This precludes any input from the user. As a further complication, each section is unique and consequently the software cannot use the tessellation of a previous section as a guide. Altogether, tessellation is by far the most difficult part of producing a SeisScape display.

The simplest technique for getting around tessellation problems is to provide the user with the option to tessellate using either an up-dip or a down-dip favoring schema (see Figure 9.8). This approach is fast and is practical for sections with limited dips. However, for most seismic sections, it is inadequate for two reasons. The first is that the simple up-dip/down-dip schemas are only valid for small dips. They cannot handle situations where the correct tessellation requires connections with samples other than to one of the two

nearest samples on the adjacent trace. The second reason is that, as was shown in Figure 9.11 and Figure 9.12, each schema favors one dip orientation and degrades the other. This makes them unsuitable for sections that have conflicting dips.

What is required is an adaptive system of tessellation that determines the correct tessellation for each point in the section. In the remainder of this section, I report on several methods that I developed to accomplish this task. First, I report on a subdivision approach that I developed early in my research and later abandoned as impractical. I report on it here for two reasons; (1) whereas it was impractical at the time, with the advent of gpu based geometry processors it will become practical in the near future; (2) I use surface normals generated via this approach, to develop a practical low-dip tessellation schema.

9.3.1 Forward Loop Subdivision

The problem of ambiguous tessellation parallels a problem that is already familiar to geophysicists, that of under sampling. The lower the frequency of sampling, especially in the spatial direction, the greater is the effect of tessellation ambiguity. For example, if we sampled the data shown in Figure 9.9 at twice the spatial and temporal frequency, we would considerably reduce the saw-tooth artifacts. The tessellation would still connect points across the apex of the events rather than along them. However, the resampling would reduce the difference in amplitude between the four connected points and therefore it would lessen the saw-tooth effect, which is pronounced on the display. This suggests that one way to reduce the effect of incorrect local tessellation is to resample the data both spatially and temporally. The optimal way to resample a data set is in the frequency domain. However, I considered that this approach was computationally excessive. Instead, I concentrated on strictly time domain approaches.

I first considered using simple averaging to resample the data. I did not implement this method, however, because I concluded, based upon Figure 9.13, that whereas averaging may reduce tessellation artifacts it would not eliminate them.

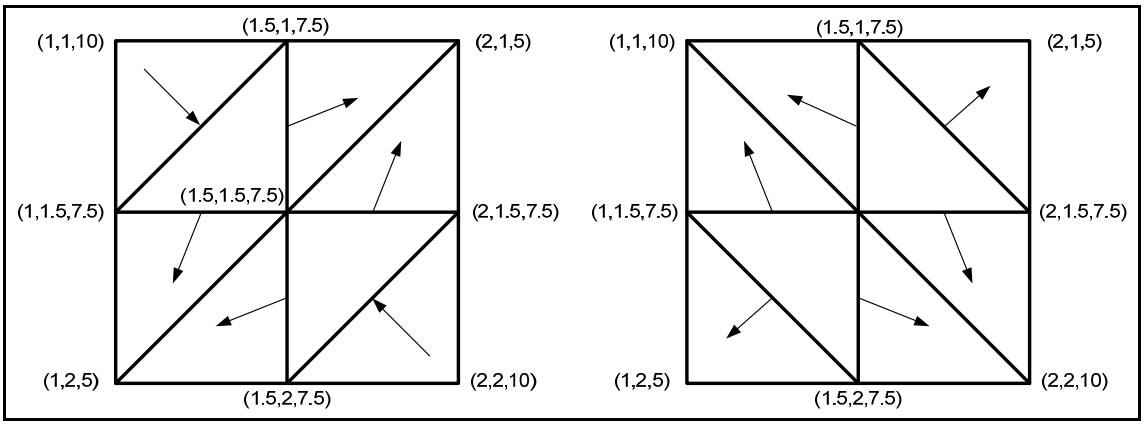


Figure 9.13: Possible tessellation schemas for a resampled mesh. The original data is the four points that I show in Figure 9.8 but with the temporal and spatial sampling frequency doubled. I calculate the values of the resampled points by simple averaging.

I base the data that I show in Figure 9.13 upon resampling the data from Figure 9.8 using averaging. It is possible to connect the original four points to produce either a trough or a ridge. The objective of tessellation is to determine which of those two alignments is correct. In the case of Figure 9.13, which depicts just two of the possible schemas for the resampled data, it is clear that whatever schema you use, the result would be neither a ridge nor a trough. Rather the schema produces a dip-oriented indentation in the event, regardless of whether the data is up-dip or down-dip. I concluded that simple averaging would degrade the appearance of all events and did not pursue it further.

To test the effect of resampling upon tessellation artifacts, I implemented a subdivision schema based upon B-splines. Subdivision is a term from computer graphics and it refers to resampling a three-dimensional surface. Forward subdivision refers to increasing the level of tessellation of the surface whereas reverse subdivision refers to decreasing the level. In geophysical terms, these are analogues to up sampling and down sampling the data respectively.

In computer graphics, controlling the level of tessellation for an object is important when dealing with large, complex scenes. When you view an object from a distance, it covers only a small number of pixels. Consequently, it is inefficient to render all the details, given that many of them overlap the same pixel. Reverse subdivision is used to

create reduced levels of geometric detail. The farther away an object is from the viewer, the lower the level-of-detail needed to render it.

In 3D graphics, reverse subdivision is generally more important than forward subdivision because objects in a scene are geometrically well known. In the case of seismic data, however, we need to increase the level-of-detail, not reduce it. To accomplish this, there are a number of time-domain approaches that produce higher detail. The one I selected for trial was Forward Loop Subdivision (Loop 1987).

Forward Loop subdivision is just one of many possible face-splitting schemes. Its primary advantage is that all faces in the input mesh must be triangular. The resulting subdivided mesh is also triangular with each input triangle split into four output triangles. The output topology of the mesh thus mimics that shown in Figure 9.13, which makes it ideal for use on seismic data. Other schemes, such as Catmull-Clark subdivision (Catmull 1978), which also use B-splines, use input quadrilaterals rather than triangles. Initially, a quadrilateral scheme looks like a better fit for subdividing seismic data because the input data to the tessellation always consists of a quadrilateral of four samples. In practice, however, Catmull-Clark and other quadrilateral based schemes unduly smooth the input samples. Loop subdivision also affects the input samples but to a lesser degree and for that reason I decided to test it.

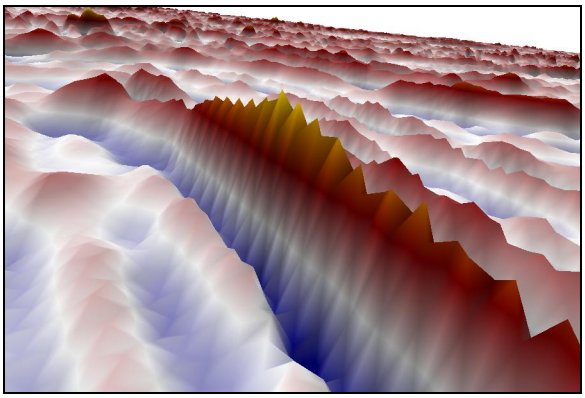


Figure 9.14: Close-up of the diamond pattern and saw-tooth tessellation artifacts produced by incorrect local tessellation.

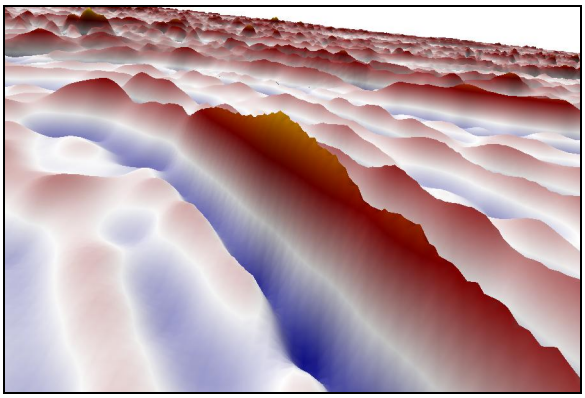


Figure 9.15: The same data shown in Figure 9.14 but after one level of Loop subdivision. Note how the subdivision has reduced the tessellation artifacts.

The Loop scheme is based upon the three-dimensional box spline, which produces C^2 continuous surfaces over regular meshes. I show an example of applying this technique in Figure 9.14 (the incorrectly tessellated data), and Figure 9.15 (the same data after one level of subdivision). The subdivided mesh is clearly smoother and provides a better visual representation of the seismic event. Nevertheless, although Loop subdivision appeared promising, I decided not to pursue this line of research. Because of that, I did not effectively test its effect on microstructure and as a result, I cannot comment on whether or not it would improve our ability to perceive conflicting dips.

My reasons for abandoning this approach were based upon hardware limitations. A typical seismic line contains millions of samples, each of which requires at least two triangles to render. Loop subdivision increased this to eight triangles per sample, which made the tessellated meshes too large for the gpu architecture then available. The subdivision was also CPU based and consequently too slow when animating a large 3D volume.

Nevertheless, I include subdivision here because it may hold possibilities for the future. Since I ran this test, gpu architecture has considerably improved and a four-fold increase in rendered triangles is no longer a serious limitation. At the time of writing, a new gpu shader, the geometry shader, is also starting to make an impact. The geometry shader can produce new triangles during rendering. Consequently, it is now possible, in theory, to perform subdivision on the fly. This would eliminate the performance issues

caused by subdividing large 3D data sets in the CPU. For these reasons, whereas real-time subdivision of seismic data it not currently practical, it may become practical in the near future and is a possible line of future research into tessellation.

9.3.2 Adaptive Tessellation

A seismic section is a complex object, which contains details on several different levels. At the beginning of this chapter, I defined two terms to describe these levels; (1) macrostructure, which refers to the prominent events that are visible on any display; and (2) microstructure that is the fine scale seismic details not visible on wiggle trace displays. Both levels of events are affected by tessellation. The most obvious effect of incorrect tessellation is the saw-tooth pattern that degrades the appearance of dipping macrostructure events. Beyond this, however, as I showed in Figure 9.12, incorrect tessellation can effectively erase microstructure events.

The challenge is to develop an efficient adaptive tessellation schema that eliminates the obvious saw-tooth artifacts and preserves conflicting-dip microstructure events. I use the term adaptive because to meet the above criteria, the schema must determine, for each seismic sample, the best way of connecting it to its neighboring trace. There is an additional caveat; any schema must be practical in terms of hardware limitations and software performance on large datasets. To those ends, I set three conditions that a schema must meet before I would consider it:

1. It must produce approximately the same number of triangles as was used to render the original data set.

2. It must preserve the original amplitude of the samples.

3. For performance issues, it must also perform any calculations on the gpu.

In the remainder of this section I report on three tessellation schemas, two that I consider appropriate for low-dip sections ($\leq$ 1 samples per trace) and one for high-dip sections ($\geq$ 2 samples per trace). I call the three adaptive schemas because they adapt the

tessellation for each sample. The low-dip schemas decide which of the two connections shown in Figure 9.8 is most appropriate for a given sample. These schemas form triangles that connect only to the two adjacent samples on the next trace. The high-dip schema lets you form triangles that connect to samples outside this range.

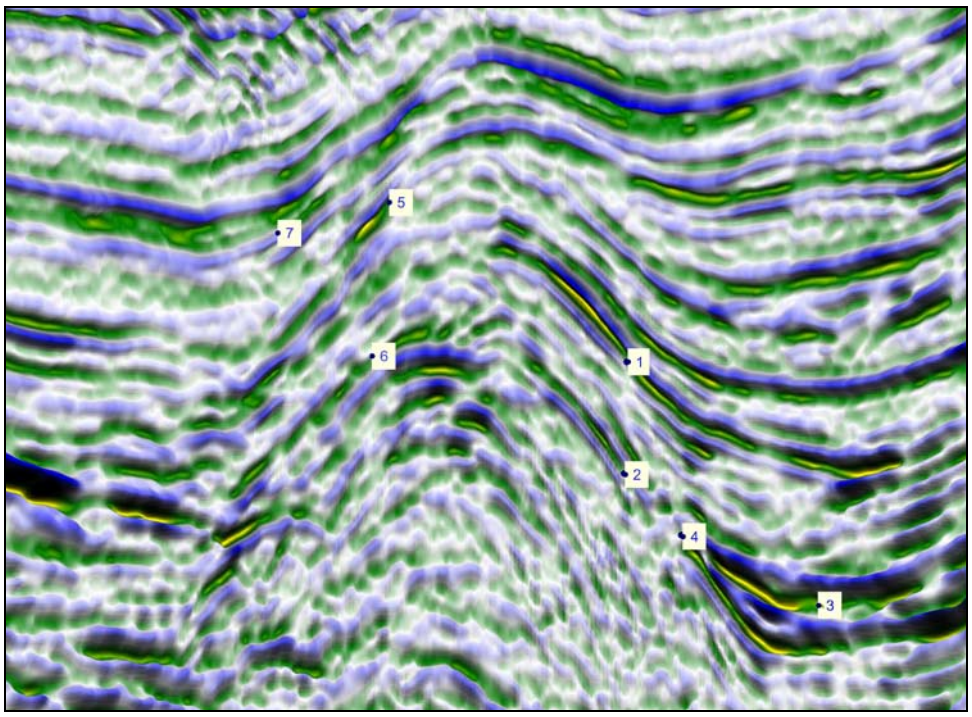


Figure 9.16: SeisScape display of seismic data over a salt dome. The area shown contains both steeply up-dipping and down-dipping events many of which exhibit tessellation artifacts. The display is oriented from left to right, consequently #'s 1, 2 & 4 are down-dip events and #'s 5, 6 & 7 are up-dip.

Figure 9.16 is an overview of the data shown in Figure 9.9 and Figure 9.10. I used this is the section of data to illustrate the effectiveness of the techniques. The numbered events have dips in the range of 1 – 1.5 samples per trace which puts them at the limit of the low-dip schemas. I evaluate each schema by analyzing how effective it is at reducing the macrostructure artifacts on these events. I will deal with the effectiveness of the techniques on microstructure in the next chapter that covers lighting.

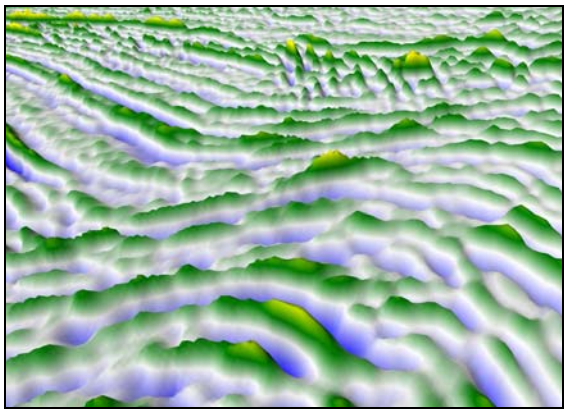


Figure 9.17: Up-dip flank of the salt dome tessellated with a down-dip schema. Note the rough, saw-tooth appearance of the event apexes.

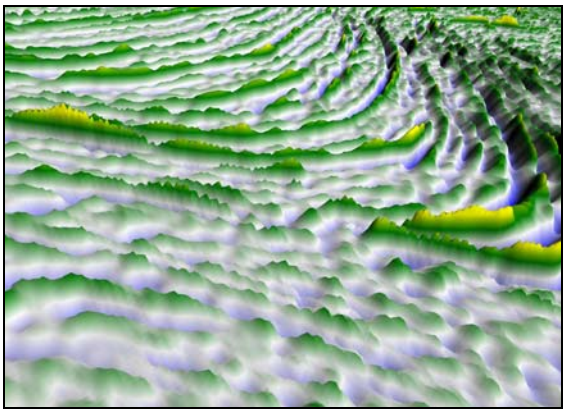


Figure 9.18: Down-dip flank of the salt dome tessellated with an up-dip schema. Again, note the rough, saw-tooth appearance of the event apexes.

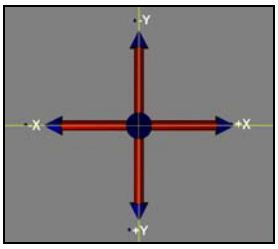
For future reference, I include images of the "worst case" tessellation scenarios for both flanks. Figure 9.17 shows the up-dip side of the salt dome tessellated with a down-dip schema. Figure 9.18 is the corresponding down-dip flank, tessellated with an up-dip schema. As one would expect, the apex of the events on both flanks are rough and have an obvious saw-tooth appearance. The objective is to use a single schema to eliminate these artifacts on both flanks.

9.3.2.1 *Low-Dip Surface-Normal Adaptive Tessellation*

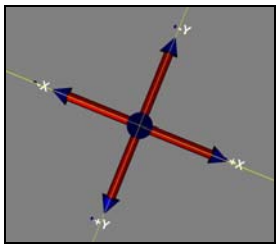
Determining the correct tessellation for any four points is a matter of determining the local dip of the seismic data at the control point. The control point is the sample that defines the upper left corner of the quadrilateral that you want to tessellate. For example, the control point in Figure 9.8 has coordinates of (1, 1, 10). Regardless of the technique used, to be practical it must be fast. Considering the number of samples in a seismic section, the technique, therefore, must be both mathematically simple and programmable on the gpu.

I based my first technique upon the surface normal. The surface normal for a given sample is a vector perpendicular to the tangent to the surface. It is used extensively in calculating the lighting for a surface and because of that I had already developed several techniques for generating the normals on the gpu (see Appendix B). Using the surface normal to determine tessellation is a typical seismic technique in that you need the

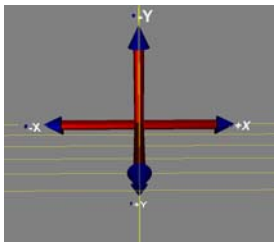
answer before you can perform the calculation. Tessellation defines the surface and hence the normal.



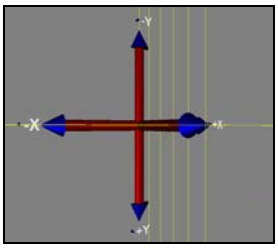
1: An idealized surface normal to an event along the x-axis.



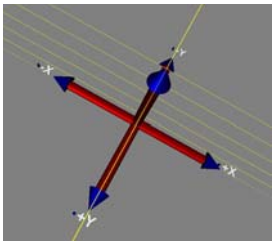
2: Z-Axis rotation (ω_z) is caused by the event dip.



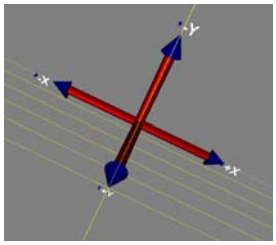
3: X-axis rotation (ω_x) results from amplitude changes across an event.



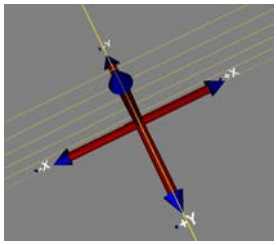
4: Y-axis rotation (ω_y) results from amplitude changes along an event.



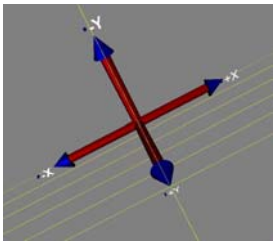
5: $\{+\omega_z, -\omega_x\}$ = ☒



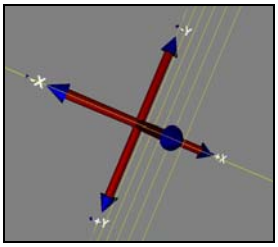
6: $\{+\omega_z, +\omega_x\}$ = ☒



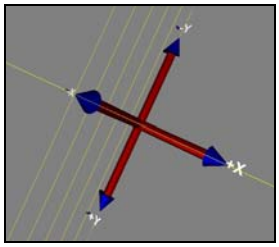
7: $\{-\omega_z, -\omega_x\}$ = ☑



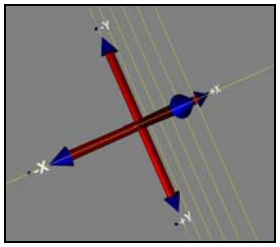
8: $\{-\omega_z, +\omega_x\}$ = ☑



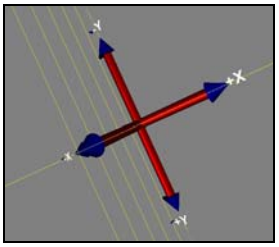
9: $\{+\omega_z, +\omega_y\}$ = ☑



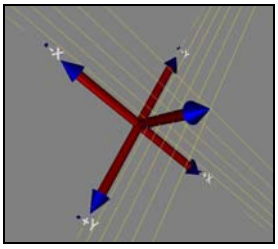
10: $\{+\omega_z, -\omega_y\}$ = ☑



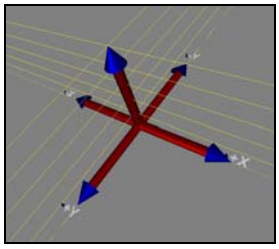
11: $\{-\omega_z, +\omega_y\}$ = ☒



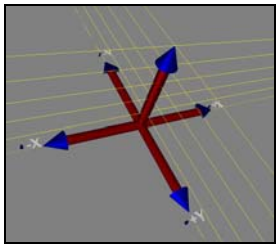
12: $\{-\omega_z, -\omega_y\}$ = ☒



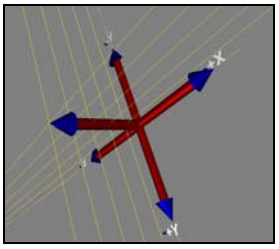
13: $\{+\omega_z, +\omega_x, +\omega_y\}$ ≅ ☒,☑



14: $\{+\omega_z, -\omega_x, -\omega_y\}$ ≅ ☒,☑



15: $\{-\omega_z, -\omega_x, +\omega_y\}$ ≅ ☒,☑



16: $\{-\omega_z, +\omega_x, -\omega_y\}$ ≅ ☒,☑

Figure 9.19: The surface normal at any sample is perturbed by three rotations, ω_z due to the dip of the event, ω_x due to amplitude changes across the event and ω_y due to changes along it.

The technique of using the normal (the product) to find the surface (the input) is similar in many respects to depth migration. Depth migration requires, as input, a detailed velocity-depth model. Determining that model, however, is why we perform depth migration in the first place. In this regard, the two techniques are analogous because they both need the answer as input. I understood this difficulty before I developed the technique. I realized, however, that the surface normal has interesting properties that were possibly useful in determining at least the direction of local dip and as a result, I decided to experiment with it.

I illustrate these properties in Figure 9.19 that shows the rotations that the surface normal experiences. Figure 9.19.1 is a conceptualized surface normal to the seismic section. The surface normal is in the z-direction (pointing out of the image), the time samples are the y-direction and the traces are in the x-direction. This normal undergoes three rotations; (1) a rotation around the z-axis (ω_z), which is caused by the dip of the event; (2) a rotation around the x-axis (ω_x), which is caused by amplitude variations across an event; and (3) a rotation around the y-axis (ω_y), which is caused by trace-to-trace amplitude variations along an event. To determine the correct tessellation for any sample we must determine ω_z, the rotation around the z-axis, at that point.

What is interesting is the effect that combining these three rotations has upon the sign of the x and y components of the surface normal vector. I show the results of combing ω_z and ω_x rotations in Figure 9.19.5 to Figure 9.19.8. These rotations correspond to a rotation of the surface normal as it moves across a dipping event. Significantly, regardless of where the normal is on the surface, the x and y components of the normal vector have the same sign (☑) if the event is up dipping ($-\omega_z$) and they have the opposite sign (☒) if the event is down dipping ($+\omega_z$). This suggests that the sign of the x and y components of the surface normal indicates the direction of local dip.

The situation, however, is complicated when you consider the y-axis rotation. In Figure 9.19.9 to Figure 9.19.12, I show the results of combing ω_z and ω_y rotations. As with the previous rotation, rotating the normal around the y-axis affects the sign of its x

and y components in a consistent manner. However, as I show in Figure 9.19.12 to Figure 9.19.15, combining all three rotations has an unpredictable effect upon the same signs. Depending on the magnitude of the ω_y rotation, the sign of the x and y components are either the same or opposite. At first, this appears to rule out using the sign of the normal as an indicator of the direction of local dip. However, for any given event, amplitude changes across the event are generally much larger than amplitude changes along it. Consequently $\omega_x >> \omega_y$ and therefore the technique still has possibilities.

The primary negative aspect of this technique is that you must tessellate the surface before you can calculate its surface normal. From the outset, I recognized that the ambiguities in tessellation as shown in Figure 9.8 potentially posed a serious problem. These ambiguities, however, only pertain to the tessellation of points and not necessarily to the surface normal at any given sample. This is because the surface normal at a given sample is the average of all of the face normals to which the sample contributes. As I show in Figure 9.20, each seismic sample forms one vertex of six faces and consequently its surface normal is the average of the six. In Figure 9.20, I show the faces for a down-dip tessellation schema, the faces are different for an up-dip schema but the vertex normal at a given sample position is still the average of six face normals. I anticipated that the averaging process would smooth out the effects of tessellation ambiguity. If it did not then there was danger that the tessellation itself would unduly influence the result.

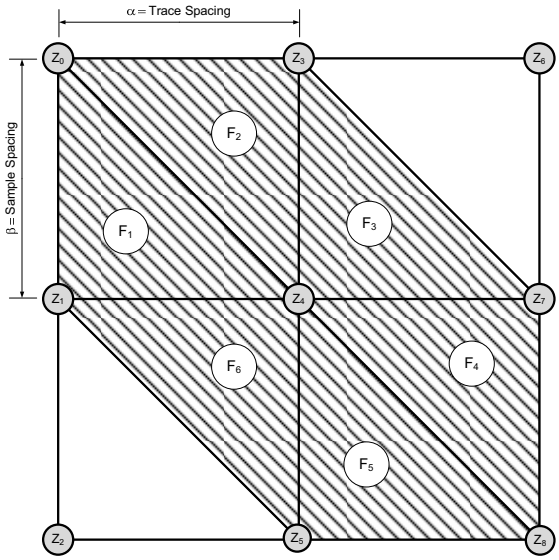


Figure 9.20: The surface normal for sample Z_4 is the average of the face normals for faces F_1-F_6.

I further reduced any inherent tessellation bias by adopting a surface normals technique based on Forward Loop subdivision. I showed in 9.3.1, that Loop's technique produces a smoother surface, albeit at the cost of quadrupling the number of triangles. In

Appendix B, however, I develop a technique that uses the same Loop technique to produce surface normals. This technique only produces normals for the original samples but I calculate them as if I had subdivided the mesh with Forward Loop subdivision.

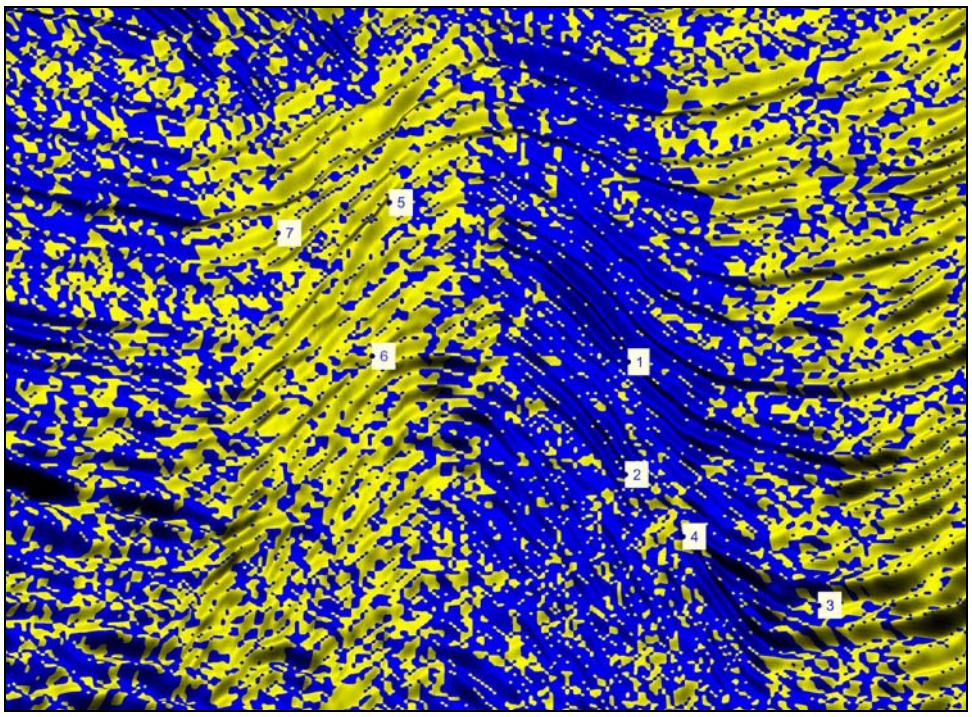


Figure 9.21: The results of tessellating the data shown in Figure 9.16 using the Loop Adaptive scheme. The coloring shows the sign of the x and y components of the surface normal. Yellow shows where the signs are the same (up-dip) and blue where they are opposite (down-dip).

Because of the uncertainties and ambiguities inherent in this technique, I had to rely on empirical evidence to decide if it was useful. To test the technique, I made a small modification to the pixel shader described in Appendix B. In that shader, I write the x and y components of the surface normal to a two channel floating point texture. I then use the texture to calculate lighting during rendering of the seismic data. The shader that I used for the test replaces the x and y components with a single value. That value is positive when the sign of the x and y components is the same and negative when they are opposite. In Figure 9.21, I show an overview of the results; positive values (up-dip) are colored in yellow whereas negative values (down-dip) are colored in blue.

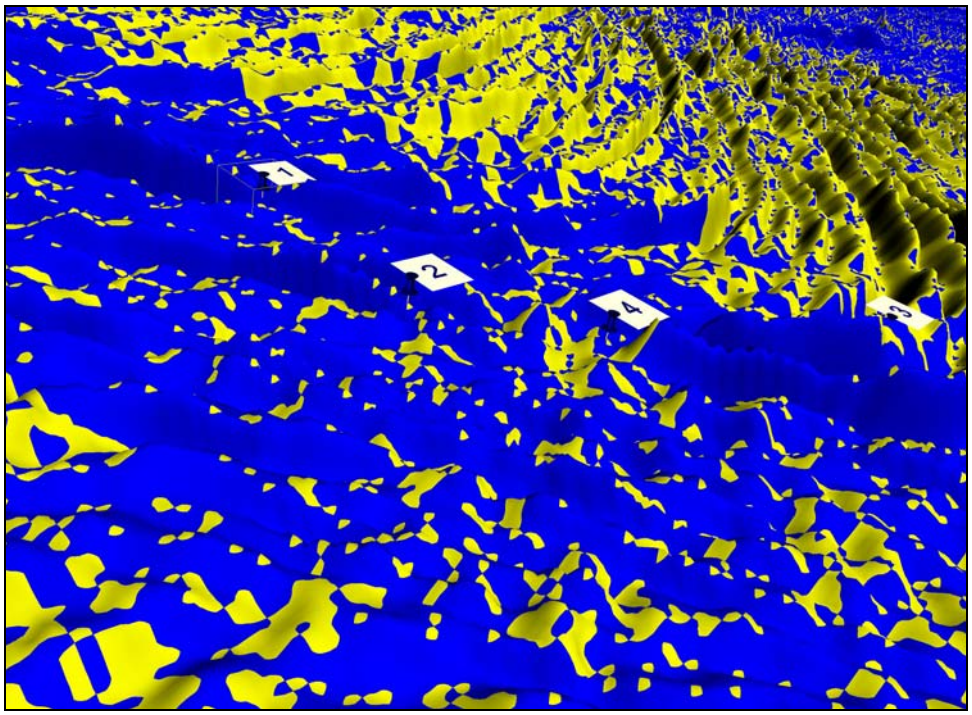


Figure 9.22: The down-dipping flank colored by the calculated dip direction. Blue indicates that the seismic is down-dip whereas yellow indicates that it is up-dip. The yellow spots at the apex of down-dip events indicates that the technique breaks down where it is needed the most.

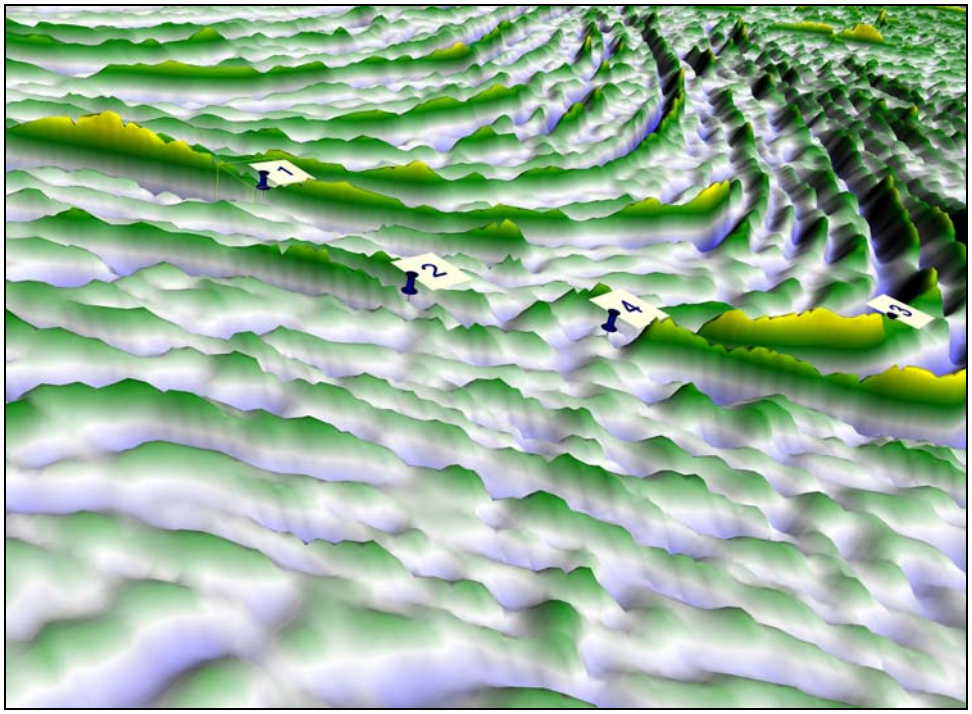


Figure 9.23: The display as Figure 9.22 but colored with HA1. Compare this image to Figure 9.9 and you will see that the down-dip events are now smoother and more continuous. However, whereas the tessellation in improved, many saw-tooth artifacts remain.

This technique indicates the direction of local dip but not the magnitude of that dip. Its only purpose is to determine whether the seismic at a given sample is up dipping or down dipping. In that context, the results are generally correct. Where the underlying seismic data is up-dipping (Figure 9.21-5, 6 &7) the color is primarily yellow and where it is down-dipping (Figure 9.21-1, 2, 3, 4) the color is primarily blue.

A casual inspection of Figure 9.21 suggests that this technique has promise. If you inspect the two flanks (down-dip: Figure 9.22 - Figure 9.23, up-dip: Figure 9.24 - Figure 9.25 you will see that this adaptive approach has, in general, improved the display. Most of the saw-tooth effects on the "worst case" images are gone and the apexes of the events on both flanks are, for the most part, continuous. However, a closer inspection shows that the calculation has an inherent weakness when calculating the dip direction at the apex of an event. For the technique to work, the sign of the x & y components must be determined by the x-axis rotation (ω_x), i.e. the rotation across an event. At the apex of the event, however, this rotation is close to zero and consequently the sign of the components is dominated by the y-axis rotation, i.e. the rotation along an event. If you recall, this rotation produces an opposite effect upon the sign of the x & y components. As a result, this approach may produce the incorrect results at the apex, where we need it the most, and the correct result along the flanks, where we need it the least.

You observe this problem on Figure 9.22 and Figure 9.23. Events 1, 2 & 3 in the images are all steeply down-dip and, as expected, they are generally blue on Figure 9.22. However, there are a significant number of places along the apex of each event where the color flips to yellow, indicating an incorrect tessellation. When you look at the corresponding locations on Figure 9.23, these errors show up as notches across the apex. The up-dip flank has the same potential apex problem. You can clearly see on Figure 9.24 that there are entire bands of incorrectly tessellated regions along the apex of many of the events. This is possibly due to the down-dip bias of the Loop schema that I used in the calculations. Paradoxically, however, when you compare Figure 9.25 with Figure 9.17, you do not see any notches and all of the events appear smooth and continuous.

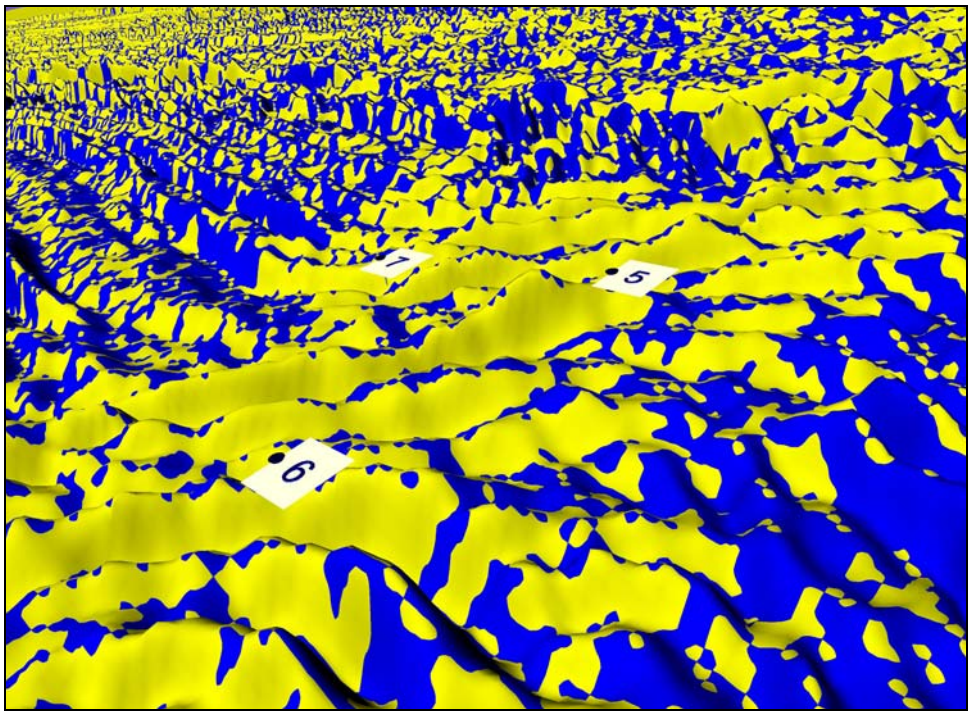


Figure 9.24: The up-dipping flank colored by the calculated dip direction. Note the presence of bands of incorrect dip calculations (blue) at the apex of the events.

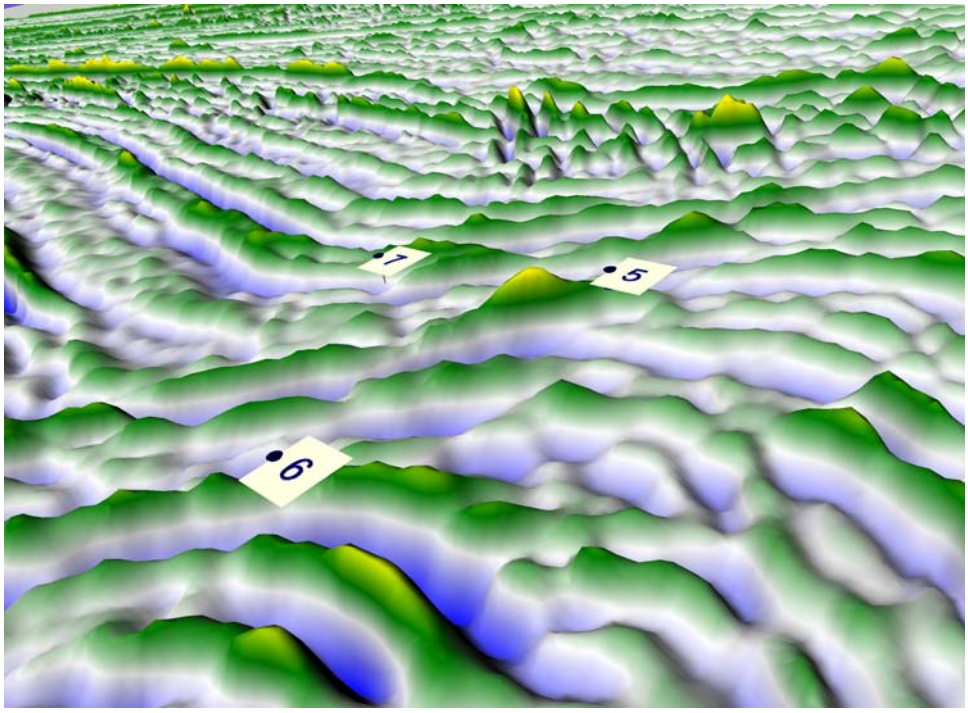


Figure 9.25: The same display as Figure 9.24 but colored with HA1. Even though the apexes of the up-dip events are incorrectly tessellated, there are very few tessellation artifacts visible. This suggests that the presence of artifacts is dependant upon both dip magnitude and frequency content.

Consider event #5; the color-coding clearly indicates a series of incorrect tessellations along the apex. Unexpectedly, these errors do not show up as notches on Figure 9.25 and the apex appears smooth. I have observed the same effect on other sections but, because of time constraints, I have not studied the exact cause of this paradox. I believe, however, that it arises from the frequency dependency of tessellation artifacts. I have observed that, for a given sample interval, higher dominant frequency events tend to have more severe visual artifacts. Looking at the overview image of this data (Figure 9.16) it appears that the up-dip events have a lower dominant frequency than the down-dip events. This is why, in my opinion, there are fewer saw-tooth patterns on the up-dip events. This is only my opinion, however, and the dependency between tessellation artifacts and the event dip and dominant frequency, remains to be determined in a future study.

This was my first adaptive tessellation schema and despite its limitations, it showed promise. Even with the previously mentioned errors, when you compare Figure 9.23 with Figure 9.9 (the same data tessellated with an up-dipping schema), it is clear the adaptive approach improves tessellation. The events are significantly smoother and most of the saw-tooth artifacts are gone. In the end, I discarded this technique in favor of the technique that I discuss next. However, even though the technique was far from perfect, it was still significant because it established that adaptive tessellation would reduce artifacts.

9.3.2.2 *Low-Dip Correlative Dip Adaptive Tessellation*

Although the Loop-Adaptive approach to tessellation substantially reduced tessellation artifacts, it was subject to errors at the apex of events. In this section, I report on a second approach to adaptive tessellation, one that does not suffer from the same defect. This technique, which I call the “Correlative Dip” schema, uses a conventional approach to determining the local dip at each sample. To determine the dip I use a small window (± 4 samples) around the sample in question and calculate the normalized cross-correlation between it and its neighboring trace. This is a real-time process and therefore efficiency was paramount. With that in mind, I only allowed a ±2 sample shift in the

correlation, and I implemented the technique in a pixel shader (see Appendix B for the actual shader details).

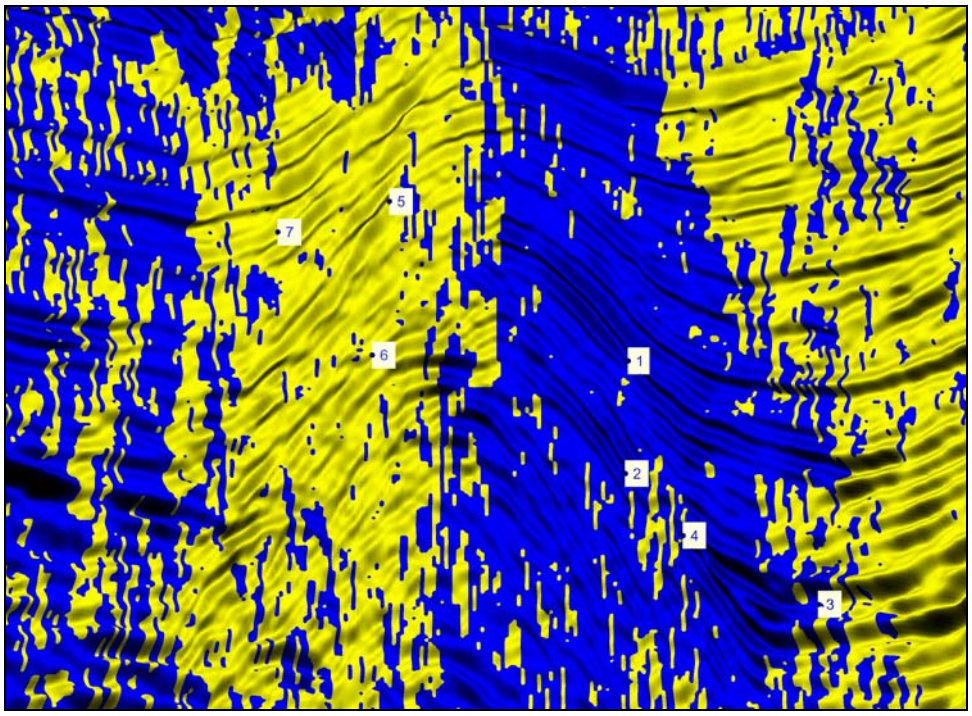


Figure 9.26: The results of tessellating the data shown in Figure 9.16 using the Correlative Dip Adaptive scheme. The coloring shows the sign of the x and y components of the surface normal. Yellow shows where the signs are the same (up-dip) and blue where they are opposite (down-dip). When compared to Figure 9.21, the correlative dip technique is clearly more consistent and robust at determining local dip than is the Loop Adaptive scheme.

Figure 9.26 shows an overview of the results of applying the correlative dip adaptive scheme. Again, as in Figure 9.21, the yellow color indicates samples where the seismic is determined to be up-dipping and blue where it is determined to be down-dipping. I expected that this approach would reduce or eliminate the previously mentioned problems with tessellating event apexes and a comparison of Figure 9.21 and Figure 9.26 shows that it does. It is clear from comparing the two that the correlative dip technique is superior. The blue and yellow colors, which indicate down-dip and up-dip areas respectively, almost uniformly follow the correct dip alignment and are more consistent across the apex of events. Clearly, the correlative dip technique is more consistent and robust at determining local dip than is the Loop Adaptive scheme.

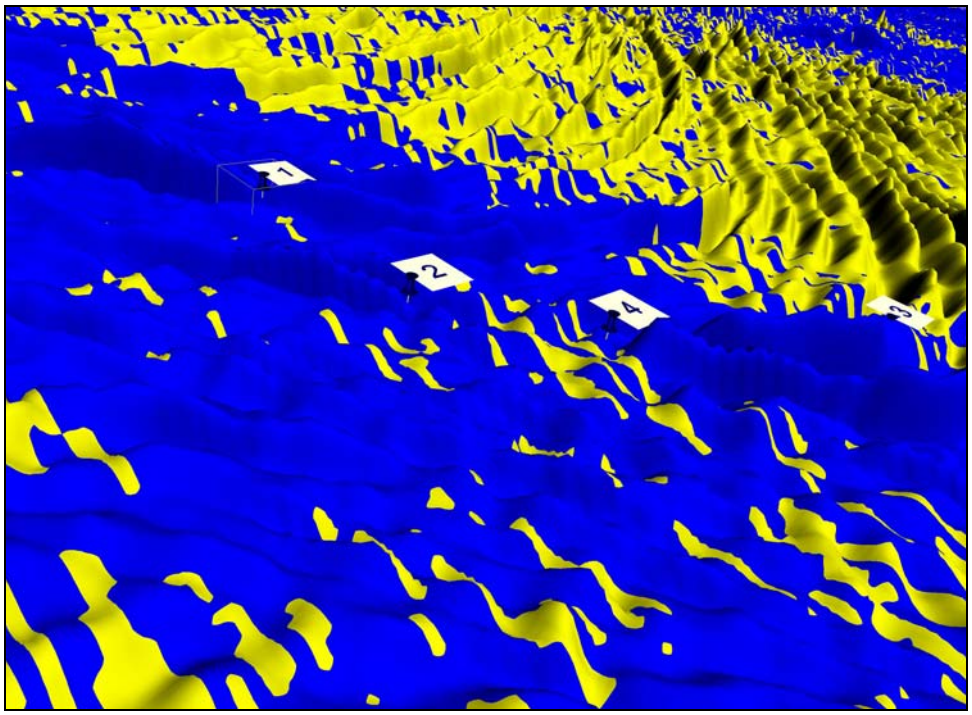


Figure 9.27: The down-dipping flank colored by the dip direction (correlative dip calculation). Blue indicates that the seismic is down-dip whereas yellow indicates that it is up-dip. Colors now correctly follow the correct dip alignment and are consistent across the apex of events.

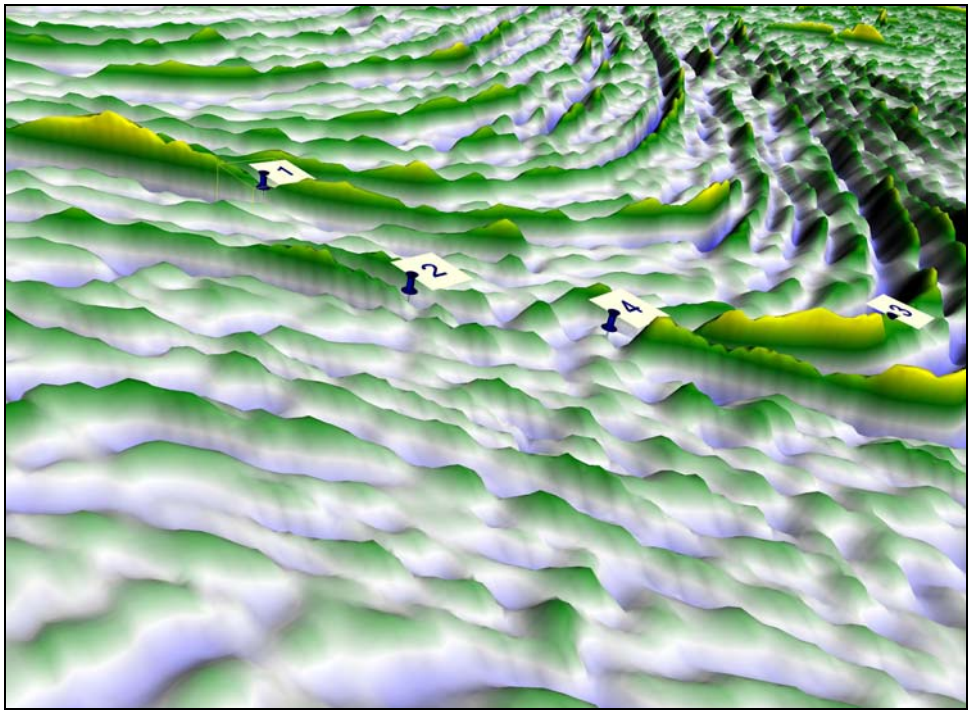


Figure 9.28: The same display as Figure 9.27 but colored with HA1. Whereas this approach calculates local dip better than the Loop schema, it does not fully remove apex tessellation artifacts. The remaining artifacts are caused by the steepness of the events, which a low-dip schema cannot handle.

This approach calculates local dip better than the Loop schema but this does not necessarily translate into a perfectly tessellated display. When you compare the dip colored images for the down-dip flank (Figure 9.22 and Figure 9.27) you see that the correlative approach follows the local dip better and has fewer apex artifacts. However, when you compare the HA1 colored images (Figure 9.23 & Figure 9.28) the improvement in dip calculation does not substantially improve the down-dip tessellation. Note that I ignore the up-dip flank here because the results of the Loop schema and the Correlative Dip schema are virtually identical.

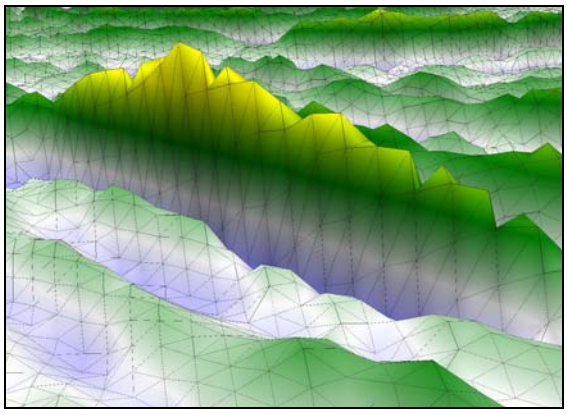


Figure 9.29: Close-up of Event #1 tessellated using the Loop schema. The wireframe overlay shows the outline of the tessellation. Note the deep notches caused by errors in the tessellation.

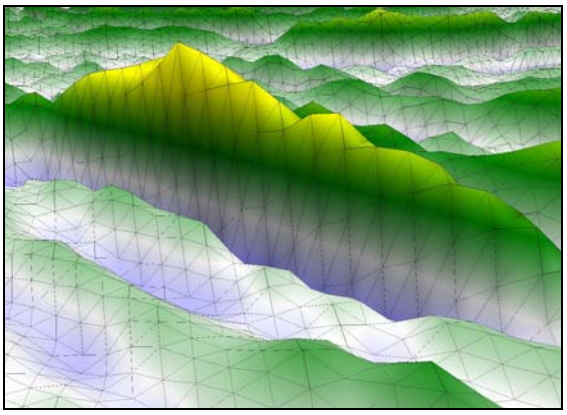


Figure 9.30: Close-up of Event #1 tessellated using the Correlative dip schema. Notice that the deep notches are missing but that the apex of the event is still not smooth.

I illustrate the advantages and limitations of this technique in Figure 9.29 and Figure 9.30, which are close-ups of event #1 tessellated with the Loop and Correlative techniques respectively. The Loop schema produces artifacts that appear as deep notches across the events. You can clearly identify these notches on Figure 9.29 but if you look closely at Figure 9.30, you will see that correlative schema has eliminated them and that you have a better perception of the event as a whole. However, whereas the notch artifacts are gone, there is still the perception that the tessellation is not perfect and that the amplitudes along the event are not representative of the true amplitude structure.

This is because the dip on the event is approximately 1.5 samples per trace, which places it at the limit of low-dip tessellation. There are two issues that are important when tessellating steeply dipping events. The first is determining the direction of local dip,

which this approach does very well. The second, however, is determining the magnitude of the dip and the technique fails in that respect. There are places where the apex of the event shown in Figure 9.30 jumps two samples between traces and this technique cannot handle that jump. For that, we need a high-dip schema, which I discuss next.

Still, for low-dip scenarios, this approach meets all of the criteria that I set out at the beginning of this section. It is robust, fast and I use it throughout the remainder of this thesis.

9.3.2.3 *High-Dip Correlative Dip Adaptive Tessellation*

The low-dip correlative dip schema is capable of effectively tessellating dips of ~1-1.5 samples per trace, which makes it ideal for stratigraphic settings. However, structural sections may have dips that far exceed that limit and in such cases, the low-dip approach is inadequate. In Figure 9.31 and Figure 9.32, I show an example of what happens to the tessellation as events become steeper and eventually exceed the one sample per trace limit. Both of these images show events on the steeply dipping flanks of a salt dome. The event dips grade from approximately 2.5 samples per trace to effectively zero samples per trace.

Figure 9.31 shows the data tessellated with an up-dip schema. As one would expect, the down-dip events on the flank are saw-toothed and it is very difficult to perceive their amplitude structure. By contrast, Figure 9.32 shows the same data tessellated with the low-dip schema previously described. If you follow any event from the upper left where the dips are steepest, to the lower right where they are the shallowest, you will see that the number of saw-tooth artifacts decreases and the amplitude structure becomes better defined. However, in the high dip regions, the events are still incorrectly tessellated and the amplitude structure is very hard to discern.

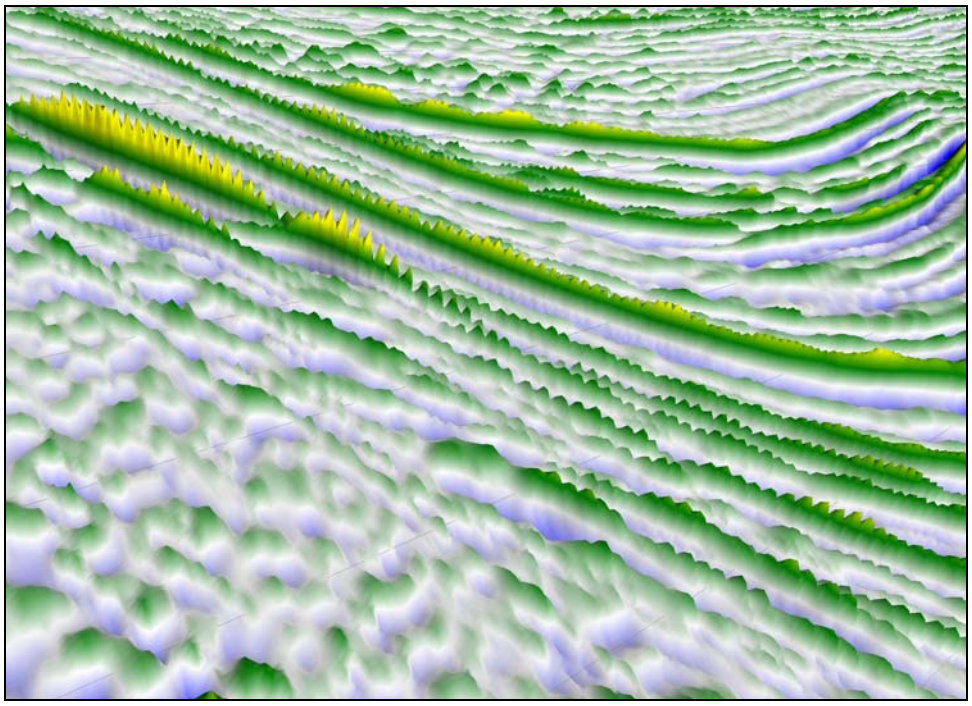


Figure 9.31: Events on the steeply dipping flank of a salt dome (data courtesy unnamed source) tessellated with the up-dip schema. Event dips grade from 2.5 samples per trace to zero samples per trace.

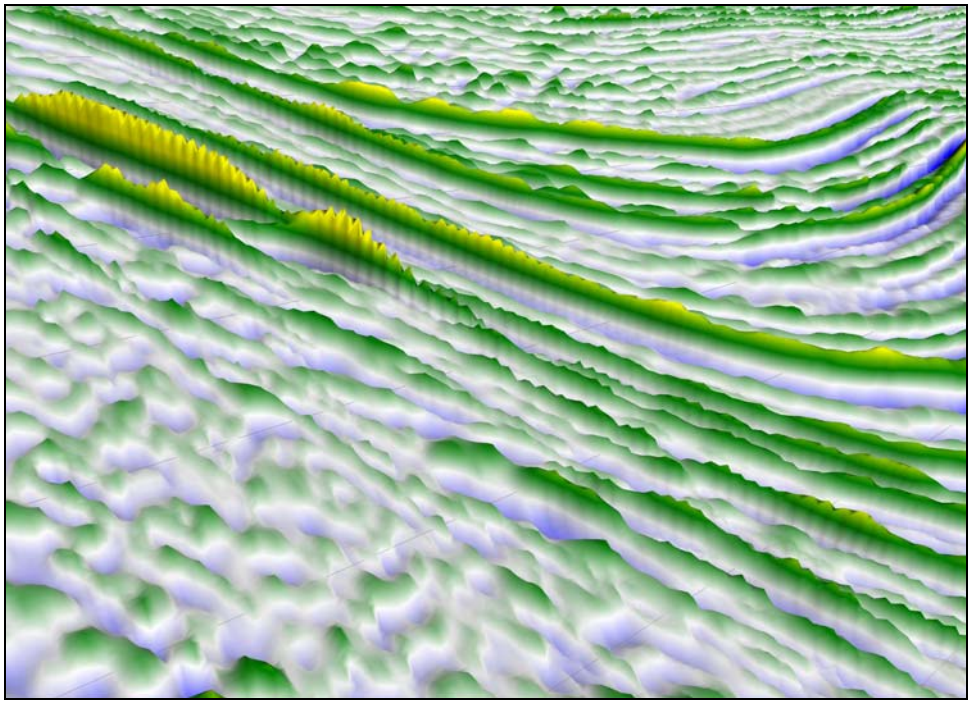


Figure 9.32: The same events shown above but tessellated with the low-dip correlation schema. Note how the tessellation improves as the events start to flatten out at the lower right of the section. In these regions, the amplitude structure of the events is better defined.

This degradation of the amplitude structure as an event becomes steeper is typical of what you see on structural data sets. In this section, I describe an experimental technique for tessellating these high-dip events, one that corrects most of the remaining problems. The technique uses the same normalized cross-correlation approach that I used in for the low-dip schema but with one significant modification. In the low-dip case, I was only interested in determining the direction of local dip and I expected that those dips would be small. Consequently, I restricted the cross-correlation to a ±2 sample shift. In this case, however, I needed to determine the magnitude of the dip and I expected that the dips would be much steeper. To that effect, I expanded the cross-correlation shift to ±4 samples. I implemented the technique in a pixel shader, which I detail in Appendix B.

In the low-dip case the output from the shader was a simple +- switch that I used to indicate dip direction. The output from this shader, however, was a signed number. The sign was the same dip direction indicator and the number was the local dip in samples per trace. Calculating the magnitude of the local dip was only the first step in this technique. The purpose of the low-dip schema was to determine which of the up-dip or down-dip schemas illustrated in Figure 9.8 to use for a given sample. In this case, however, neither schema was appropriate because the dip magnitude can be greater than one and this causes problems with the tessellation.

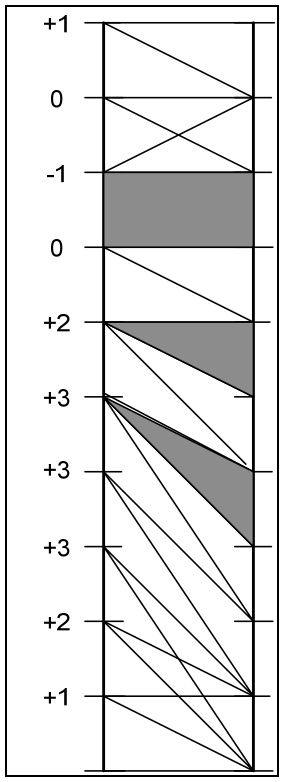


Figure 9.33: Errors in high-dip tessellation. Numbers indicate the magnitude of local dip.

In Figure 9.33, I illustrate what can happen to the tessellation as the dip magnitude changes. The numbers on the left side of the image represent the dip magnitude as output by the shader. As these numbers change from sample to sample, two problems can occur. The first is that gaps can appear, as I show by the gray areas. The second is that the tessellation from one sample can overlap the tessellation from the previous sample(s). In the examples that follow, I attempted to correct these problems in the cpu code. However, the only way to

correct them completely is to either add or remove triangles, as needed, which I could only do in a geometry shader. Unfortunately, geometry shaders are very new and I decided to exclude them from this work. Consequently, the problems that I just mentioned are visible in the images that I show of this technique.

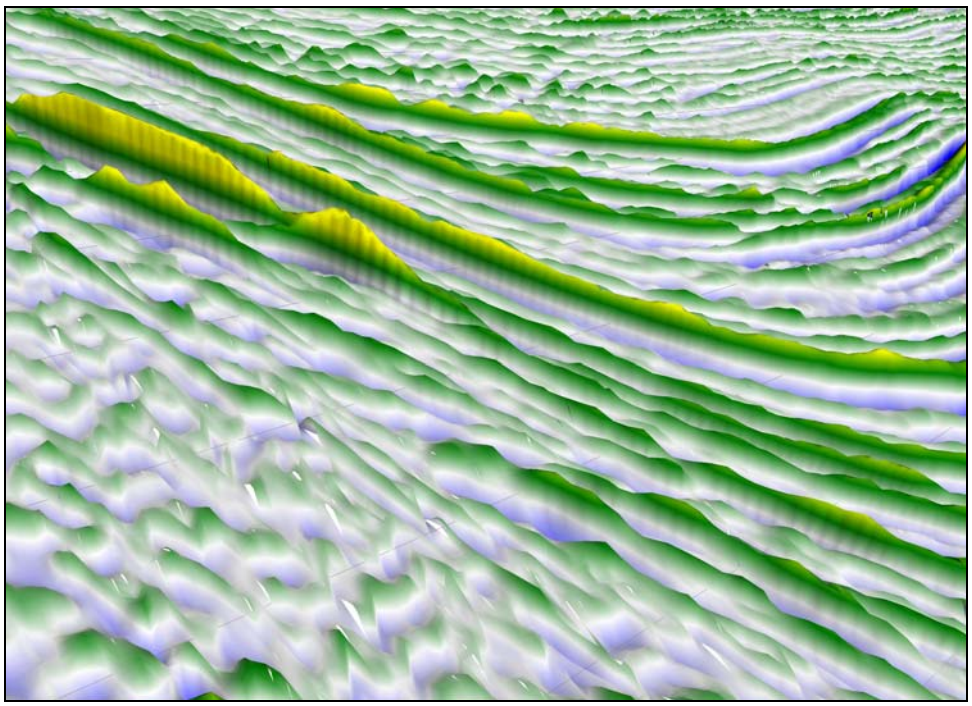


Figure 9.34: The same data shown in Figure 9.32 but tessellated with the High-Dip schema. The saw-tooth artifacts visible on the previous images of this data are gone and the amplitude structure of the events is now clear. The holes in the data occur where the sample-to-sample dip magnitude changes.

In Figure 9.34, I applied the high-dip schema to the same salt dome events that I showed previously. As you can see, the saw-tooth artifacts are now gone and it is possible to follow the amplitude structure of the events from the high dip regions to the low dip regions. I further illustrate the improvement in tessellation in Figure 9.35 and Figure 9.36, which are close-ups of the uppermost events. I overlaid the wireframe tessellation mesh upon the surfaces so that you could see exactly how the two techniques have gone about determining the surface.

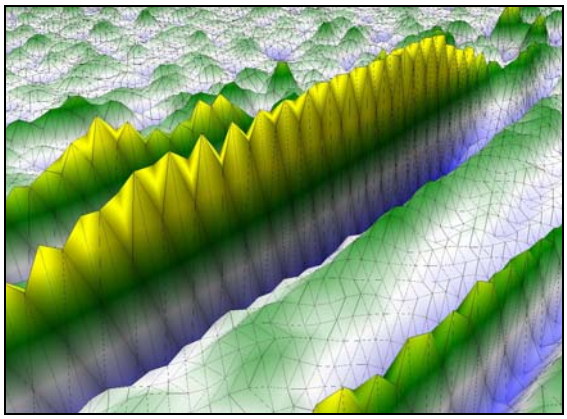


Figure 9.35: Close-up of the uppermost events on Figure 9.32. Wireframe outlines the tessellation.

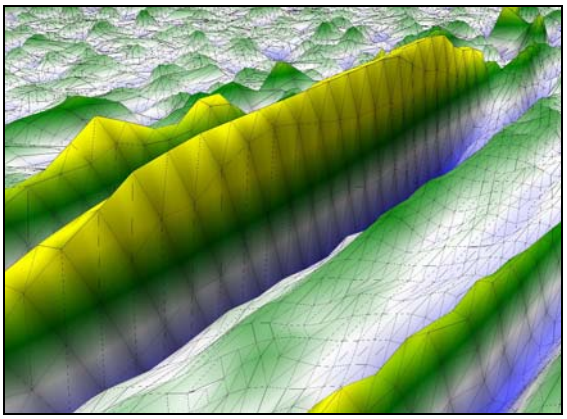


Figure 9.36: Close-up of the uppermost events on Figure 9.34.

The dip on these events is ~ 2.5 samples per trace and as a result, the low dip schema shown on the left is unable to reproduce the amplitude structure along the ridge of the events. The high dip schema, shown on the right, however, has almost perfectly reproduced the structure. This proves that we can correctly define the seismic surface even in the presence of steep dips.

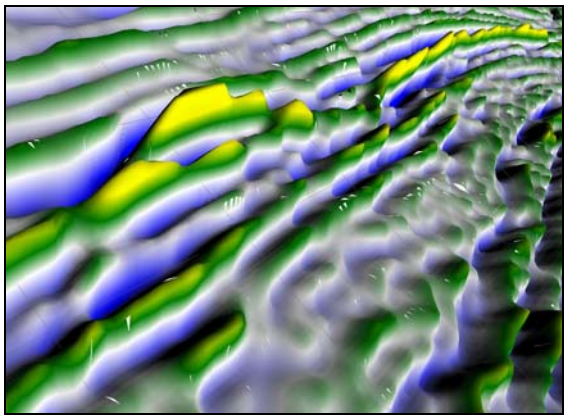


Figure 9.37: Section of data showing visible holes caused by errors in high-dip tessellation.

This technique will remain experimental until I implement it in the geometry shader. Although it has correctly defined the high-dip events, the holes and overlaps in the tessellation, which I illustrate in Figure 9.37, make it unsuitable for general use. The holes in the structure are the most obvious feature of this image but if you look closely, you will also see that there are places along the bottom of the troughs where there are also overlaps in the triangles. Surprisingly, all of these artifacts are more pronounced at low-dip regions, possibly because they are generated where the dip changes simply from positive to negative. This occurs most frequently, of course, for low-dip regions.

Correcting the errors in this technique is a job for the future. Even so, the results prove that you can correctly define the amplitude structure of steeply dipping macrostructure events. To put it into practical use, all that remains is to implement the technique in the geometry shader.

9.3.3 Future Work

One of the stated goals of this thesis is to determine the sciences behind visualization and the directions of research that we must follow to improve our ability to communicate seismic information. In this chapter, I introduced the subject of forming the seismic surface, which is the first of the practical sciences of seismic visualization. This introduction was necessarily brief and we must do much more work before we can successfully tessellate complex, conflicting dip seismic surfaces.

In particular, I discussed two techniques that we must program on a geometry shader before we can use them in a real world setting. The first of these was subdivision, which is the process of resampling the data to provide a smoother surface. The second was the high-dip tessellation schema that I could only program in a pixel shader. Both of these techniques generate new triangles and/or drop existing triangles, processes that require the geometry shader. As the geometry shader becomes available over the next year, I will develop techniques that both subdivide and perform high-dip tessellation on an "as needed" basis. The ultimate tessellation schema will be one that analyzes each sample in context with its neighbors and decides the level of subdivision and the tessellation schema to apply. Before we can develop that schema, however, we must do more research on the nature of the seismic surface itself.

Another line of research is the effect that tessellation has upon microstructure. In this chapter I focused on the effect that tessellation has upon macrostructure and in particular on the effect that it has on defining the apex or ridge of major seismic events. When we observe macrostructure events, we primarily focus on the amplitudes along the ridge and as I showed, tessellation has a major impact on our ability to perceive amplitude structure. The effect of tessellation on microstructure, however, is harder to define. In

many cases, you can only observe microstructure events as perturbations that cross the macrostructure events. To observe them we must concern ourselves not with just the apex of the macrostructure events but with their flanks as well. In the following chapter, I discuss lighting and its effect on our ability to perceive microstructure. Lighting, however, is based upon the definition of the surface, which is itself based upon tessellation. Tessellation, then, defines both our ability to perceive macrostructure and microstructure and we must do much more work before we fully understand its affect on the latter.

9.4 Macrostructure Examples

In 9.1.3 I stated that visualization has two objectives, the first is to reveal microstructure, the second is to reveal the amplitude structure of macrostructure events. Macrostructure events are, by definition, the events that we can see on a wiggle trace displays. We can see them but our perception of them usually comes from the zero crossings and we generally perceive simple monochromatic blobs. The peak amplitudes along the ridge of an event define the amplitude structure of an event but these amplitudes typically overlap. Consequently, even when we see the amplitude structure, we only see it over a very limited amplitude range and only over a very few traces.

Variable density displays are capable of showing amplitude variations better than wiggle trace displays. However, I show in Chapter 12, that human color perception is very poor (and very personal) and that no combination of colors innately defines high and low. Using color, we can see gross amplitude changes but unless we know the specific palette and how we map it to amplitude, we can never know what those changes represent. Moreover, variable density displays never let us form percepts of amplitudes and percepts are the desired goal of visualization.

In the following section, I present a series of macrostructure comparisons between SeisScape displays and wiggle trace displays. The purpose of these comparisons is twofold. My first purpose is to overcome the unfamiliarity of SeisScape displays. Wiggle trace displays are already familiar (and comfortable) to any experienced geophysicist but

SeisScape displays are new and very different. They are different and consequently they are challenging because any new technology has a necessary learning curve. We know what seismic data looks like as a wiggle trace display but we have to learn what it looks like as a SeisScape display and that poses a challenge. These comparisons address the learning issue because the wiggle trace displays all show familiar seismic scenarios. I have inserted numbered reference points into them so that you can relate what you see on the new SeisScape displays back to what they looked like on the familiar wiggle trace displays. In this way I hope that the reader will begin the process of learning what particular seismic expressions look like on SeisScape displays.

The second purpose of these comparisons is to show how much better you perceive macrostructure on SeisScape displays. They highlight just how much amplitudes actually change along events. In addition, they show how much better one can perceive low-amplitude events when they are surrounded by high-amplitude events. Finally, they show how much more continuous low-amplitude events are on SeisScape displays. In all of the comparisons the purpose is to focus on the physical structure of the section, the lighting and coloring are irrelevant.

Typically, wiggle trace displays are displayed flat on computer monitors and that is how I show them here. You can rotate SeisScape displays to any viewing angle, however, and depending upon the angle you can see different features of the data. This is one of the major advantages of SeisScape displays but one that I cannot effectively reproduce on paper. In these examples, I use a wide range of viewing angles so that the viewer can develop a sense of what seismic looks like from different visual perspectives. However, for each SeisScape display, the orientation that I chose is not necessarily the best one for the particular data set. Other orientations may show the section better.

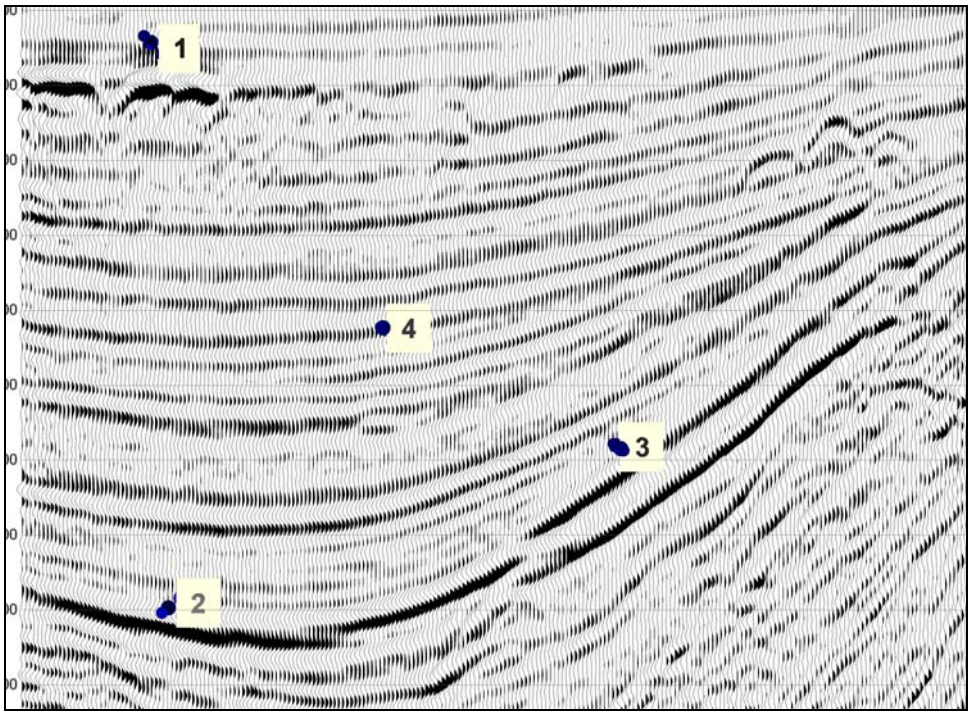


Figure 9.38: An unmigrated seismic section showing both high and low amplitude events. Note how high amplitude events appear as featureless monochromatic blobs. Lower amplitude events show more amplitude features but the amplitude relationship between high and low amplitude events is obscured.

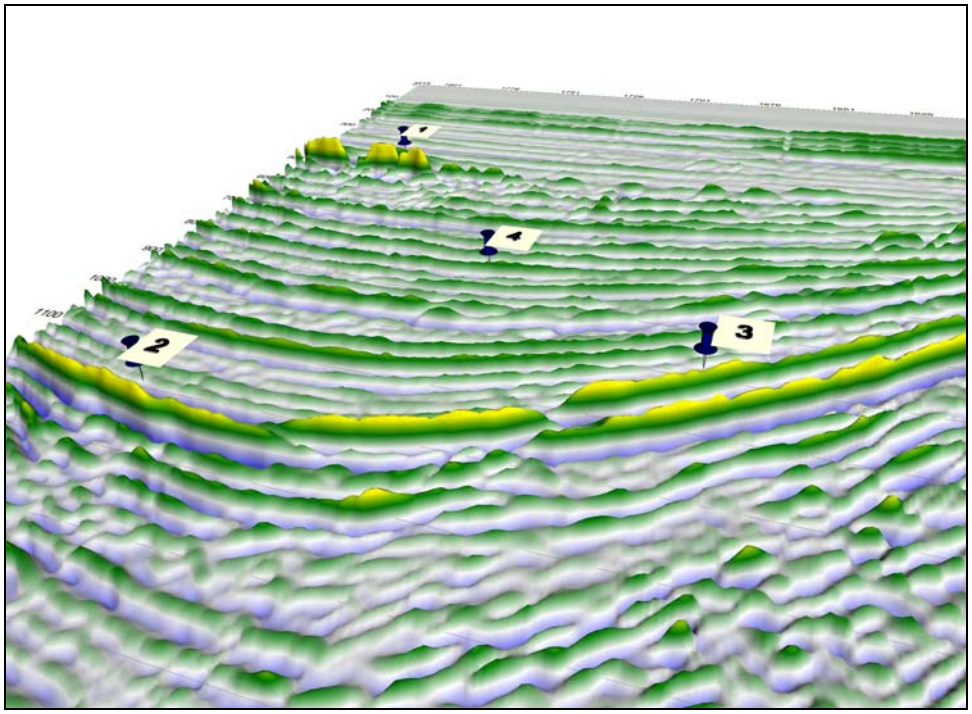


Figure 9.39: On the SeisScape display, you clearly perceive the amplitude structure of both the major and the minor events. Neither class of events dominates, amplitude changes on the high amplitude events are just a clear as they are on the low amplitude events.

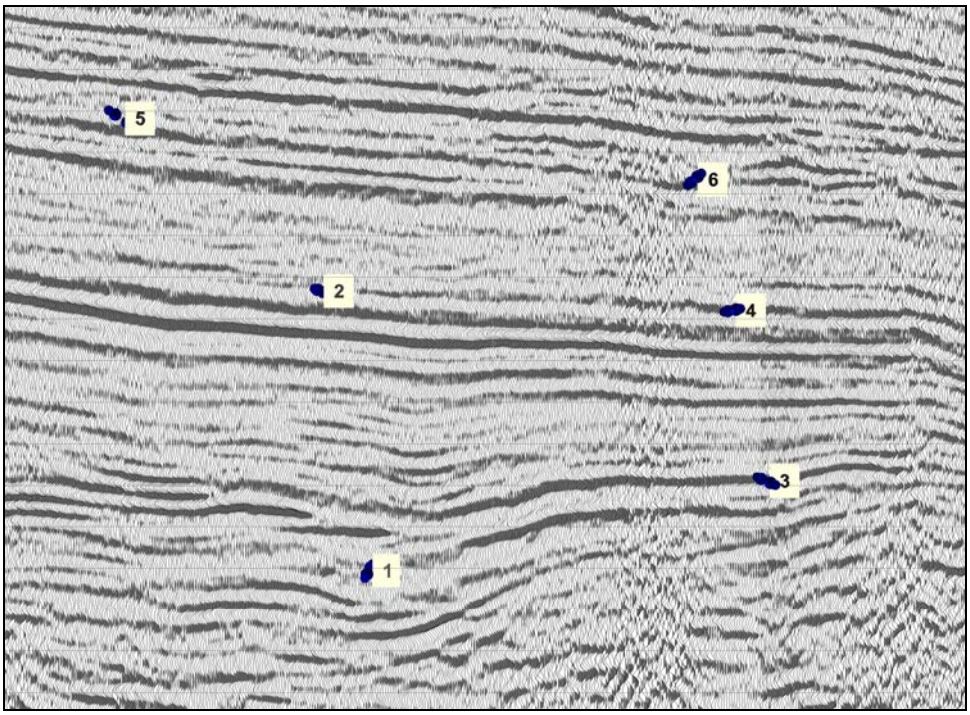


Figure 9.40: Section of noisy data from the Huallaga area of Peru (data courtesy PeruPetro). There is considerable amplitude contrast between the major and minor events and consequently I had to use a higher trace excursion (3.5) to show the low amplitude events.

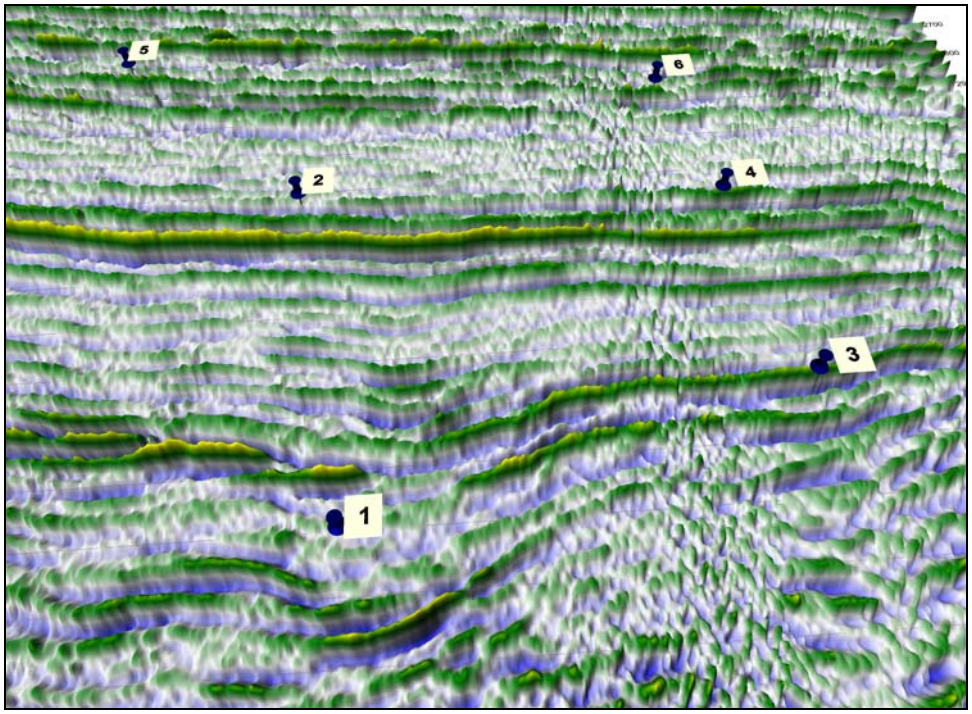


Figure 9.41: This section contains significant levels of noise, the degree of which is more apparent on the SeisScape image. The amplitude structure is also clearer, especially between markers 1 and 3.

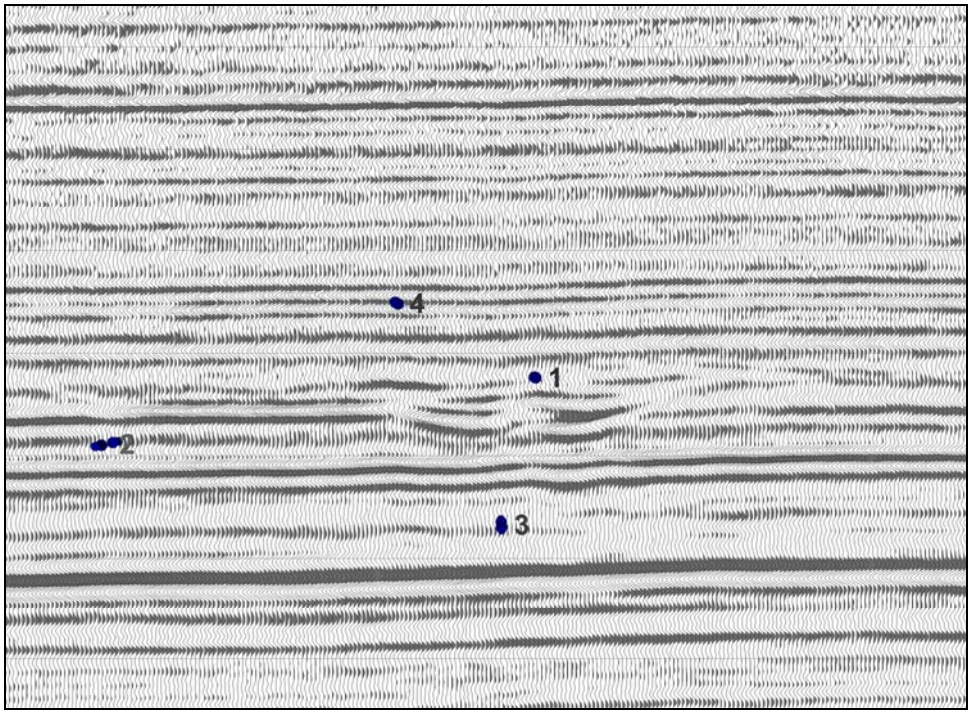


Figure 9.42: An example of a small channel embedded in typical Alberta “railroad track” data (data courtesy unnamed source). This display shows one of the strengths of wiggle trace displays, they are excellent for pattern recognition. Even so, the low amplitude zone around marker 3 is relatively featureless.

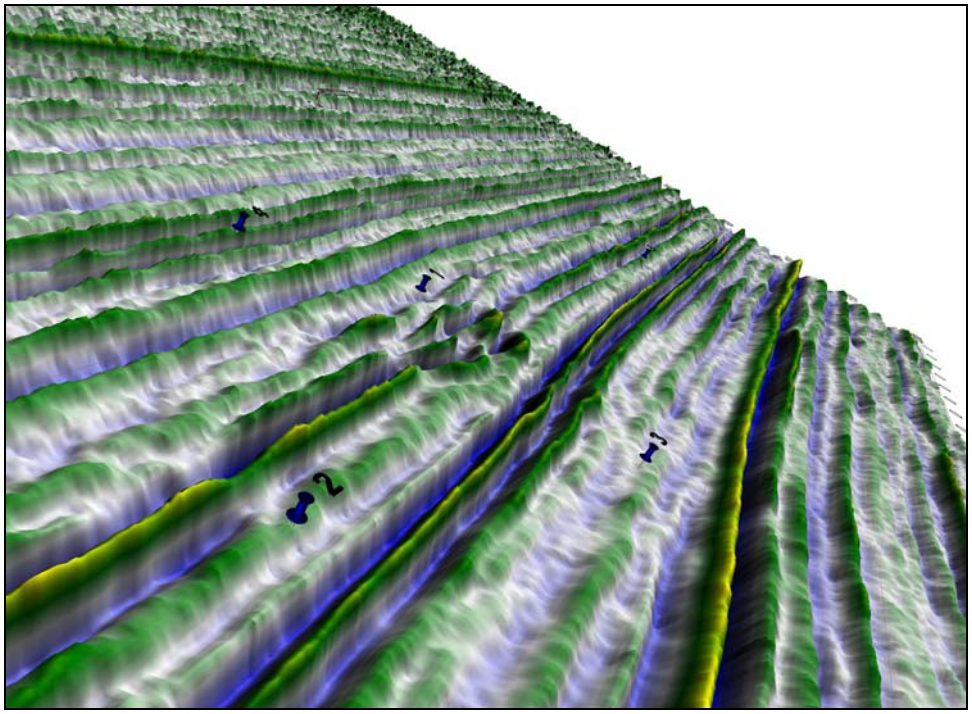


Figure 9.43: The SeisScape display shows a different perspective of the channel. I oriented the display to highlight the low amplitude zone around marker 3. On the SeisScape display, you can follow the events in this zone just as clearly as you can follow the higher amplitude events above and below it.

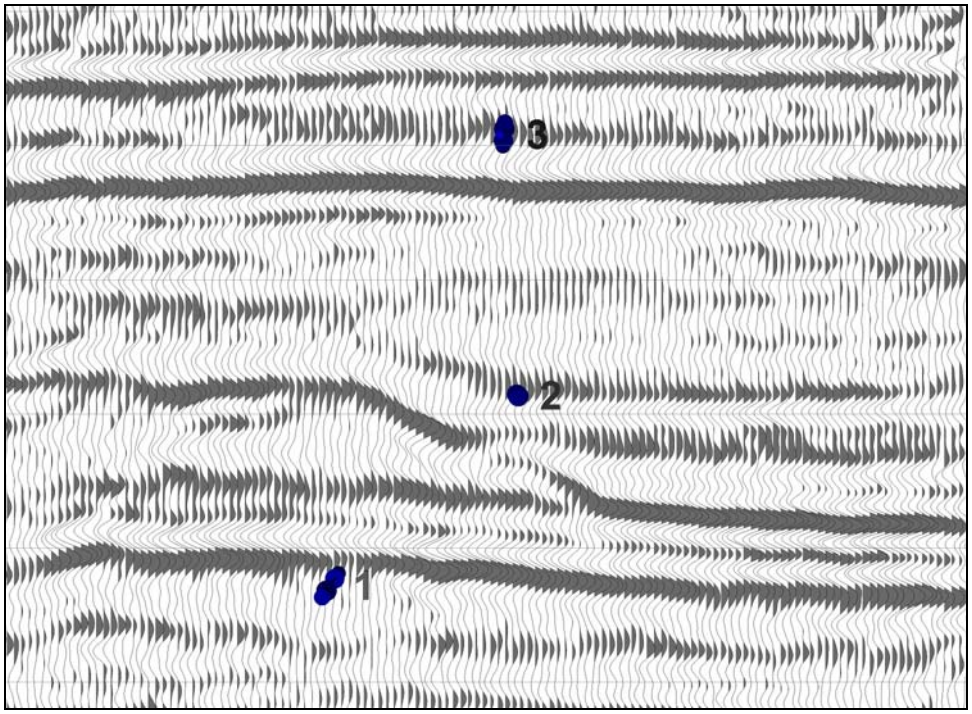


Figure 9.44: Display of the edge of a Leduc reef, data courtesy Divestco Inc. The amplitude of both the platform and the top of the reef decreases at marker 1 but this is not evident on this display.

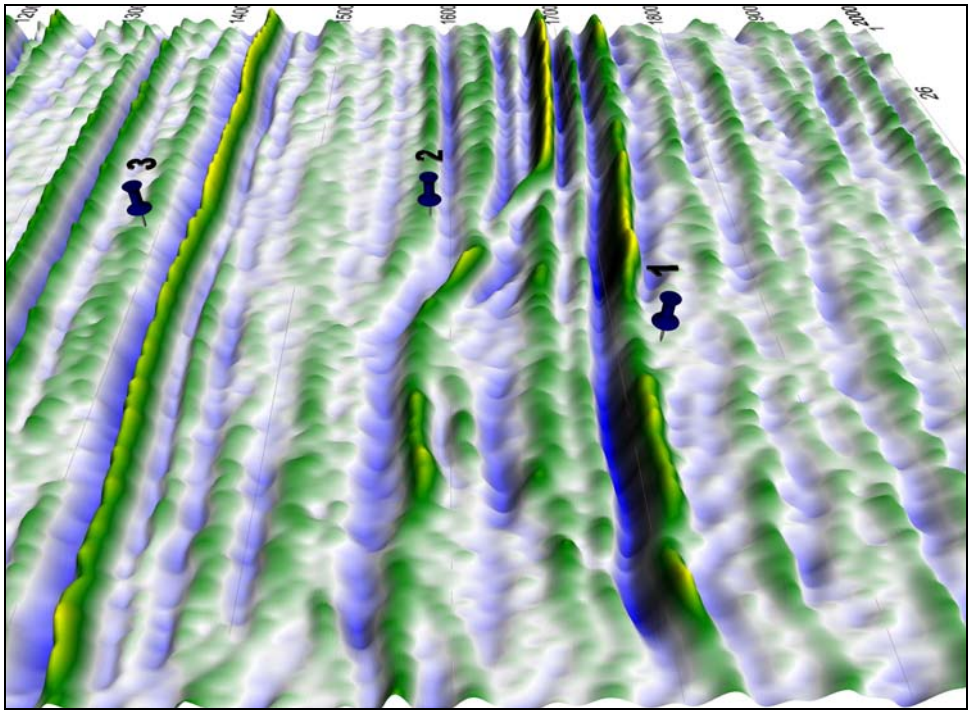


Figure 9.45: The amplitude decrease at marker 1 is obvious on the SeisScape display as is the amplitude structure along the top of the reef. You also perceive the general structure of the low amplitude region between markers 2 and 3 better on this display than you do on the wiggle trace image.

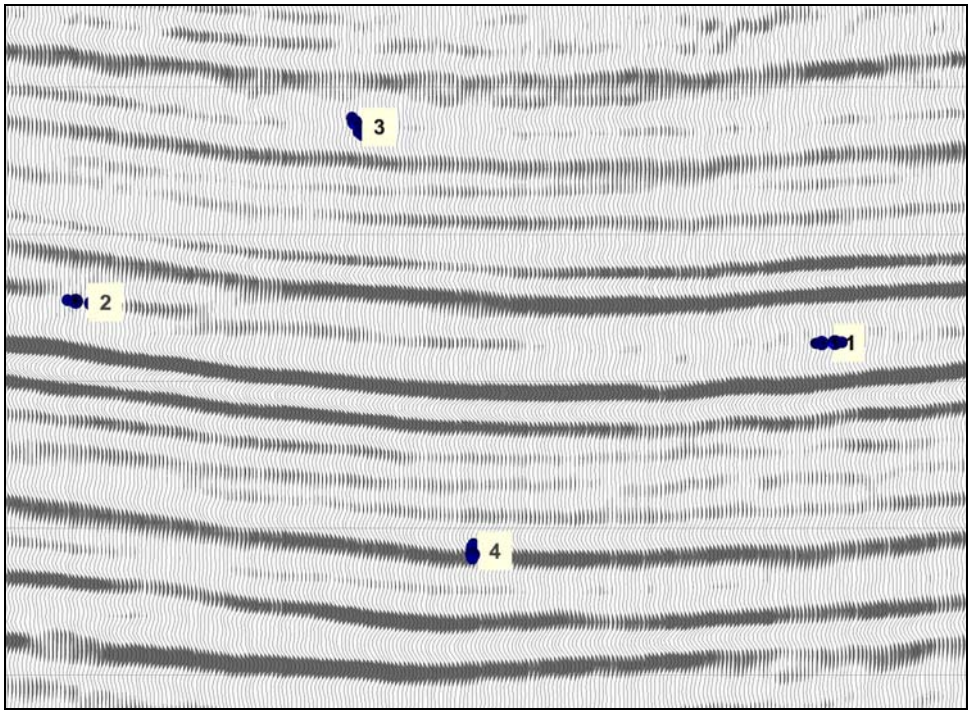


Figure 9.46: Wiggle trace display of a section of Trujillo data. Locations 1, 2 & 3 mark low amplitude features surrounded by higher amplitude events.

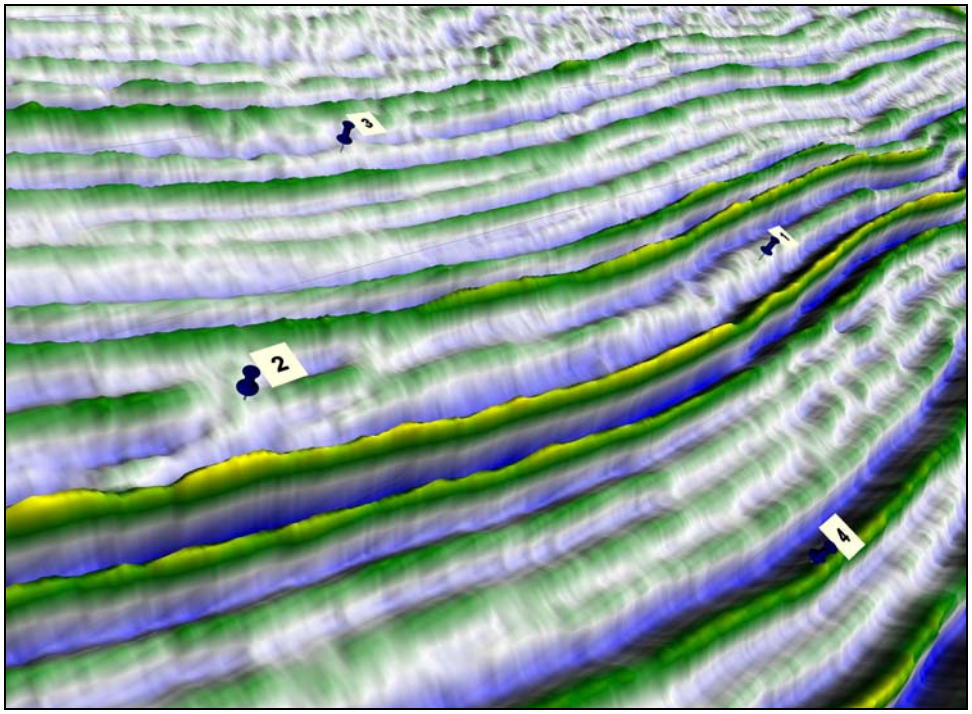


Figure 9.47: Details of the seismic structure is a great deal easier to see on this display, especially between markers 1 and 2 and around marker 3. This zone is surrounded by high amplitude events whose amplitudes are also a lot easier to follow on the SeisScape image than they are on the wiggle trace image.

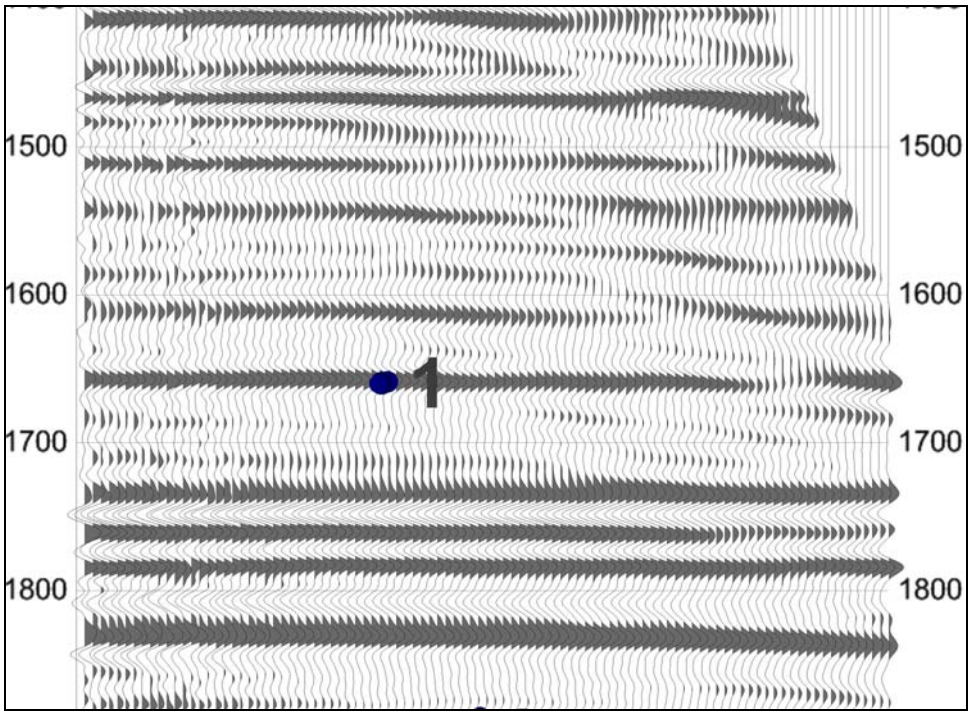


Figure 9.48: Typical common offset record containing a series of low amplitude multiples. In the region between 1570 ms to 1740 ms, there is a series of multiples that are much lower amplitude than the primaries and consequently they are hard to follow

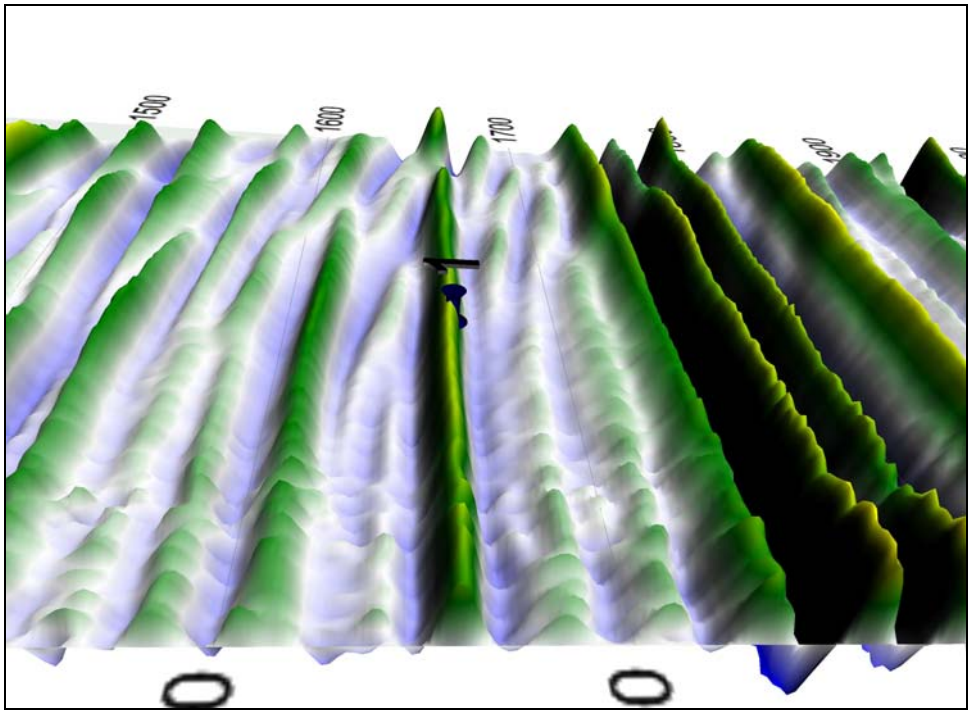


Figure 9.49: Even though the multiples are much low-amplitude than the primaries they are just as easy to see on the SeisScape display. The effect of the multiples as they cross the primaries is also a great deal more noticeable on this display.

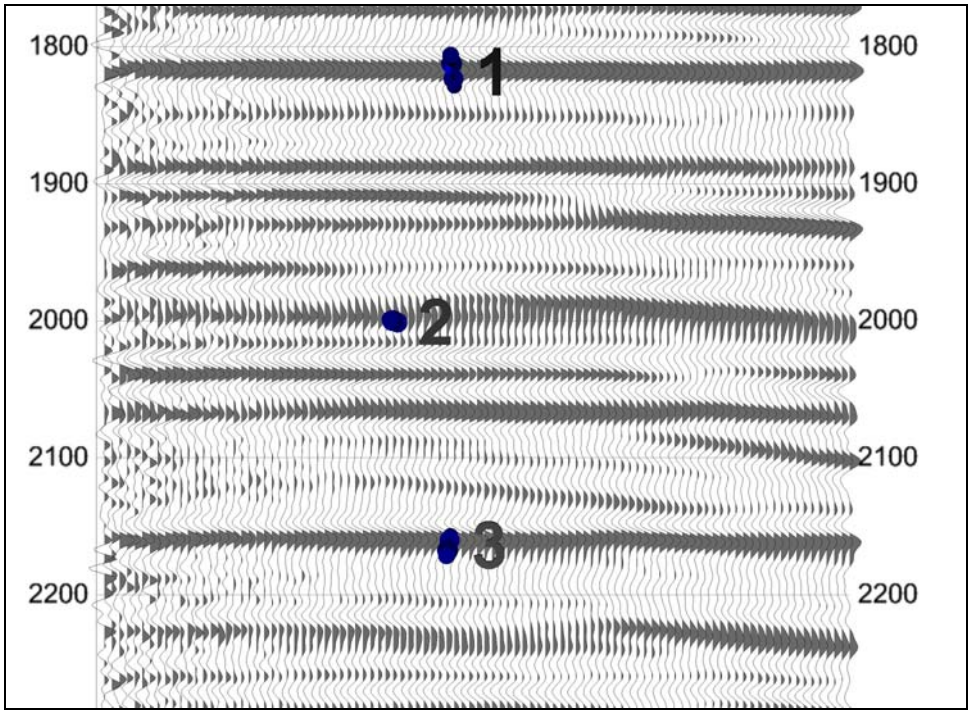


Figure 9.50: A common offset record that shows a series of events with complex offset related amplitude changes. Common offset displays are one of the principal tools that we use to detect and analyze AVO and other offset related effects.

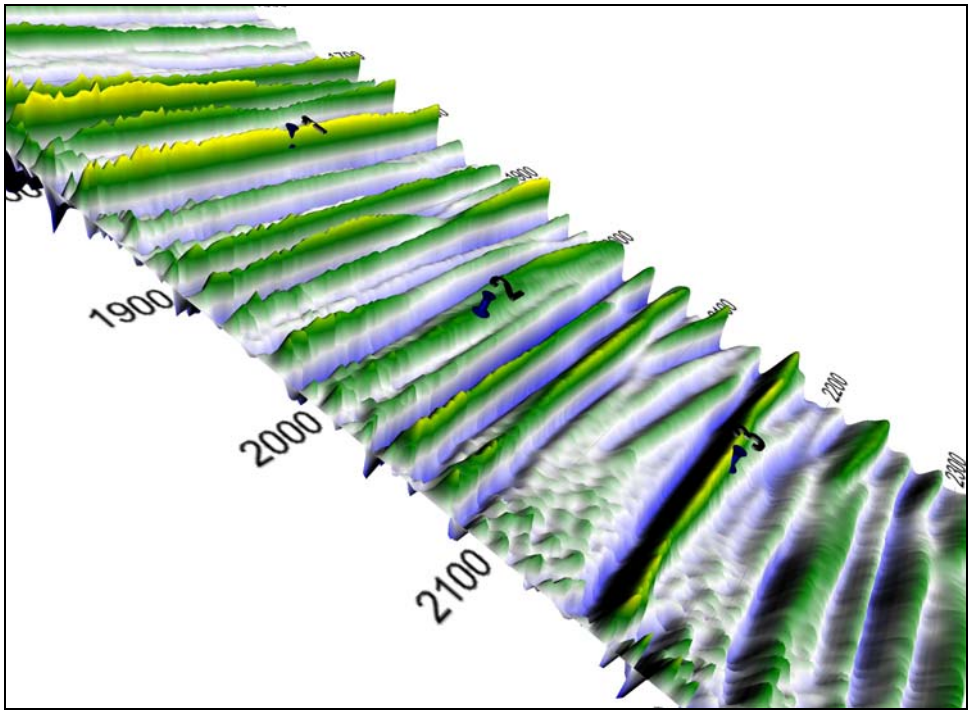


Figure 9.51: The amplitude changes with offset between markers 1 and 2 are far more obvious and easy to understand on the SeisScape display. In addition, the multiples between markers 2 and 3 are a great deal more noticeable as is their effect on the primary amplitudes.

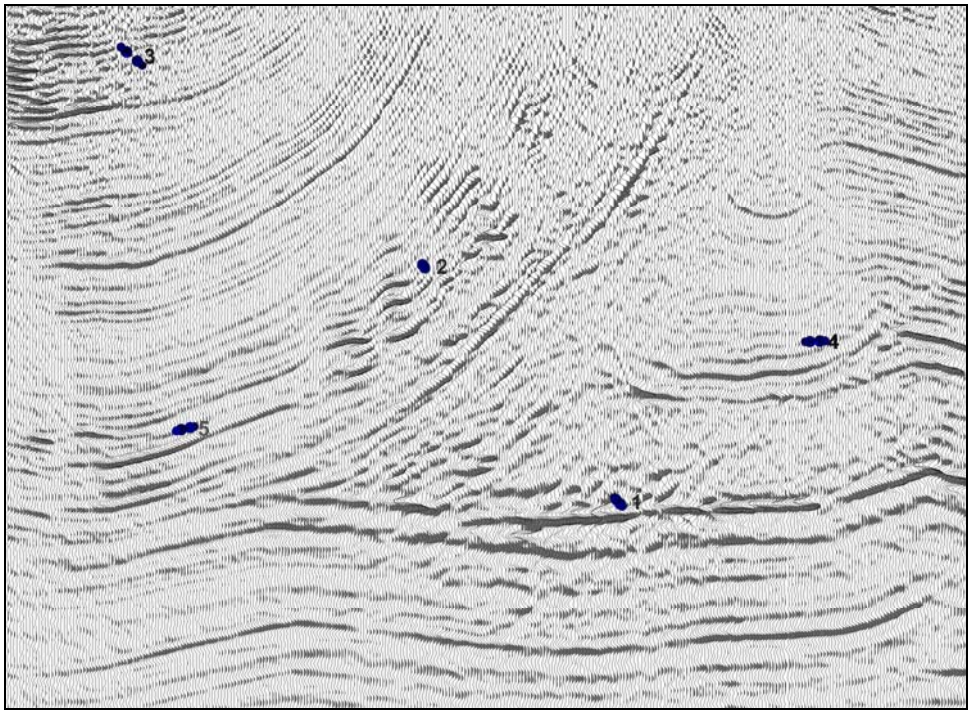


Figure 9.52: A noisy, prestack time migrated Alberta foothills section (data courtesy Husky Oil). Markers 1, 2, 3 and 5 show zones with high amplitude events whose amplitudes are difficult to put into perspective with the surrounding low amplitude events.

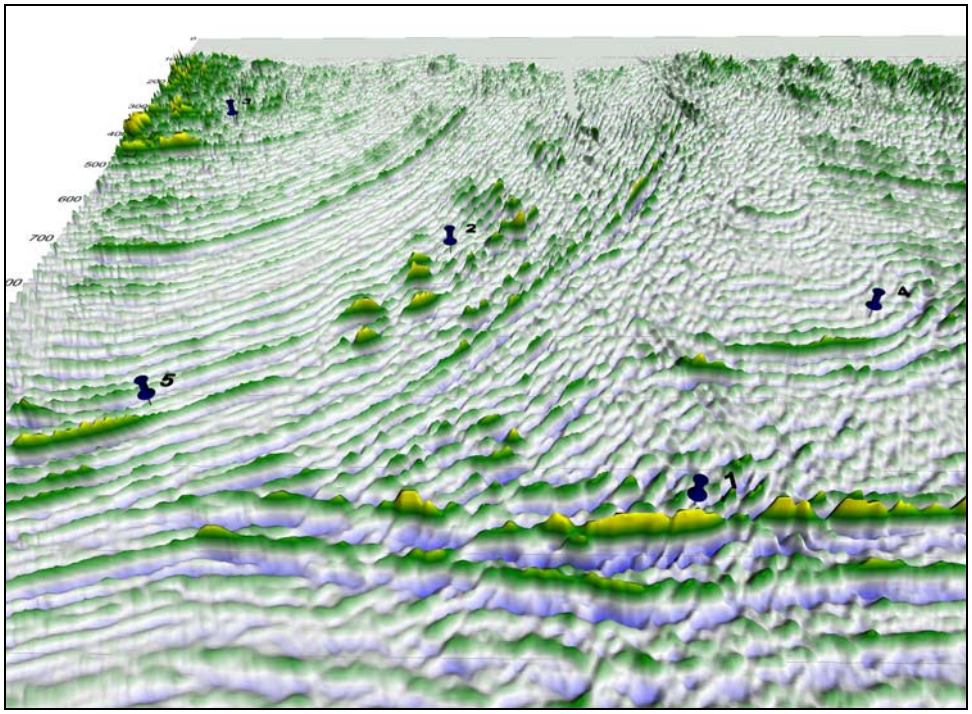


Figure 9.53: It is easier to understand just how high the amplitude of the marked events is on the SeisScape display. The zone below marker 1 also shows more detail on the SeisScape display than it does on the wiggle trace display.

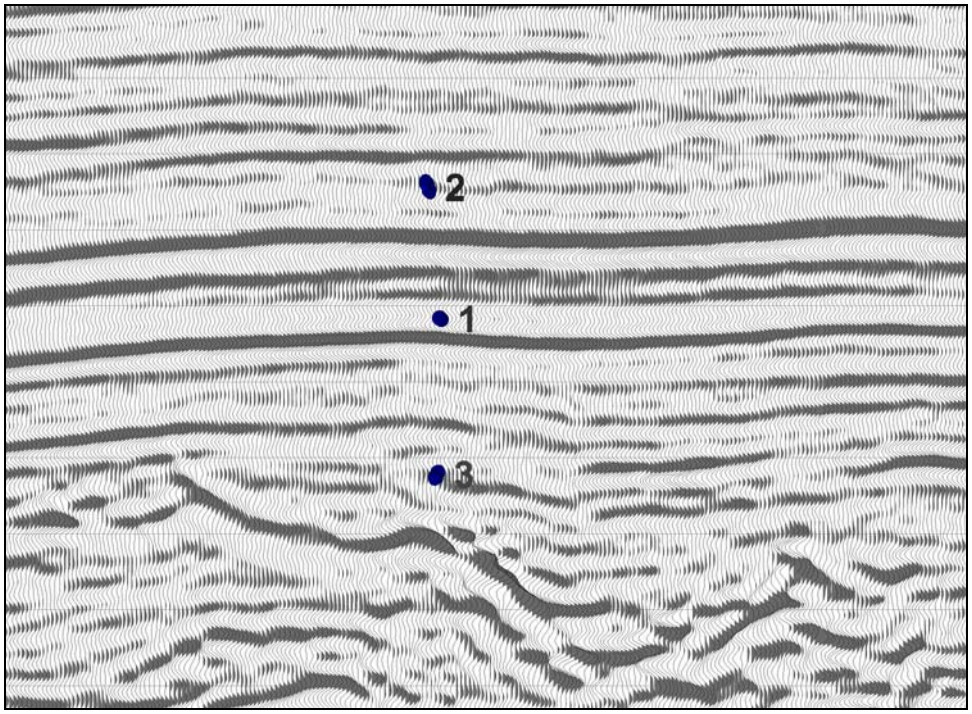


Figure 9.54: An example of relatively noise free data from the Tambo area of Peru, data courtesy PeruPetro. There is nothing specific to look for in this image. I present it as just a typical seismic section, one that contains both structural and stratigraphic changes.

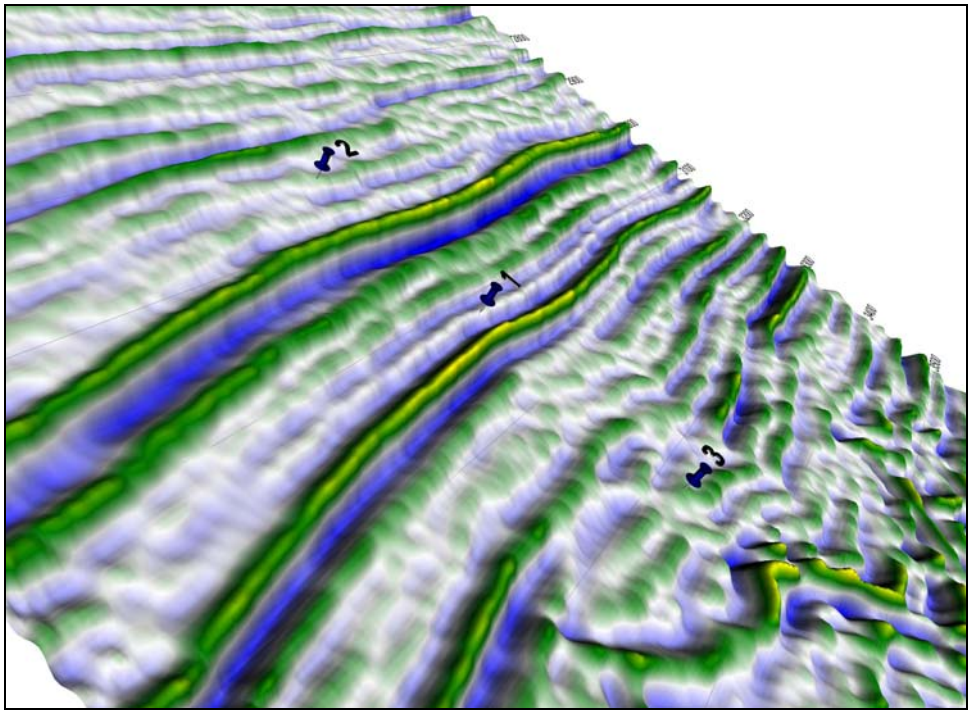


Figure 9.55: This is a typical orientation for a SeisScape display. It is the orientation that I use the most often when viewing seismic. I present it here just to show how a typical seismic section normally appears on a SeisScape display.

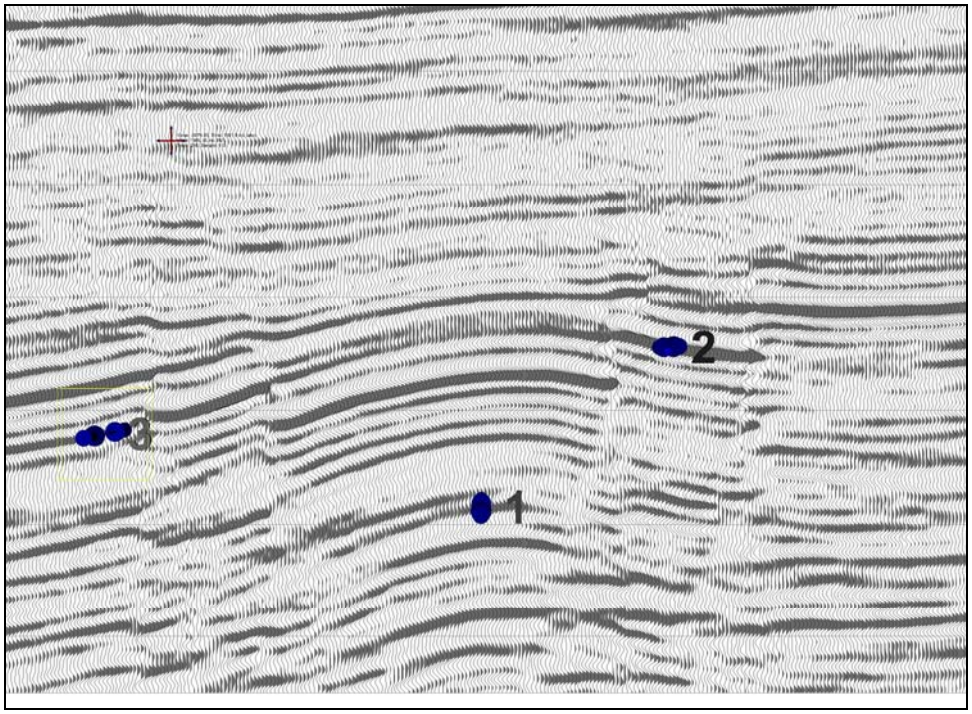


Figure 9.56: A Horst and Graben structure from China (data courtesy unnamed source).

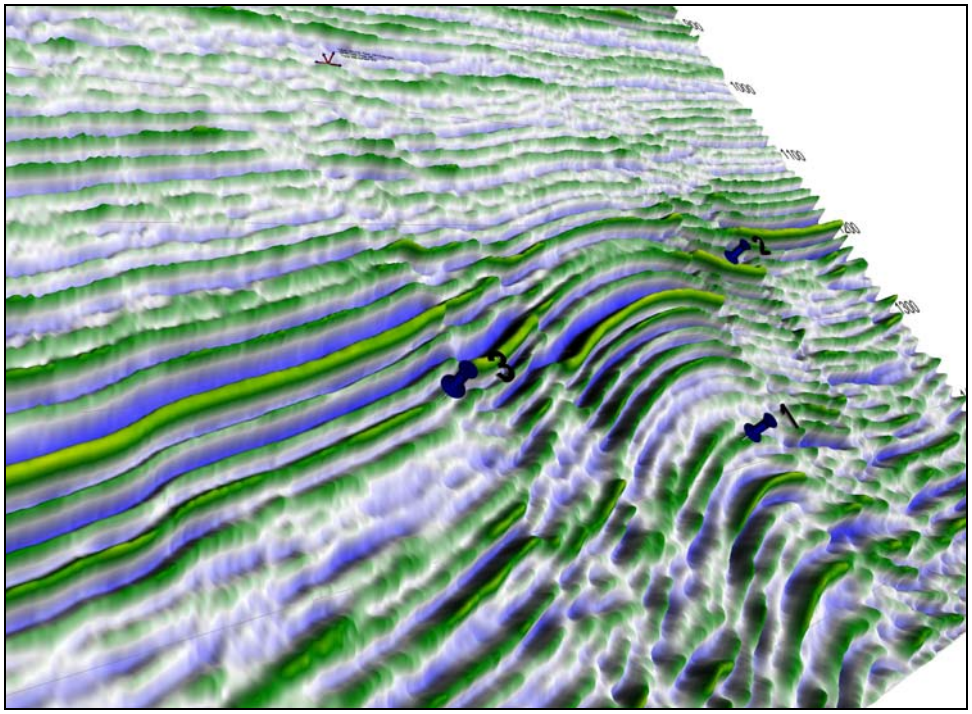


Figure 9.57: As with the previous example, I show this image just to highlight what a familiar seismic display looks like as a SeisScape section.

CHAPTER TEN: ILLUMINATING SEISMIC DATA

We have learned to whittle the Eden Tree to the shape of a surplice-peg,
We have learned to bottle our parents twain in the yelk of an addled egg,
We know that the tail must wag the dog, for the horse is drawn by the cart;
But the devil whoops, as he whooped of old: "it's clever, but is it Art?"

The Conundrum of the Workshops
Kipling

10.1 Introduction

In the fall of 1980, I was living on the outskirts of Calgary, very close to a major dog park called Edworthy Park. I was injured and all I could do at the time was walk so I started hiking through the park on a daily basis. I took different routes through the park but I always ended up walking back along one particular trail. I hiked that park for several months before I realized that it was selfish to walk through a dog park without a dog. To fix the problem I bought a Golden Retriever and named him Captain Kirk. That was his official name but his breeder nicknamed him "Kirk the Jerk", and that name stuck for the rest of his life. For the first year of his life, we walked through Edworthy Park every day, always taking different routes but always returning along the one familiar trail.

After the first year, I realized that Kirk was getting lonely because I had to leave him for most of the day. To make him less lonely (and to protect my furniture), I bought a second retriever and called him Mr. Spock, a name which turned out to be grossly inappropriate for a retriever. For the next two years, we walked through the park on a daily basis and no matter what route we took, we always returned along that same familiar trail. Late in 1983, I realized that I was lonely and to remedy that I met a young woman called Jan, who fortunately took more to the dogs than she did to me. She eventually married them but in a stroke of luck, I got to go on the honeymoon and she decided to keep me. For the next two years, there was one constant element in out lives; everyday the four of us would walk through that park. It did not matter how foul the weather was, everyday we walked through Edworthy park, taking different trails each time but always returning along that one well trodden path.

Late in 1985, we moved to a small community on the outskirts of the foothills and for the next five years, we took our walks out there. Then in 1990, we moved to Kelowna and we lived there for the next eleven years, finally returning to Calgary in 2001. When we came back, we moved into the same general area where we had lived in the 80's. Once again, since we were close to Edworthy Park, we took up our daily routine of walking along the various trails and back along the old familiar one. Jan and I were dog less at the time and just as I did in 1980, we realized that since we were walking through a dog park we might as well get a dog. We bought an English springer spaniel, called him Yoda and he is lying on my feet as I write this. That was in 2003 and almost every day since then we have walked him through the same park, along the same trails and back along the same familiar trail.

And in all the years the we hiked through that park, along that same trail, nothing remarkable ever happened!

Nothing remarkable ever happened, except for one day late in the summer of 2006. On that day, I was walking Yoda back along that same trail. Being late in the day, the light was low to the horizon and as I walked over a small rise in the trail, there right in front of me, illuminated by the late evening sun, was the clear and unmistakable outline of a partially buried stone circle. For those of you who don't know what stone circles are, they are the remains of Indian encampments. The indigenous people in this area lived in teepees and when they erected them, they placed large rocks around the base to hold them down. When they moved on, they simply rolled the rocks off the teepee and left them there, boulders being in no short supply in Alberta. This left behind a circle of stones to mark where they had camped and it is estimated that there are over one million of them in Alberta alone.

Given its proximity to the Bow River, it should come as no surprise that there is at least one stone circle in the park. What is a surprise is that nobody ever noticed it before. Edworthy Park has been a dog-walking park for decades and hundreds of people walk through it every day. I have walked through it and over where the circle lays literally thousands of times. But no one ever noticed it and I would never have noticed it …

Until that one day when the light was at just the right angle and at just the right elevation above the horizon … and then you just couldn't miss it!

This chapter, which covers the subject of lighting and the following chapter, which covers the uses of color, are the final two chapters in my study of the science of visualization. I proved, in Chapter 2, that when primates look at a scene they do not interpret it as a whole. Rather, our visual processing system splits it into parallel neural streams; one that contain purely achromatic information and two that contain purely chromatic information. In Chapters 6, 7 and 8, I showed how we form those streams, what their properties are and how we use them to establish the form and color of an object.

When it comes to determining the form of an object, our main source of information is the achromatic channel. We use the chromatic channel and it is essential to our understanding of form. However, it only provides us with secondary information and even then, the information that it provides is subjective. Primates establish most of their color perception via contrast between the L and the M cones but the distribution of these cones differs greatly between individuals. The chromatic separation between the two types of cones also varies between people and in some people (the author included), one or the other of the cone types is missing entirely. What this means is that from the very lowest levels, color is subjective. What one person sees in a chromatic display another may miss entirely. Therefore, if we attempt to communicate information using purely chromatic displays, we cannot guarantee that the viewer will see it the way that we do, or even that they will see it at all.

By contrast, we establish achromatic information through summation. Our visual processing system combines the inputs from the L and the M cones together along lines of orientation. Therefore, the relative abundance of the two cones or the chromatic separation between the peak of their absorption spectra is irrelevant. Consequently, achromatic information is objective because an achromatic display will produce a similar

stream of information to most viewers. This is not to suggest that all viewers will form the same percepts from a given display. Percepts are a function of both the input streams of information and personal experience and because of that, visualization is ultimately subjective. However, when comparing chromatic input and achromatic input, at least with the latter we all start with the same information.

In the natural world, the terms achromatic information and reflectivity[1] are synonymous. A number of factors, the principals of which are the shape of the object and the direction of the incident light, determine the reflectivity of any given surface. Our achromatic neural circuitry has evolved to interpret this reflectivity. It is so proficient at it that we do not have to see the physical shape of an object to determine what it is. This is a simple but important concept. Consider a simple monochromatic cylinder. If you look at it, side-on then you cannot see the physical shape at all. Nevertheless, you can still tell that it is cylindrical. You can even tell it is a cylinder if you cover one eye and block your stereoscopic vision. That is because the information that you use is the pattern of how the light changes in intensity across the surface. The achromatic system knows what this pattern represents and can construct the percept of a cylinder from it alone; it does not need to see the physical shape at all.

This is why shaded relief (Batson, 1975) and bump mapped (Blinn, 1978) displays effectively communicate form. Both are strictly two-dimensional and yet, as I showed in Chapter 2, they are capable of producing striking percepts. What the percepts are, however, is interesting because different people may form different percepts from exactly the same achromatic visual information. In Chapter 2, I reported on the results of a simple visual experiment where I sent a purely chromatic image and a purely achromatic image of the same object (the Crowsnest Pass) to a number of viewers from different fields. Almost everyone formed some sort of percept from the achromatic image but interestingly, the percept that formed was linked to his or her personal experience.

---

[1] The ratio of the energy of light reflected from a surface to the energy of the incident light.

Viewers who were familiar with satellite imagery identified the image as a mountain range whereas accountants identified it as crumpled paper. This illustrates a fundamental point of perception; that point being the visual system must be trained. Visual information is often ambiguous. I do not know if we are born with a prepackaged set of percepts but I do know that as we mature and encounter new objects; our visual system learns how to interpret them.

This is perhaps why the geophysical community has been so slow to recognize the importance of lighting. There are almost no papers on the subject in the literature. Besides the author's own work there is only one other reference, that being Barnes in a short note to Geophysics (Barnes, 2003). Even in that one reference, Barnes concluded that shaded relief was useful for time slices but was of no value for vertical sections. I show in this chapter, that this conclusion is wrong. Bearing in mind my point about the visual system needing to be trained, in Barnes's defense I believe that he simply arrived at his conclusion too quickly and before his visual system had time to learn how to interpret what it was seeing. I know this from personal experience and I caution the reader not to be hasty when you view the examples at the end of the chapter. I have worked with SeisScape displays for a number of years and yet, as I searched through my data sets for examples of microstructure, I had a hard time finding any. This was because I was working too quickly. I was quickly scanning large amounts of data and I was only looking at each image for a few seconds. Once I slowed down and gave my visual system time to train itself, then the things that I was looking for literally popped out of the screen.

This chapter deals with two subjects. The first is the mechanics of generating and using reflectance seismic images. Reflectance images are just shaded relief images. They are simple to generate but because of the nature of seismic data they are difficult to use effectively and a large part of this chapter describes the techniques that I have developed to make them useful. The second subject that I consider is microstructure, which is the subtle seismic information that you can only see by using lighting. This is where visual training comes into play. By definition, microstructure falls below the visual resolution of

conventional seismic displays. This means that you have never seen it before and consequently your visual system has not had the opportunity to learn how to form percepts from it. Please take time when you view the images and let your visual system train itself. Once you do, you will be surprised at what you see.

10.2 Calculating the Reflectivity of the Seismic Surface

When you simulate the light that reflects from a surface there are three forms of lighting to consider. The first of these is diffuse lighting, which is the form of lighting that has the most connection to physical reality. It is based upon Lambert's Law, which states that for surfaces that are ideally matte (i.e. without shininess), the reflected light is calculated as the cosine between the light vector and the surface normal. The second form of lighting is specular lighting, which makes a surface look shiny by creating highlights. These highlights help the viewer understand the curvature of a surface and make an object appear more realistic. The third form of lighting is ambient lighting, which provides a constant amount of illumination for a surface. It is often used to provide some light to the underside of surfaces that otherwise would not receive any light at all.

Of these three, I only consider diffuse lighting in this thesis. I do not use ambient lighting because I apply the diffuse component to both the front and the back surfaces of SeisScape displays. I also do not use specular lighting because I experimented with it early in my research and determined that it did not add any significant information but that did it introduce a significant level of visual clutter. I will reexamine specular lighting in the future but in the all the displays that follow, I treat the seismic surfaces as being totally matte and without shininess.

10.2.1 Diffuse Lighting

When I calculate diffuse lighting, I make several assumptions about both the light and the surface. These assumptions are:

- The light rays are parallel, which means that the light source is infinitely distant from the surface as in the sun.
- The light rays strike all parts of the surface equally, which means that there are no shadows and that troughs receive the same illumination as peaks.
- The surface is totally matte and has no shininess.
- The backside of the surface receives the same light as the front side.

The intensity of the diffuse lighting is calculated from Lambert's Law, which is given by:

$$i_{diff} = \vec{n} \bullet \vec{l} = \cos\theta \tag{10.1}$$

Where: $\vec{n}$ is the surface normal and $\vec{l}$ is the light vector.

At first glance (10.1), which is simply the dot product between the normalized light vector and the normalized surface normal, appears to hold the record for being the simplest equation in geophysics. As with all things seismic, however, if it looks simple then you simply are not looking hard enough. There are parallels with the previous subject of tessellation and whereas the equation itself is simple, calculating and correctly scaling, the surface normal is not.

10.2.2 Calculating the Surface Normal

The surface normal at any point on the seismic surface is the vector perpendicular to the tangent to the surface at that point. When you calculate this normal there are two factors that you must consider. The first is that surface must be tessellated correctly. As I showed in Chapter 9, the tessellation defines the surface. Therefore, all of the factors that

I discussed in Chapter 9 also come into play when calculating the normal. The second is that a seismic sample may contribute to a maximum of six triangles. The surface normal for the seismic sample is the average of the face normals from these triangles.

In Chapter 9, I established that for low-dip sections, there are two ways to tessellate the same four points. One schema favors up-dip events whereas the other favors down-dip events. The adaptive tessellation techniques that I developed, determined for each sample, which of these two schemas was locally appropriate. To calculate the normal for a given sample it is necessary to calculate the face normals for each of the triangles to which it contributes. Depending upon whether the tessellation is up-dip or down-dip at the sample, the faces averaged in the normals calculation will change. In Appendix B, I derive two sets of equations for simplifying the calculation of the surface normal, one each for the up-dip and down-dip schemas.

For the down-dip tessellation schema (see section B.1 for the derivation and pixel shader code), the normals are calculated as:

$$\begin{aligned} N_0x &= \beta(2z1 + z0 + z5 - (z3 + 2z7 + z8)) \\ N_0y &= \alpha(2z3 + z0 + z7 - (z1 + 2z5 + z8)) \\ N_0z &= 6\alpha\beta \end{aligned} \tag{10.2}$$

For the up-dip tessellation schema (see section B.2 for the derivation and pixel shader code), the normals are calculated as:

$$\begin{aligned} N_0x &= \beta(2z1 + z3 + z2 - (z5 + 2z7 + z6)) \\ N_0y &= \alpha(2z3 + z6 + z1 - (z2 + 2z5 + z7)) \\ N_0z &= 6\alpha\beta \end{aligned} \tag{10.3}$$

Where:

α is the trace spacing.

β is the sample interval.

z0 – z8 are the samples surrounding the target sample, which is given as *z4*.

Both of these calculations reduce the calculation of the averaged face normals down to a pair of simple algebraic expressions, one each for the x and y component of the normal vector (the z component is constant). These two calculations are readily implemented in pixel shader code and I include that code along with the derivation.

Equations (10.2) and (10.3) are both based upon broader assumptions about the local dip than are the adaptive tessellation schemas developed in Chapter 9. If we consider each set of four adjacent points as a cell then the adaptive tessellation schemas tessellate each cell independently. When calculating the surface normal, however, each sample contributes to four cells, each of which may be tessellated differently. For simplicity, I assumed that the local dip in each of the cells was the same. In practice, this means that in regions where the direction of local dip changes rapidly, that the surface normals might not accurately correspond to the tessellated surface.

In Chapter 9, I also developed a tessellation schema for high-dip situations. That technique remains experimental, however, and consequently I have not yet developed an effective schema for generating its corresponding normals. Once I implement the high-dip schema in the geometry shader, I will develop an appropriate technique for calculating the normals.

10.3 Scaling the Normals

Equations (10.4) and (10.5) are easily programmed in the pixel shader and can be output to a two-channel floating point texture for subsequent use in lighting calculations. By themselves, however, they represent the x and y components of the “raw” normals. The three axis of the display are all in different units, the x-axis is in distance units, the y-axes is in milliseconds and the z-axis is in amplitude. Using the normals as calculated, therefore, results in lighting that varies greatly between sections and which may be washed out or too dark. Before you can use the normals, they must be modified to compensate for the three axes being different.

10.3.1 Amplitude Scaling

The first compensation is to normalize the seismic amplitudes during the calculation of the normals. Different seismic sections can have vastly different amplitude ranges. Figure 10.3 shows a tie between two seismic lines that have different amplitude ranges. The section on the left side of the display has amplitudes in the range ±32768 whereas the section on the right has amplitudes in the range of ±12608. Left uncompensated, the lighting effect for both lines would be significantly different.

To correct for this, during the calculation of the normals, I divide each seismic amplitude by five times the standard deviation of the data set. I use this value, rather than the maximum value of the data, because seismic sections often contain anomalously high amplitudes that are not representative of the data as a whole. Consequently, normalizing to the maximum amplitude often produces sections whose lighting appears washed out. As the reader can see, using the standard deviation as the basis for the normalization results in sections whose lighting is much better balanced.

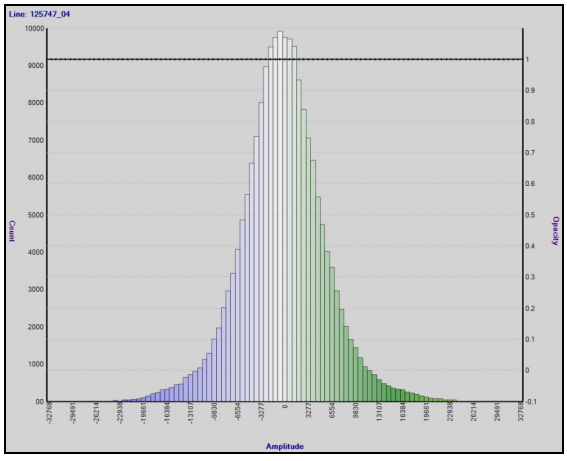


Figure 10.1: Amplitude statistics for the seismic line on the left side of Figure 10.3.

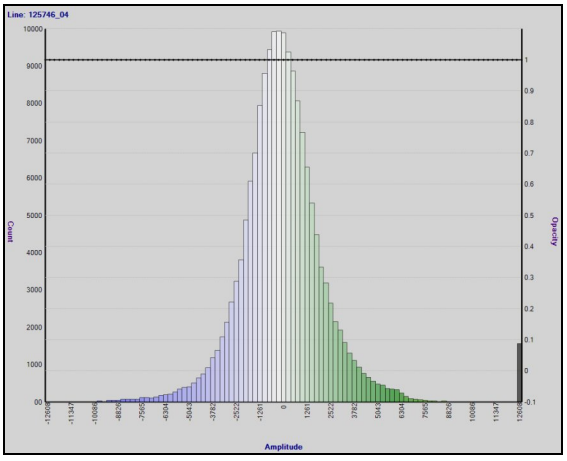


Figure 10.2: Amplitude statistics for the seismic line on the right side of Figure 10.3

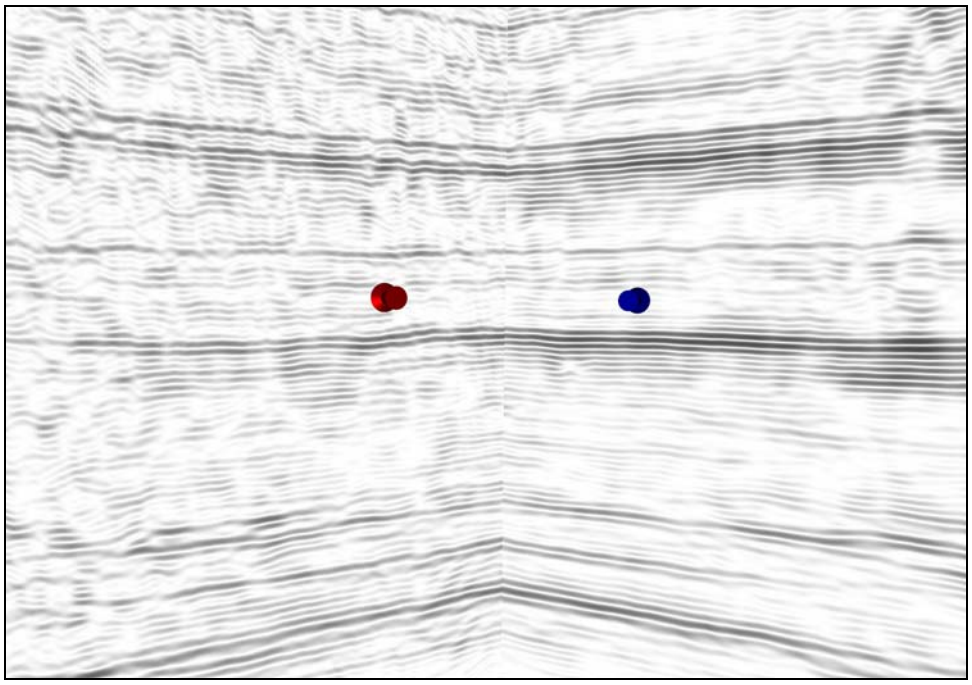


Figure 10.3: Lighting only display of a tie between two seismic lines (apparent height of 10). Both lines have the same processing, sample interval and trace spacing. Amplitudes for the section on the left (red pin) are approximately three times higher than those of the section on the right (blue pin). Even though the amplitudes are significantly different, the amplitude normalization used when calculating the normals has effectively balanced the lighting between the two.

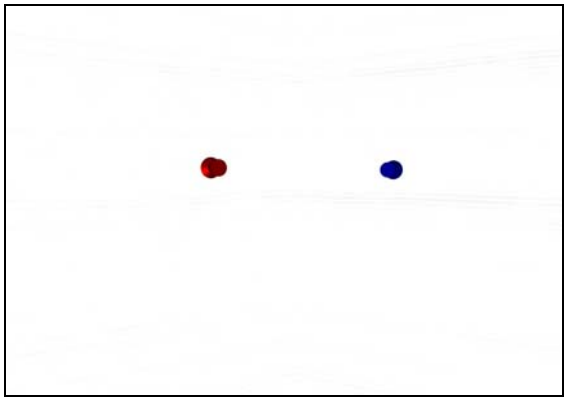


Figure 10.4: The same display shown in Figure 10.3 but using an apparent height of 1. Note that the lighting is washed out.

The amplitude correction normalizes amplitudes to the range of approximately ±1. The resultant normals, if used directly, produce a washed out display similar to Figure 10.4. To correct for this, I apply a multiplier to the x and y components of the normals during the calculation of the lighting. This multiplier is the "apparent height" of the data. Applying it makes the lighting look as if it was calculated from a surface with maximum height equal to the multiplier. The reason I apply it during the calculation of the lighting rather than during the calculation of the normals is so that the user can change it to make the lighting appear more or less pronounced.

10.3.2 Time Axis Scaling

Equations (10.2) and (10.3) define the complete normal vector but this is not how I calculate them in the pixel shader. I calculate the normals by ignoring the sample interval and trace spacing. The resultant equations for the x and y components then become:

Down-Dip Normals:

$$\begin{aligned} N_0x &= 2z1 + z0 + z5 - (z3 + 2z7 + z8) \\ N_0y &= 2z3 + z0 + z7 - (z1 + 2z5 + z8) \end{aligned} \tag{10.4}$$

Up-Dip Normals:

$$\begin{aligned} N_0x &= 2z1 + z3 + z2 - (z5 + 2z7 + z6) \\ N_0y &= 2z3 + z6 + z1 - (z2 + 2z5 + z7) \end{aligned} \tag{10.5}$$

The x and y components are thus calculated as if we were displaying the seismic data on a unit grid. We cannot use these normals “as is” because different seismic lines have different trace spacings and sample intervals. Ignoring these, which (10.4) and (10.5) do, leads to lighting effects that change dramatically from line to line. The x, y and z components of the normals must be scaled by the trace spacing and the sample interval; that much is obvious. What are not obvious is what the sample interval is and what effect it has upon the reflectivity of the surface. Consider Figure 10.5 and Figure 10.6, which are reflectivity (vertical lighting) SeisScape displays of the same data.

Figure 10.5: Reflectivity SeisScape display of a section of Trujillo data. The numbers are for reference only. Lighting is vertical and the vertical exaggeration equals one.

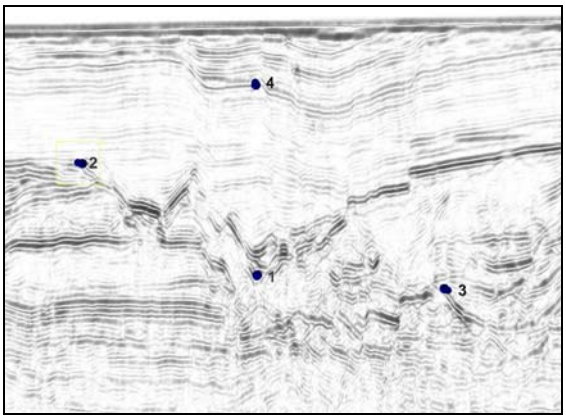


Figure 10.6: Exactly the same display shown in Figure 10.5 but with a vertical exaggeration of five.

These two displays are identical in all but one respect. Figure 10.5 has a vertical exaggeration of one whereas Figure 10.6 has a vertical exaggeration of five. I have included the displays to illustrate a critical element of SeisScape displays; their effectiveness is tied to the vertical exaggeration. Figure 10.5 is the "natural" display of this data set. One millisecond along the vertical axis occupies the same screen space as one meter in the horizontal direction. I consider that this display is ineffective because it is too compressed. Figure 10.6, which has a vertical exaggeration of five, is far more effective. It looks more natural and we can identify features, and in particular dipping features, far more clearly.

The reason for this is found in the effect that the vertical exaggeration has upon the normals. Recalling equations (10.2) and (10.3), both the x and the z component of the normal are affected by the sample interval. The sample interval, in this case, is normally considered as the sample interval of the data. For SeisScape displays, however, if we just used the data sample intervals then the calculated reflectivity would not change as the vertical exaggeration changed. This is clearly undesirable because the reflectivity should reflect the visual appearance of the section. For us to calculate the reflectivity correctly, so that it matches the visual appearance of the section, then we must modify the sample interval thus:

$$\beta_{eff} = \beta \times V_e \tag{10.6}$$

Where:

β_{eff} is the effective sample interval.

V_e is the vertical exaggeration of the display.

This is an almost trivial, but vitally important, correction to apply to the normals. One of the most important effects of lighting is that it highlights dipping microstructure events. These events are not visible unless you apply (10.6). The reason for this is that changing the effective sample interval changes the relative strength of the x and y components of the surface normal. As you increase the vertical exaggeration, the x component of the normal, which is derived from trace to trace amplitude changes, becomes increasingly dominant.

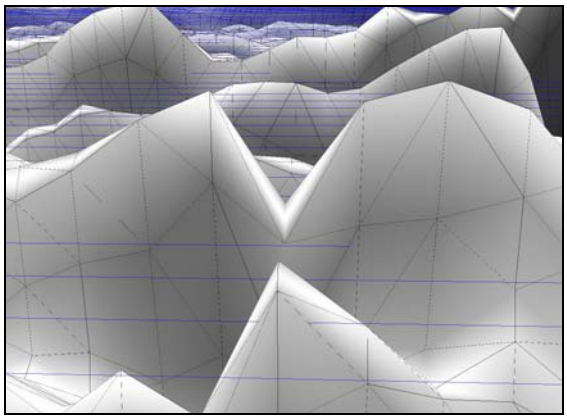


Figure 10.7: Achromatic SeisScape display of a notch in a seismic event. The blue lines are 10ms timing lines. Vertical exaggeration equals one. The trace-to-trace amplitude change across the notch is approximately the same as the sample-to-sample amplitude change across the event. The reflectivity change is less pronounced, however, because the amplitude change is spread over a greater effective distance.

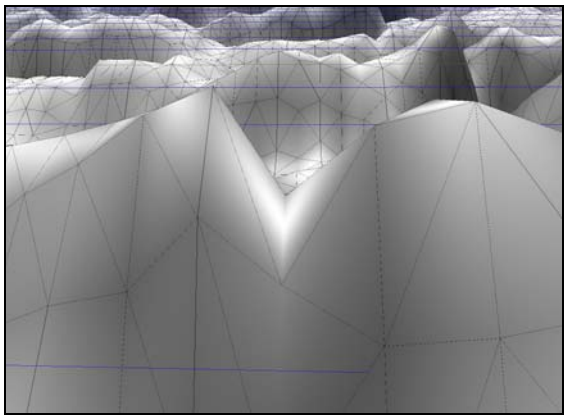


Figure 10.8: The same display shown in Figure 10.7 but with a vertical exaggeration of five. The reflectivity change across the notch is more pronounced in comparison to the sample-to-sample change because both are now spread over approximately the same effective distance.

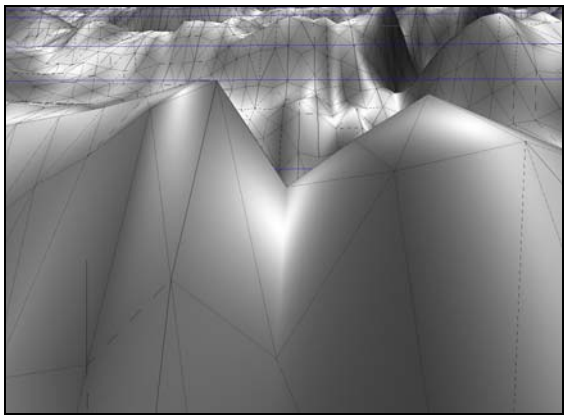


Figure 10.9: The same display shown in Figure 10.7 but with a vertical exaggeration of ten. The reflectivity change across the notch is now much more pronounced than the sample-to-sample change.

Dipping microstructure events are generally lower in amplitude than macrostructure events and you can only observe them as trace-to-trace perturbations. As the vertical exaggeration increases, the effect of the perturbations becomes increasingly pronounced. I illustrate this effect in Figure 10.7 to Figure 10.9 inclusive. Figure 10.7 is a reflectivity SeisScape display ($V_e = 1$) of a notch in a macrostructure event. The amplitude change across the notch is roughly the same as the sample-to-sample amplitude change across the event. However, the reflectivity change across the notch is less pronounced because the sample-to-sample change occurs over a much shorter distance on the display.

You can see in Figure 10.8 ($V_e = 5$) and Figure 10.9 ($V_e = 10$), however, that as the vertical exaggeration increases; the reflectivity change across the notch approaches and eventually exceeds the change across the event. This makes dipping features more pronounced as I illustrate in Figure 10.10 and Figure 10.11. Figure 10.10 is a reflectivity display with $V_e = 5$. Immediately to the left of the light source indicator is a clearly visible, sharply dipping event. This event is completely missing on Figure 10.11, which is exactly the same display but with $V_e = 1$.

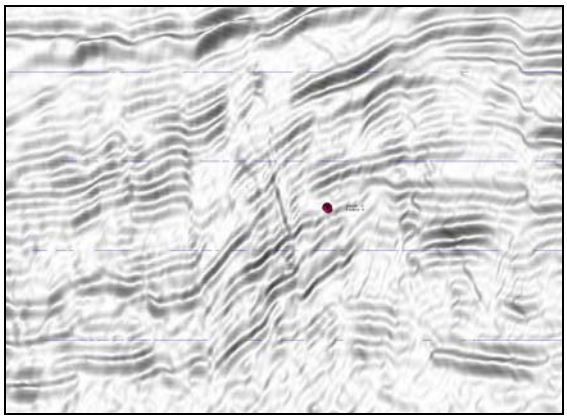


Figure 10.10: Reflectivity display ($V_e = 5$). Blue lines are 100ms timing lines. Immediately to the left of the light source indicator is an unmistakable, sharply dipping feature.

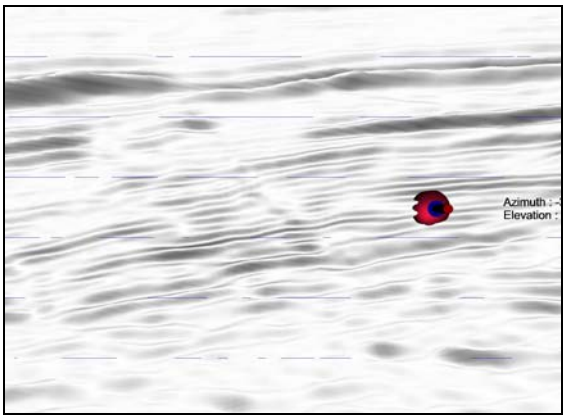


Figure 10.11: Reflectivity display ($V_e = 1$) of the same data shown in Figure 10.10. The dipping feature is completely missing in this display.

10.3.3 Combined Normal Scaling

Increasing the vertical exaggeration has two effects upon the surface normal. The first is that it changes the relative strength of the x and y components of the vector and thus makes dipping events more prominent. The second effect is that it also increases the z component of the vector and this has the effect of reducing the overall lighting effect. To maintain the same relative level of reflectivity as the sections V_e is changed, the z component must be scaled by the original sample interval and not the effective sample interval.

The final scaling applied to the normal during rendering then becomes:

$$\begin{aligned} Nx &= N_0 x \times \beta_{eff} \times Z \\ Ny &= N_0 y \times \alpha \times Z \\ Nz &= 6\alpha\beta \end{aligned} \tag{10.7}$$

Where Z is the apparent height of the display as discussed in 10.3.1.

10.4 Reflectance vs. Frequency Content

The objective of applying the scaling given by (10.7) is to balance the overall level of reflectance between seismic lines with different amplitude ranges, trace spacing's and sample intervals. What this scaling cannot do is balance the reflectance between seismic sections with different frequency contents. Figure 10.13 is a fence diagram of a tie between two seismic lines. The line on the left is a 2D line and I show its amplitude spectrum (600ms to 1450 ms) as the blue line in Figure 10.12. The line on the right is part of a crossline from a 3D and I show its amplitude spectrum as the black line in Figure 10.12.

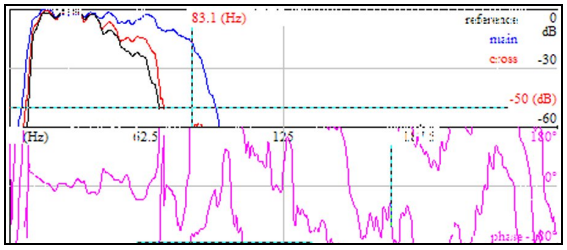


Figure 10.12: Frequency spectrum for the line on the left of Figure 10.13 (blue) and for the line on the right (black).

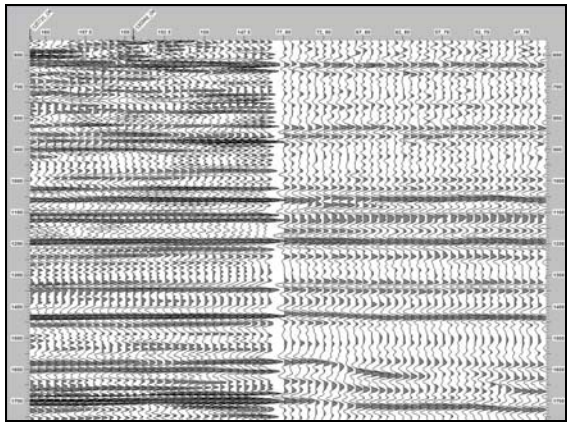


Figure 10.13: A fence diagram between a 2D line (left side) and an inline from a 3D (right side).

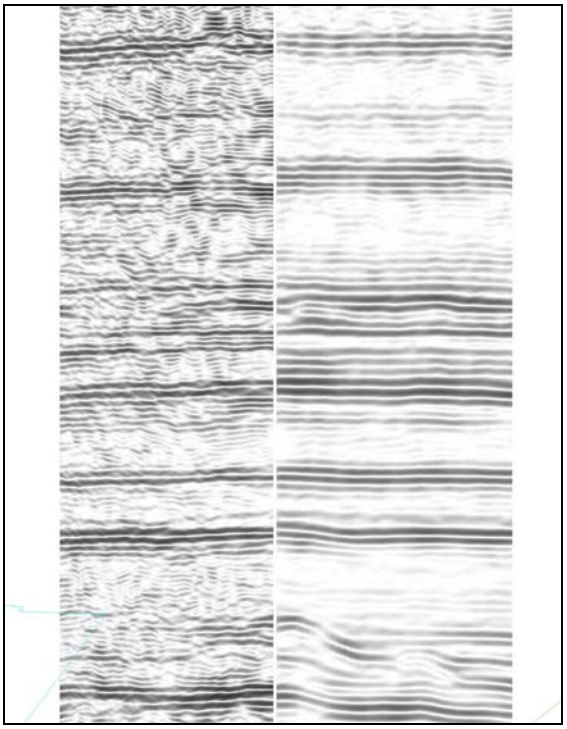


Figure 10.14: A reflectance display of the same data shown in Figure 10.13. Note that the overall magnitude of the reflectance is lower for the 3D line on the right.

Both the amplitude spectra and the wiggle trace display show that the 2D line is higher frequency than is the 3D crossline. The spectrum for the 3D line starts to drop off at around 60 Hz whereas the 2D line does not start to decline significantly until approximately 83 Hz. This change in frequency content has an affect on the overall level

of the reflectance on the two displays. The reflectance of the 2D line is more pronounced than for the 3D line, which in some places appears to be washed out. This is not a critical effect but it is one that you will run into when you work with seismic lines of different vintages.

There is a simple reason why the overall level of the reflectance decreases with frequency. As the dominant frequency of an event decreases, the event becomes more spread out in time and this decreases the slope across the event. Since the surface normals are perpendicular to the slope, this has the effect of flattening out the normals, which reduces the impact of the lighting. This is not a serious effect because you can compensate for it by simply increasing the apparent height of the lighting on lower frequency sections. I only discuss it here to explain why this manual correction is often necessary.

10.5 Seismic Microstructure and the Corrigan Effect

The purpose of this chapter is to develop techniques to expose seismic microstructure. Microstructure constitutes an entire level of seismic detail and it only becomes visible with the appropriate lighting. The reflectance from a seismic surface depends primarily upon two things. The first of these is the surface normal that I discussed in the previous section. The second is the light vector, which is the direction of the light relative to the surface. Although the surface normal is important in the reflectance calculation (10.1), it is a constant and by itself, it does not reveal microstructure. The most important element for revealing microstructure is the direction of the light vector, which I discuss in this section.

Consider Figure 10.15 and Figure 10.16 which are a wiggle trace display and a SeisScape display respectively of an area of obviously faulted events (data courtesy PeruPetro). There are two major faults in particular; the first is immediately to the right of the #1 whereas the second is immediately to the left of #2. There are no obvious fault plane reflections visible on either display. This is not surprising for the wiggle trace display because, as I discussed in Chapter 4, wiggle trace displays have very low

apparent resolution. What is surprising, however, is how little difference there is between the apparent resolution of it and the corresponding SeisScape display of Figure 10.16.

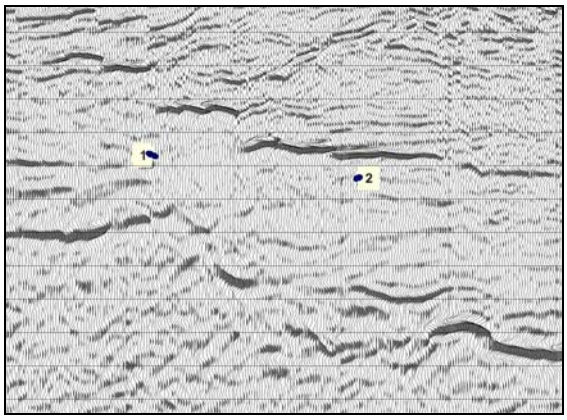


Figure 10.15: Wiggle trace display of a faulted data set from Trujillo. There are numerous major faults, in particular to the right if #1 and to the left of #2. On this display, however, there is little evidence for a fault plane reflection off either fault.

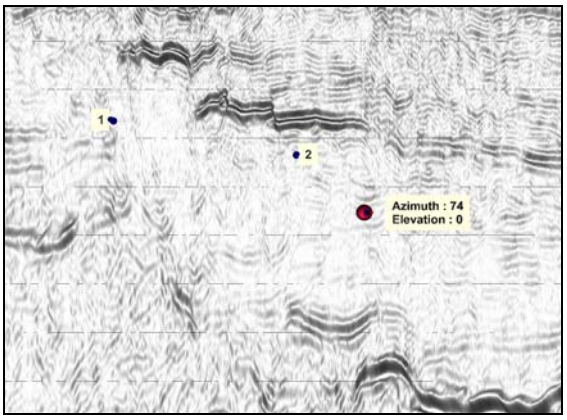


Figure 10.16: A reflectance only SeisScape display of the data that I show in Figure 10.15. The light vector is at an angle of zero degrees to the vertical and whereas it effectively highlights the macrostructure there are no obvious fault plane reflections.

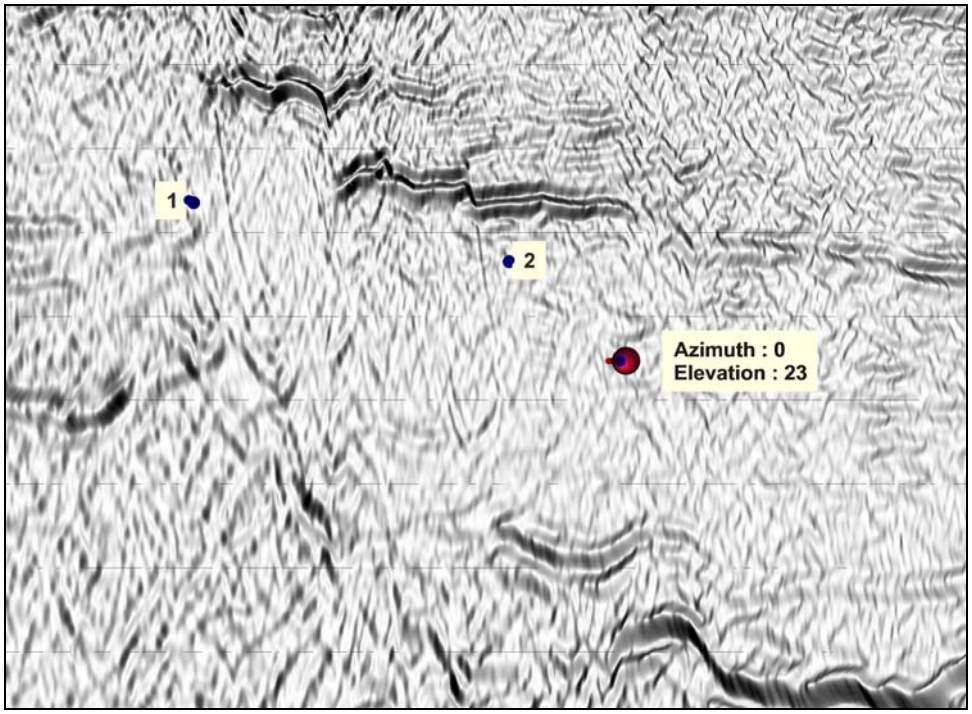


Figure 10.17: Reflectance only SeisScape display but with the light vector at an angle of 23 degrees to the normal and directed from left to right. You can now clearly identify fault plane reflections to the right of #1 and to the left of #2 (as well as other features).

Given that the microstructure of a section is, by definition, the collection of seismic events that are not visible on wiggle trace displays, you would have to say that the SeisScape display is revealing very little of it. Almost everything that you can see on the SeisScape display is also visible on the wiggle trace display and vica-versa. By comparison, however, you can definitely see fault plane reflections on Figure 10.17 as well as a myriad of other, less meaningful features. The only difference between the two SeisScape displays is the direction of the light vector.

On the first display, I directed the light along the z-axis (zero degrees to the vertical). This simulates midday lighting conditions and it illuminates all surfaces equally, irrespective of their time-dip. Midday lighting produces flat and relatively featureless reflectance. For any given seismic event, its up-slope and down-slope sides receive exactly the same illumination regardless of the dip of the event. The reflectance contrast across an event is then determined only by the steepness of the event, which is directly related only to its amplitude. What this means is that under midday lighting conditions low amplitude features become almost invisible because the change in reflectance across them is very low. We perceive high-amplitude macrostructure events but the subtleties escape us.

By contrast, I directed the light vector on Figure 10.17 at an angle of 23 degrees to the vertical and an azimuth of zero degrees to the x-axis (i.e. from the left to the right). The two fault plane reflections are now clearly visible as both appear as prominent dark stripes. There is nothing new here because it is a well-known feature of shaded relief displays that lighting highlights features perpendicular to it. The reflectance calculation of (10.1) is simply the dot product between the normal and the light vector. Given that both vectors are normalized, the calculation reduces to being the cosine of the angle between the two. From spherical geometry, it is obvious this cosine is at a maximum when the z-axis rotation of the normal is the same as the rotation of the light vector and minimum when it points the other way (proof left to the reader).

What this means in practice is that our ability to see microstructure depends directly upon the orientation of the light vector. Microstructure is, by definition, extremely subtle

and we start out by not seeing it or knowing that it is there. Only by carefully rotating the light vector around the z-axis can we make it visible. To the best of my knowledge, this tendency of lighting to highlight perpendicular events does not have an official name, at least in geophysical circles. In a seismic context, I have used the term the "Corrigan Effect" to describe it since 2001. I chose this name after Mike Corrigan, a geophysical consultant, first showed me how important the light direction was in identifying multiples on Ostrander gathers.

10.6 Microstructure Classification

To this point in this chapter, I have discussed how to generate and scale the surface normals that we use when calculating the surface reflectance. I have also shown how our ability to detect microstructure is dependant upon the orientation of the lighting. These are the technical considerations of lighting and they are important. However, they are not as important as the microstructure that they reveal. That seismic microstructure exists is undeniable. If you compare any of the SeisScape sections that I have shown with its corresponding wiggle trace display, you will quickly see that the SeisScape display reveals far more of the subtle seismic detail. However, how much of this detail is geologically significant and how can we determine what is relevant and what is noise? I address the first of these questions in this section by introducing three broad classifications of microstructure, specifically:

1. Noise Trains.
2. Enhanced Resolution.
3. Fault Plane Reflections.

10.6.1 Noise Trains

For any given seismic line, as you manipulate the light source in a SeisScape view the things that will attract your attention first are the noise trains. Recalling the discussion on how primates use their achromatic visual system to determine form, we are very proficient at detecting linear features. Our ability to do that is one of the principal reasons

why SeisScape displays are effective. Unfortunately, the most prominent linear features on any migrated seismic section turn out to be byproducts of the migration algorithms and if you do not identify them correctly, they can undermine the validity of an interpretation.

Consider Figure 10.18, which is a wiggle trace display of a subbasement section of data. You can clearly identify the basement structure at the top of the display but beneath it, there is nothing but multiple energy and noise trains. Throughout the subbasement area you can see steeply dipping linear noise trains but on this display they are relatively short in duration and they do not appear to extend through the basement and into the main section. Contrast this with Figure 10.19, which is the SeisScape display of the same data. The lighting has done an excellent job of improving our ability to follow these noise trains. Not only can you see more of them but you can also trace their path for a much longer distance and in many cases, you can trace them as they cross the basement structure and extend into the valid data region.

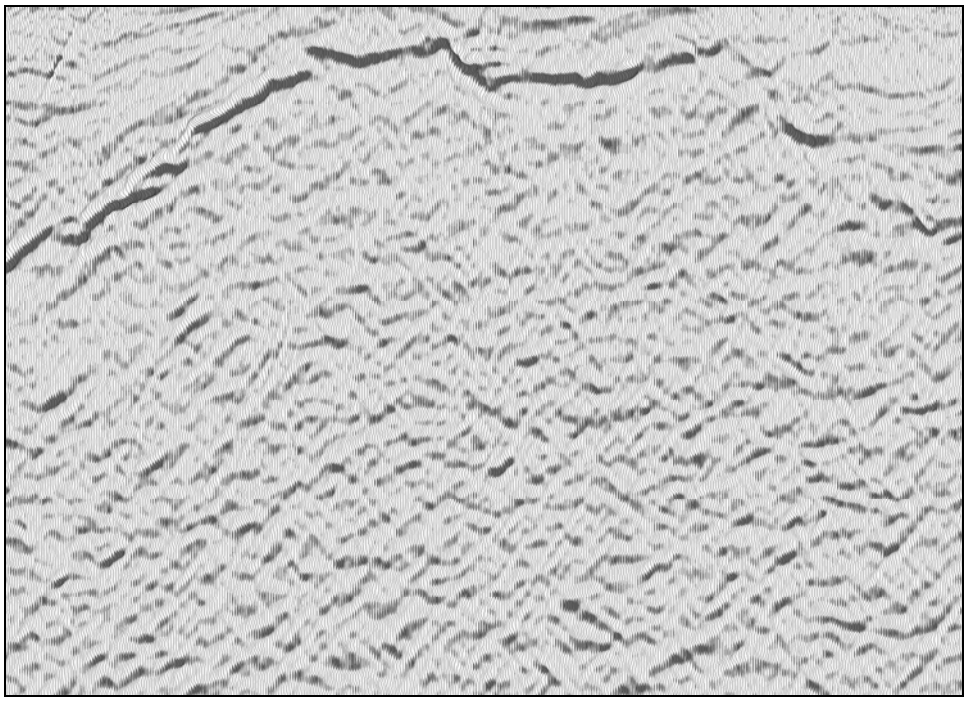


Figure 10.18: Wiggle trace display of subbasement migration artifacts.

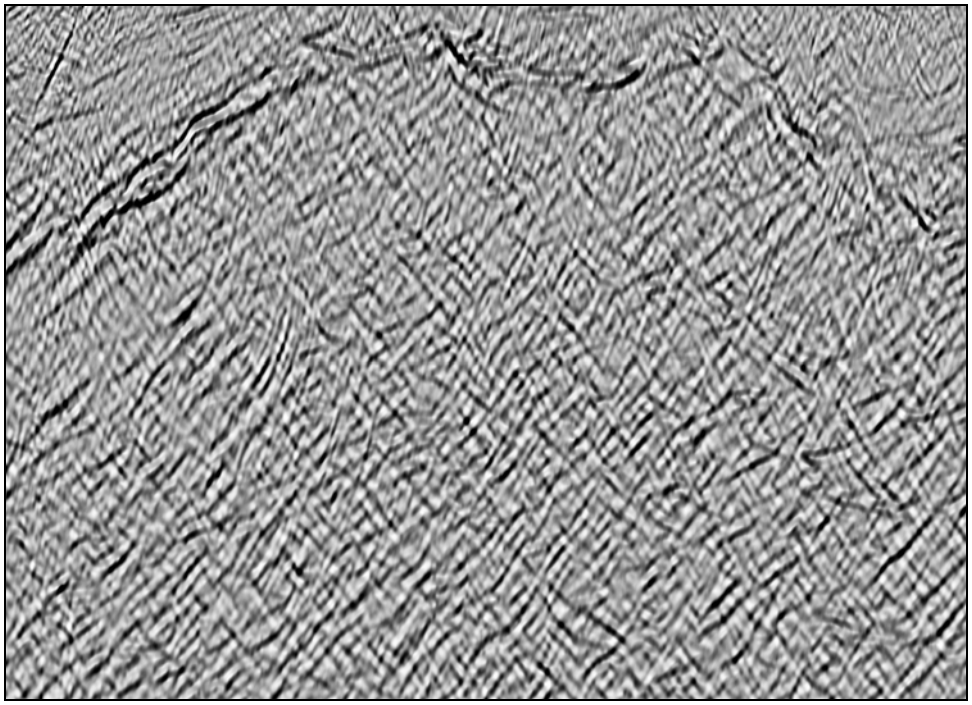


Figure 10.19: SeisScape display of the same data shown in Figure 10.18. Notice that the artifacts are more prominent on this display and that you can follow them as they extend upwards through the basement and into the relevant data.

Although migration artifacts are often the most prominent linear features that you observe through lighting, they are not the only type of linear noise affecting a section. Consider Figure 10.20, which is a wiggle trace display of a section of data immediately beneath the sea floor and Figure 10.21, which is the SeisScape display of the same section. The wiggle trace display is relatively free of apparent linear noise trains. Conversely, the SeisScape display shows there are a regular series of linear noise trains extending downwards from the sea bottom and right through the display. These noise trains are not the result of migration because on close inspection they start at the sea bottom and extend downwards. What they are exactly is hard to determine but I suspect that they are an artifact of the acquisition. Regardless of what they are, however, it is undeniable that they exist. Whereas Figure 10.21 shows that they are obvious noise, they are not as obvious once one looks deeper in the section.

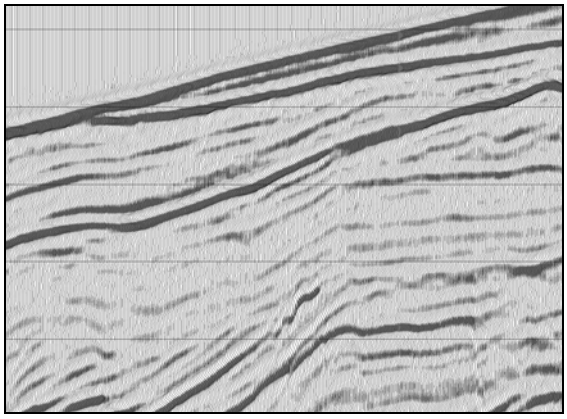


Figure 10.20: Wiggle trace display of a section of data immediately beneath the sea floor.

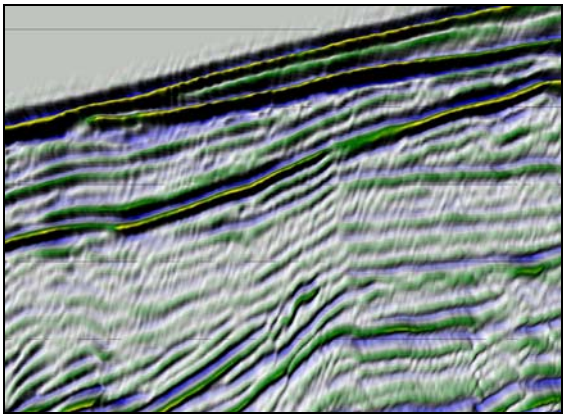


Figure 10.21: SeisScape display of the same data shown in Figure 10.20. Note the steeply dipping linear noise trains that extend downwards from the sea floor.

These two sets of examples introduce the most important aspect of examining microstructure. All seismic sections carry within them low-amplitude, linear noise trains many of which are below the visual resolution of conventional displays. Distinguishing these noise trains from geologically relevant microstructure is often difficult but it is always necessary.

10.6.2 Enhanced Resolution

One of the main points of this thesis is that the display ultimately establishes seismic resolution. In Chapter 3, I defined two terms, the first was absolute resolution, which is the resolution of the data itself. The second was apparent resolution, which is a property of the display. The main point of this chapter is that lighting is the principal tool that we use to bring the two together. More than any other single element, the reflectance of the surface enhances the apparent resolution of the display.

For example, consider Figure 10.22, which is a wiggle trace display of a downlapping sequence and Figure 10.23, which is a SeisScape display of the same data. In the center of the wiggle trace display is an area that is relatively free of major reflections. This is downlapping depositional sequence and despite the absence of clear reflections you can, nevertheless, infer several potentially downlapping events. There is no inference, however, on the SeisScape display because the downlapping events are clear and the zone is richer in significant seismic detail. The events that the SeisScape display reveals are consistent with the known geology and so I feel confident in classifying them as being geologically significant.

There are two ways in which the display can decimate resolution and this example shows both. The first way is that the amplitude of an event can be below the visual resolution of a given display. This is the case for many of the seismic features that you see on the SeisScape image. They are very low in amplitude and below the visual resolution of the wiggle trace display. The second way that the display degrades resolution is that the form of the display may not allow you to develop a true percept of the data. You observe this effect on the dipping, right edge of the feature. Here, the amplitude of the events is high enough that they are visible on the wiggle trace display. However, the form of the wiggle trace display does not provide the visual system with enough information and even though we can see the gross structure, we lose most of the details.

Enhanced resolution is the second classification of microstructure events.

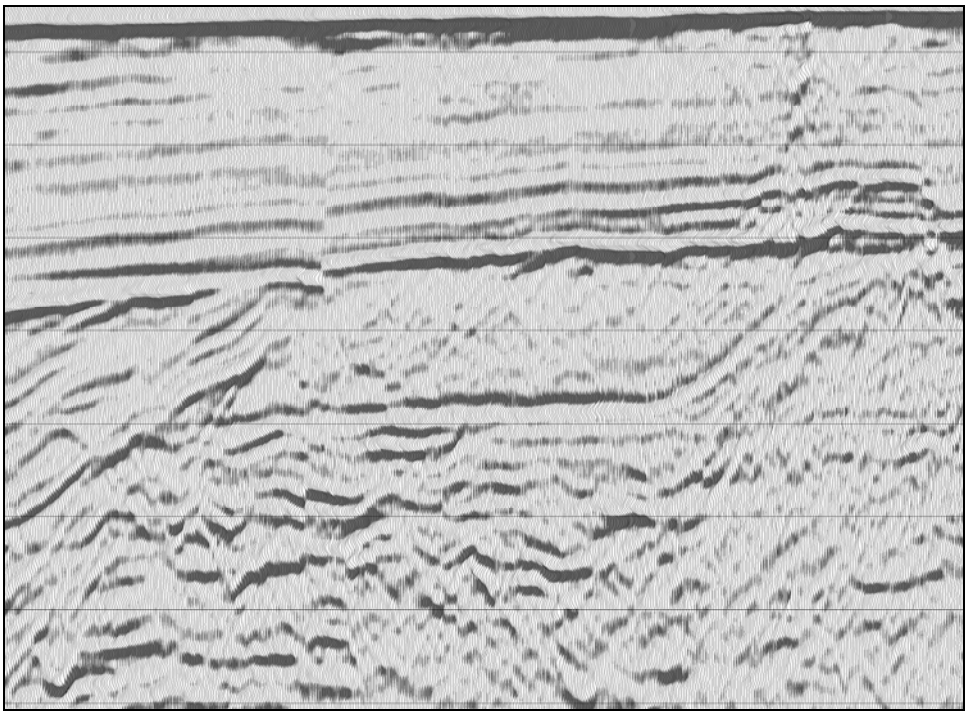


Figure 10.22: Wiggle trace display of a downlapping sequence (data courtesy PeruPetro). The area of interest is the low reflection zone in the middle of the section.

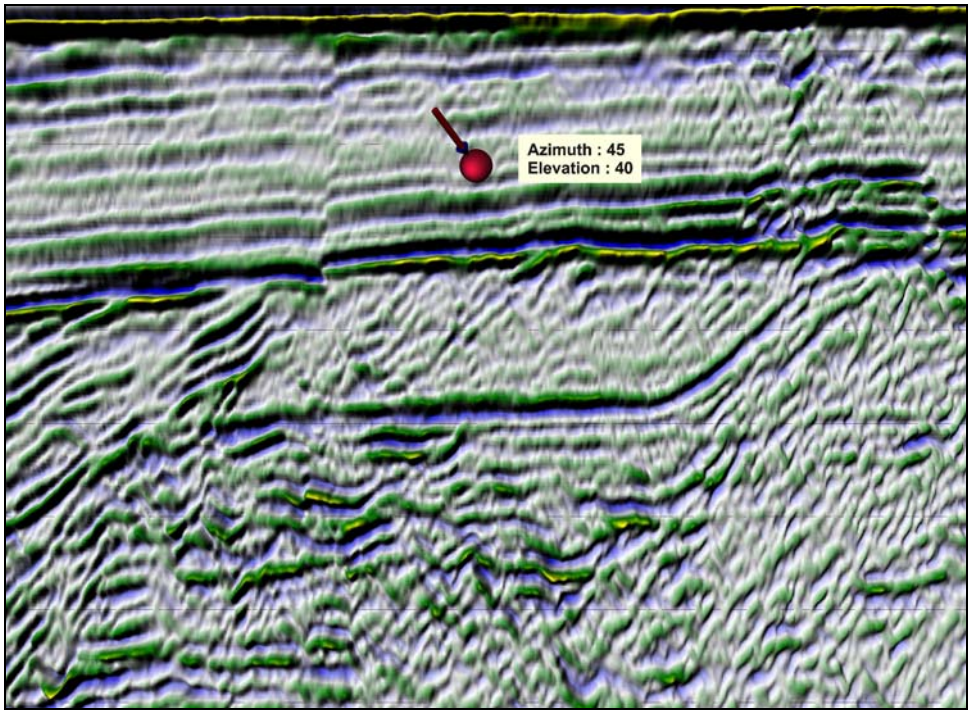


Figure 10.23: SeisScape display of the same section of data shown in Figure 10.22. The microstructure reveals significant details in the low reflection zone but how much of it is real?

If we applied a fast agc to the data then many of the features apparent on Figure 10.23 would also be apparent on Figure 10.22. Fast agc's, however, are not generally applied to sections because they destroy relative amplitude information. Consequently, whereas a fast agc may make microstructure more perceptible on a wiggle trace display it would, at the same time, destroy the amplitude structure of the macrostructure events.

10.6.3 Fault Plane Reflections

One of the most startling aspects of lighting is its ability to reveal fault plane reflections. Despite their prevalence on many sections, reflections from fault planes are often difficult to observe and in most cases, you imply faults from the data but you do not directly observe them. Seismic attributes such as coherence are often used to enhance our ability to observe faults directly. Coherence, however, is generally regarded as being a 3D seismic attribute and it is also not particularly effective for viewing faults on vertical sections, which is where we would like to see them. One of the reasons why fault plane reflections are difficult to observe is because the amplitude of any given reflection depends upon the contrast in rock properties across the boundary. In the case of faults, this contrast is continually changing because the throw on the fault brings different types of rock into contact. In some places along the fault, the impedance contrast across the fault may be high, in some places, it may be low or negative and in some places, where the throw was less than the bed thickness, it may be virtually zero.

Figure 10.24 is a wiggle trace display of a series of high angle faults and it illustrates the points that I just made. Most of the faults are obvious but we only recognize them by implication. We see that the major markers are broken and offset and we imply the faults from them. What we do not see on this display are reflections from the fault planes themselves. I have marked several places where the data is obviously faulted and where, in my opinion, you should see reflections from the fault plane. That they are not visible is a limitation of the display.

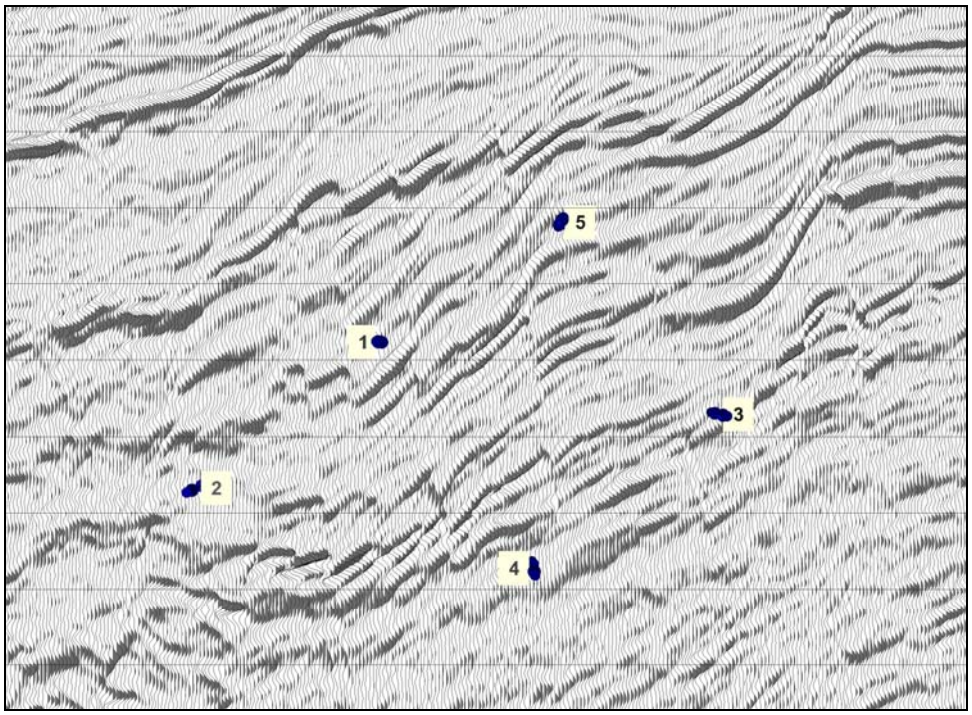


Figure 10.24: Wiggle trace display of complex faulting (data courtesy PeruPetro). The numbers mark fault locations.

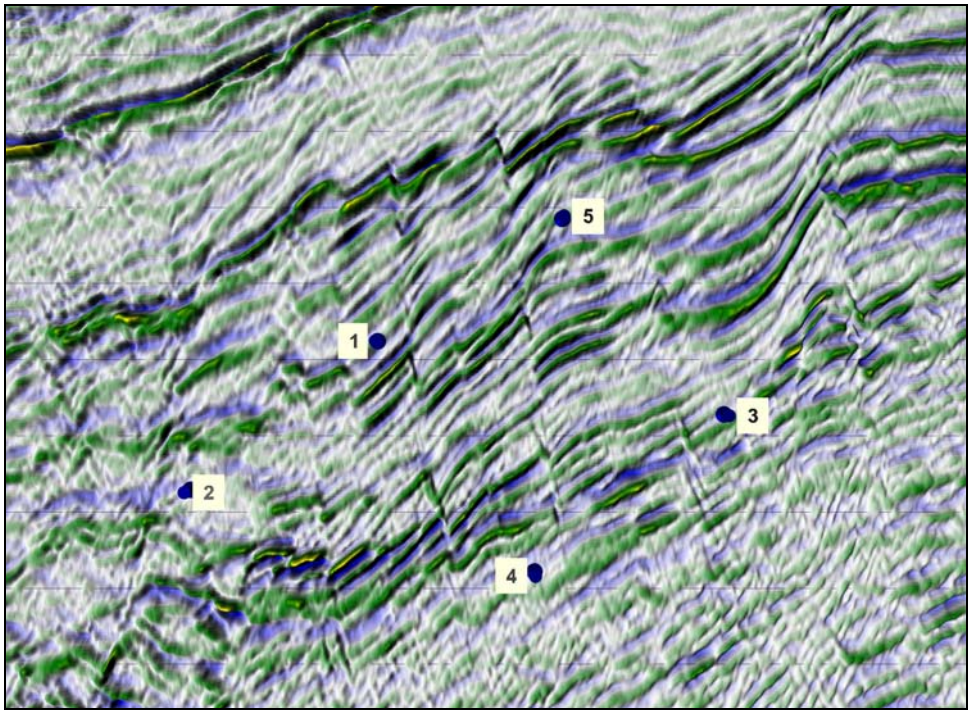


Figure 10.25: SeisScape display of the same data shown in Figure 10.24. The marked locations indicate areas where there are noticeable fault plane reflections.

By contrast, you can clearly see fault plane reflections on the SeisScape display (Figure 10.25). In some places (left of 3 and right of 4), the reflections are distinct and continuous. In other places such as to the left of #5, the reflections are fragmentary but in all of the marked cases, if you look carefully enough you will see evidence of reflections from the fault planes themselves.

Fault plane reflections are the third classification of microstructure but often they are the most nebulous. In the previous example, the reflections are obvious and in most cases, it is difficult to confuse them with anything else. There are numerous distinct events in the stratigraphic sequence and the throws across each fault are sufficient to reposition events with significantly different impedances so that they are adjacent across the fault. There should be reflections from the fault planes and Figure 10.25 shows that there are. However, what about situations where the impedance contrast across the fault is small?

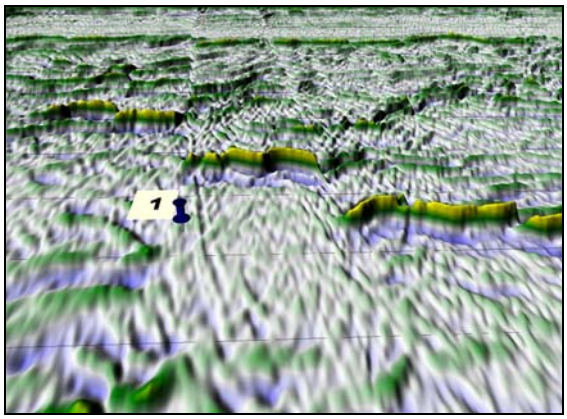


Figure 10.26: The same fault 1 shown in Figure 10.17 but viewed from beneath.

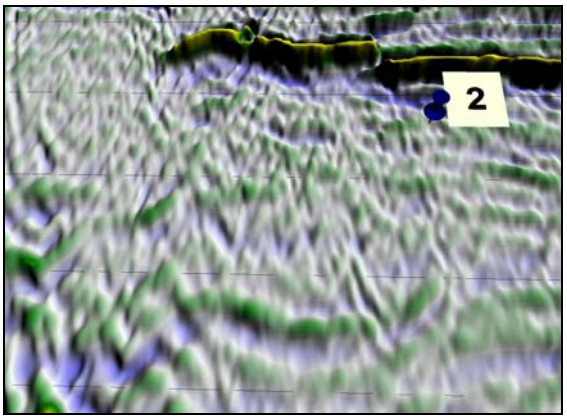


Figure 10.27: The same fault 2 shown in Figure 10.17 but viewed from beneath. Notice the conflicting dips on the microstructure.

I show an example of just such a fault in Figure 10.26, which is the same fault plane reflection that I showed in Figure 10.17. To produce this image I rotated the display around the x-axis so that you could see the upper events in the section. When you look at the display in this orientation, it is clear the identified event is a fault plane reflection because it extends downwards from a fault that continues all the way to the surface. The

reflection has very low amplitude but even so, when you see it in its geological context, there is little doubt that it is a fault plane reflection.

Contrast this with the proposed fault plane reflection that I show in Figure 10.27. The reflection is to the left of 2 and considering that it connects to an obviously faulted event it is tempting to classify it as geologically significant feature; but is it? In this case, it is not as clear that the event indentified as a fault plane reflection actually is one. This section of the data contains numerous high angle noise trains and many of them dip in the opposite direction to it. The dips on the noise events conflict and many of them have the same general amplitude as the event we are trying to assess. Given the contamination can we reliably conclude that the identified fault plane is what we think it is or is just conveniently located noise?

This example shows the primary difficulties that you encounter when you push the limits of apparent resolution. In many cases, especially when you are examining faults, sometimes obvious and sometimes subtle noise trains contaminate the seismic events that you examine. In the next section I discuss a series of criteria that I have developed to help me assess whether a given microstructure event is geologically relevant or not.

10.6.3.1 *Coherency*

The reader should not take from this analysis the suggestion that we should deprecate the use of coherence to detect faults. Coherency is a well-established technique for detecting faults and edges in 3D seismic data. There are also techniques available for 2D data as well; however, they are not as effective as their 3D counterparts are. Coherency displays, as do other seismic attribute displays, suffer from the problem that they lack context. Once we switch to a coherence display, we lose the view of the seismic amplitudes that place the detected faults and edges into necessary context.

Early in my research, I investigated co-rendering coherency and seismic amplitudes. I produced displays that used coherence for the lighting component of the display and seismic amplitude for the color. This display, which I show an example of in Figure 11.17, effectively places the faults and edges detected by the coherency into context.

These co-rendered displays are very effective, however, more work needs to be done in comparing the effectiveness of coherency and lighting at detecting subtle faulting, especially on 2D data. Because this work concentrates primarily on visualizing the seismic surface directly, I decided not to pursue any further research into co-rendering of different attributes, including coherency.

10.6.4 Place Features into Context

In this section, I established that you can classify microstructure into three broad categories. The first of these is noise trains and they are often the most prominent and readily identifiable features of a SeisScape display. It is generally simple to visually identify and dismiss most of these events. However, as you zoom into a display and as you look for increasingly more subtle features, it can become increasingly difficult to determine with a sufficient degree of confidence whether or not a given feature is geologically relevant.

As a general guideline, you should always place an observed feature into context before making a decision on what it is. As you zoom into a display, you generally start to see increasingly more subtle features and sometimes it is easy to become fooled into thinking that they are relevant when they are not. Consider Figure 10.28, which is a close-up of a section of data with complex micro faulting. To the left of #3 are a series of dipping events that look as if they are related to the faulting. If this were your only display, you might erroneously think that these events are real. When you zoom out though and place the display into context, as in Figure 10.29, you see that these events are part of a series of high angle noise trains that extend to the surface.

These events are not real and even though they look interesting, they add nothing to the interpretation and you should ignore them. You could only determine that by zooming out of the display and looking at it in context. Figure 10. and Figure 10.31 show another example of the advantage of placing a feature into context. In the first example, zooming out of the display proved that the observed features were simple noise. In this example, it proves that the feature is, in fact, a true fault plane reflection.

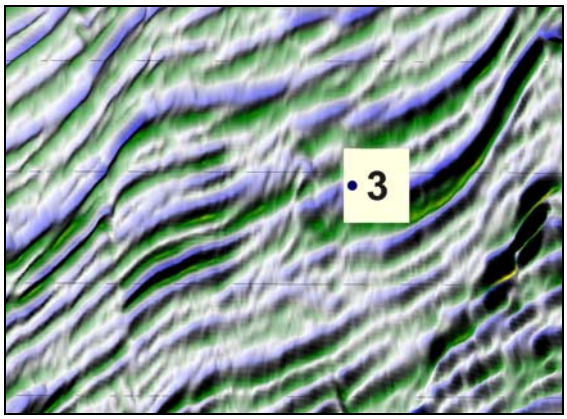


Figure 10.28: At first glance, given the micro faulting of the data, the dipping events to the left of #3 appear geologically relevant.

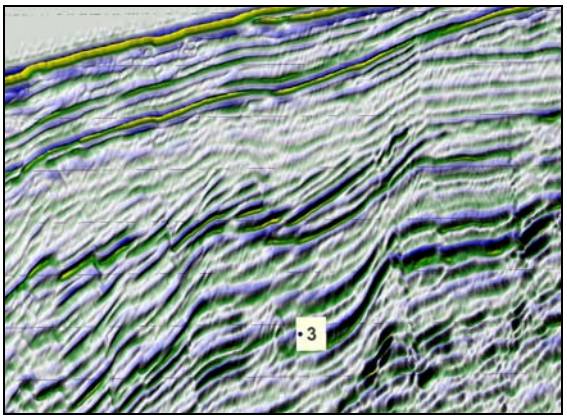


Figure 10.29: The same section shown to the left but zoomed out to show the upper part of the section. The events in question are now clearly part of a noise train that extends to the surface.

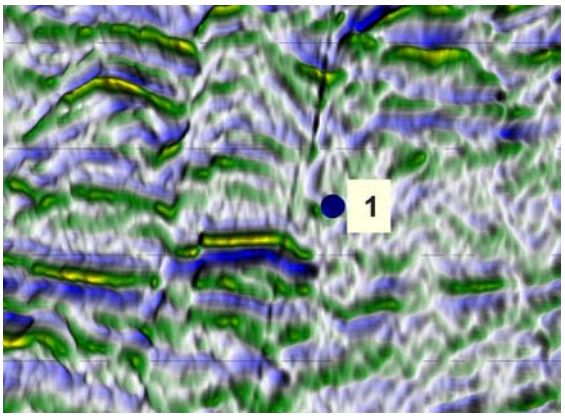


Figure 10.30: SeisScape display of a potential high angle fault plane reflection. Is this real or is it noise.

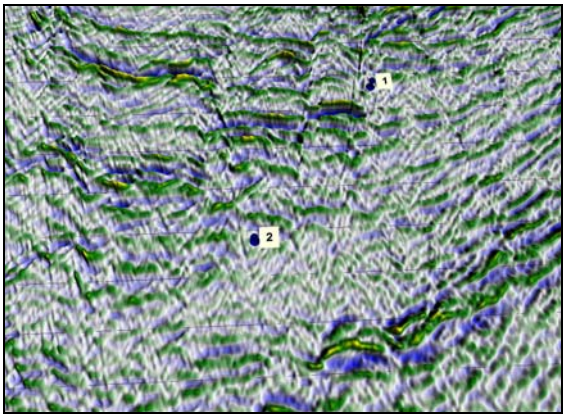


Figure 10.31: Zoomed out display of the same fault shown in Figure 10.. In this context the event is clearly a fault plane reflection.

Immediately to the left of #1 in Figure 10. is what appears to be fault plane reflection. It is not obvious that it is, however, because the events are not clearly offset across it. Furthermore, there appears to be an offset of the fault itself, which makes you wonder if the feature is valid or if it is noise. When you zoom out of the display and place it into context, however, the event clearly becomes a fault plane reflection. The context display shows that the feature is not part of a series of noise trains, which you would expect if it were a migration artifact. It is also continuous and it appears to connect to a basement fault. The context display also shows other similar features such as the feature to the left of #2, which also appears to be a fault plane reflection. In this case, placing the feature

into context with the rest of the display has confirmed it as a valid geologically based feature.

10.7 Summary and Examples

In Chapter 9, I described a seismic section as having two types of structure. The first is the macrostructure, which consists of the major events that you identify on any seismic display. The second is the microstructure, which consists of those events or signals that for one reason or another are below the visible resolution of wiggle trace displays. The study of microstructure is in its infancy and for most sections, it represents a potential rather than a reality. By definition, it falls below the visible threshold of the displays that were used when the data was processed. Consequently, if a section contains pertinent microstructure, it is by accident, not design.

When I started this thesis, I did not know that microstructure existed and it was not until I began studying the Trujillo data set that I recognized it for what it was. Trujillo is a unique combination of geology, seismic acquisition and seismic processing. By accident, all of the sections in the basin are saturated with microstructure of all three classifications. Trujillo represents what seismic can accomplish but we are a long way from achieving it on a regular basis. In our processing, we regularly obliterate relevant microstructure with migration noise. Migration artifacts are the bane of seismic visualization. They are, in fact, the bane of seismic interpretation in general. As much as I was impressed with the microstructure in Trujillo, I was shocked by the migration artifacts in my other data.

Migration is a necessary component of enhancing seismic resolution and almost every practicing geophysicist understands this. What we generally do not realize, however, is how damaging the byproducts of migration can be to an interpretation. One of the most important aspects of any interpretation is to assess risk. Central to assessing this risk is determining how well we can trust the seismic data itself. As I will show, lighting is a crucial element in this process. Sometimes all it shows is the level and

impact of artifacts but as we push the limits of resolution and as we start to chase increasingly more subtle targets, knowing these things is vital.

Ultimately, for seismic to reach its potential, we have to come to grips with noise trains. In a recent conversation with Mike Perz, the Manager of Research and Development for GeoX Processing, he told me that the processing industry recognized the seriousness of migration artifacts and that reducing their impact was an active area of research. This is a welcome development and, hopefully, SeisScape displays and, in particular, lighting can play a part.

Of the three elements of a SeisScape display, this chapter only considers lighting and what we learn from it. To that end, I show the following SeisScape displays without color and without any structure. I do this so that the readers can assess from themselves how effective lighting is at enhancing microstructure. As you compare these images with their wiggle trace equivalents, keep two things in mind. The first is that your visual system must train itself to see all of the nuances (microstructure) in the displays. The second thing is that lighting displays are dynamic and that the static images that follow only show a subset of the total microstructure. Rotating the light to a different orientation would highlight different features.

Unless otherwise noted, all of the timing lines in the following displays are at 100ms intervals.

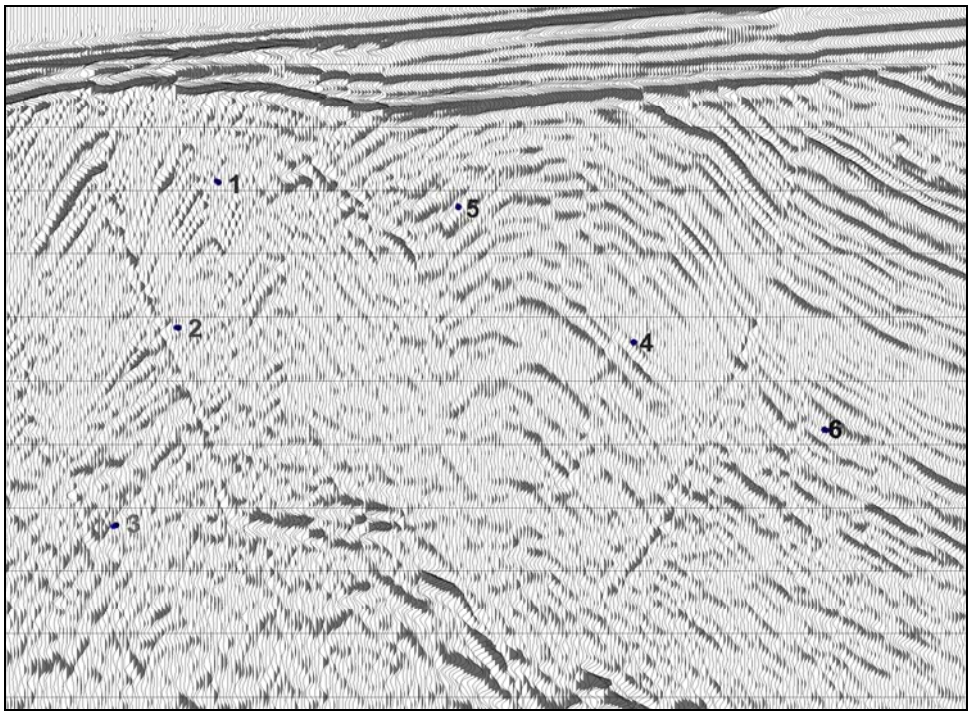


Figure 10.32: Wiggle trace display of a section of Trujillo data. Most of the faults are pronounced, even so, there are few fault plane reflections and the display, especially around 1 and 5, lacks visual resolution.

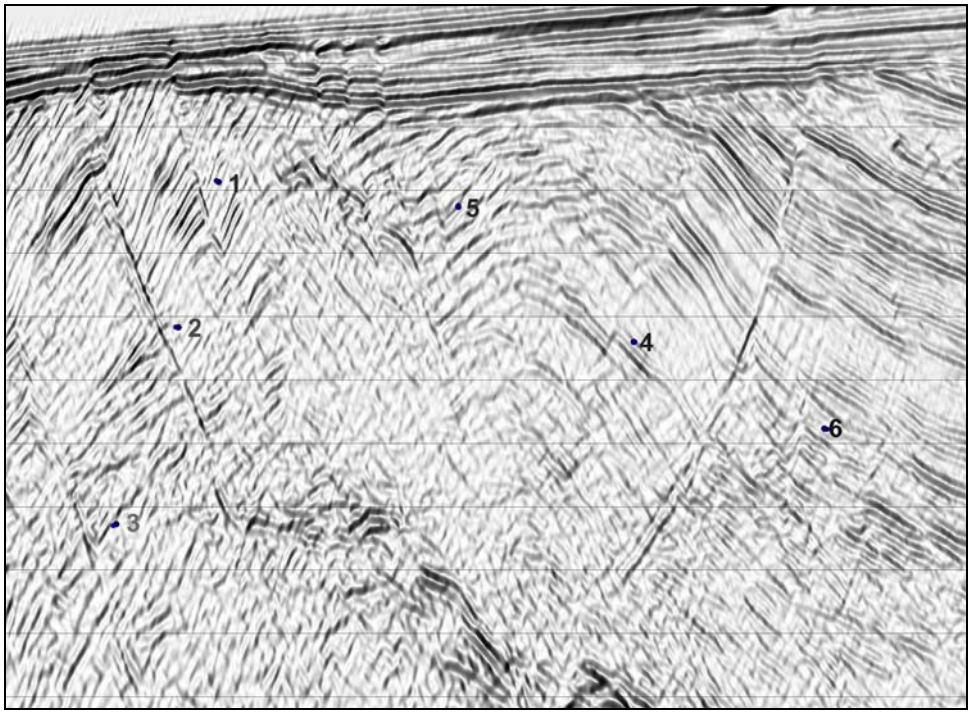


Figure 10.33: Reflectance display of the same data shown in Figure 10.32. Note the increased visual resolution around #1. Also note that the fault plane reflections that are visible on Figure 10.32 are much clearer on this image and that numerous other fault plane reflections (i.e. at #5 and #6) are now visible.

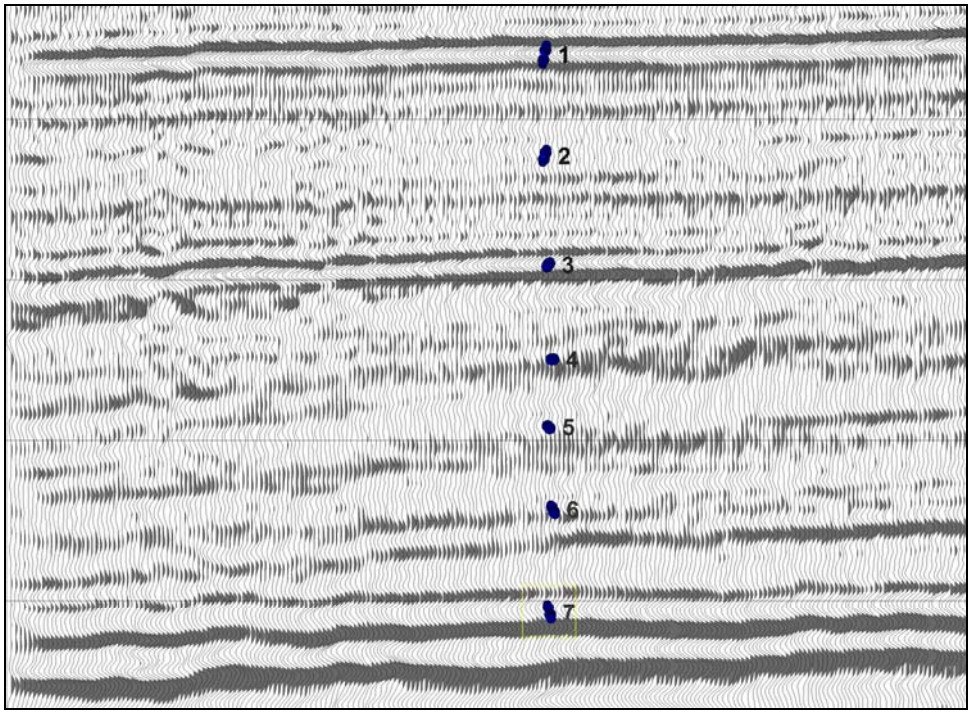


Figure 10.34: Wiggle trace display of noisy Alberta data (data courtesy Divestco Inc.). Numbers are for reference. The noise appears random and it is difficult to follow the events between the major markers.

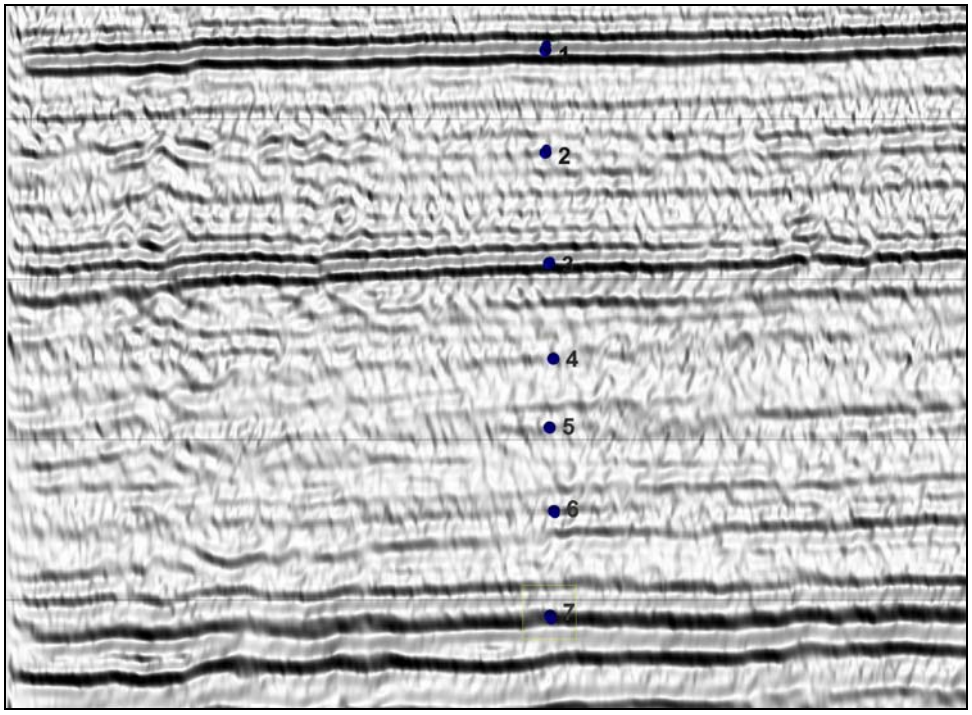


Figure 10.35: Reflectance SeisScape display. This example highlights both noise based microstructure and enhanced resolution. The lighting shows that the noise is structured and likely the result of a poor migration. Notice, also, the increased continuity of the low amplitude events between the major markers.

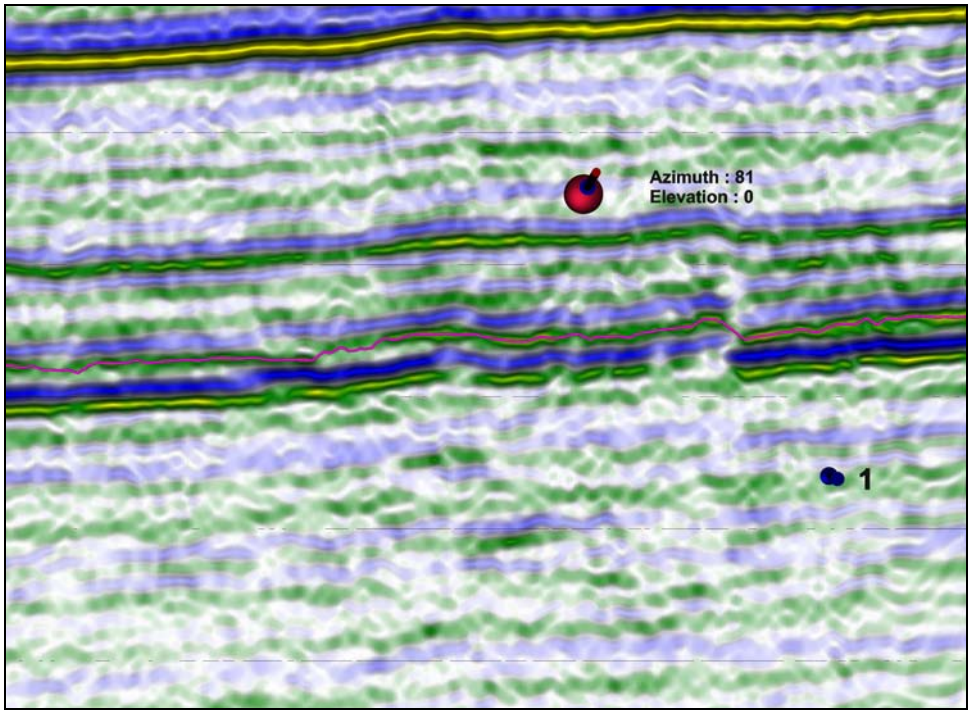


Figure 10.36: SeisScape display of an Alberta section (data courtesy Divestco Inc.). The marked event is the faulted basement. The exploration target is subtle, amplitude based and related to the faulting.

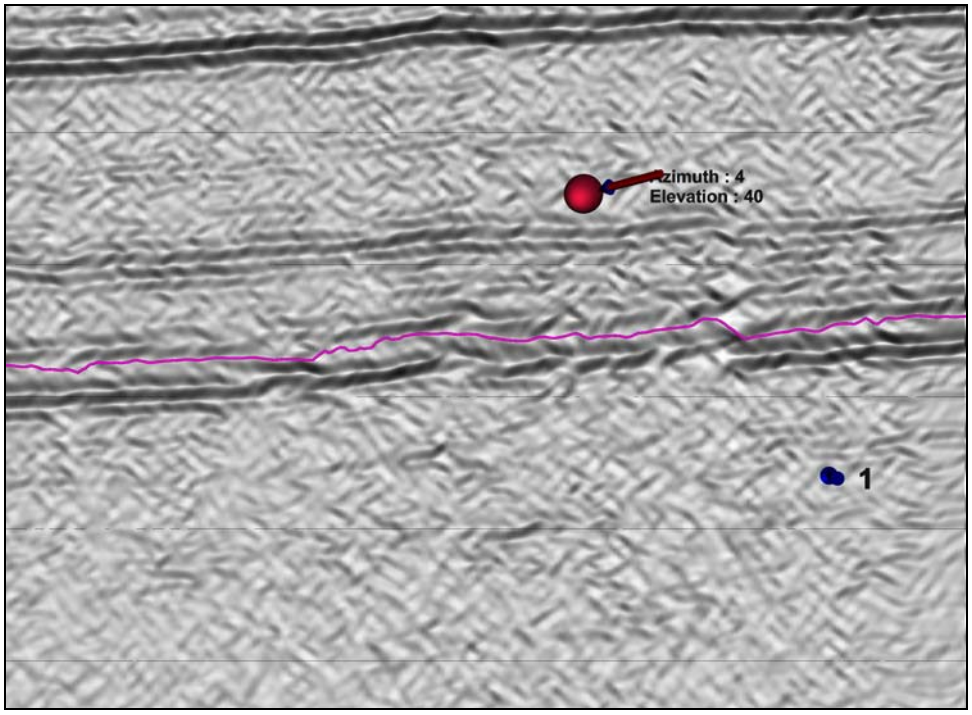


Figure 10.37: Reflectance display of the date shown in Figure 10.36. The lighting shows that the target area is saturated with migration artifacts from both the edge of the section (#1) and from other, deeper locations. Given that the target is subtle and amplitude related, how confident can we be of an interpretation knowing the level of artifact contamination?

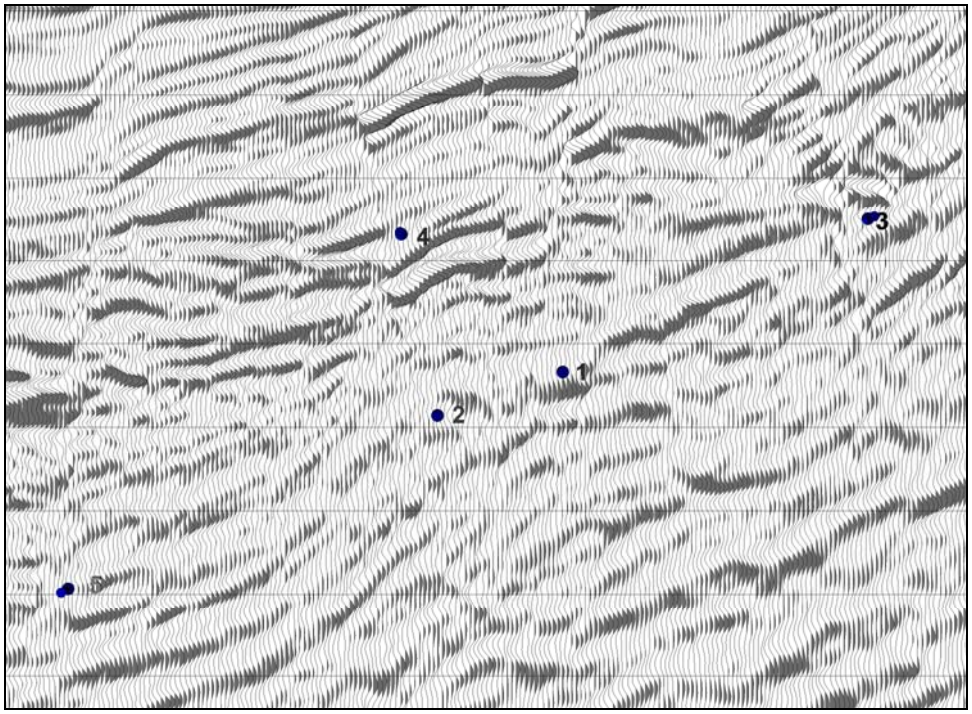


Figure 10.38: Wiggle trace display of faulted Trujillo data. The quality of the entire Trujillo data set is generally exceptional. However, it was migrated post-stack and consequently it is inundated with artifacts.

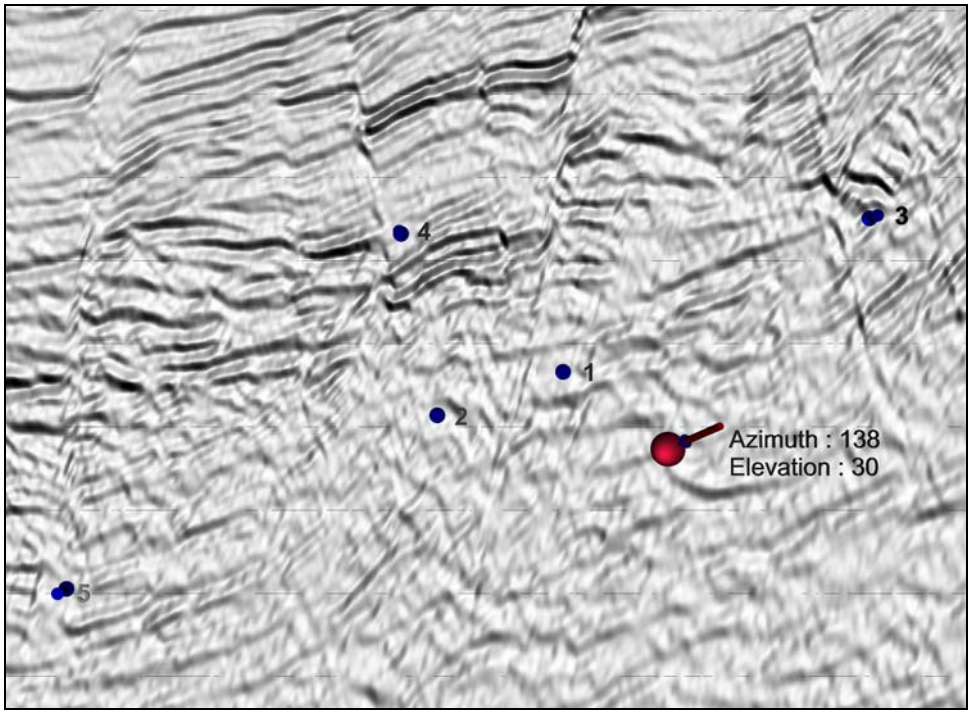


Figure 10.39: Reflectance display of the same data shown in Figure 10.38. Despite the contamination by migration artifacts, which are evident on this display, you can still identify (numbered locations) distinct fault plane reflections that are not clear on the wiggle trace display.

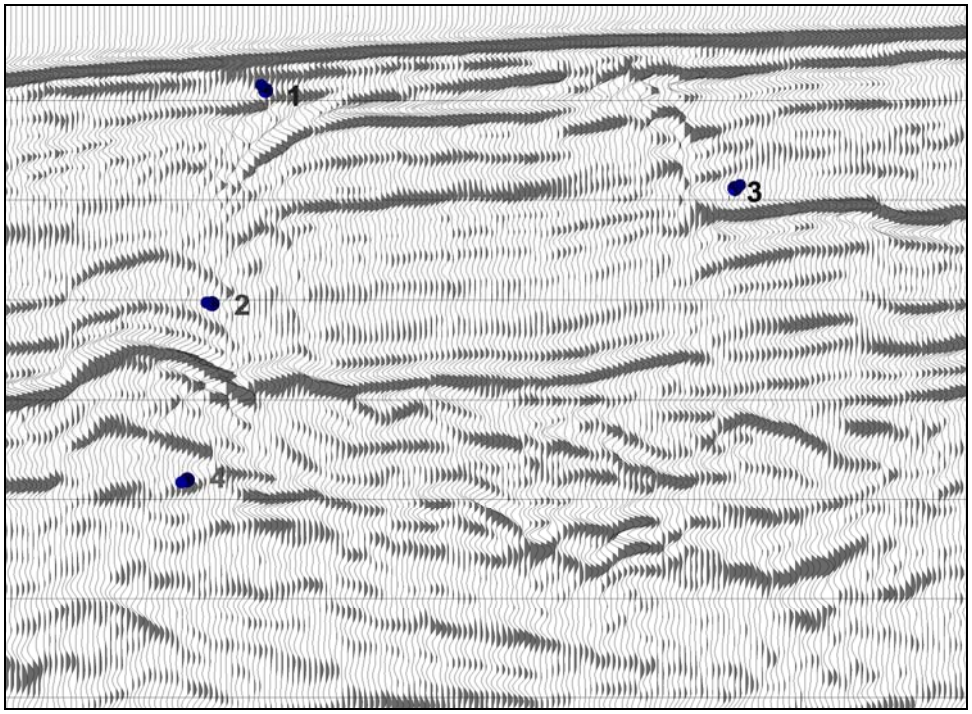


Figure 10.40: Wiggle trace display of a shallow section of post-stack migrated data from Trujillo.

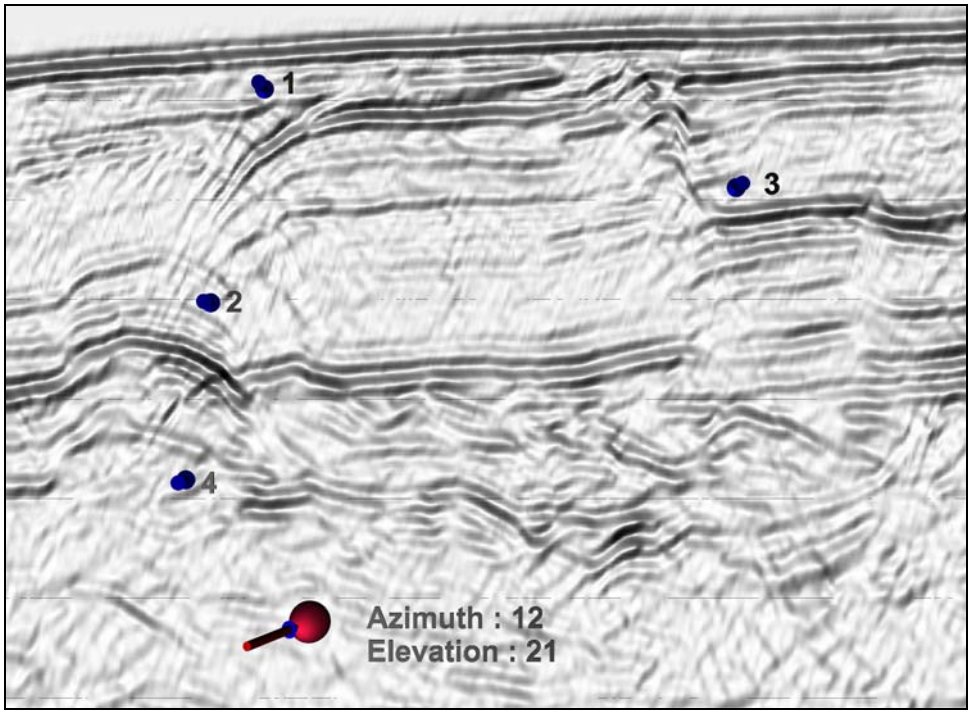


Figure 10.41: Reflectance display of the same data shown in Figure 10.40. The steeply dipping marked events are not faults. Rather, they are uncollapsed diffractions caused by under migrating the upper parts of the section.

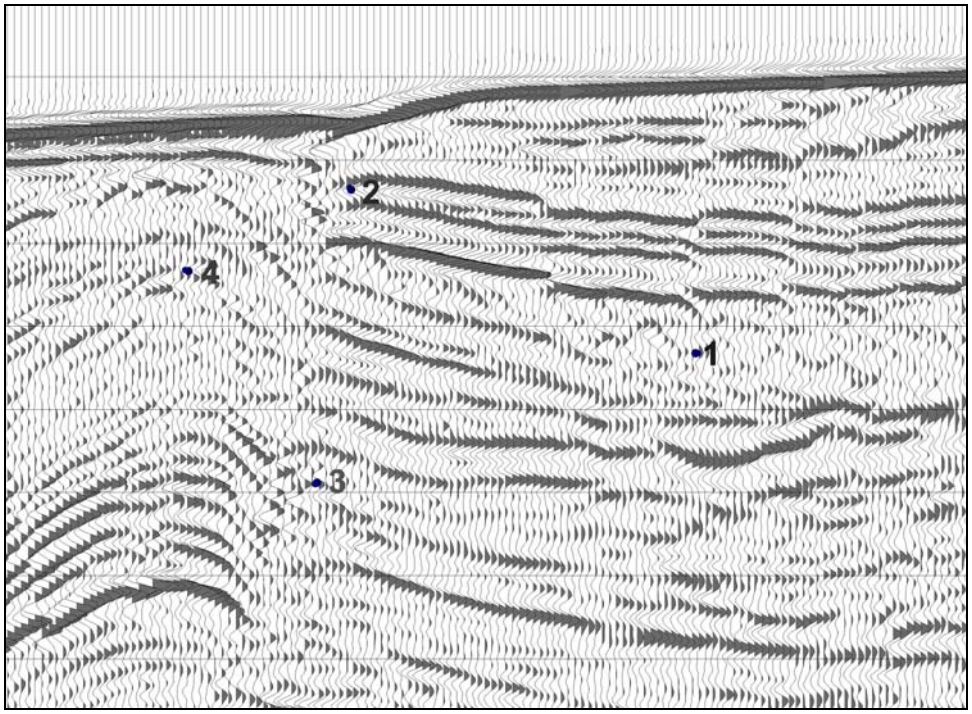


Figure 10.42: Wiggle trace display of shallow Trujillo data. Note how difficult it is to perceive events as they overturn (2 and 3) and how little information is visible in the low amplitude regions around 1 and 4.

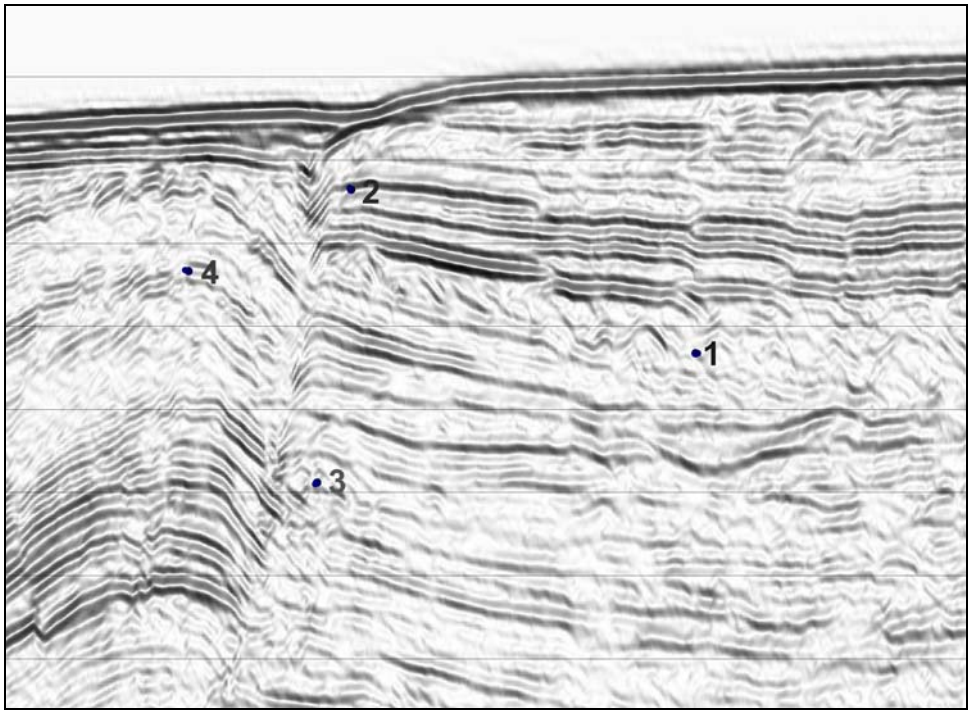


Figure 10.43: Reflectance display of the same data shown in Figure 10.42. The dipping events at 2 and 3 are clearer and there is detail that is much more visible in the low amplitude regions around 1 and 4.

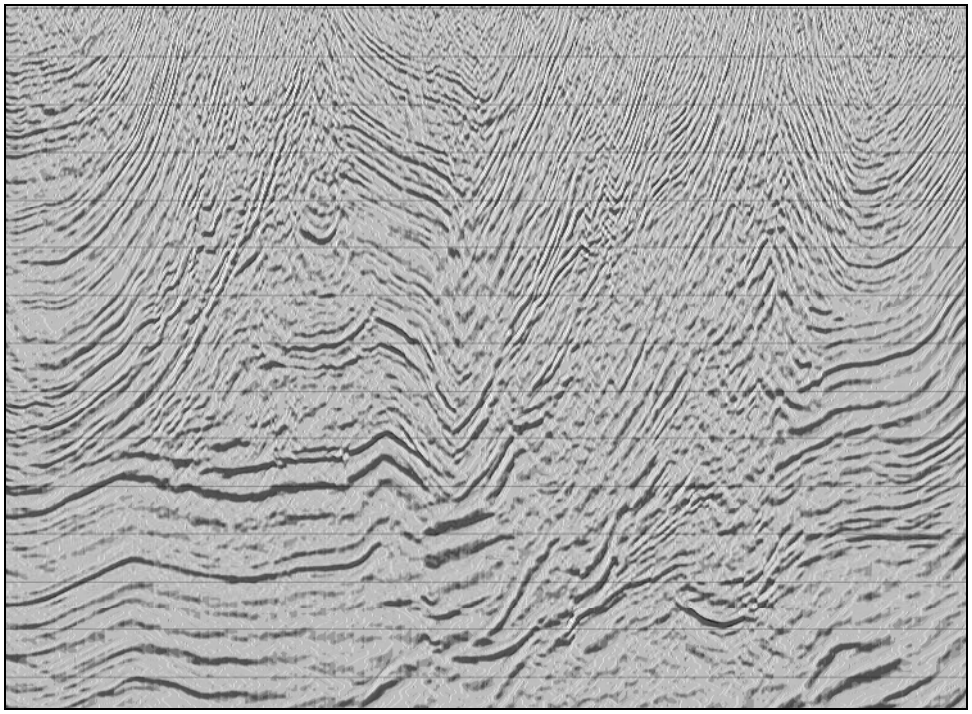


Figure 10.44: Prestack migrated section of a Canadian foothills line (data courtesy Husky Oil, processing from GeoX). There are ~1500 traces in the display, only every 2nd trace is displayed.

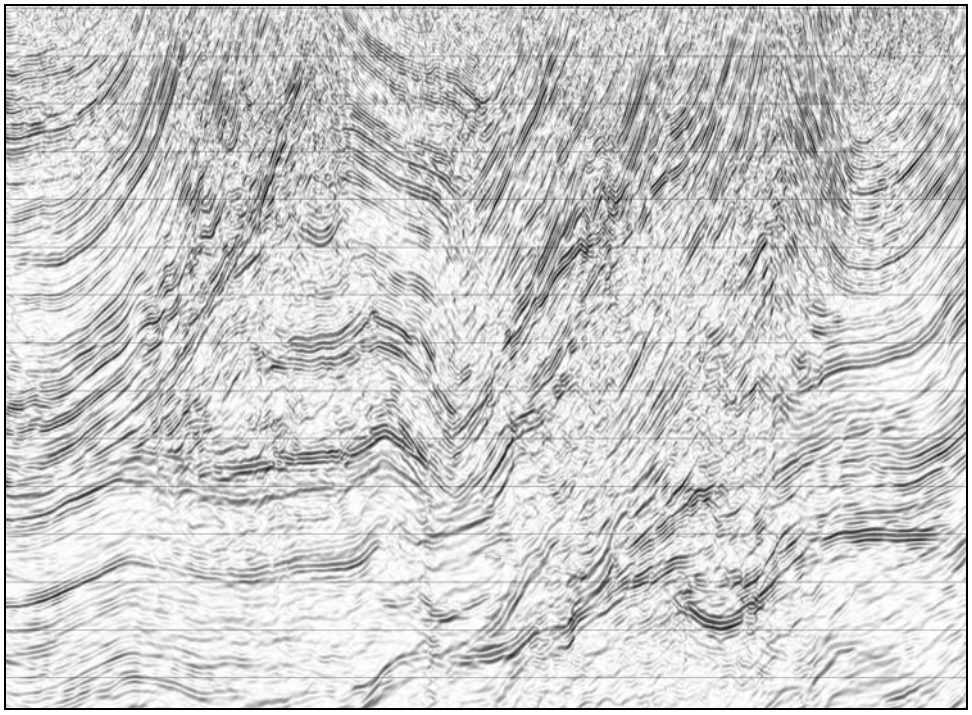


Figure 10.45: Reflectance display of the same data shown in Figure 10.44. This comparison shows how effective lighting is on large-scale displays. Despite the fact there are almost 1500 traces, there is no loss of resolution and fine scale details can still be observed.

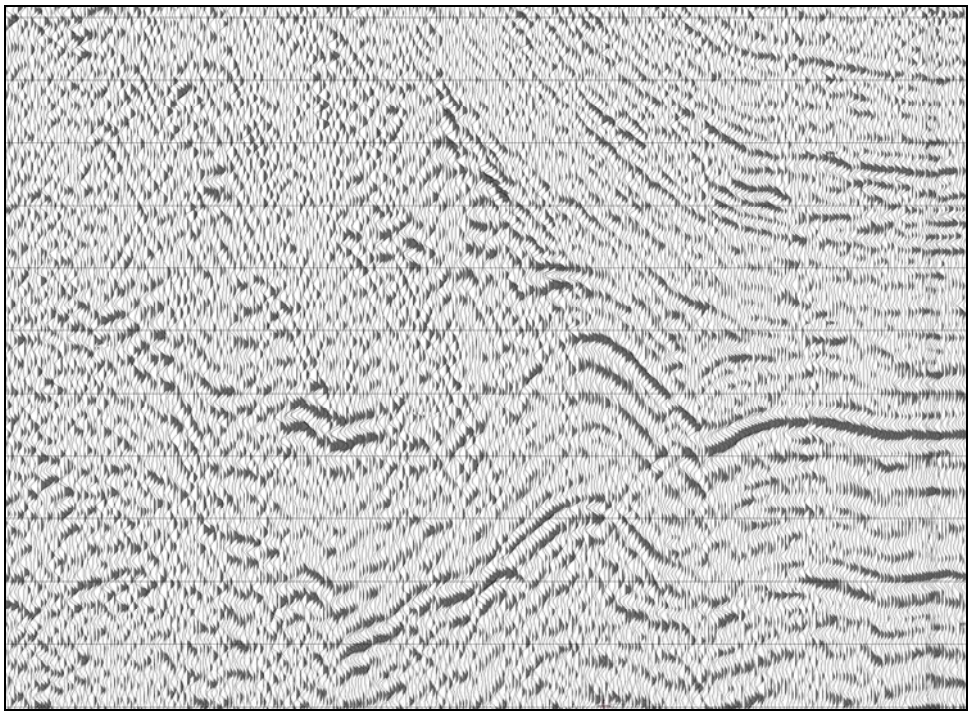


Figure 10.46: Wiggle trace display of an unmigrated SW Alberta Foothills line (data courtesy of Divestco Inc.). This is 1977 vintage data and processing and both are of questionable quality.

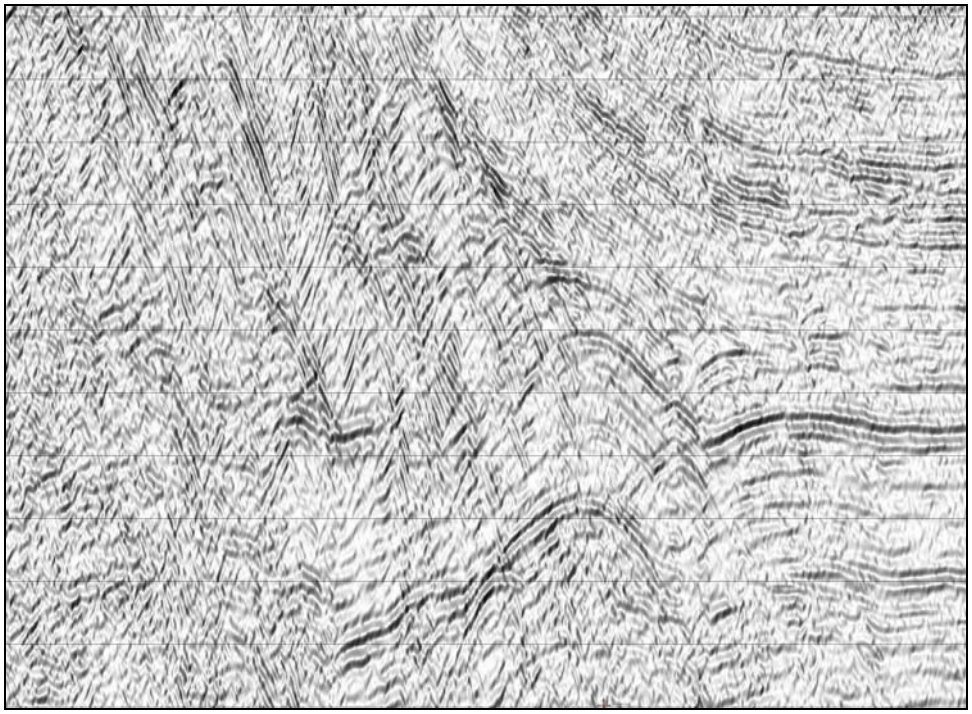


Figure 10.47: SeisScape display of the same data shown in Figure 10.46. This data predates the routine use of time migration. This comparison illustrates that lighting can also be effective at enhancing the visibility of signals in noisy data. The events you see are, by strict definition, macrostructure. However, the processing and data are such poor quality that you cannot observe them on the wiggle trace display.

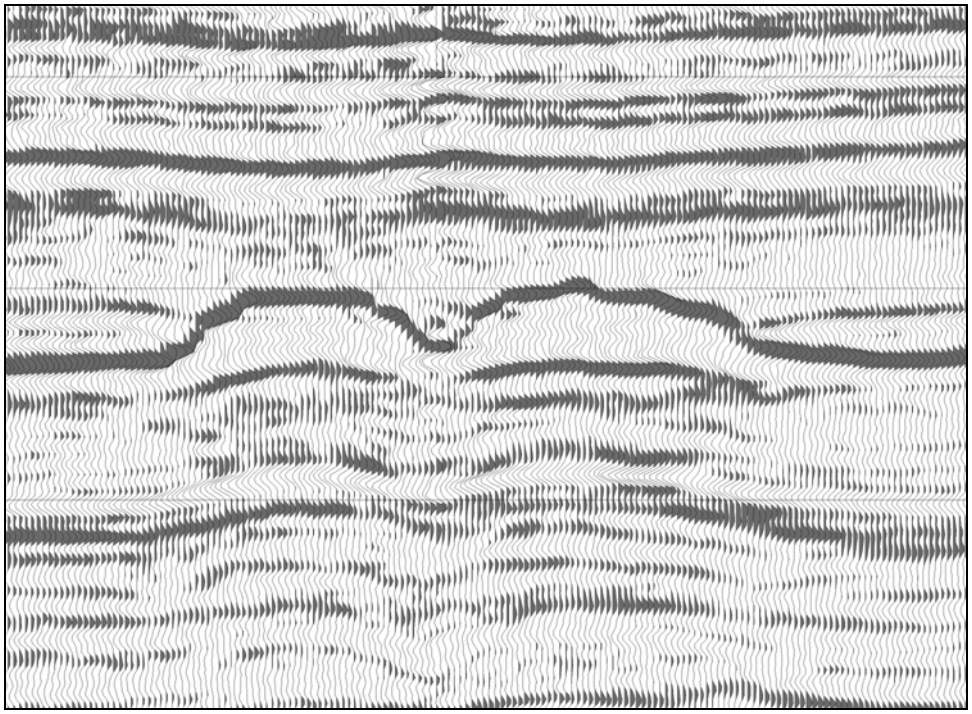


Figure 10.48: Wiggle trace display of Winnipegosis reefs (data courtesy unnamed source). One of the limitations of wiggle trace displays is that negative amplitudes (wiggle only) do not form visual percepts.

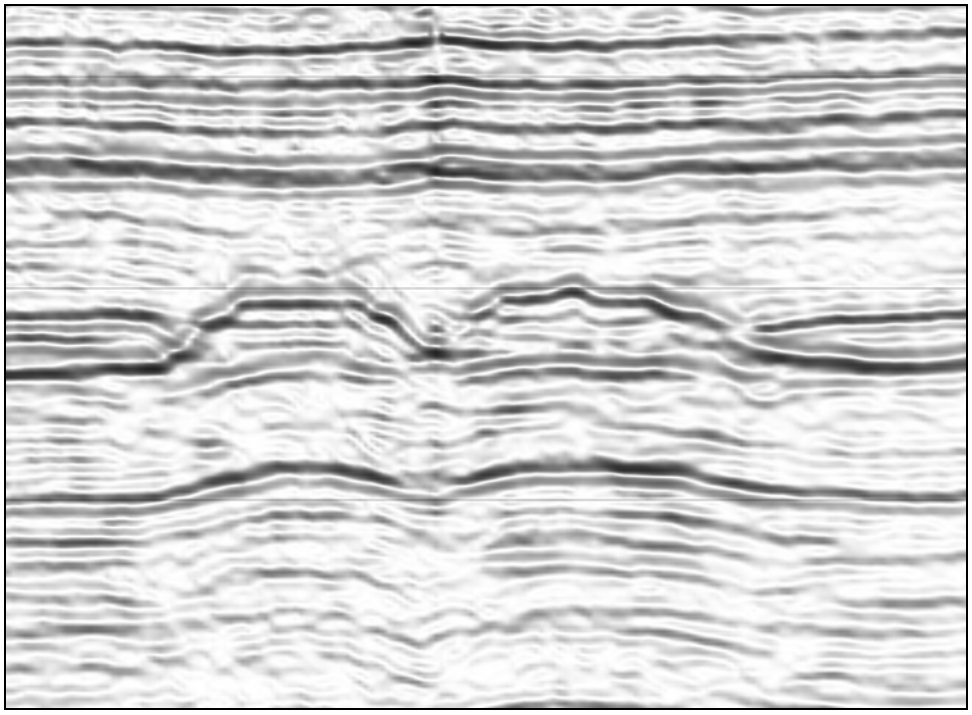


Figure 10.49: Reflectance display of the same data shown in Figure 10.48. Wiggle trace displays only form visual percepts from positive (filled) amplitudes. Consequently, there is no perception of structure within the reefs. Reflectance, however, responds to the change in amplitude, which in this case reveals microstructure within the reef.

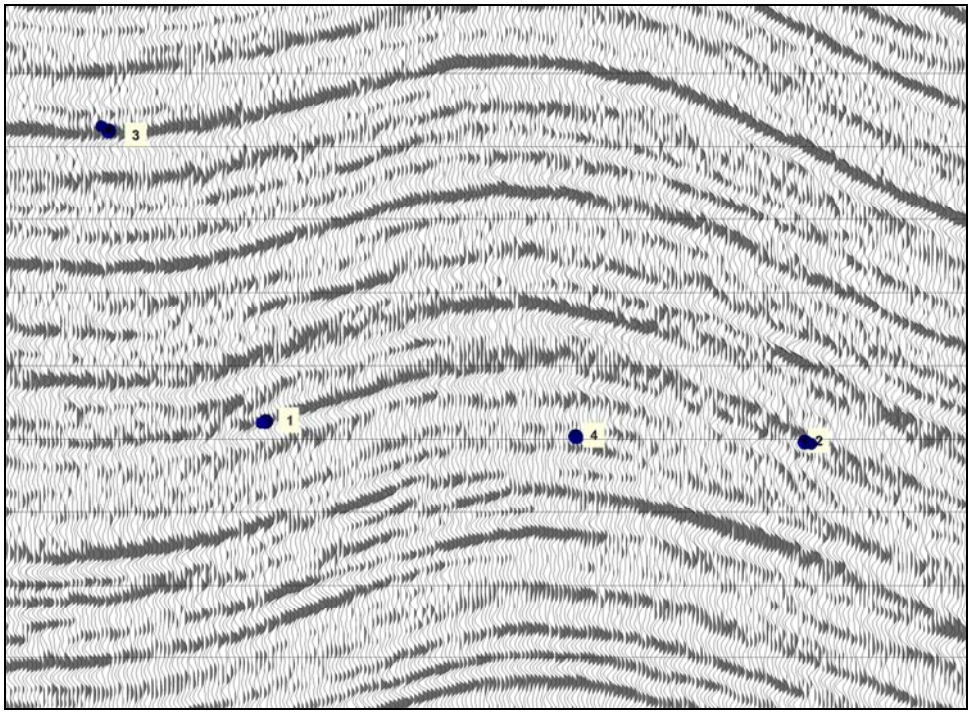


Figure 10.50: Wiggle trace display of data from the Huallaga area of Peru (data courtesy PeruPetro). Again, the noise appears random and the trace to trace ampitudes along the major events are erratic.

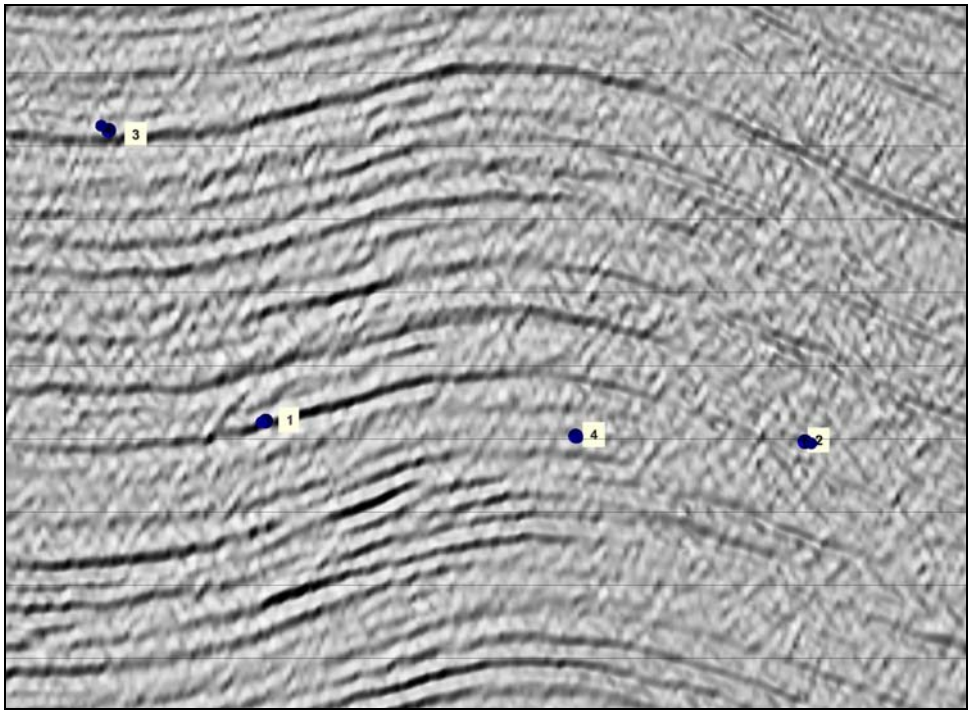


Figure 10.51: Reflectance display of the data shown in Figure 10.50. Once again, the noise is structured and primarily linear. The events to the left of #1 and #4 show that the section may be faulted and there are microstructure features that might be fault plane reflections. These events look similar to the noise trains, however, and cannot be interpreted with confidence.

CHAPTER ELEVEN: COLOR AND SEISMIC DATA

They fought and they talked in the North and the South,
they talked and they fought in the West,
Till the waters rose on the pitiful land, and the poor Red Clay had rest –
Had rest till that dank blank-canvas dawn when the dove was preened to start,
And the Devil bubbled below the keel: "It's human, but is it Art?"
"The Conundrum of the Workshops"

11.1 Introduction

In 1991, my wife and I packed our very young children into a canoe and set of on a 120 km expedition around the Bowron Lakes in central British Columbia. After many arduous days of paddling through storms, up rivers and across beaver dams, we arrived very mysteriously back were we started. Bowron Lakes is a rectangular series of interconnected lakes and you take out from the final lake almost where you put in to the first.

This seemed unfair and wholly anticlimactic, to work so hard just to get back to the start. And we had to work hard because when we started out we had no idea how to canoe (without an outboard engine), where we were going or how long it would take to get there. We reasoned though, that it was hard to go the wrong way on a lake and as long as we went downstream in the rivers, we would eventually find our way where we were going.

We arrived back to the vast and obvious pleasure of the Park Rangers, who all thought that they would have to come and rescue us. We arrived back where we started, tired, hungry and desperately in need of beer, (the only thing we did not pack). We arrived back where we started, much younger than when we set out, much more experienced than when we set out and infused with both joy and enthusiasm for the things we had seen and the things we had learnt. It was a lifetime experience!

Little did I know when I started this thesis back in 2003 how much my journey through it would mirror that trip. I started out with only a vague idea of where I was going and with no idea of the skills that I would have to develop to get there. Along the

way, I had to face my deepest fears and I had to develop skills on the fly or founder. It took months of planning before my wife and I took our first step on the first portage around Bowron Lakes and it took years of planning before I took my first step in this thesis. In both cases, both my wife and I knew that with those first steps, there was no way back. There was never another option but to see both journeys all the way through to their very end.

And once again, just as I did back then, I arrive at the very end and back at the very start.

This thesis began with a discussion on the purposes of art and science and that is how it ends. The first point that I made in Chapter 1 was that when you create an image and place it in front of someone; you both engage and inform them. The former is art; the latter is science and learning how to focus on the science has been my primary focus for the past five years.

There was a problem with my early SeisScape displays, they were too engaging. People who looked at them described them with terms such as wonderful and spectacular. As positive as that sounds, I was not looking for that response. I knew that SeisScape displays revealed far more seismic detail than did conventional displays. That and only that is what I wanted people to focus on. With that in mind, if the images in this thesis are spectacular because of the seismic detail they reveal, then I have succeeded but if they are only spectacular because of how they engage you, then I have wasted my time.

The purpose of this final chapter is the same as the first. It is to convince you that visualization is not art and that liking an image is irrelevant as long as it shows what it needs to show. Everything I have done is to the purpose of exposing the sciences of visualization and illustrating their affects upon seismic resolution. The focus of this chapter is to show how it can all be undone by the use of a careless color palette.

This chapter is concerned with color and consequently it was the hardest chapter for me to write because I am colorblind. The term colorblind is a euphemism and strictly

speaking, it is incorrect. I see colors very well; I just do not see them as you do. There are many types of colorblindness, mine is called deuteranopia. A person is deuteranopic either because their M-cones are missing or because the spectral sensitivity of their M-cones is very close to that of their L-cones. Consequently, deuteranopes have a moderately reduced ability to distinguish reds and greens. I can distinguish between reds and greens but to me they are a very similar color. Surprisingly, given that my M-cone is deficient, I perceive green very well but red very poorly. In tests that compensate for deuteranopia, I perceive green as being almost unchanged whereas red becomes a completely different color, one that I have never seen before.

In one sense, my color blindness was a detriment to writing this chapter. Because I have difficulty separating reds and greens, I could not empirically establish which color combinations worked the best for seismic. Consequently, I have only developed one series of palettes, which I called HA1 and HA2. On the other hand, the same colorblindness was an asset because it forced me to consider color from a theoretical perspective. Color serves a particular purpose to primates; we did not develop trichromacy to appreciate the sunset. Rather, we developed it for sound evolutionary reasons. Because I am colorblind, I was forced to search out those reasons.

11.2 The Evolution of Primate Trichromacy

As I showed in Chapter 6.5, the earliest vertebrates had four classes of cones that are designated SWS1, SWS2, MWS and LWS. In Eutherian mammals (which includes primates), this complement is reduced to two classes, LWS and SWS1. This is most likely because of the nocturnal lifestyle that dominated much of their evolutionary history. What this means is that because they only retain two of the original four-color receptors, all Eutherian mammals are genetically dichromatic.

Catarrhine primates', who include the old world monkeys, apes and man, are unique among mammals because they have evolved a form of trichromacy. This trichromacy arises from their possessing two copies of the LWS gene (M & L). These two genes are very similar in their spectral sensitivity and provide a fine scale ability to separate red-

green hues. Obviously, at some point in our evolution, the ability to separate between colors at this end of the spectrum was significant to our survival.

Why anthropoid primates evolved trichromacy is still a contentious subject and I am sure that at this stage the reader does not want me to delve into the controversy in any detail. For a detailed examination of this subject, I refer the reader to the following source:

Anthropoid Origins New Visions, pp 615-635 and pp 665-688, Kluwer Academic/Plenum Publishers, New York, ed. Calum F. Ross and Richard F. Kay.

What is generally accepted is that anthropoid primates evolved trichromacy for dietary purposes. Polyak (Polyak 1957) proposed that trichromacy evolved to detect red-to-yellow fruits against a background of green leaves. In this scenario, trichromacy and fruit color coevolved (Darwin, 1859; Mollon, 1989) and there is quantitative evidence that trichromacy does aid in the identification of red-to-yellow fruits against vari-luminant and vari-chromatic backgrounds.

By contrast, Dominy and Lucas (2001) have proposed an alternative scenario. They hypothesized that catarrhine trichromacy evolved to detect young, nutritious leaves rather than ripe fruit. They found that the reddening of young leaves is highly correlated with their nutritional value. This reddening would be hard for a dichromat to detect against a background of green foliage. Dichromats, however, could still detect yellow fruits against the same green foliage, which lends credence to their argument. On a personal note, as a dichromat that is exactly what I experience. I can readily detect yellow against a green background but find it much more difficult to detect red against that same background.

For the purposes of this chapter, either scenario is equally as relevant. Regardless of whether we evolved trichromacy to select red-to-yellow fruits or to select red foliage, what is important is that the background for both is the same green foliage. The catarrhine primate visual system is tuned to detect red and yellow objects against a green background. What this means for the development of a color palette is that primates see

green very well but it does not attract our attention. On the other hand, reds and yellows strongly attract our attention and this leads directly to the subject of why conventional color palettes are unsuitable for SeisScape displays.

11.3 Conventional Seismic Color Palettes

In Figure 11.1, I show a series of six images of the same section of Trujillo data. The first image is an achromatic reflectance image whereas the other five are all chromatic variable density images. The chromatic images employ palettes that are actively used in interpretation. At this point, the colors used in each individual palette are not particularly important. What is important is that each one is significantly different from any of the others. Each palette was developed by a different interpreter as his or her own personal favorite. Importantly, none of them were developed using any objective principals.

These palettes illustrate how we typically use color. The first thing they show is that color perception is always subjective. Each of these palettes represents someone's personal preference. Despite how different they all are, somebody "likes" them best. With this in mind, given how different the palettes are from each other, it is clear that even trichromats cannot agree on how to communicate chromatically. One must wonder if color can ever be a universal tool.

The second thing they show is that left to their own devices, primates prefer vivid, contrasting colors. The palettes are saturated with reds, yellows, oranges, and blues and other colors that I cannot recognize. What are lacking are the greens and browns that humans (and presumably other primates) perceive as dull and uninteresting. I did not choose these palettes specifically because they lacked those colors. They are just a representative collection of "personal preference" palettes. The other palettes that I have also lack the same dull background colors.

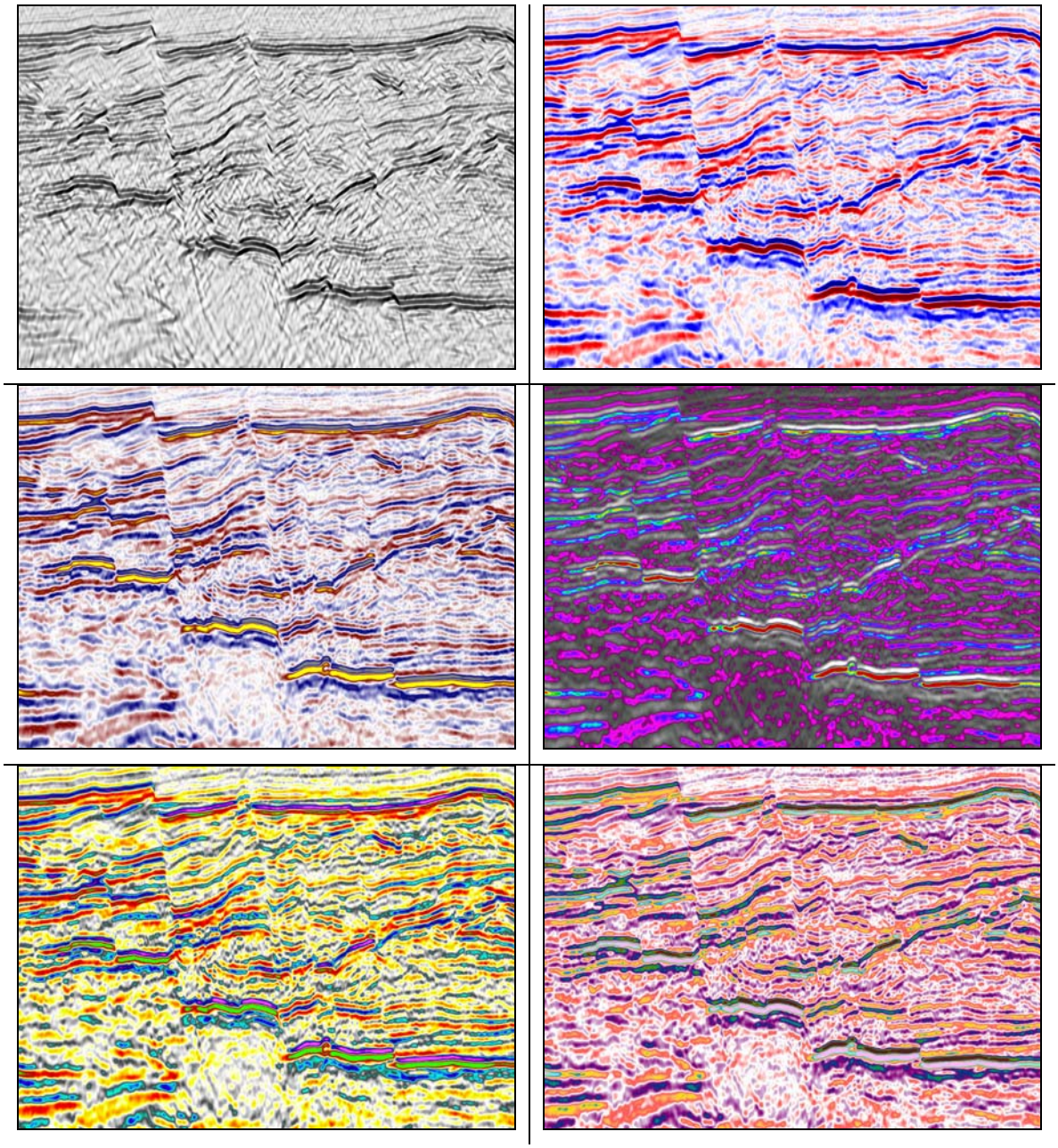


Figure 11.1: Six images of the same section of Trujillo data. The top left image is the achromatic reflectance image whereas the other five are chromatic variable density images. The palettes used in the chromatic images are all actively used in interpretation but none of them conveys any sense of perception.

The final point is that whoever designed these palettes expected too much from them. Each palette was developed with the expectation that it would convey a full range of seismic information. This includes events with different amplitude ranges, amplitude changes along events, positive and negative amplitude events, faults, diffractions and noise trains. This seismic information, taken together, is the form of the data and each palette is attempting to communicate all of it. That is why most of the palettes are so "busy". The interpreter was trying to see as much in the data as he or she could but as we now know, they were all doomed to fail.

They were doomed to fail because as I showed in Chapter 2, primates do not establish form through purely chromatic information. Used correctly, color attracts our attention and contributes to our determination of form but by itself, it does not produce any sensation of perception. You can prove this by comparing any of the chromatic images with the plain, dull reflectance image. None of the variable density images communicates as much information as the reflectance image and none of them produces percepts.

This is an important point in understanding how to use color. As people begin to use SeisScape displays, it will be natural for them to use their old familiar palettes for the color. That is, after all, what I did when I first started working with the displays. I simply took the palettes that I was already using for my variable density displays and applied them. The point of this discussion is to begin to show that this approach may be inappropriate. Conventional variable density palettes were designed to show what we now show much better using reflectance. They were also designed for use in a sensation free environment and I will have more to say on this in a later section. For now, however, simply consider this; color modifies perception. Given what we know about why we developed our color vision, are these palettes appropriate to that purpose?

11.4 Color, Seismic Polarity and the Determination of Form

In this section, I begin to discuss the appropriateness of conventional seismic color palettes. As I discussed in the previous section, most conventional color palettes were designed for use in a perception free environment in which we relied purely upon color to communicate information. This section first discusses what happens when we use those same palettes in a perceptive environment and then second, it discusses how we use color to modify perception.

The only "official" standards for how to relate color to seismic data pertains to polarity. The basic seismic color scheme uses blue which grades to white at zero amplitude, which then grades to red. The question is which of these colors should represent positive amplitudes and which should represent negative amplitudes. Two standards relate color to polarity; both apply to zero phase data. The first standard is the American polarity, which states that:

> *An increase in impedance yields a positive amplitude normally displayed in blue. A decrease in impedance yields a negative amplitude normally displayed in red.*

The second standard is the European (or Australian) polarity, which states that:

> *An increase in impedance yields a negative amplitude normally displayed in red. A decrease in impedance yields a positive amplitude normally displayed in blue.*

Although the standards differ in what yields positive and negative amplitudes, they both state that you should display positive amplitudes in blue and negative amplitudes in red. These are the accepted standards and even though most interpreters use more detailed color palettes than the simple blue-white-red, they still apply. This is because almost all of the palettes in use today are variants upon the original blue-white-red theme. If you look again at the palettes that I show in Figure 11.1, you will see that all of them use colors from the short wave end of the spectrum (blue) to represent one polarity and colors from the long wave end of the spectrum (red) to represent the other.

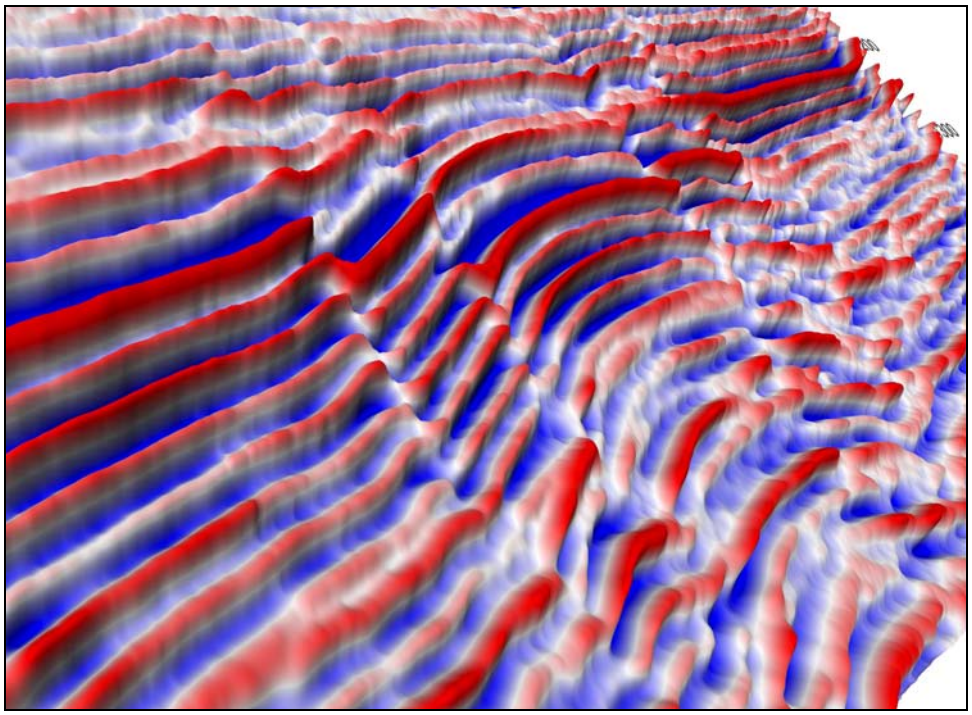


Figure 11.2: SeisScape display using a conventional blue-white-red color palette. Blue is used to color troughs, which have negative amplitudes. Red is used to color the peaks that have positive amplitudes.

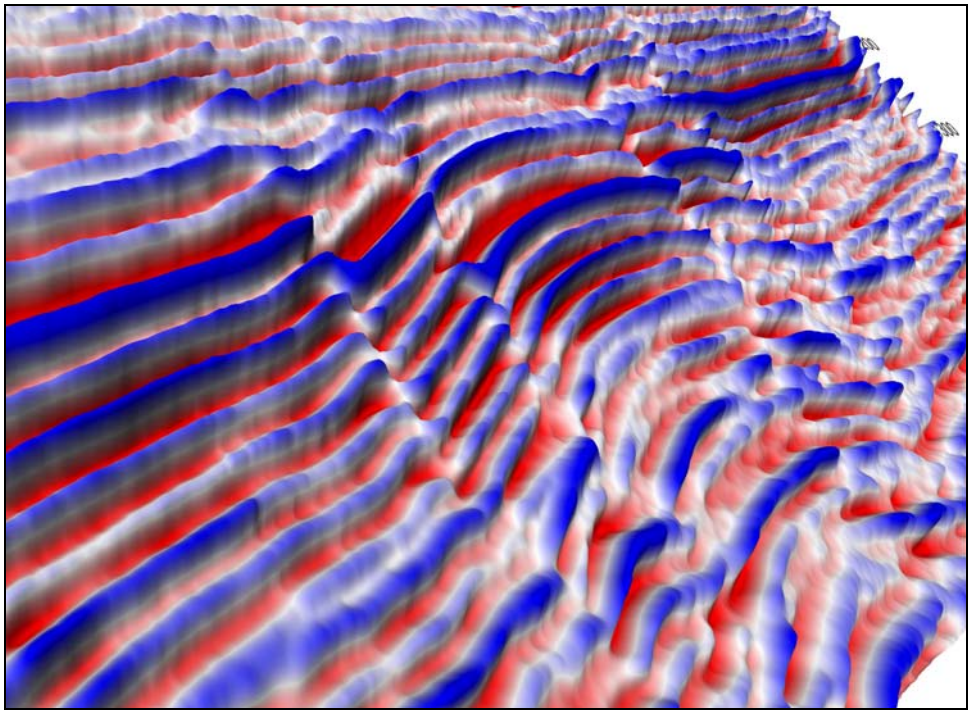


Figure 11.3: Same display show in Figure 11.2 but using blue to color the positive amplitude peaks and red to color the negative amplitude troughs. This palette orientation produces an uncomfortable optical illusion, as the visual system erroneously perceives the red as the high amplitude peaks.

Having a standard is important because there is nothing in the palettes themselves that states “positive” or “negative”. If you use them on variable density displays, you could switch the polarity of the palettes and the viewer would never know. Unfortunately, this does not apply when you use the same palettes to color SeisScape displays. Figure 11.2 and Figure 11.3 shows what happens to our perception when we reverse the color polarity. For the first image, I used red for the peaks and blue for the troughs and the display looks exactly the way you would expect it to. On the other hand, your perception of the second image is anything but normal. For this image, I reversed the colors and used blue for the peaks and red for the troughs. This color reversal produces an optical illusion because you erroneously perceive the red troughs as peaks and the blue peaks as troughs.

This optical illusion is a problem because SeisScape displays are similar to wiggle trace displays in that they have both peaks and troughs. In keeping with the convention for wiggle trace displays, I use peaks to display positive amplitudes and troughs to display negative amplitudes. I followed this convention because peaks dominate both displays. Wiggle trace peaks are solid and therefore they are visually dominant. SeisScape peaks are also visually dominant because they often obscure the troughs. By convention, a normal polarity wiggle trace display shows positive amplitudes as peaks. I adopted the same convention for SeisScape displays so as not to confuse viewers who were comparing the two types of displays.

Using the standard conventions, I should display the peaks in blue and the troughs in red but when I do, I produce the optical illusion that you see in Figure 11.3. I have tested this effect using numerous palettes and I have discovered that the illusion is not the result of using strictly blue and red. Rather, it is more general. The illusion is always generated when you use short wavelength colors for peaks and long wavelength colors for troughs. It never occurs when you use the opposite configuration, i.e., short wavelength colors for the troughs and the long wavelength for the peaks.

It is important to understand how this illusion is generated because it gives us some insight into how we should use color in the future. From earlier chapters, we know that the visual system splits the images into streams of achromatic and chromatic information and processes them separately but in parallel.

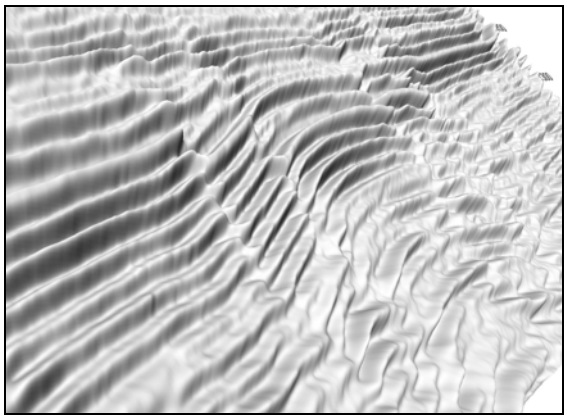


Figure 11.4: Reflectance display of the same data shown in Figure 11.2.

Figure 11.4 is the reflectance image, which is processed by our primary achromatic circuit. Looking at this display, we correctly perceive the peaks and troughs but interestingly, the display is still rather ambiguous. The reflectance along both the peaks and the troughs is similar and by itself, it cannot tell us, which is high and which is low. This implies that our visual system must use other clues, such as the occlusion of one event behind another, to form its achromatic perceptions of the peaks and troughs.

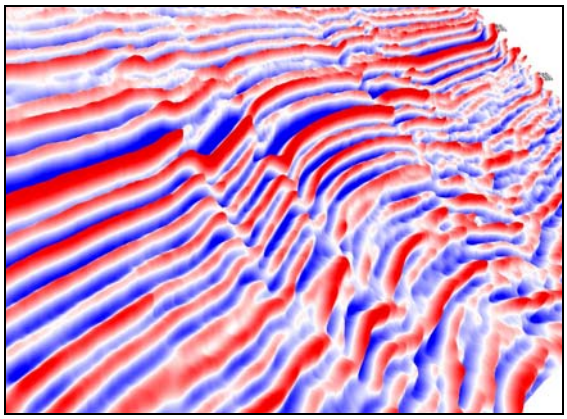


Figure 11.5: The same display shown in Figure 11.2 but without lighting.

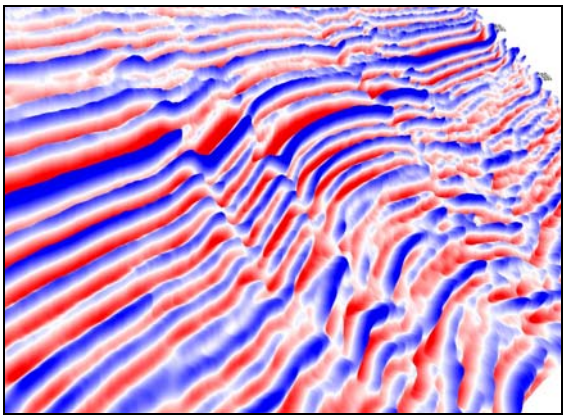


Figure 11.6: The same display shown in Figure 11.3 but without lighting.

Figure 11.5 and Figure 11.6 are the corresponding chromatic images of Figure 11.2 and Figure 11.3 respectively. We process these with our secondary chromatic neural circuitry. Significantly, we observe the same optical illusion in them that we observed on the earlier images that also contained lighting information. Figure 11.5 uses red for the peaks and the visual system perceives it correctly, Figure 11.6 uses blue for the peaks and

the visual system gets it wrong. These two images prove that the illusion is not dependant upon achromatic information and that it arises from our chromatic neural circuitry alone.

Why this illusion is important is because it gives us some insight into what is going on behind the scenes. Interestingly, it shows that despite our establishing form primarily from reflectance, chromatic information can modify and dominate it. Consider that Figure 11.3 provides the visual system with two starkly different perceptions of the same object. Our achromatic channel perceives it correctly (Figure 11.4), whereas the chromatic channel perceives it incorrectly (Figure 11.6). It is the latter incorrect perception that we end up with. This implies that our base perception is formed first from the achromatic channel and then subsequently modified by the perception from the chromatic channels.

Figure 11.2 and Figure 11.3 also show that there is a visual "pecking order" for color, which can only be understood by looking deeper into our chromatic circuitry. To that end, look again at either Figure 11.3 or Figure 11.6 and force your attention onto the blue events. If you do, you will see that the optical illusion goes away. If, however, you let your attention wander it naturally refocuses itself on the red events and the illusion returns. It is almost as if the visual system were generating two separate and conflicting perceptions, a dominant one from the red and a submissive one from the blue. This, of course, is exactly what is happening.

Recalling Figure 2.5, which illustrates the Hering Theory of Opponent Color Vision, perception is actually fed by two parallel channels of color information, an opponent blue-yellow channel and an opponent red-green channel. This explains what is happening to produce this illusion. The red in these images is processed by the red-green channel while the blue is processed separately by the blue-yellow channel. I cannot explain why the perception produced by the red channel is wrong but from everything that I have learnt about the visual system, the red-green channel should be dominant. Ultimately, the blue channel does produce the proper perception, which is why your perception corrects when you focus on the blue. This does not matter though because the red channel, which gets it wrong for some reason, overrides it.

This illusion is more than just an interesting anomaly because it shows that as we move into perceptive environments such as SeisScape, we cannot be as casual with how we use color. By convention when we display seismic data, we consider blue and red to be of equal importance. That was acceptable when we were using perception free variable density displays but we cannot think that way anymore. As soon as we start to talk about perception then we have to take into account the fact that the visual system responds more strongly to long wave colors than to short wave colors. If we ignore that fact then all sorts of curious things are likely to happen.

To conclude this section, since peaks are normally associated with positive amplitudes then by convention they should be colored blue. Fortunately, most interpreters do not follow this convention because they usually associate wiggle trace peaks with red instead. This makes the transition to SeisScape displays smoother because as I have just shown, using blue or other short wavelength colors for peaks is not an option.

11.5 The Annoying, Engaging Palettes

And so, I finally arrive at the very last point of this entire dissertation. So far, I have put visualization into context and differentiated it, from art. I have placed it into context with respect to seismic resolution and shown that it is the display and not the data that ultimately establishes resolution. I have explained how our visual processing system, which is the first principal of visualization, has evolved and what its various properties are. I have followed that line of research to the very limits of human knowledge in an effort to establish how primates determine form and color. I have proven that there is far more relevant information in a seismic section than we are used to seeing and I have developed techniques to expose it.

Nevertheless, I will have wasted all of that effort if I cannot do something about the first point that I raised in Chapter 1. That point is the engaging nature of SeisScape displays, which I illustrate in Figure 11.7.

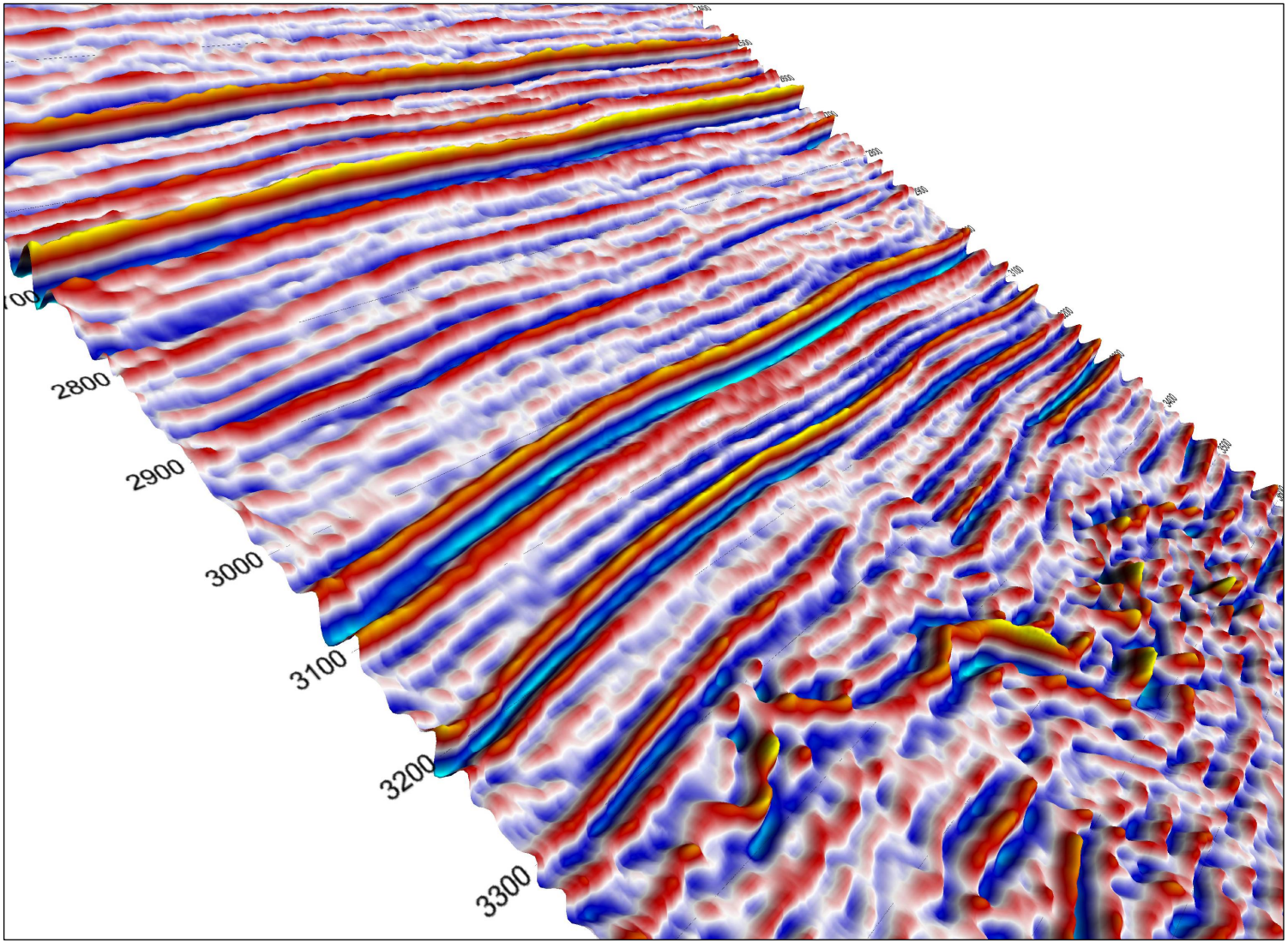


Figure 11.7: SeisScape display of an “engaging” SeisScape scene. The palette uses deep reds and yellows, both of which are colors that attract our attention and provide us with positive feelings. The display is attractive and appealing but it is busy and confusing because we are attracted to everything in it.

At this point, the reader may begin to wonder what I am complaining about. Figure 11.7 is a truly wonderful image of a seismic section. Of all of the images that I have included in this thesis, this is probably the one that I like the best. There is nothing particularly unique about it; it is just one of the many SeisScape displays that I studied over the years. I hope you like it because you are meant to. To produce it, I used a vibrant color palette with deep reds and yellows, which we now know are very attractive to primates. The display is striking, distinctive but it is almost completely uninterpretable.

My first complaint with this image is that the palette makes almost everything stand out and attract your attention. This makes it very difficult to interpret because it is hard to focus on any one event; they are all visually commanding. My second complaint is that this display lacks a dispassionate nature. When I look at this image, I want to stare at it and admire it. This is where the art motif comes in. This display engages me on an emotional level and whereas I do not consider it to be art per se, I would much rather admire it than work with it.

Science, and geophysics is no exception, requires a dispassionate presentation of fact. Wiggle trace and variable density displays are the essence of this. Neither produces strong perceptions and when we work with them, we are free to concentrate on the data itself and not the display. Figure 11.7, by contrast, is the exact opposite of this. It is the antithesis of dispassionate and consequently we are more likely to focus on the display rather than the data it shows. This defeats its purpose, and I realized very early in my research that I had to manage their appealing nature if I were to succeed.

Fortunately, managing the appealing nature of the displays is quite easy; I have been doing it all the way through this thesis. The solution itself is almost trivial and I am tempted to simply state it and end. However, there is the larger issue of whether or not our existing color palettes are appropriate to consider. Because of that, I will spend a few more paragraphs exploring the theoretical aspects of the solution.

Consider our original color palettes. From a strictly theoretical perspective, blue-white-red seems to be a good color combination. As I showed in the previous section,

primates use separate neural processes to establish blue and red and even dichromats do not confuse the two. This makes the combination appear to be ideal for seismic data, which requires distinctive colors to identify positive and negative amplitudes. Everything that I have just said though also applies to a blue-white-green palette but interestingly, very few people ever use it instead. Why we use red instead of green is obvious. Green is simply a very boring color and given a choice between the two, most primates prefer red.

That green is boring and that we prefer red is not just my personal opinion, it is based on why we evolved trichromacy. As I showed earlier in this chapter, our trichromatic vision evolved during the Eocene-Oligocene as our arboreal, insectivorous ancestors were becoming frugivors and folivors. The world that they lived in was predominantly green but their food was predominantly red and yellow. Trichromacy evolved so that they could locate it against a backdrop of green foliage. This leads to the question of what happened when they found it. It is not too much of an intellectual stretch to believe that a hungry ape, upon locating a food cache, would feel very pleased with himself. If he were hungry enough, he might even feel quite emotional.

Besides other things, emotions are one of the ways that our senses communicate with us. It is well established that humans have an emotional attachment to the color red. It is generally considered the color of love but is this because someone, lost in the mists of time, decided it should be or is it because it evokes a strong, primordial, emotional response. I believe it is the latter and that our response to the color red evolved in concert with our trichromacy, as a signal to alert us to our proximity to food. I do not know of any research that supports this concept so you are free to disagree with me. However, it is logical to assume that if we evolved trichromacy to find food that we might also evolve the emotions that let us know when we had succeeded.

Even though this gives a possible explanation to why primates prefer red to green, it does not explain why using a red-yellow palette on a SeisScape display produces a pseudo-emotional response when using the same palette on a variable density display does not. To understand that we have to consider SeisScape displays in a larger context.

SeisScape displays are not just a new form of seismic display; they represent a quantum change in technology. That change is the evolution of technology from being sensation less to being sensational. In this context, the term sensation is synonymous with perception. In geophysics, conventional seismic displays are two-dimensional and for all of the reasons that I have discussed in this thesis, they produce only limited perceptions and sensations. SeisScape displays, on the other hand, are three-dimensional and as such, the visual system cannot distinguish them from reality. They engage all of our visual process and consequently, they produce a full range of perceptions; they are, therefore, sensational.

I hope that I have already shown that engaging all of our visual processes produces measurable and significant benefits. Engaging them, however, produces a chain response. Perception produces sensation. Sensation produces emotion. Emotion can be directed by color. These three things are inevitable and perfectly natural but they also a problem for scientific visualization.

As I stated previously, science requires a dispassionate presentation of fact. In the perception – sensation – emotion chain, these dispassionate facts are found in the first link, the perceptions. Most of these perceptions are defined by the dispassionate reflectance and if we could ignore color and stop there, we would have no problems. We cannot stop there, however, because as I have shown in this chapter, color is too important to our final perceptions to be ignored.

We have to use color but we have to think about it more carefully than we have in the past and we cannot base our color choices on what we “like” anymore. In a percept free environment, color does not really affect us and so we are free to choose colors based upon our personal preferences. In sensational environment, however, for perfectly natural reason, those same colors can be expected to produce an emotional response that is simply out of place.

One of the questions that I had to answer was whether we could continue to use our existing color palettes. It seems like I have done an awful lot of work simply to prove that we cannot.

11.6 HA1 and HA2

The final task that I have to complete is to propose a new approach to using color. Because of my often-mentioned dichromatism, this section will be necessarily brief. I have developed two prototype palettes, HA1 and HA2, both of which reduce visual clutter and calm the displays. Viewers with a better sense of color will probably do much better. To aid you in this, here is a series of observations to keep in mind when designing palettes to use in SeisScape environments.

- Existing seismic palettes were developed for use in two-dimensional, percept free environments. They were also developed to communicate the maximum level of seismic form. They are inappropriate for use in three-dimensional environments, which primarily communicate form via achromatic contrast.

- Use foliage type colors for the majority of the display, this makes the highlight colors stand out and reduces visual clutter and confusion.

- Avoid using vibrant colors such as reds, yellows and oranges, for the majority of the palette. Use them sparingly to highlight specific features.

- Do not try to do too much with color. Reflectance and the physical shape of the surface communicate the majority of the seismic information. Use color to distinguish positive and negative amplitudes clearly and to highlight specific amplitudes.

- Blue, and other short wavelength colors, are appropriate for troughs but not for peaks.

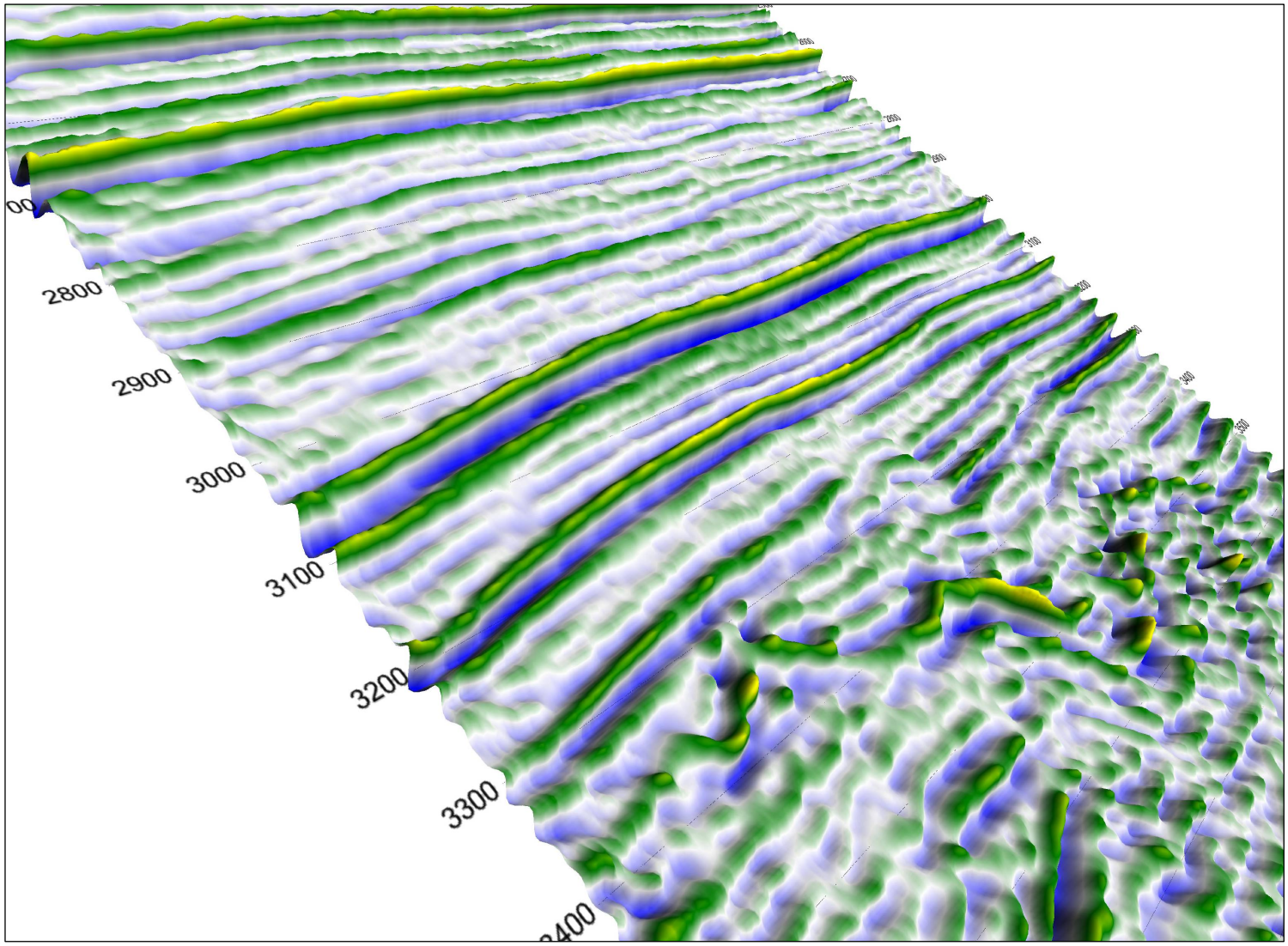


Figure 11.8: The same display shown in Figure 11.7 but using palette HA1. This palette uses yellow for the highlight color and is appropriate for dichromats.

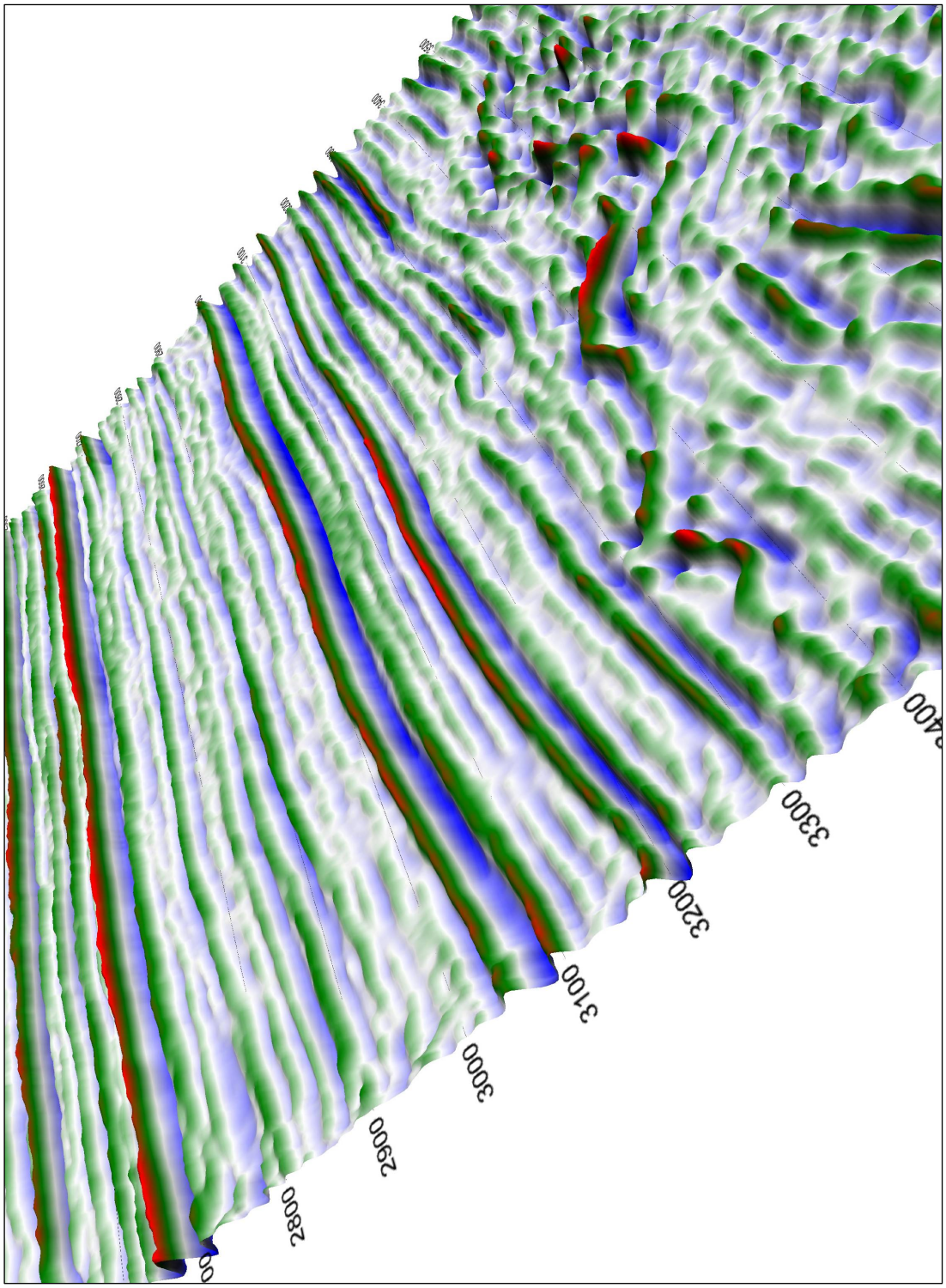


Figure 11.9: The same display shown in Figure 11.7 but using palette HA2. This palette uses red for the highlight color and is appropriate for trichromats.

Figure 11.8 (HA1) and Figure 11.9 (HA2) are both images of the same section of data that I showed in Figure 11.7. In both palettes, I color the negative amplitudes in blue, the zero crossings in white and the background positive amplitudes in dark green. The difference between the two is that HA1 uses yellow for the highlight color whereas HA2 uses red. The use of yellow as the highlight color makes HA1 more universal because even dichromats can distinguish the highlights. HA2, on the other hand, is a better choice for trichromats because it makes use of the unique primate capabilities at distinguishing reds and greens.

If you compare either image with Figure 11.7 you will see that the use of the green foliage background has reduced the visual clutter and made it easier for you to follow events and distinguish the highlights.

As I was finishing this chapter, I decided to explain all this to my wife. I am not sure what I expected her to think but I hoped she would be as impressed and amazed by my insight and intelligence as I was. As one might expect, after 23 years of marriage it takes considerably more to impress her than I showed here. Instead of worshiping at my feet, she took a step backwards and looked at me with a mixture of abject horror and thinly disguised contempt. "Do you mean to tell me", she spluttered, "that after working on this for five years, all you have proven is that green is boring and that red makes us emotional". "Not exactly", I stammered, "it is not quite as simple as that and I did learn a few other things along the way". She stepped back a few more paces and looked me up and down as if I had just beamed in from another planet. Then she walked over to me, hit me on the back of the head and said, "Red is emotional and green is boring – like that wasn't blatantly obvious to begin with."

And then she shook her head, walked away and left me to do the dishes.

T.A.F.

11.7 Examples

I decided to use palette HA2 in the following examples rather than the HA1, which I used throughout this thesis. I did this with the assumption that the reader is trichromatic and that this color palette works best for you. Normally, I would vet these images through some of the young female trichromats that sit nearby. Unfortunately, I have bothered them so much lately that they are starting to suspect that my motives are not purely chromatic. To get around the problem of my not being able to see this palette very well, I setup the images using HA1 and then switched to HA2 for the output. I can only hope that the effects that I am looking for survive the transition.

I present these examples, virtually without comment, to see if I have accomplished what I set out to do. My early displays impressed people but for the wrong reasons. They impressed because of their stunning visual appeal, which by now you should know was never my goal. My single goal was to impress with the seismic detail that my displays exposed. To that sole purpose, alone I present the following images.

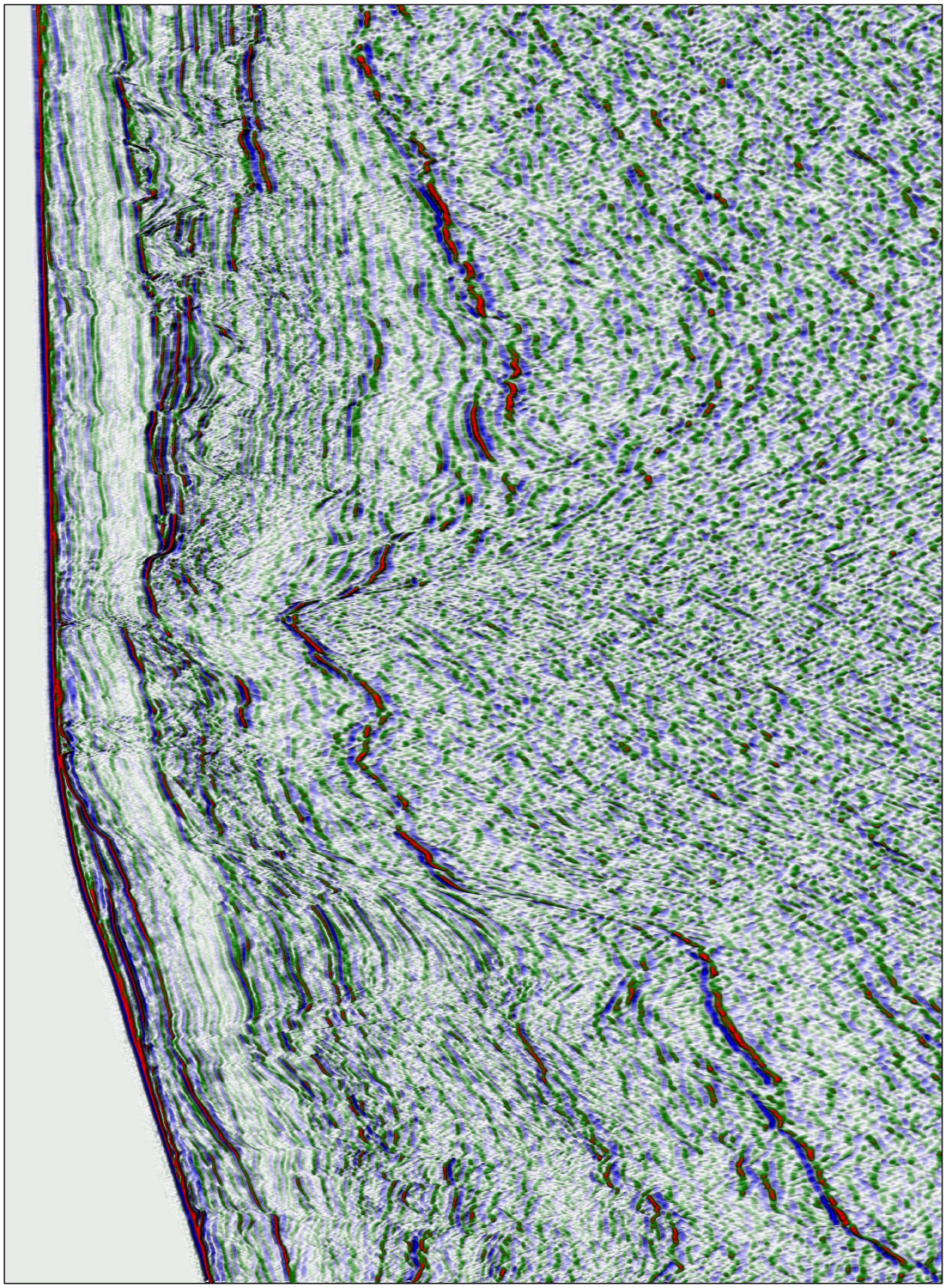


Figure 11.10: SeisScape display of a large section (>3000 traces) of Trujillo data.

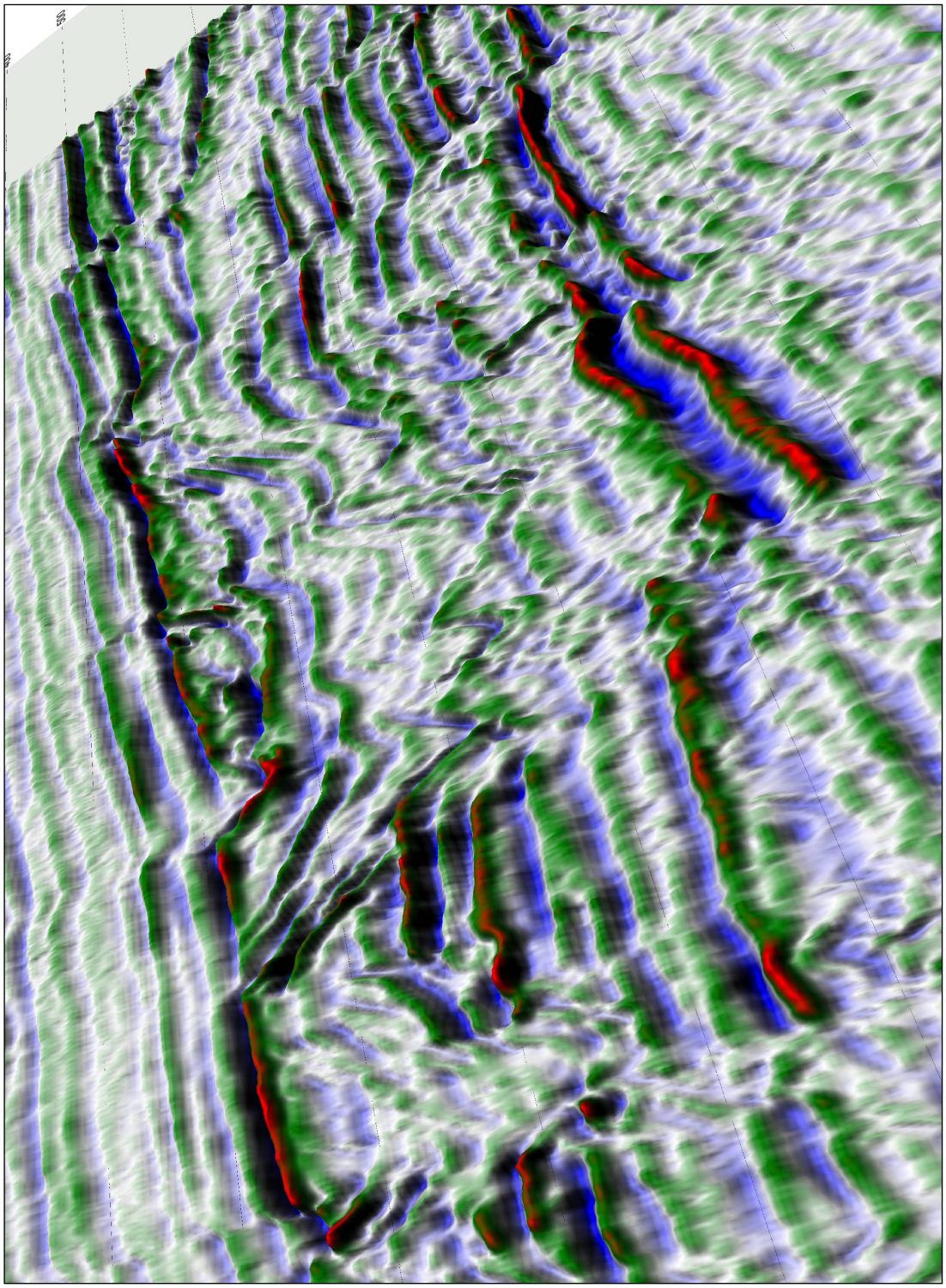


Figure 11.11: SeisScape display of the channel located upper right in Figure 11.10.

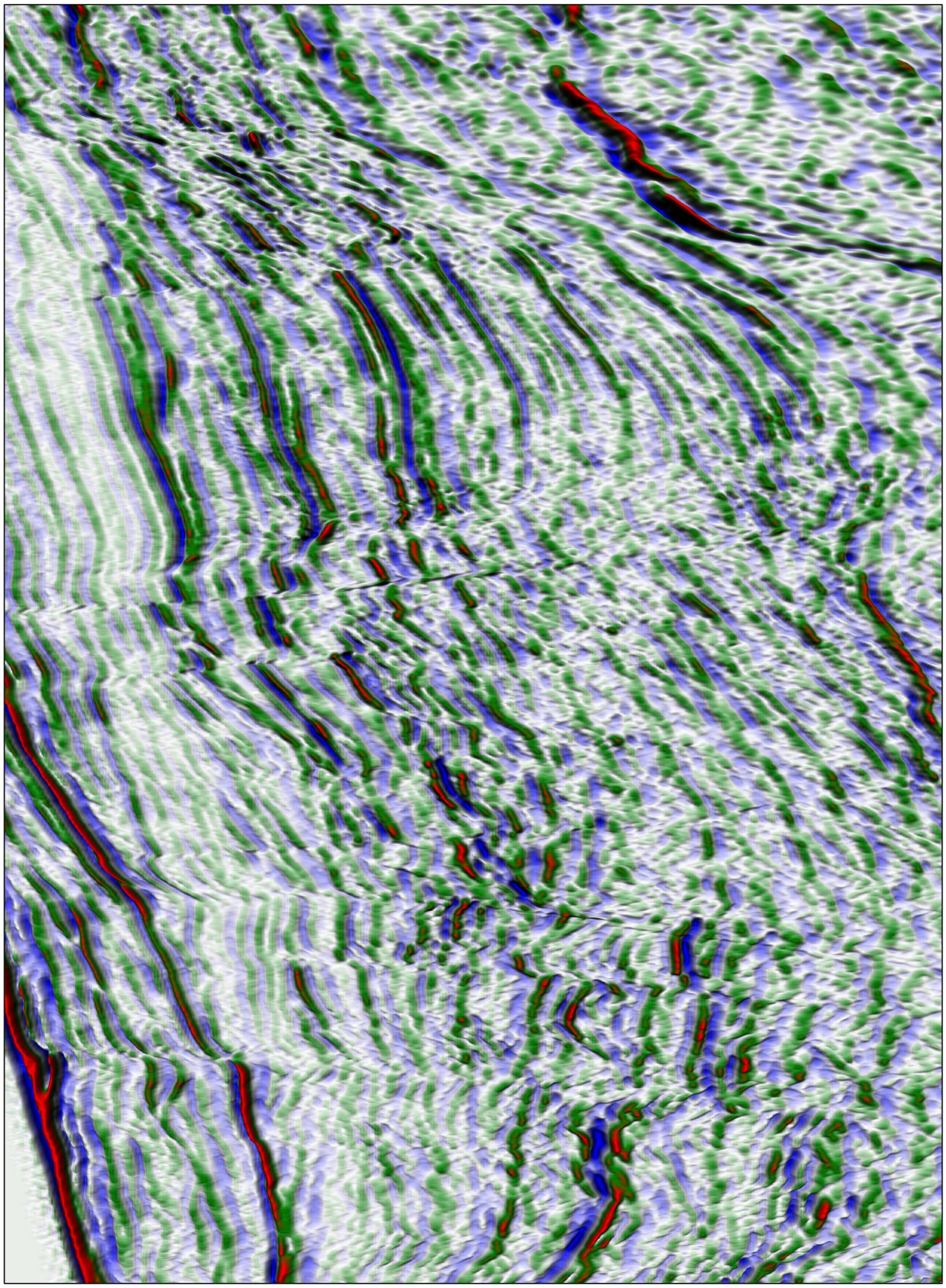


Figure 11.12: SeisScape display of the complext faulting located at the upper left of Figure 11.10.

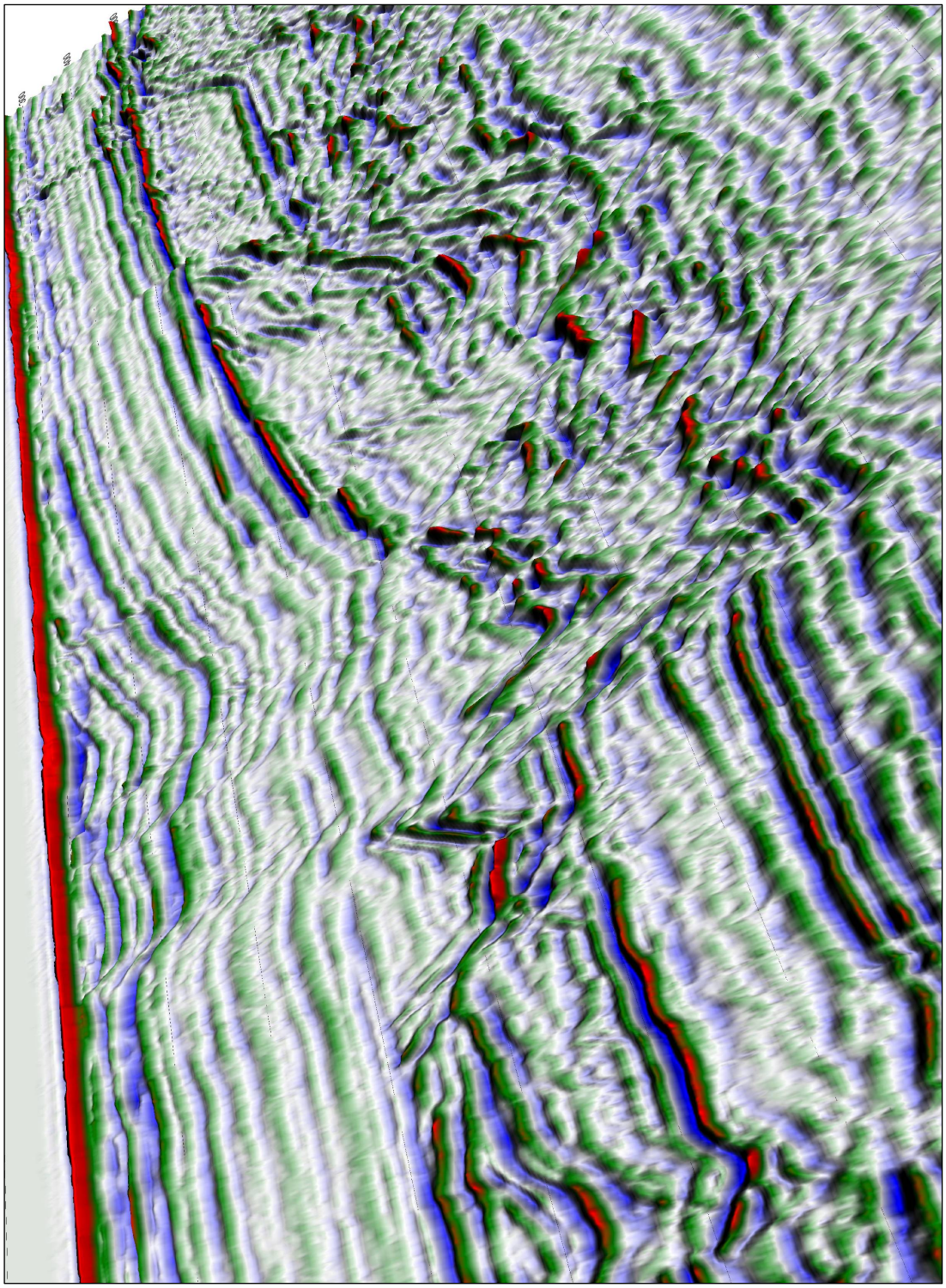


Figure 11.13: SeisScape display of a turbidite section from Trujillo.

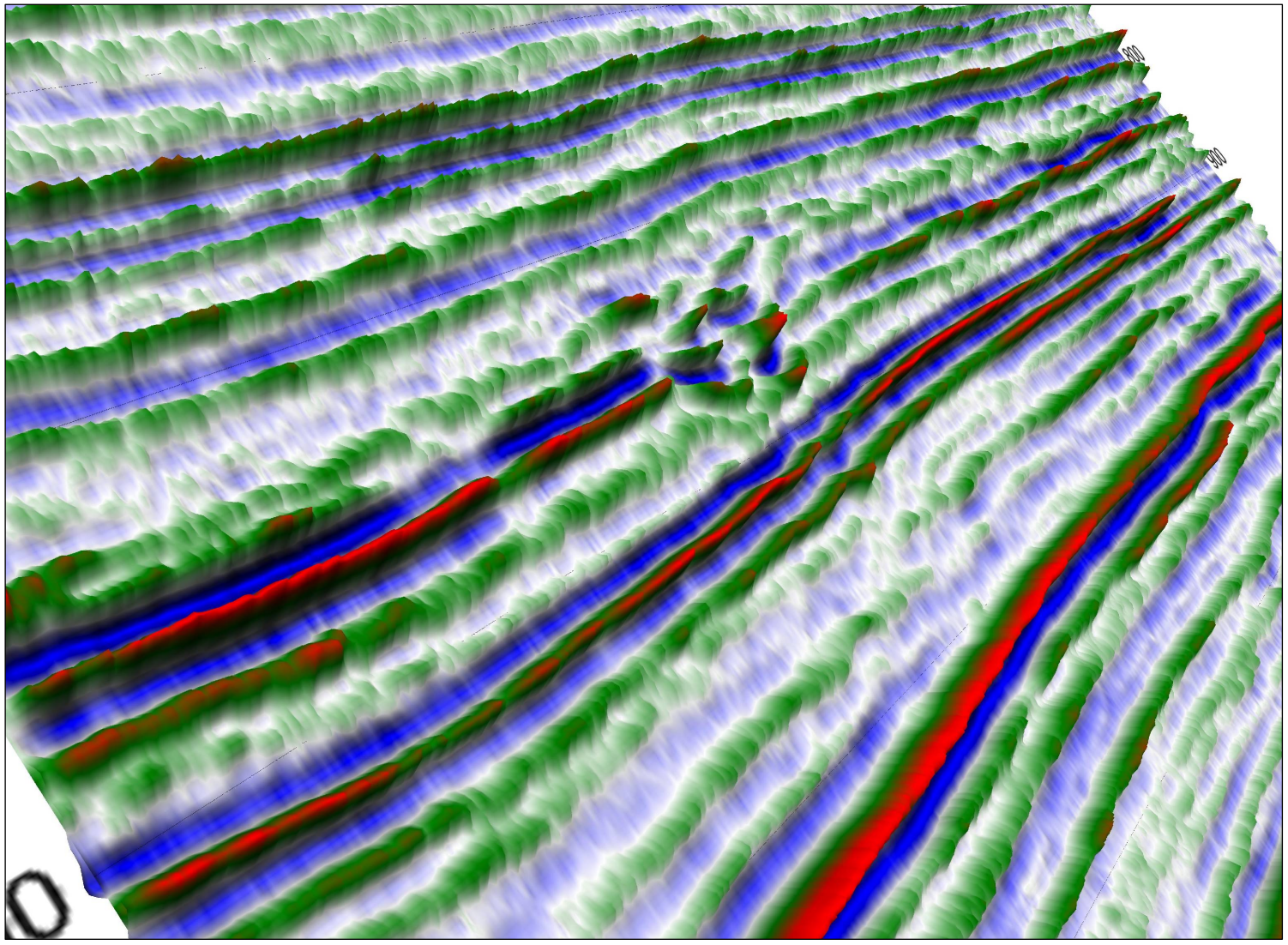


Figure 11.14: SeisScape display of a Southern Alberta channel. The display uses a “sea level” composite density display to show negative amplitudes.

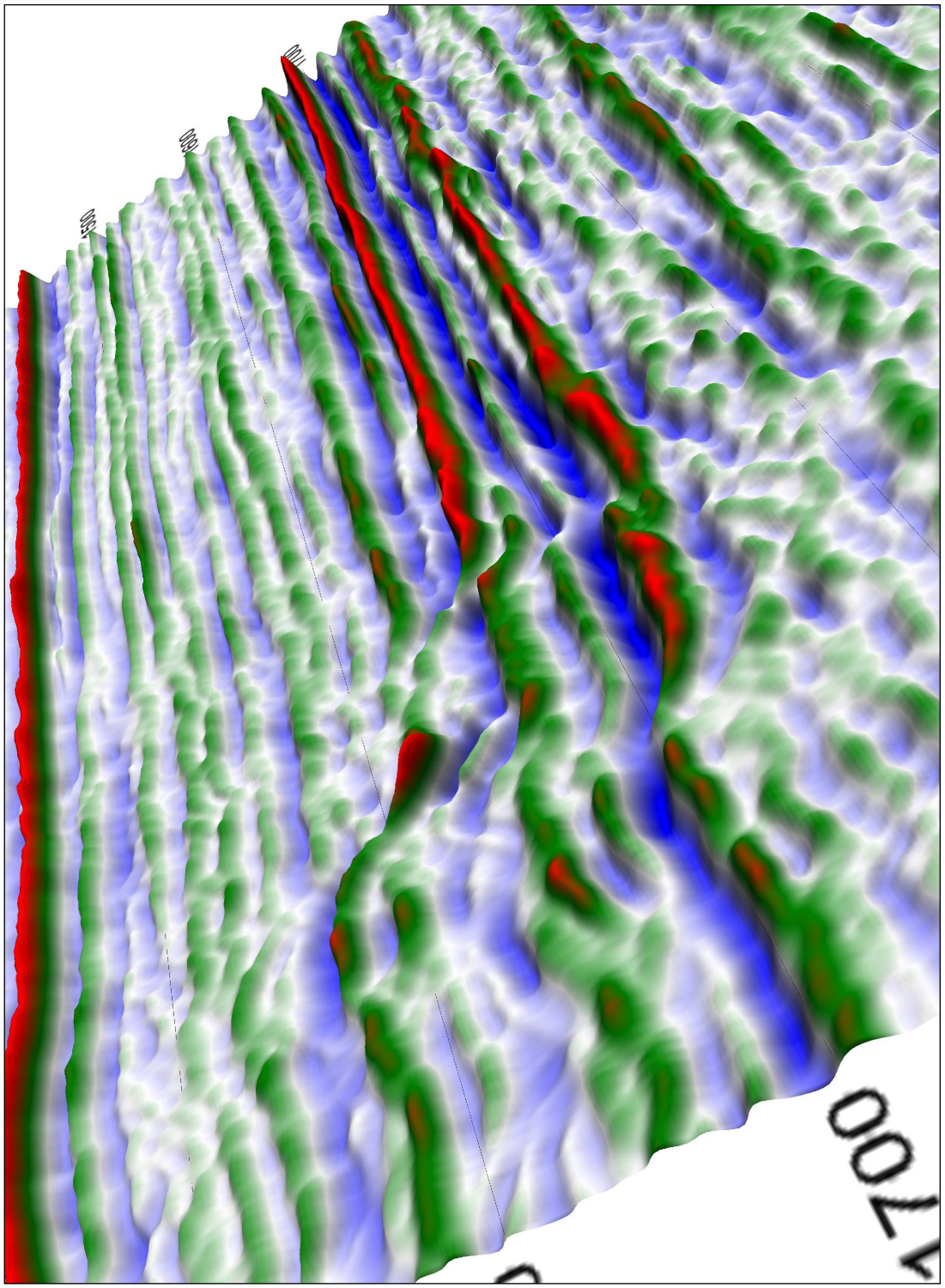


Figure 11.15: SeisScape display of a Leduc Reef.

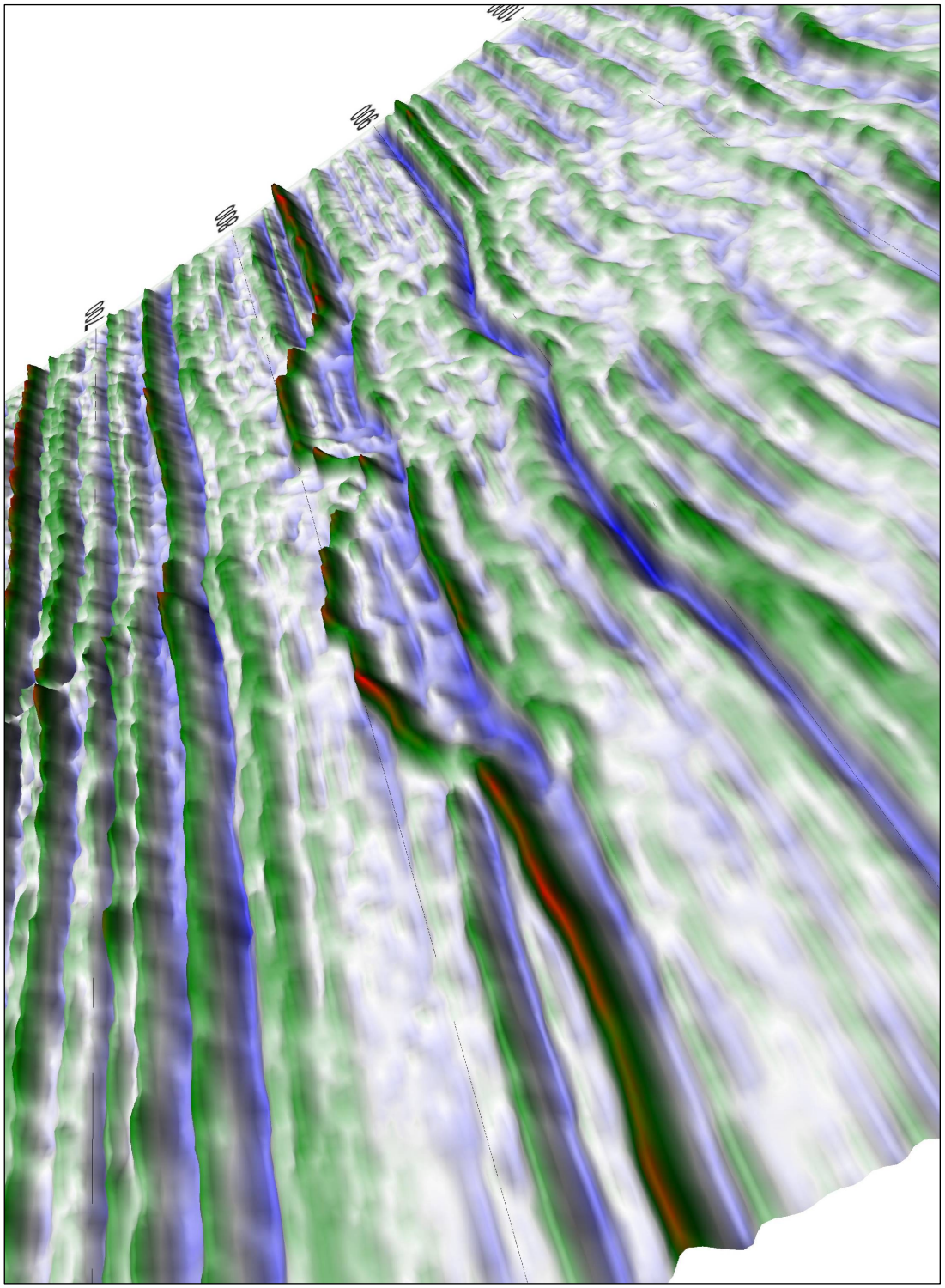


Figure 11.16: SeisScape display of detail within Winnepegosis reefs.

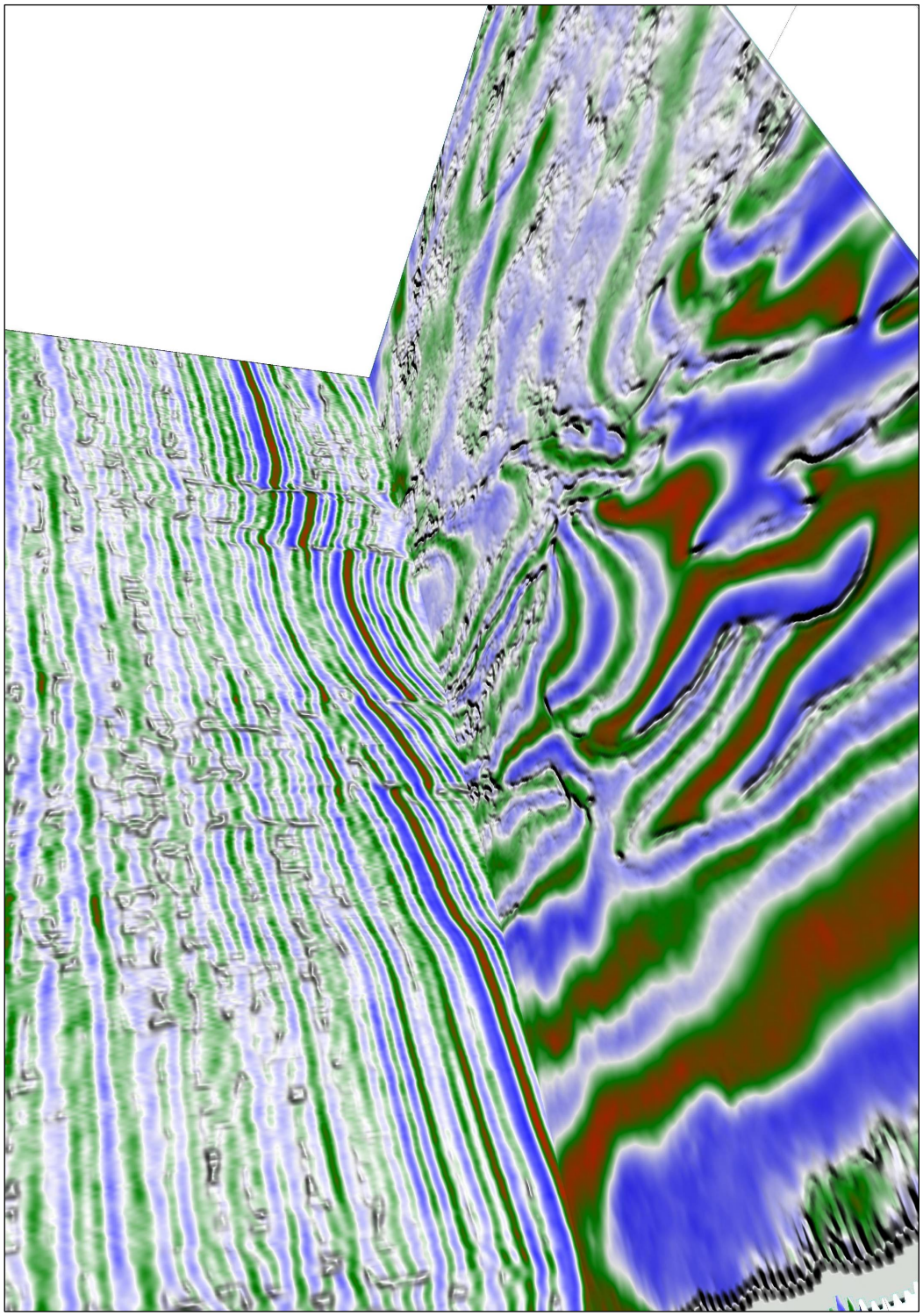


Figure 11.17: An example of co-rendering seismic amplitude as color and coherency as reflectance, which puts the coherency into context.

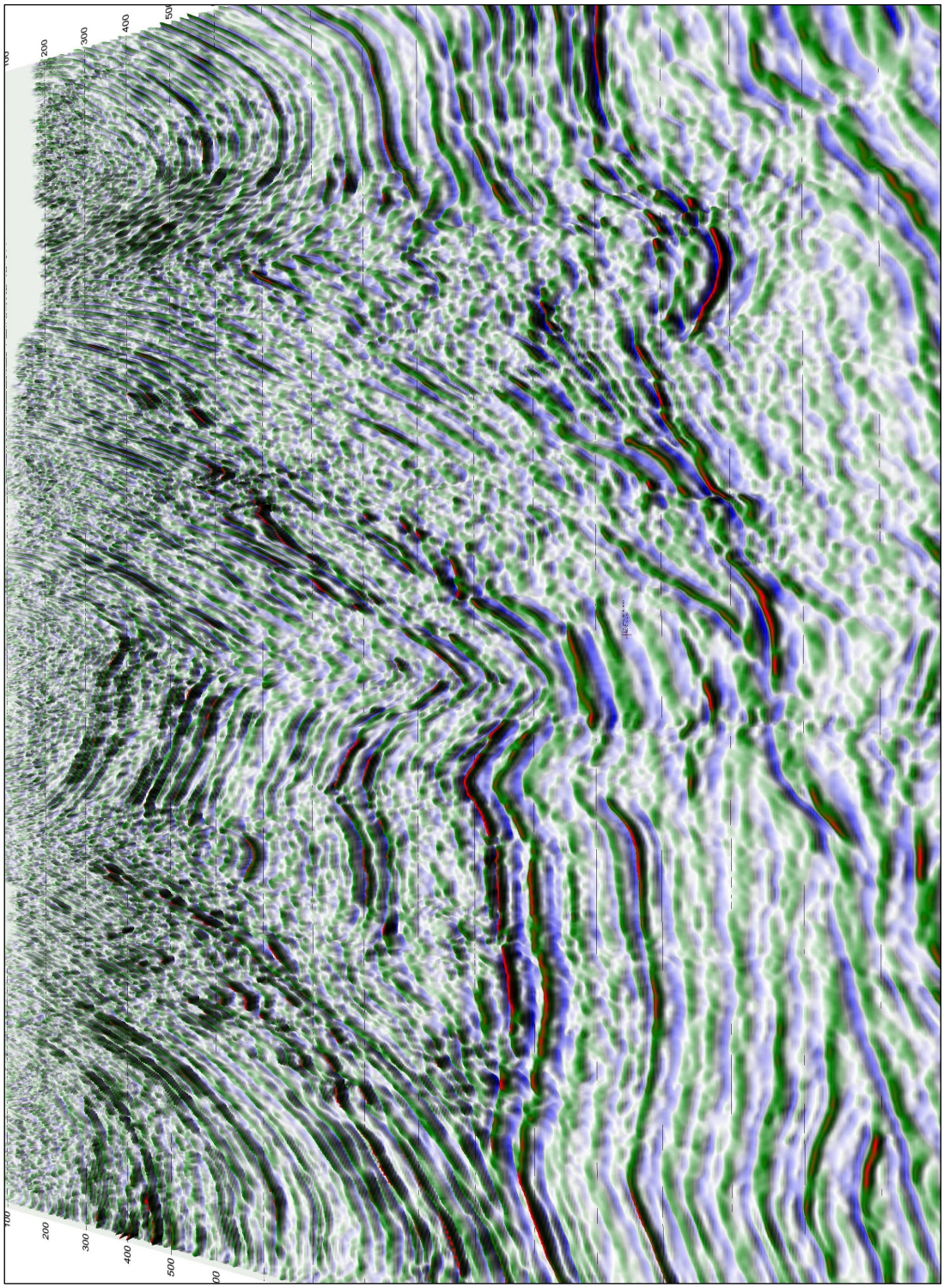


Figure 11.18: SeisScape display of prestack migrated Alberta Foothills line.

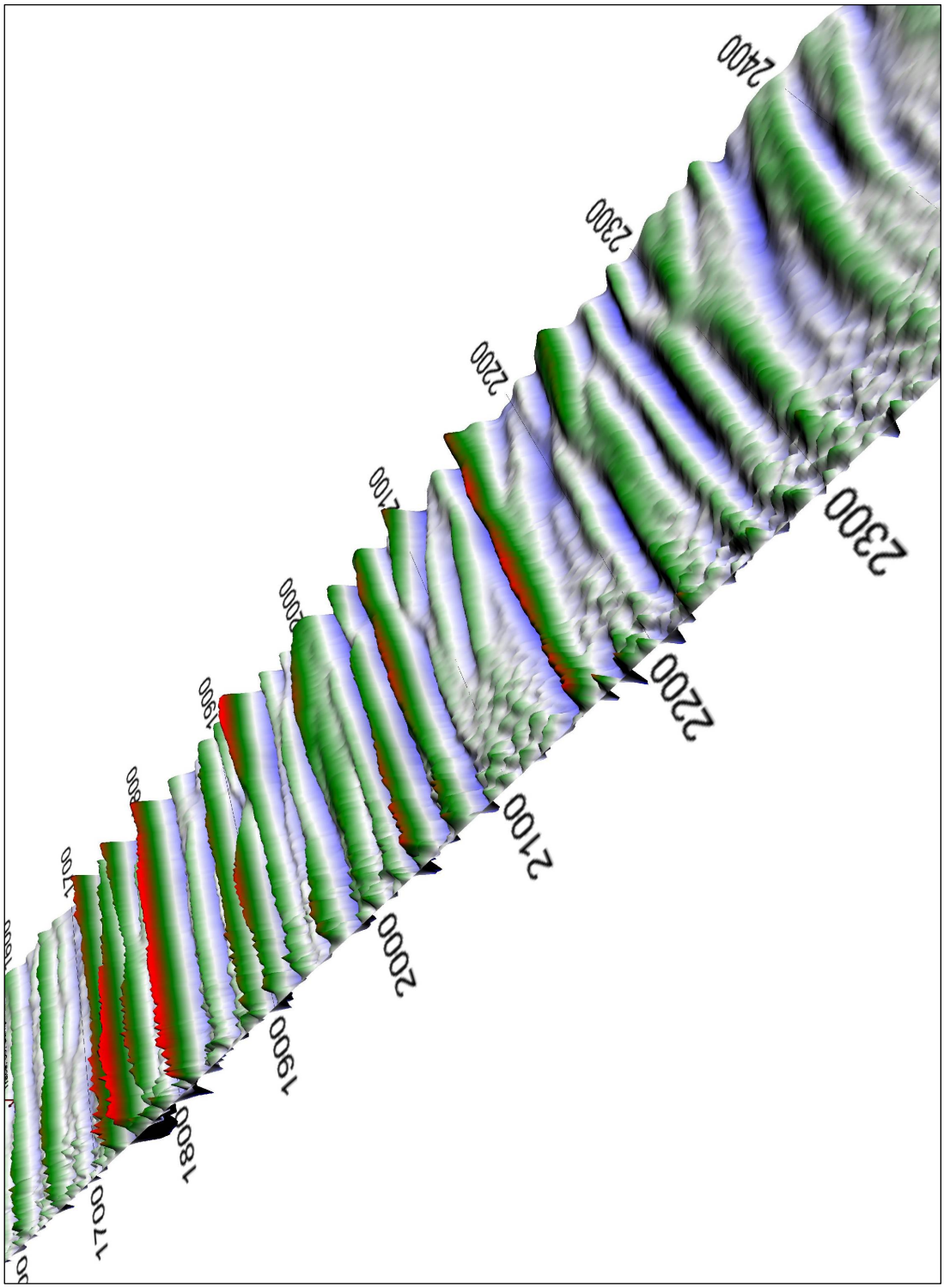


Figure 11.19: SeisScape display of multiples and amplitude anomalies within a common offset record.

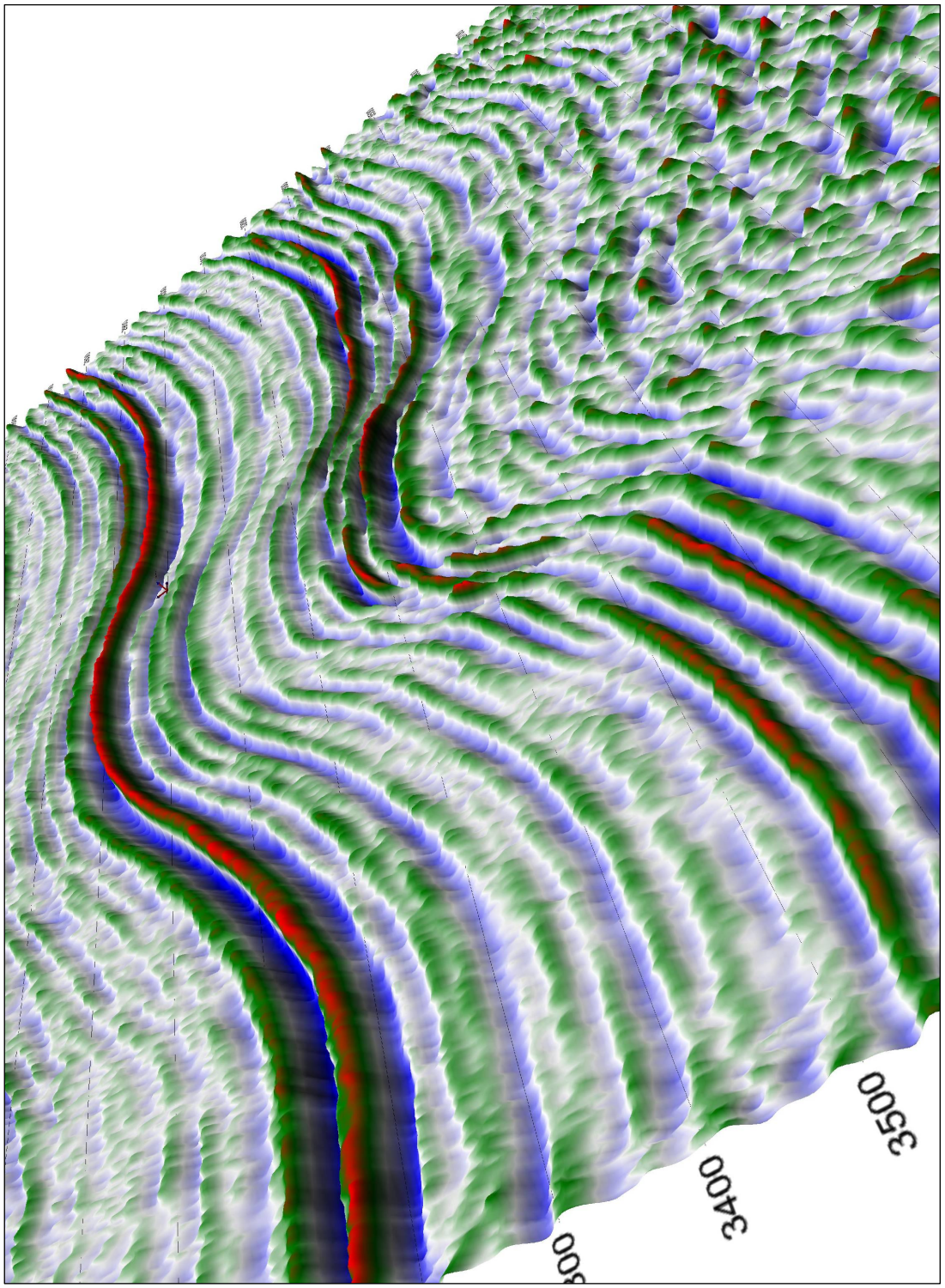


Figure 11.20: SeisScape display of a structure from the Huallaga area of Peru (data courtesy PeruPetro).

CHAPTER TWELVE: SUMMARY, CONCLUSIONS AND FUTURE WORK

When the flush of a new-born sun fell first on Eden's green and gold,
Our father Adam sat under the Tree and scratched with a stick in the mould;
And the first rude sketch that the world had seen was joy to his mighty heart,
Till the Devil whispered behind the leaves "It's pretty, but is it Art?"
"The Conundrum of the Workshops"
Kipling

12.1 Summary and Conclusions

12.1.1 Art vs. Science

I started this final chapter by quoting the lines from Kipling that I used to open my first chapter. They set the motif for the entire work by asking the question "It's pretty, but is it art?" It is time now to answer that question because I have carried the motif right through the text. The answer to whether or not SeisScape displays are art is that in my opinion, they are not. To me, they are pretty, but they are definitely not art! Art has its own principles, which SeisScape displays do not follow. They are pretty and they can be engaging but I knew before I started this thesis that they were not art in a classical sense. With that in mind, you might well ask why I carried the art - science motif all the way through this dissertation.

This thesis concerns itself with the visualization of seismic data. The first thing that I had to do was to define what visualization meant in a geophysical context. This is not as obvious a task as it first appears because visualization has two sides and I used the art - science motif as a metaphor for them. SeisScape displays are not just a new form of seismic display; they represent a quantum change in technology. That change is the evolution of technology from being sensationless to being sensational. In this context, the term sensation is synonymous with perception. In geophysics, conventional seismic displays are two-dimensional and produce only limited perceptions; they are virtually sensationless. SeisScape displays, on the other hand, are three-dimensional and produce a full range of perceptions; they are, therefore, sensational.

This is where the two sides of visualization come from. Perceptions produce sensations but sensations produce emotions or feelings. The first side of visualization is the perceptions that it produces. The second side is the emotions and feelings that are attached to the perceptions. I used the art – science motif to illustrate this dichotomy, science is the perceptions, art is the emotions that arise from the perceptions.

To understand why I had to make this distinction, you only have to consider my early SeisScape displays. They were new and unique and I found it very easy to impress with them. The viewers, though, were mostly impressed for the wrong reason. I produced visually stunning displays that engaged the viewer. The displays were impressive but in a scientific setting such as seismic interpretation, we require a dispassionate presentation of facts. Feelings are generally irrelevant, unwanted, distracting and inappropriate. Everyone really liked my early displays; however, they liked them because they were pretty. They considered them engaging curiosities but because they were not dispassionate, they did not recognize their scientific potential.

I realized that before I could make any progress I had to define the goal of scientific visualization as clearly as possible. Ultimately, I defined visualization as the science of making visual displays as dispassionate as possible while making them as informative as possible. I defined everything else as art and irrelevant.

12.1.2 Seismic Data and Perception

Vision is one of our five classical senses, all of which translate their sensory stimulations into an experience called perception. The sensory experience itself is called a percept but unlike the percepts from our other senses, we rarely note visual percepts. That is because unlike the other senses, the visual system is never quiescent. We notice other senses as part of the world around us but the visual system **is** the world around us.

The percepts of vision are the objects that surround us and interact with us. It is the purpose of the visual system to identify these objects and discover their properties. Given that visualization is the science of communicating information via the visual processing system, to succeed its end products must also be percepts.

A seismic section is a complex mosaic of events, faults and noise trains, which visualization must convert into visual percepts. Conventional variable density displays, however, produce virtually no perceptions, which imply that they do not fully engage our visual processing system.

The modern model for how humans (and other primates) see color is called "The Stage Theory". The first part of the model is called the Trichromatic Theory of Color. It can be considered as the receptor stage, and it consists of the three photopigments (the blue, green and red cones). The second part is called the Hering Theory of Opponent Color Vision. Hering hypothesized that the trichromatic signals from the cones were subject to subsequent neural processing. He proposed two major opponent classes of processing, a spectrally opponent process, which provides our ability to separate hues and a spectrally non-opponent process, which provides our achromatic vision.

Opponent processing produces three streams of visual information, each of which is processed by a separate neural circuit in the visual cortex and each of which makes its contribution to perception. The primary circuit is called the Achromatic Neural circuit. It processes our achromatic information and it dominates perception. The secondary circuit, which is called the Chromatic Neural Circuit, has two channels. One processes Red-Green opponent information and the other processes Blue-Yellow opponent information. This processing of the Trichromatic cone signals into three channels of information is known as Trivariant color vision and among mammals; it is unique to Old World primates.

Trivariant color vision explains why purely chromatic variable density displays do not produce any sensation of perception. Primates primarily produce visual percepts via their achromatic channel. The chromatic channels are important but they act upon our base achromatic perceptions. Since chromatic variable density displays lack any achromatic information, there are no percepts to act upon and the resultant displays do not produce visual sensations. Because variable density displays do not convert seismic objects into percepts, they cannot be considered as true seismic displays.

12.1.3 The Relevance of Visualization to Seismic Resolution

Conventionally, we consider that there are two principal forms of resolution; temporal which is the ability of the seismic wavelet to resolve reflections (in time) from thin beds and spatial which is the ability of the wavelet to resolve closely spaced geological details. The mathematically derived limits of both are well known and understood. These limits, however, are simply the theoretical limits of resolution and do not pertain to the empirical limits that are observed on any seismic display.

Ultimately, the limits of resolution are determined by what we can see in the data. Consequently, we must consider that there is a third form of resolution, namely visual resolution, which acts upon the previous two. To help understand the difference between the theoretical and empirical limits of resolution I defined two terms. The first was "absolute resolution", which is a product of the conventional processes of enhancing spatial and temporal resolution and is a quality of the data itself. The second term was "apparent resolution", which is a product of the visual system and the display, and is the subset of the absolute resolution perceived on any given display.

In the context of these two definitions, the display serves as a filter upon resolution. The data fed into the filter is the absolute resolution; the output is the apparent resolution. Unlike a conventional filter whose output is a modified set of data, the output of the display filter is a set of perceptions, all of which occur in the mind. Conceptually, the display filter has two stages. The first stage is the physical display, which serves to produce the input to the second stage. The second stage is the primate visual system, which produces, as the final output, a set of perceptions. In this context, visualization becomes the process of designing display filters such that they have minimal effect. Its goal is to make the apparent resolution equal to the absolute resolution.

Visualization is the third science of seismic resolution. Whereas deconvolution and migration establish the limits of absolute resolution, visualization establishes the limits of apparent resolution.

12.1.4 Conventional Seismic Displays

SeisScape displays are a fundamentally new type of seismic display. This does not mean, however, that they are ideal for all circumstances. There are three conventional seismic displays, (1) wiggle trace displays, (2) achromatic (gray-scale) variable density displays, and (3) chromatic variable density display. Geophysicists generally recognize that none of these displays is ideal for all circumstances. I showed in Chapter 2, that purely chromatic displays do not fully engage the visual system and should not be used as the primary seismic display. Given the nature of SeisScape displays, what place do the other two conventional displays have in the future?

Of all of the seismic displays, wiggle trace displays have the lowest apparent resolution. Consequently, we should not use them in the future as the primary seismic display. However, since we construct the display purely from lines and edges, they feed our primary pattern recognition engine better than any other display and therefore they remain an essential component of seismic visualization.

In terms of apparent resolution, amplitude mapped gray-scale displays are higher resolution than wiggle trace displays. However, they are lower resolution than the shaded-relief component of SeisScape displays and they produce less well-defined percepts. Given that the visual system processes amplitude mapped gray-scale displays and shaded-relief displays identically, we should deprecate the former in favor of the latter.

12.1.5 The Origins of Primate Vision

One of the goals of this thesis is to lay the foundation for a more scientific approach to visualization. To that end, it is important to understand the physiological processes that give rise to our perceptions. It is important to know how we see but it is even more important to know what we have evolved to see. Among mammals, catarrhine primates, the clade to which we belong, possess a unique form of vision. Our vision evolved for specific dietary reasons and despite the fact that we have moved beyond our original

environments, it has not. We retain much of the same visual processes of our early ancestors.

Humans are classified as order *Primates*, suborder *Haplorrhini*, infraorder *Simiiformes*, parvorder *Catarrhini*, superfamily *Hominoidea*, family *Hominidea*. Primates are essentially an arboreal species and they are well adapted for a life of climbing, leaping and running in trees. They differ from other mammals in the organization of the neocortex. The olfactory sense is the primary sensory modality in all nonprimate mammals. In primates, however, the olfactory system has been severely reduced and replaced with a dominating tactile and visually dominant sensory system. Whereas all other mammals are dominated by their sense of smell, primates are dominated by their senses of touch and sight.

A significant development in the history of anthropoid vision was the development of the fovea, an area of the retina with a high density of visual receptors and a low ratio of photoreceptors to ganglion cells. The fovea improves visual acuity in a small area of the visual field and among vertebrates is found in fishes, reptiles and birds that are diurnal visual predators. This suggests that haplorrhines evolved a fovea to increase acuity for diurnal visual predation of insects. This occurred early in anthropoid evolution while they were still predominantly insectivorous, and as they were converting from nocturnal to diurnal habits.

12.1.6 General Organization of the Primate Visual System

A reader familiar with seismic will recognize that it can be split into three broad areas, each with links to the others; these are acquisition, signal processing and interpretation. The visual system can also be split into these three same areas. The reader is cautioned not to take the analogies too far because the feedback systems between the three are infinitely more complex than their geophysical equivalents. Nevertheless, it is possible to consider the visual system in terms of acquisition, processing and interpretation.

The goal of the visual processing system is perception, perceptions being loosely defined as "*the internal representations of the external world*". In terms of its geophysical equivalents, the visual acquisition system captures the photons that enter the eye and converts them into streams of neural impulses. It essentially projects the three-dimensional physical world onto a two-dimensional plane. The visual signal processing system then modifies the neural impulses leaving the photoreceptors and converts them into parallel streams of information that travel the optic nerve. Finally, the visual cortex interprets these streams of information and attempts to reconstruct, in the mind, the percepts of the original three-dimensional world.

12.1.7 The Visual Acquisition System

The retina is the part of the eye that receives the light from the external three-dimensional world and converts it into chemical energy. This chemical energy activates nerves that transmit the light energy out of the retina and onto the higher functions of the visual processing system and the brain. There are ten layers of cells in the retina that can be seen with a microscope. These are arranged into four primary layers; (1) the pigment epithelium; (2) the photodetector layer containing the rods and cones; (3) a layer of nerve cells called the bipolar cells; (4) a layer of neurons called ganglion cells, which transmit the visual information out of the eye along the optic nerve fibers.

The retinal layers are reversed from what one would expect. The epithelial layer, which contains the photodetectors, is at the very back of the eye and the ganglion layer is at the front. Therefore, for light to sensitize a photoreceptor it first has to travel through all of the non-photosensitive layers of the retina. In the human retina there are approximately 6,400,000 cones and between 110,000,000 and 150,000,000 rods. Although there are a vast number of individual photoreceptors, there are only approximately 1.2 million axons or neural fibers in the optic nerve. This indicates that the raw visual signals undergo preprocessing before being passed onto the brain.

Cones, which are responsible for our daylight vision, are not equally spaced throughout the retina. Cone density peaks in the area of the retina called the fovea

centralis and rapidly falls off to an even density outside of it. Catarrhine primates are the only mammals to develop a fovea and it provides them with most of their visual information. It is only about 1500 μm in diameter and covers roughly six deg of visual arc. The fovea is only slightly larger than a pinhead, however, when the gaze is fixed on an object the lens and the fovea are in direct alignment and so it provides us with most of our critical visual information. Within the fovea, cone densities exceed 100,000 cones/mm^2, consequently anthropoid visual resolution matches theoretical optical resolution.

12.1.8 The Visual Signal Processing System

Catarrhine primates have three types of cones, which are known colloquially as the blue-cone, the green-cone and the red-cone. In name, these cones evoke similarities to the RGB colors of a computer monitor and it is logical to assume that we form colors and images in the same way. However, despite their names, the photoreceptors do not emit a color-coded signal. The raw signals leaving each cone merely indicate if the amount of light being received by the cone has increased or decreased; it says nothing about the wavelength of the light being received. This raw information bears no relationship to what we are conscious of seeing. Our conscious vision is constructed out of multiple streams of information that are extracted from this raw chemical energy and all of these streams, in some way, depend upon contrast.

There never exists, anywhere in the brain, a single image that bears any relationship to what we are conscious of seeing. This is a critical point for the science of visualization. Each image that we produce is ultimately segregated into multiple, parallel streams of information, and each stream is interpreted separately and for different purposes. Knowing what those streams are is vital to understanding how we communicate visual information.

The visual signal processing system produces streams of information for detecting achromatic contrast, chromatic contrast(s), movement, position and orientation. Of particular interest to visualization are the “private line” streams of information that

directly connect each cone in the fovea centralis to the brain. The processing of the raw visual signal into these “private line” streams primarily occurs in three locations; (1) the bipolar cell layer of the retina; (2) the ganglion cell layer of the retina; and (3) the lateral geniculate body.

12.1.9 The Visual Interpretation System

There are two primary pathways of information from the lateral geniculate nucleus into the primary visual cortex. The first is the parvocellular pathway, which arises in the midget (PC) ganglion cells of the retina and is processed by the ventral stream of the visual cortex. The second is the magnocellular pathway, which has its origins in the parasol (MC) ganglion cells. It is processed by the dorsal stream of the visual cortex. The ventral stream is of particular importance to visualization. It is called the “What Pathway” because it is associated with form recognition, object representation and the storage of long-term memory.

By the time it reaches the visual cortex, the raw chemical energy leaving each cone has been converted into a single opponent signal that contains two signals, a chromatic signal and an achromatic signal, multiplexed together. These two signals are demultiplexed in the ocular dominance columns that are located in area V1 of the visual cortex. The achromatic signals are extracted by combing the L and M cone signals along predefined axis of orientation. This system is specifically tuned to detect borders of energy contrast; it is essentially an edge detection system.

Chromatic contrast is harder to determine and requires a larger mosaic of cones. This system of determining chromatic contrast has the disadvantage that because it integrates over a large area of visual space it loses the ability to detect fine scale changes. Because of this it posses far fewer orientation selective channels than does the achromatic system. On the other hand, it produces four chromatic contrast detectors for each unit of chromatic space. These detectors are highly sensitive to any change in the spectral characteristics of light falling upon the mosaic of cones but are relatively insensitive to a change in the total energy being received.

In the real world, when an object is viewed, the striate cortex produces two streams of visual information that are processed for form by separate neural circuits. Of these two, the achromatic stream is the sharper because a higher number of orientation-selective cells, each of which has a very small receptive field, detect it. The purely chromatic stream is lower resolution with less well-defined edges because only the low-resolution orientation-selective chromatic contrast detectors detect it.

In the case of purely achromatic contrast, a sharply defined but achromatic object is perceived; in the case of purely chromatic contrast, a blurry but colored object is perceived. When both types of contrast are present, it is believed that they suppress each other; the chromatic detecting system suppresses the achromatic color and the sharply defined achromatic border suppresses the fuzzy chromatic border.

12.1.10 The Objectives of Seismic Visualization

A seismic section is a complex mosaic of overlapping and often conflicting signals, some of which are geologically or seismically relevant and some of which are noise. Of the relevant signals, some have strong amplitudes and we can see them on all seismic displays. Some, however, have weak amplitudes and are superimposed on the stronger events. I consider that these two levels of events form different structures within the seismic as a whole. The strong amplitude, major events, forms the seismic macrostructure whereas the weak amplitude events form the seismic microstructure.

I defined these terms because this thesis is dedicated to developing techniques to improve seismic visualization. Those techniques generally fall into one of two classes depending upon which type of structure they are designed to enhance. Some techniques enhance seismic macrostructure whereas some enhance seismic microstructure. This leads to the twin purposes of seismic visualization. Specifically, the primary purpose of seismic visualization is to reveal seismic microstructure. In terms of absolute and apparent resolution, this equates to minimizing the difference between the two. The secondary purpose of visualization is to reveal the amplitude structure of macrostructure events.

12.1.11 Tessellating a Seismic Surface

A SeisScape display is a three-dimensional representation of seismic data and is composed of three elements; (1) a tessellated mesh of points that form the mosaic of the surface; (2) a lighting component that illuminates the surface; and (3) a variable density color display that is draped over the surface. Tessellation is the process of converting a series of points into a triangulated surface. Given that seismic data occurs on a regular grid, it appears that constructing a tessellated seismic mesh should be simple. However, tessellating four points is always ambiguous because the same four points can be connected two ways. One way favors up-dipping seismic events whereas the other favors down-dipping events.

Before you can correctly tessellate a series of points, apriori knowledge of the surface is essential. Rendering a model in three-dimensions requires two sets of data; (1) a set of vertices that define the points in the model; and (2) a set of indices that define the vertices for each triangle. Whereas the vertices define the general outline of the model, it is the indices that give it shape. Vertices are just points in space; indices form surfaces out of those points. Tessellation determines what those indices should be.

Tessellation affects both macrostructure and microstructure. The most obvious effect of incorrect tessellation is to produce a saw-tooth pattern that degrades the appearance of dipping macrostructure events. Beyond this, however, incorrect tessellation can effectively erase microstructure events. The challenge is to develop an efficient adaptive tessellation schema that eliminates the obvious saw-tooth artifacts and preserves conflicting-dip microstructure events.

To that end, I developed several schemas that determined, for each seismic sample, the optimal way to connect the sample to its neighbors. Of these, the best approach used a local cross-correlation between the two traces being connected to determine whether the data was up-dipping or down-dipping at a given sample. This approach was fast and robust and I eventually used it in production.

12.1.12 Illuminating Seismic Data

One of the fundamental components of a SeisScape display is the diffuse lighting component. Diffuse lighting is based upon Lambert's Law, which states that for surfaces that are ideally matte (i.e. without shininess), the reflected light is calculated as the cosine between the light vector and the surface normal. The surface normal at any point on the seismic surface is the vector perpendicular to the tangent to the surface at that point. When you calculate this normal there are two factors that you must consider. The first is that surface must be tessellated correctly. The second is that a seismic sample may contribute to a maximum of six triangles. The surface normal for the seismic sample is the average of the face normals from these triangles.

Reflectance is the primary means by which SeisScape displays reveal microstructure. There are three types of microstructure to consider. The first of these, and often the most prominent microstructure events on a display, are high angle noise trains that are the byproducts of the migration algorithms. The second and probably the most important is enhanced resolution. This enhanced resolution is made up of weak, and often steeply dipping, events that are below the visual resolution of wiggle trace displays. The third form of microstructure is fault plane reflections. Fault plane reflections are prevalent on many sections but they are often difficult to observe. Reflectance can effectively reveal these events, especially if the direction of lighting is perpendicular to the direction of the fault.

12.1.13 Color in Relation to Seismic Data

Catarrhine primates' are unique among mammals because they have evolved a form of trichromacy. It is generally accepted that anthropoid primates evolved trichromacy for dietary purposes. Polyak proposed that trichromacy evolved to detect red-to-yellow fruits against a background of green leaves. By contrast, Dominy and Lucas hypothesized that catarrhine trichromacy evolved to detect young, nutritious leaves rather than ripe fruit. They found that the reddening of young leaves is highly correlated with their nutritional value. Regardless of which scenario is correct, catarrhine trichromacy evolved to locate

and identify red-yellow objects against a green foliage background. For seismic purposes, either scenario is equally as relevant because both state that our visual system is tuned to detect red and yellow objects against a green background.

Most conventional seismic palettes are based upon a blue (short wave) – white - red (long wave) motif. They were developed to communicate the maximum level of seismic form and they were developed for percept free, two-dimensional environments. They are inappropriate for use in three-dimensional environments, which primarily communicate form via achromatic contrast. SeisScape displays produce percepts and those percepts produce sensations, which in turn, produce color-controlled emotions. As a result, the high concentration of reds and yellows in conventional color palettes evoke strong emotional responses when used on SeisScape displays. These responses are inappropriate in situations such as seismic interpretation, which requires a dispassionate presentation of fact.

To reduce the engaging effect of SeisScape displays, color palettes should use a dull green or another foliage color for the majority of positive seismic amplitudes. Primates see greens just as well as they see reds, yellows and oranges but their attention is not drawn to them. Reds and yellows, which attract our attention, should be used as highlight colors only and should be used sparingly. Negative amplitude troughs should be displayed using a short wave color such as blue. Short wave colors should never be used to color peaks because they produce an uncomfortable optical illusion.

12.2 Future Work

I began this thesis five years ago with the aim of discovering the sciences behind visualization. I wanted to lay the foundation for a more scientific approach to seismic visualization and I believe that for the most part, I have accomplished that goal. I knew when I began, however, that I would only briefly touch upon most of the topics and as I finish, I realize that is all I have done.

There is much more work to do on a number of fronts. One of the main areas that needs more work is tessellation. I did develop an efficient low-dip tessellation schema but

it is not appropriate for high-dip scenarios. To handle high-dip situations, the data itself will have to be subdivided and triangles will have to be added and dropped on the fly. This will require the use of the geometry shader, which is not yet generally available on all machines. Once it is, developing high-dip tessellation schemas will be a main priority.

Another area that needs more work is the subject of reflectance. Because of time constraints, I only examined the use of diffuse lighting on seismic data. I did some work on specular lighting but did not carry the research very far and I did not report on it here. I am convinced, however, that it is possible to use specular highlights to co-render different attributes efficiently. Using an attribute such as coherency to provide specular highlights on top of diffuse lighting derived from amplitudes might provide insights that would otherwise be missed. This area needs much more work.

One of the things that I regret about this research is that I was able to do so little with color. Color is a major part of any display, SeisScape displays especially, but because I am colorblind, I was not able to explore it to the depth that I wanted to. I can only hope that someone in the future picks up the theme of this work and carries it forward.

At times, my sojourn through the visual system seemed to be yielding very few concrete results. As I reached the end, however, I discovered several intriguing facts that indicated directions for future research. First among these was the discovery that our visual system is specifically tuned to detect lines and edges. This was a fascinating discovery because it provided proof of why wiggle trace displays, which are all lines and edges, were so effective in certain situations. I had hoped to research techniques that would introduce sharp edges into SeisScape displays. My plan was to make SeisScape displays as effective as wiggle trace displays for pattern recognition. Because of time contraints, I could not pursue this line of research. However, I still believe that introducing sharp edges may help with pattern recognition and it is an area that should be researched in the future.

One of the most intriguing aspects of the entire work was my discovery of seismic microstructure and I have the Trujillo data to thank for that. Microstructure constitutes an

entire new level-of-seismic-detail but it suffers from the fact that most sections are severely contaminated with migration noise. This noise can obscure and confuse relevant microstructure and I believe it is of major importance that we learn how to control it. Towards the end of this work I examined several seismic lines that had been migrated both pre and poststack and I noticed that the prestack migrated sections always contained fewer artifacts. This is a potential solution to the problem but it must wait upon further research to determine if it is.

If I could only prove one thing with this thesis, it would be that seismic data contains far more detail than we realize. We are engaged in a grand quest for seismic resolution but the visible resolution of conventional displays is far below the theoretical resolution of the data itself. This opens the door to the possibility that because we could not detect microstructure on our displays, we have obscured it in the processing. Examining this possibility is a fascinating topic and one I would love to explore.

But if you will forgive me, I will do that on another day!

12.3 Final Thoughts

This brings me to an end. The thought of this thesis first came to me when I saw the very first SeisScape image in December of 1999. I wanted to discover the sciences that lay behind it and I knew that this was the only way I could do it. More than that, however, I have always loved seismic data and the window that it provides to the past. To me, seismic data has always been a time machine and every section has transported me back through the years to an ever more fascinating past. A seismic section is the ultimate history lesson and if exposing science motivated me to pursue this, exposing more of our history kept me enthused.

There are many other subjects that I researched but did not cover in the text. I have not covered co-rendering attributes or “sea level” displays or even composite density displays. In truth, I have not covered any single subject to the depth I would have preferred. I am satisfied, though, that I have accomplished all of the things that I set out to accomplish five years ago. With that in mind, I think we will all be a lot happier now if I call it a day.

REFERENCES

Ahnelt, P.K., Hokoc, J.N., and Röhlich, P., 1995, Photoreceptors in a primitive mammal, the South American opossum, Didelphis marsupialis aurita, Vis. Neurosci., **12**: 793-804

Ahnelt, P.K., Kolb, H. and Pflug, R. (1987) Identification of a subtype of cone photoreceptor, likely to be blue sensitive, in the human retina. J. Comp. Neurol. **255**: 18-34

Akenine-Möller, Tomas and Eric Haines, "Real-Time Rendering", A.K.Peters

Balch, A.H., 1971, Color Sonograms: A New Dimension in Seismic Data Interpretation, Geophysics, **36**: 1074-1098

Barnes, Arthur E., "Shaded relief seismic attribute", Short Note, Geophysics **68**:1281-1285.

Batson, R. M., Edwards K. and Eliason, E.M., 1975, Computer generated shaded relief images: J. Research, U.S. Geol. Surv., 3, No 4, 401-408.

Baylor, D.A., Fuortes, M.G.F. and O'Bryan, P.M. (1971) Receptive fields of the cones in the retina of the turtle. J. Physiol. (Lond.) **214**: 265-294.

Beard, K. C., Krishtalka, L., Stucky, R. K., 1991, First skulls of the Early Eocene primate Shoshonius cooperi and the anthropoid-tarsier dichotomy, Nature **349**: 64-67

Beard, K. C., Qi, T. ,Dawson, M., Wang, B., Li, C., 1994, A diverse new primate fauna from middle Eocene fissure-fillings in southeastern China , Nature **368**: 604-609

Beard, K. C., Tong, Y., Dawson, M. R., Wang, J., Huang, X., 1996, Earliest complete dentition of an Anthropoid Primate from the Late Middel Eocene of Shanxi Province, China, Science **272**: 82-85

Benton, M.J., 1999 Early origins of modern birds and mammals: molecules vs. morphology, BioEssays **21**: 1043-1051.

Bisti, S., and Maffei, L., 1974, Behavioural contrast sensitivity of the cat in various visual meridians, J. Physiol. (Lond.), **241**: 201-210

Blinn, James, "Simulation of wrinkled surfaces", Computer Graphics (SIG-GRAPH '78 Proceedings), pp 286-292, August 1978

Boire, D., Dufour, J.S., Théoret, H., and Ptito, M., 2001, Quantitative analysis of the retinal ganglion cell layer in the ostritch, Struthio camelus, Brain Behav. Evol., **58**: 343-355

Bone, R.A., Landrum, J.T., and Tarsis, S.L., 1985, Preliminary identification of the human macula pigment, Vision Res, **25**: 1531-1535

Bova, L.M., Wood, A.M., Jamie, J.F., and Truscott, R.J.W., 1999, UV filter compunds in human lenses: Invest, Opthamol., Vision Res., **25**: 1531-1535

Boycott, B.B., and Wässel, H., 1991, Morphological classification of bipolar cells of the primate retina, Eur. J. Neurosci., **3**: 1069-1088

Boycott, B.B., and Wässel, H., 1999, Parallel processing in the mammalian retina, Invest. Ophthalmol. Vis. Sci., **40**: 1313-1327

Brohman, L., Phillips, M.J., and Penny, D., 1999, Growing up with dinosaurs: Molecular dates and the mammalian radiation, Tr. Ecol. Evol. **14**: 113-118

Buttery, R.G., Haight, J.R., and Bell, K., 1990, Vascular and avascular retinae in mammals, Brain, Behav. Evol., **35**: 156-175

C. F. Ross, C. F., 1995, Muscular and osseous anatomy of the primate anterior temporal fossa and the functions of the postorbital septum, Am. J. Phys. Anthropol. **98**: 275-306

Cajal, S.R. (1892) The Structure of the Retina. (Transl. Thorpe, S.A. and Glickstein, M.), Thomas, Springfield, Il., 1972.

Calderone, J.B., and Jacobs, G.H., 1999, Cone receptor variations and their functional consequences in two species of hamster, Vis. Neurosci., **16**: 53-63

Castenholz, A., 1984, The Eye of Tarsius, in: Biol J. Linn. Soc., **5**: 377-387

Catmull, E. and Clark, J., 1978, Recursively generated B-spline surfaces on arbitrary topological meshes. Computer-Aided Design, **10(6):** 350-355

Cavonius, C.R., and Robbins, D.O., 1973, Relationship between luminance and visual activity in the rhesus monkey, J. Physiol. (Lond.), **232**: 239-246

Collin, S.P., Knight, M. A., Davies, W. L., et al., 2003, Ancient color vision: multiple opsin genes in the ancestral vertebrates, Curr. Biol., **13**: R864-865

Cooper, G.F., and Robson, J.G., 1969, The yellow color of the lens of the grey squirrel, J. Physiol. (Lond.) **232**: 239-246

Cowey, A., and Ellis, C.M., 1967, Visual acuity in the rhesus and squirrel monkeys, J. Comp. Physiol. Psychol., **64**: 80-84

Cronin, H., 1993, The Ant and the Peacock: Altruism and Sexual Selection from Darwin to Today, Cambridge University Press, Cambridge.

Curcio, C. A., Sloan, K. R., Packer, O., Hendrickson, A. E. and Kalina, R. E. (1987) Distribution of cones in human and monkey retina: individual variability and radial asymmetry. Science **236**: 579-582.

Curico, C.A., Sloan, K.R., Kalina, R.E., and Hendrickson, A.E., 1990, Human photoreceptor topography, J. Comp. Neurol., **292**: 497-523

Dacey, D., Packer, O.S., Diller, L., Brainard, D., Peterson, B. and Lee, B. (2000) Center surround receptive field structure of cone bipolar cells in primate retina. Vision. Res. 40, 1801-1811.

Darwin, C, 1859, The Origin of Species by Means of Natural Selection, Random House (Reprint 1993), New York.

Daw, N., 1968, Colour-coded ganglion cells in the goldfish retin: Extension of their receptive fields by means of new stimuli, J. Physiol., **197**: 567-592

De Bonis, L., Jaeger, J.J., Coiffait, B., Coiffait, P.E., 1988, C., R. Acad. Sci. Paris, Ser. II **306**: 929

Derrington, A.M., and Lennie, P., 1984, Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque, J. Physiol. (Lond.), **357**: 219-240

Dominy, N. J., and Lucas, P. W., 2001, Ecological importance of trichromatic vision to primates, Nature **410**: 363-366.

Dow, B.M., 1974, Functional classes of cells and their laminar distribution in monkey visual cortex, J. Neurophysiol., **37**: 927-946

Drasdo, N. and Fowler, C. W. (1974) Non-linear projection of the retinal image in a wide-angle schematic eye. Br. J. Ophthal., **58**: 709-714.

Eizirik, E., Murphy, W. J., and O'Brien, S.J., 2001, Molecular dating and biogeography of the early placental mammal radiation, J. Hered. **92**: 212-219

Eizirik, E., Murphy, W. J., Springer, M. S. and O'Brien, S.J., 2004, Molecular Phylogeny and the dating of early primate divergences, Anthropoid Origins New Visions, 45-64, Kluwer Academic/Plenum Publishers, New York

Feagin, Frank J., "Seismic data display and reflection perceptibility", Geophysics **46**: 106-120.

Fernald, R. D., 2001. The Evolution of Eyes: Where Do Lenses Come From? Karger Gazette 64: "The Eye in Focus" (http://www.karger.com/gazette/64/fernald/art_1_4.htm)

Fite, K.V., and Rosenfield-Wessels, S., 1975, A comparative study of deep avian foveas, Brain Behav. Evol., **12**: 97-115

Fleagle, J.,Kay, R. F., 1987, The phyletic position of the Parapithecidae , J. Hum. Evol. **16**: 483-532

Franco, E.C.S., Finlay, B.L., Silveira, L.C.L., Yamada, E.S., and Crowley, J.C., 2000, Conservation of absolute foveal area in New World monkeys, Brain Behav. Evol., **56**: 276-286

Franzen, J. L., in Anthropoid Origins, J. G. Fleagle and R. F. Kay, Eds. (Plenum, New York, 1994), pp. 99–122;

Garber, P. A., 1992, Vertical clinging, small body size, and the evolution of feeding adaptations in the Callitrichinae, Am. J. Phys. Anthropol. **88**: 469-483.

Gatesy, J., Milinkovitch, M., Waddell, V. and Stanhope, M., 1999, Stability of cladistic relationships between Cetacea and higher level artiodactyl taxa, Syst. Biol. **48**: 6-20

Gibson, J.J., 1950, The perception of the visual world, Boston: Houghton Mifflin

Gingerich, P., in Evolutionary Biology of the New World Monkeys and Continental Drift, R. Ciochon and B. Chiarelli, Eds. (Plenum, New York, 1980), pp.123–138.

Godinot, M., Mahboubi, M., 1992, Earliest known simian primate found in Algeria, Nature **357**: 324-326.

Gorgels, T.G.M.F., and van Norren, D., 1992, Spectral transmittance of the rat lens, Vision Res., **32**: 1509-1512

Gouras, P. and Kruger, J., 1979, Responses of cells in the foveal visual cortex of the monkey to pure color contrast, J. Neurophysiol., **42**: 850-860

Gouras, P. and Zrenner E. (1981) Color vision: a review from a neurophysiological perspective. Prog. in Sens. Physiol. **1**,139-179.

Gouras, P., 1974, Opponent colour cells in different layers of foveal striate cortex, J. Physiol. (Lond.), **238**: 583-602

Gouras, Peter, Color Vision,, Webvision, http:webvision.med.utah.edu

Govardovskii, V.I., Röhlich, P., Szél, A., and Khokhlova, T.V., 1992, Cones in the retina of the Mongolian gerbil, *Meriones unguiculatus:* An immunocytochemical and electrophysical study, Vision Res., **32**: 19-27

Grossling, B.F., 1969, Color Mimicry in Geology and Geophysics, Geophysics, **34**: 249-254

Groves, C. P., 1989, A Theory of Human and Primate Evolution, Clarendon Press, Oxford.

Halder, G., Callaerts, P., Gehrig, W. J., 1995, New perspectives in eye evolution, Current Opinion in Genetcis and Development, **5**: 602-609

Hering, E. (1964) Outlines of a theory of the light sense. Translated by L.M. Hurvich & D. Jameson. Cambridge, MA: Harvard University Press.

Herman, L.M., Peacock, M.F., Yunker, M.P., and Madsen, C.J., 1975, Bottlenosed dolphin: Double-slit pupil yields equivalent aeriel and underwater diurnal acuity, Science, **189**: 650-652

Hofer, H., Carroll, J., Neitz. J., Neitz, M. and Williams, D.R. (2005) Organization of the human trichromatic cone mosaic. J. Neurosci. **25**: 9669-9679

Hofstetter, R., in Evolutionary Biology of the New World Monkeys and Continental Drift, R. Ciochon and B. Chiarelli, Eds. (Plenum, New York, 1980), pp. 103–122.

Hubel, D. H., and Wiesel, T.N., 1968, Receptive fields and functional architecture of monkey striate cortex, J. Physiol. (Lond.), **195**: 215-243

Hubel, D.H., 1988, Eye, Brain and Vision, Scientific American Library, NY: W.H. Freedman and Co.

Hughes, A., 1977, The Topography of Vision in Mammals of Contrasting Life Style: Comparative Optics and Retinal Organisation, in: Handbook of Sensory Physiology: The Visual Sytem in Vertebratesm F. Crisitelli, ed., Springer-Verlag, New York, pp 613-756.

Humanski, R.A. and Wilson, H.R. (1992) Spatial frequency mechanisms with short-wavelength-sensitive cone inputs. Vision Res. **32**: 549-560.

Hunt, D.M., Wilkie, S.E., Bowmaker, J.K., et al., 2001 Vision in the ultraviolet, Cell. Mol. Life Sci., **58**: 1583-1598

Hurvich, L.M. (1981) Color Vision. Sunderland, MA: Sinauer Assoc.Inc.

Juliusson, B., Bergström, A., Röhlich, P., Ehninger, B., van Veen, T. and Szél, A., 1994, Complimentary cone fields of the rabbit retina, Invest. Opthalmol. Vis. Sci., **35**: 811-818

Kallweit, R.S., and Wood, L.C., 1982, The limits of resolution of zero-phase wavelets: Geophysics, **47**: 1035-1046

Kaneko, A. (1970) Physiological and morphological identification of horizontal, bipolar and amacrine cells in goldfish retina. J. Physiol. (Lond) **207**: 623-633.

Kaplan, E. and Shapley, R. M. (1986) The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. Proc. Natl. Acad. Sci. USA. **83**: 2755-2757.

Kaplan, E., Purpura, K., and Shapley, R.M., 1987, Contrast affects the transmission of visual information through the mammalian lateral geniculate nucleus., J. Physiol., **391**: 267-288

Kass, J.H., Guillery, R.W., and Allman, J.M., 1972, Some principals of organisation in the dorsal lateral geniculate nucleus, Brain Behav. Evol. **6**: 253-299

Kass, J.H., Huerta, M.F., Weber, J.T., et al., 1978, Patterns of retinal terminations and laminar organisation ot the lateral geniculate nucleus of primates., J. Comp. Neurol., **182**: 517-554

Kay, R. F., Cartmill, M., 1977, Cranial morphology and adaptations of Palaechthon nacimienti and other paromomyidae (Plesiadapoidea, ? primates), with a description of a new genus and species, J. Hum. Evol. **6**: 19-53

Kay, R. F., Ross, C. and Williams, B. A., 1997, Anthropoid origins, Science, **275**: 797-804

Kirk, C.E., Kay, R.F., 2004, Anthropoid Visual Acutiy, in Anthropoid Origins New Visions, 540-602, Kluwer Academic/Plenum Publishers, New York

Klug, K., Herr, S., Tran Ngo, I., et al, 2003, Macaque retina contains an S-cone OFF midget pathway, J. Neurosci., **23**: 9881-9887

Kolb, H. and DeKorver, L. (1991) Midget ganglion cells of the parafovea of the human retina: A study by electron microscopy and serial section reconstructions. J. Comp. Neurol. **303**: 617-636.

Kolb, H. and Marshak, D. (2003) The Midget Pathways of the Primate Retina. Doc. Ophtal. In press.

Kolb, H., 1970, Organisation of the outer plexiform layer of the primate retina: Electron microscopy of Golgi-impregnated cells, Philos. Trans. R. Soc. Lond. B, **258**: 261-283

Kolb, H., 2003, How the retina works, American Scientist, 91: 28-35

Kolb, H., Linberg, K. A. and Fisher, S. K. (1992) The neurons of the human retina: a Golgi study. J. Comp. Neurol. **318**: 147-187.

Kolb, H., Zhang, DeKorver, L. and Cuenca, N. (2002) A new look at calretinin-immunoreactive amacrine cell types in the monkey retina. J. Comp. Neurol., **453**: 168-184

Krauskopf, J., Williams, D.R., and Heeley, D.W., 1982, Cardinal directions in colour space, Vision Res., **22**: 1123-1131

Land, M. F., Fernal,d R. D., 1992, The evolution of eyes. Annu. Rev. Neurosci., **15**: 1-29.

Land, M.F., and Nilsson, D.E., 2002, Animal Eyes, Oxford University Press, Oxford

Lavergne, A., Douzery, E. Stichler, T., Catzeflis, F. M., and Springer, M. S., 1996, Interordinal mammalian relationships: Evidence for paenungulate monophyly is provided by complete mitochondrial 12S rRND sequences, Mol. Phylogenet. Evol. **6**: 245-258

Lee, B.B., Kremers, J., and Yeh, T., 1998, Receptive fields of primate retinal ganglion cells studied with a novel technique, Vis. Neurosci, **15**: 161-175

Lennie, P., Trevarthen, C., Van Essen, D., et al., 1990, Parallel processing of visual information, in Visual Perception: The Neurophysiological Foundations (eds J. Spillman and J. Werner), Academic Press, San Diego, pp 103-128

Li, Y. Wang, Y. Wang, & C. Li. 2000. A new family of primitive mammal from the Mesozoic of western Liaoning, China. *Chinese Science Bulletin* **46(9)**: 782–785. (2001))

Lindsey, J.P., 1989, The Fresnel zone and its interpretive significance: The Leading Edge, **10**: 33-39

Lines, R.L., and Newrick, R.T., 2004, Fundamentals of Geophysical Interpretation 13, Society of Exploration Geophysicists, pp 49-52

Livingstone, M., Hubel, D., 1984, Anatomy and physiology of a color system in the primate visual cortex, J. Neurosci, **4**: 309-356

Livingstone, Margaret, Hubel, David, J Neurosciences **7**: 3416 (1987)

Livingstone, Margaret, Hubel, David, Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology and Perception, Science, New Series, Vol 240, No. 4853, pp 740-749

Long, K.O., and Fischer, S.K., 1983, The distribution of photoreceptors and Ganglion cells in the California ground squirrel, *Spermophilus beecheyi*, J. Comp. Neurol., **221**: 329-340

Loop, C.T., 1987, Smooth subdivision surfaces based on triangles. Master's Thesis, Department of Mathematics, University of Utah.

Lynch, Steven, "Ancient Evenings, Seismic Visualization using very old techniques", CSEG Recorder, October 2000

MacLeod, D. I. A. (1972) Rods cancel cones in flicker. Nature **235**: 173-174

Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, R., et al., 2001, Parallel adaptive radiations in two major clades of placental mammals, Nature **409**: 610-694.

Marc, R.E. (1982) Chromatic organization of the retina. In "Cell Biology of the Eye" (M. LaVail and J. Hollyfield, Eds), Academic Press, New York. pp 435-473.

Mariani, A. P. (1983) Giant bistratified bipolar cells in monkey retina. The Anatomical Record, **206**: 215-220.

Mariani, A.P. (1984) Bipolar cells in monkey retina selective for cones likely to be blue-sensitive. Nature **308**: 184-186.

Marr, D., 1982, Vision, San Fransisco: W.H. Freedman & Co

Martin, P.R., 1998, Color processing in the primate retina: Recent progress., J. Physiol. (Lond), **513**: 631-638

Martin, P.R., Lee, B.B., White, A.J.R., et al, 2001, Chromatic sensitivity of ganglion cells in the peripheral primate retina, Nature (Lond.), **410**: 933-936

Martin, R.D., 1990, Primate Origins and Evolution: A Phylogenetic Reconstruction, Chapman and Hall, London.

Michael, C.R., 1987,Comparative study of the color cells in layer Ivcb and in the blobs of the monkey's striate cortex. Soc. Neurosci. Abstr., 13: 2

Michael, C.R., 1987a, Color vision mechanisms in monkey striate cortex: dual opponent cells with concentric receptive fields, J. Neurophysiol., **41**: 572-588

Michael, C.R., 1987b, Color vision mechanisms in monkey striate cortex: simple cells with dual opponent color receptive fields, J. Neurophysiol., **41**: 1233-1249

Michael, C.R., 1987c, Color-sensitive complex cells in monkey striate cortex, J. Neurophysiol., **41**: 1250-1266

Mollon, J.D., 1989, "Tho' she kneel'd in that place where they grew …": The uses and origins of primate color vision, J. Exp. Biol, **146**: 21-38

Müller, B., and Peichl, L., 1989, Topography of rods and cones in the tree shrew retina, J. Comp. Neurol., **282**: 581-594

Murphy, W. J., Eizirik, E., Johnson, W.E., Zhang, Y. P., Ryder, O. A., and O'Brien, S.J., 2001a, Molecular phylogenetics and the origins of placental mammals, Nature **409**: 614-618.

Murphy, W. J., Eizirik, E., O'Brien, S.J., Madsen, 0., Scally, M., Douady, C.J., et al., 2001b, Resolution of the early placental mammal radiation using Bayesian phylogenetics, Science **294**: 234-2351.

Naka, K.-I. (1976) Neuronal circuitry in the catfish retina. Invest. Ophthal. **15**: 926-935.

Nathans, J., Thomas, D., Hogness, D.S., 1986, Molecular genetics of human color vision: The genes that encoding blue, green and red pigments, Science **232**: 193-202.

Nicol, J.A.C., 1981, Tapeta Lucida of Vertebrates, in: Vertebrate Photoreceptor Optics, J.M. Enoch and F.L. Tobey, eds, Springer-Verlag, Berlin, pp 401-431.

Nussbaum, J.J., Pruett, R.C., and Delori, F.C., 1981, Macular yellow pigment: The first 200 years, Retina **1**: 296-310.

Osterberg, G. (1935) Topography of the layer of rods and cones in the human retina. Acta Ophthal., suppl. **6**: 1-103.

Perry, V.H., and Cowley, A., 1985, the ganglion cell and cone distributions in the monkey's retina: Implications for central magnification factors, Vision Res. **25**: 1795-1810

Pirie, A., 1966, The Chemistry and Structure of the Tapetum Lucidum in Animals. in: Aspects of Camparative Opthamology, O. Graham-Jones, ed., Pergamon, London, pp 57-68.

Polyak, S., 1957, The Vertebrate Visual System, University of Chicago Press, Chicago.

Polyak, S.L. (1941) The Retina. University of Chicago Press, Chicago.

Quigley, H. A., Addicks, E. M. and Green, W. R. (1982) Optic nerve damage in human glaucoma: III Quantitative correlation of nerve fibre loss and visual defect in glaucoma ischemic neuropathy and toxic neuropathy. Arch. Ophthal., **100**: 135

Rasmussen, D. T., Simons, E. L., 1992, Int. J. Primatol. **13**: 1 -25.

Regan, D. (ed.), 1991, Binocular Vision. Vol. 9 Vision and Visual Dysfunction. London, Macmillan.

Reid, R.C., and Shapley, R.M., 1992, Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus, Nature (Lond.), **356**: 716-718

Rodieck, R.W. (1991) In "From pigments to perception" (Eds. Valberg, A. and Lee, B.B.), Plenum, New York, pp 83-94.

Roorda, A. and Williams, D.R. (1999) The arrangement of the three cone classes in the living human eye. Nature **397**: 520-522.

Ross, C. F., 1996, Adaptive explanation for the origins of the anthropoidea (primates), Am. J. Primatol. **40**: 205-230

Ross, C.F., 2000, Into the Light: The origin of Anthropoidea, Annu. Rev. Anthropol., **29**: 147-194

Salvini-Plawen Lv, Mayr E, 1977, On the evolution of photoreceptors and eyes. Evol So, **10**: 207-263.

Schmitz, J. Ohme, M., and Zischler, H. 2001, SINE insertions in cladistic amalyses and the phylogenetic affiliations of Tarsius bancanus to other primates, Genetics **157**: 777-784.

Schusterman, R.J., and Balliet, R.F., Conditioned vocalizations as a technique for determining visual acuity thresholds in sea lions, Science, **169**: 498-501

Sherman, S.M., and Guillery, R.W., 1996, Functional organisation of the thalamocortical relays., J. Neurophysiol., **76**: 1367-1395

Sherman, S.M., and Guillery, R.W., 1998, On the actions that one nerve cell can have on another. Distinguishing "drivers" from "modulators", Proc. Natl. Acad. Sci. USA, **95**: 7121-7126

Sherman, S.M., and Guillery, R.W., 2001, Exploring the Thalamus, Academic Press, San Diego

Shoshoni, J. and McKenna, M. C., 1998, Higher taxonomic relationships among extan mammals based upon morphology, woth selected comparisons of results from molecular data, Mol. Phylogenet. Evol. **9**: 572-584

Silveira, L.C.L., Grünert, U., Kremers, J., Lee, B.B., Martin, P.R., 2005, Comparative Anatomy and Physiology of the Primate Retina, in The Primate Visual System: A Comparative Approach, (ed Jan Kremers), John Wiley and Sons Ltd, West Sussex, pp 127-160

Silveira, L.C.L., 1996, Joint entropy loci of M and P cells: A hypothesis for parallel processing in the primate visual system, Braz. J. Biol., 56(Suppl 1): 345-367

Simons, E. L., Rasmussen, D. T., 1989, Cranial morphology of Aegyptopithecus and Tarsius and the question of the tarsier-anthropoidean clade**,** Am. J. Phys. Anthropol. **79**: 1-23.

Simons, E. L., Rasmussen, D. T., 1994, A whole new world of ancestors: Eocene anthropoideans from Africa, Evol. Anthropol. **3**: 128-139.

Simons, E. L., 1992, Diversity in the early tertiary Anthropoidean radiation in Africa, Proc. Natl. Acad. Sci. U.S.A. **89**: 10743-10747

Simons, E. L., 1995, Egyptian Oligocene Primates: A Review, Yearbook. Phys. Anthropol. **38**: 199-238

Simpson, G. G., 1945, The principals of classification and a classification of mammals. Bull. Am. Mus. Nat. Hist. **85**:1-350

Sinex, D.G., Burdette, L.J., and Pearlman, A.L., 1979, A psychophysical investigation of spatial vision in the normal and reeler mutant mouse, Vision Res., **19**: 853-857

Smith, R.B., Warnock, J.E., Stanley, W.D., Cole, E.R., 1972, Computer Graphics in Geophysics, Geophysics **37**: 825-838

Springer, M. S., Murphy, W. J., Eizirik, E., O'Brien, S. J., 2003, Placental mammal diversification and the Creataceous-Tertiary boundary, Proc. Antl. Acad. Sci. USA **100**: 1056-1061

Stockman, A., MacLeod, D.I.A. and DePriest, D.D. (1991) The temporal properties of the human short-wave photoreceptors and their associated pathways. Vision Res. **31**: 189-208.

Suthers, R.A., 1966, Optomotor responses by echolocator bats, Science, **152**: 1102-1104

Switkes, E., Bradley, E., and DeValois, K.K., 1988, Contrast dependence and mechanisms if masking interactions among chromatic and luminance gratings, J. Opt. Soc. Am., **5**: 1149-1162

Synder, A.W., and Miller, W.H., 1977, Photoreceptor diameter and spacing for highest resolving power, J. Opt. Soc. Am., **67**: 696-698

Synder, A.W., Bossomaier, T.R., and Hughes, A., 1986, Optical image quality and the cone mosaic, Science, **231**: 439-501

Synder, A.W., Laughlin, S.B., and Stavenga, D.G., 1977, Information capacity of eyes, Vision Res., **17**: 1163-1175

Szalay, F. S., 1977, Bull. Am. Mus. Nat. Hist. 156

Szél, A., Röhlich, P. ,Caffé, A.R., Juliusson, B., Aguirre, G. and van Veen, T., 1992, Unique topographic separation of two spectral classes of cones in the mouse retina, J. Comp. Nerol., **325**: 327-342

Taner, M.T., Koehler, F., Sheriff, R.R., "Complex Seismic Trace Analysis", Geophysics **44**: 1041-1063

Thomson L.R., Toyoda, Y., Langner, A., Delori, F.C., Garnett, K.M., Craft, N., et al., 2002, Elevated retinal zeaxanthin in prevention of light-induced photoreceptor cell death in quail, Invest. Opthal. Vis. Sci., **43**: 3538-3549

Timney, B. and Keil, K., 1992, Visual acuity in the horse, Vision Res., **32**: 2289-2293

Toyoda, J.I., (1972) Membrane resistance changes underlying the bipolar cell response in the carp retina. Vision Res. **12**: 283-294

Ungerleider and Mishkin (In: Ingle DJ, Goodale MA and Mansfield RJW (Editors), *Analysis of Visual Behavior* MIT Press, Boston, 1982)

Van Hof, M.W., 1967, Visual acuity in the rabbit, Vision Res., **7**: 749-751

Walls, G. L., The Vertebrate Eye and Its Adaptive Radiation (Hafner, New York, 1942).

Wandell, B.A. (1995) Foundations of Vision. Sinauer Associates, Inc. Sunderland, Massachussets.

Ward, G, 1994, A Contrast-Based Scalefactor for Luminance Display, Graphics Gems IV, Ed. by P. S. Heckbert, pp 415-421,

Wässel, H., and Boycott, B.B., 1991, Functional architecture of the mammalian retina. Physiol. Rev., **71**: 447-480

Wässel, H., Grünert, U., Martin, P.R., et al., 1994, Immunocytochemical characterization and spatial distribution of midget bipolar cells in the macaque monkey retina, Vision Res., **34**: 561-579

Welland, Michael, Donnelly, Nick and Menner, Tammy, "Are we properly using our brains in seismic interpretation", Leading Edge, February 2006, 142-144

Werblin, F.S. and Dowling, J.E. (1969) Organization of the retina of the mudpuppy, Necturus maculosus. II. Intracellular recording. J. Neurophysiol. **32**: 339-355.

Westheimer, G., 1981, Visual hyperacuity, Prog. Sensory. Physiol., **1**: 1-30

Widess, M., 1973, How thin is a thin bed?: Geophysics, **38**: 1176-1254

Williams, B. A, Covert, H. H., 1994, New early Eocene anaptomorphine primate (Omomyidae) from the Washakie Basin, Wyoming, with comments on the phylogeny and paleobiology of anaptomorphines, Am. J. Phys. Anthropol., **93**: 323-340

Williams, D.R., MacLeod, D.I.A. and Hayhoe, M. (1981) Punctate sensitivity of the blue-sensitive mechanisms. Vision Res. **21**: 1357-1375.

Yamada, E. (1969) Some structural features of the fovea centralis in the human retina. Arch. Ophthal., **82**: 151-159

Yaoming Hu, Jin Meng, Yuanqing Wang, Chuankui Li. 2005. Large Mesozoic mammals fed on young dinosaurs. *Nature*, **433**: 149–152

APPENDIX A: THE TECHNOLOGY BEHIND SEISSCAPE

"Any sufficiently advanced technology is indistinguishable from magic"
Arthur C. Clark

"Change is the law of life. And those who look only to the past or present are certain to miss the future"
John F. Kennedy

A.1 Introduction

A few years ago, I watched a documentary on the inventions of Leonardo DaVinci. One of the points made was that surprisingly, despite his great genius and despite the practicality of many of his inventions, very few of them found there way into general use. We remember DaVinci for has art and his inventiveness and yet he made far less impact on the world than we would expect from such brilliance today. The documentary went on to discuss the reason why this happened, why so many of his inventions languished despite the fact that it was technologically feasible to produce them. The reason why, it explained, was that in DaVinci's time, change was not expected. What had been, had been; what had been, was; and what had been, would be. Change was neither expected nor sought after and there was very little pressure on DaVinci to "productize" his inventions.

This thought pattern is very different from todays but it was the dominant thought pattern for most of human history. For most of human history, the pace of technological change was so slow that most people never experienced it once during their lifetime. That is not the case today. Since the Industrial Revolution, the pace of change has accelerated to the point that we experience technological revolutions almost on a daily basis. Consequently, anyone buying something new expects it to be substantially better than the one they bought before. Before the Industrial Revolution, revolutionary changes occurred once per century, now they occur once per day and we expect to see those changes in our lives as soon as possible.

Nowhere is this more in evidence that in the field of computer science and within that field nowhere is it more in evidence that in the field of computer graphics. Although computer graphics has been around since the 70's, it was the development of the first graphic processing unit (gpu) in 1999 that kick started it into life. The first gpu was capable of performing calculations at roughly one gigaflop. Seven years later, the current state of the art gpu's perform calculations at a sustained rate of almost 600 gigaflops and within the next year, teraflop gpu's should become the norm. Along with the explosive growth in computational power, there has been an equally explosive growth in the ability to program it. Programmable gpu's first saw use in 2003 and since then, the languages used to program them have gone through four major revisions, each a quantum leap ahead of its predecessor.

Trying to keep up with the rapid pace of these technological leaps is both exhilarating and frustrating. It is exhilarating because each day brings something new, something else learnt and applied. It is frustrating because it takes time to understand and learn the nuances of any new technology and with four major revisions in four years; by the time you understand anything it is already obsolete. As an example of this, in late 2004 I started a course project to develop a subdivision scheme for seismic data. I was working on a state of the art gpu but it was not capable of handling the number of samples in a seismic section and I decided to experiment with various schemes for reducing the volume. I failed in this project because you cannot subdivide seismic data without running into aliasing. However, during the course of the project, the state of the art changed. By the time I concluded that subdivision was not possible, I was already working on a new gpu that was so fast that subdivision was unnecessary. Change in graphics occurs just that fast. What was impossible yesterday is possible today and will be trivial tomorrow.

Programmable gpu's are the underlying technology behind both the SeisScape display itself and my concepts of visualization. As such, given that they are fundamentally different than cpu's, in this chapter I provide a very brief overview of what they are, how they have developed and what they can be used for. For a more

comprehensive review, I direct the reader to either the NVIDIA website or the ATI website.

A.1.1 Caveat

One of the problems with graphics programming is that as soon as you become proficient in the current capabilities of both the graphic cards and the graphic languages, they become obsolete. Shader technology is less than five years old but in that time; it has gone through four major revisions. At the start of this thesis, DirectX was at version 8 and the shader languages at version 1.1. I base all of the techniques that I demonstrate upon what has been state of the art technology for several years, DirectX 9.0c and shader version 3.0. Unfortunately, with the release of Windows Vista, both of these versions are once again obsolete. The current state of the art is DirectX 10 and shader 4.0, both of which are substantially enhanced and modified.

A.1.2 Coding Conventions

Throughout the remainder of this thesis, I will use the following conventions for displaying gpu based shader code and cpu based application code. I show shader code using the following style:

```
void TimeVariantScaling( float time, inout float zValue )
{
    zValue *= pow( time / 1000, TimeVariantScaler );
}
```

The calling app code is the code used to configure the shader's various input variables and data streams. It is compiled into the executable and I will show it in the following style:

```
BasicEffect.SetValue("TimeVariantScaler", TimeVariantScaler);
```

A.2 The Graphic Processing Unit

As much as this thesis is about visualization it is also about technology and in particular, it is about the emerging technology of computer graphics. The field of computer graphics is not new but over the past seven years, it has received a tremendous boost from the development of graphical processing units or gpu's. Before the development of gpu's the best we could do was produce static images of simple three-dimensional scenes; but now, after their development, we can produce dynamic displays of complex scenes and interact with them in a real time setting.

The graphic processing unit is a highly parallel structure designed and optimized to perform the types of calculations and operations associated with rendering three-dimensional graphical objects. Modern workstations and pc's generally have one or more gpu's each of which is external to the computers central processing unit and possesses its own dedicated high-speed memory. The original purpose of the gpu was to offload transform[1] and lighting[2] calculations from the cpu. Modern gpu's, however, are capable of much more and we use them now as high-efficiency computational units.

A.2.1 History

The first use of the term Computer Graphics dates back as far as 1960. It is credited to William Fetter, an employee of Boeing who used it to describe his work creating the first computer model of the human body. Despite the fact that this first use of computer graphics was in science, from its inception, subsequent developments in the field have been driven primarily by the art and entertainment industry. The first use of computer animation was in the 1976 film Futureworld where it was used to produce an animation

---

[1] Transform refers to the task of converting the spatial coordinates of three-dimensional objects in a virtual world into a two-dimensional view.

[2] Lighting refers to the task of calculating the color of objects as lights of given types and orientations are directed towards them.

of a human face and hand. In subsequent years, the development of Pixar's RenderMan software made it possible to produce complex rendered scenes and as a result, computer graphics became common in movies. Nevertheless, it was not, until 1996 when the first full-length computer-animated film "Toy Story" was released that its full potential began to be recognized. Since then, full-length computer-animated movies have become commonplace.

The release of Toy Story was a groundbreaking event for the computer graphics industry. A typical scene in the movie contained over two million triangles and took five minutes to render[3]. This illustrates one of the fundamental principals of computer graphics; as much as the result is artistic, the underlying methodology is mathematical. Objects in a three-dimensional scene are mathematical representations of the real world. The calculations for rendering them are relatively simple and include such things as matrix transformations, dot products and linear interpolations. However, whereas the scope of the calculations is simple the scale of them is not. Producing each Toy Story frame required billions of very simple, cpu based mathematical calculations.

The success of Toy Story not only influenced the development of other computer-animated movies, it also spawned a new era in computer gaming. Game developers were quick to realize the potential of producing games with the same photo-realism as Toy Story but they immediately ran into a problem. Taking five minutes to render a single frame was not a problem for the movie industry but it was for the gaming industry that required real time interaction between the user and the characters and objects in the game. To achieve photo-realistic games each scene needed to be animated and rendered 20-30 times a second. In the late 1990's this was beyond the capabilities of even the most sophisticated pc or workstation.

---

[3] Rendering is the process of generating, by computer programs, a two-dimensional image from a mathematical model of a three-dimensional scene.

What was required was hardware-accelerated 3D graphics. Hardware-accelerated 2D graphics had been available since the development of S3 Graphics single-chip accelerator, the S3 86C911, in 1991 but what was needed were graphic cards capable of accelerating the calculations required for rendering 3D scenes. The first generation of these cards, NVIDIA's RIVA TNT and TNT2 cards, ATI's Rage card and 3dfx's Voodoo3 cards were all released early in 1999. Although hyped as 3D graphic accelerators, none of these cards was. Their main function was to rasterize pre-transformed vertices and as a result, most of the calculations required for rendering were still done in the cpu.

We now recognize that the first true graphical processing unit was NVIDIA's GeForce 256, which they released in late 1999. The GeForce 256 was the first mass-produced, low cost card capable of performing the transformation and lighting calculations needed to render a 3D scene. NVidia designed it to off-load transform and lighting calculations plus the texture-mapping calculations from the cpu. The card produced the images in my first talk on SeisScape displays (see Chapter 1). The GeForce 256, originally released in August of 1999, was followed by the release of ATI's Radeon 7500 card and S3's Savage 3D card. All of these cards had similar capabilities and were capable of rendering approximately 15 million triangles per second and 480 million pixels per second. The main problem with them was that whereas they were capable of performing the basic transformation and lighting of vertices in general they lacked any form of real programmability.

NVidia and ATI addressed this lack of programmability in Feb of 2001 with the release of their GeForce 3 card and Radeon 8500 card respectively. These cards introduced the dual concepts of programmable vertex and pixel shaders that, for the first time, allowed developers the freedom to write small programs that would operate on every vertex and every pixel as they were rendered. These initial programmable gpu's were very limited in the number and type of instructions that they could execute. Both cards were capable of executing small vertex shaders and were capable of configuring (but not programming) pixel shading. As a result, whereas in theory they were

programmable, I can attest from experience that in practice the programs were very limited.

In my opinion, the most significant thing that happened with the introduction of these two cards was that it sparked an arms race between ATI and NVIDIA, the arms in this case being the capabilities of the shaders. The next and most significant generation of cards, ATI's Radeon 9700 card and NVIDIA's FX cards, both of which came out in late 2002, were the first generation of cards that were capable of true vertex and pixel shading. With their introduction, it finally became possible to perform looping operations in both the vertex and pixel shaders. In addition, both shaders became capable of performing lengthy floating-point calculations and of performing texture operations orders of magnitude faster than the cpu.

This latter functionality, the increased speed of texture operations, became even more significant with the introduction of floating point textures. Typically, we think of a texture as an image or a picture and the initial use of textures in computer graphics was to drape images over three-dimensional surfaces. Floating-point textures, however, are different. As the name suggests they are nothing more than arrays of floating point data. Being able to send them directly to the gpu opened up many interesting possibilities. These possibilities were made even more enticing by the advent of a second major gpu enhancement, off-screen render targets. A render target is simply the place where the results end up and normally that would be the screen. Typically, the gpu processes vertices and textures and normally it sends the output to the computer monitor for viewing. An off-screen render target, however, is just another floating-point texture, which means that with the introduction of the 9700 and the FX cards you could send arrays of floating point data to the gpu, process the data and get it back via a second texture. This moved the gpu into another realm, in one-step it went from being a simple 3D accelerator to being a full-fledged mathematical processing unit capable of complex high-speed operations.

It is with the introduction of those capabilities that the field of computer graphics started to explode. At the time of writing, the competition between ATI and NVIDIA

shows no sign of slowing down. Consequently, hardware capabilities have increased almost exponentially as has the ability to write longer and more complex shader programs. As of writing, the most advanced NVIDIA hardware is the 8800 series gpu, which has a memory bandwidth of 88GB/sec and is capable of rendering 36,800 million pixels per second. Given that the original GeForce 256 could render only 480 million pixels per second this translates to an astounding increase in capability of 76X in just seven years.

To put this into perspective, in 1985 while I was working at Western Geophysical in Calgary we purchased four STAR array processors whose combined speed rating was one gigaflop. The four together cost over five million dollars. NVIDIA's current flagship gaming card, the 8800 GTX has a theoretical speed rating of 520 gigaflops (NVIDIA GeForce 8800 Architecture Technical Brief) and costs less than $1,000 or roughly $2 per gigaflop. In the past 20 years then the price performance ratio has improved by a factor roughly 2.5 million times. As impressive as that comparison is another comparison is equally as impressive. I was initially attracted to visualization by the discovery, in late 1999, that NVIDIA's first card, the RIVA TNT2 card had a speed rating of roughly 0.5 gigaflop. The current state of the art 8800 card, which was release seven years later, has more than one thousand times the raw computing power.

A.2.2 The Programmable Graphics Pipeline

The architecture behind a gpu is fundamentally different from that of a cpu. The simplest way to think of a gpu is as a pipeline. The cpu feeds vertices, commands and textures into the pipeline. The various elements of the gpu then operate on them and sends the final rendered pixels either to the frame buffer for display or to an off-screen render target that is accessible by the cpu. This pipeline is known as the programmable graphics pipeline and I show it in schematic in Figure A.1.

There are two typical input streams to this pipeline. The first is the set of untransformed vertices that represent the locations, in real world coordinates, of the various points in a 3D model. In the case of seismic data, these vertices are the seismic

samples themselves. The second input stream is a series of indices that tell the gpu how to assemble the 3D surface from the input vertices. Regardless of their geometrical complexity, all 3D objects are made up of a series of simple triangles, each of which has three vertices. The index stream is just a list of which vertices are used to assemble each triangle. You can define simple models without indices but in general, most complex surfaces are defined using them.

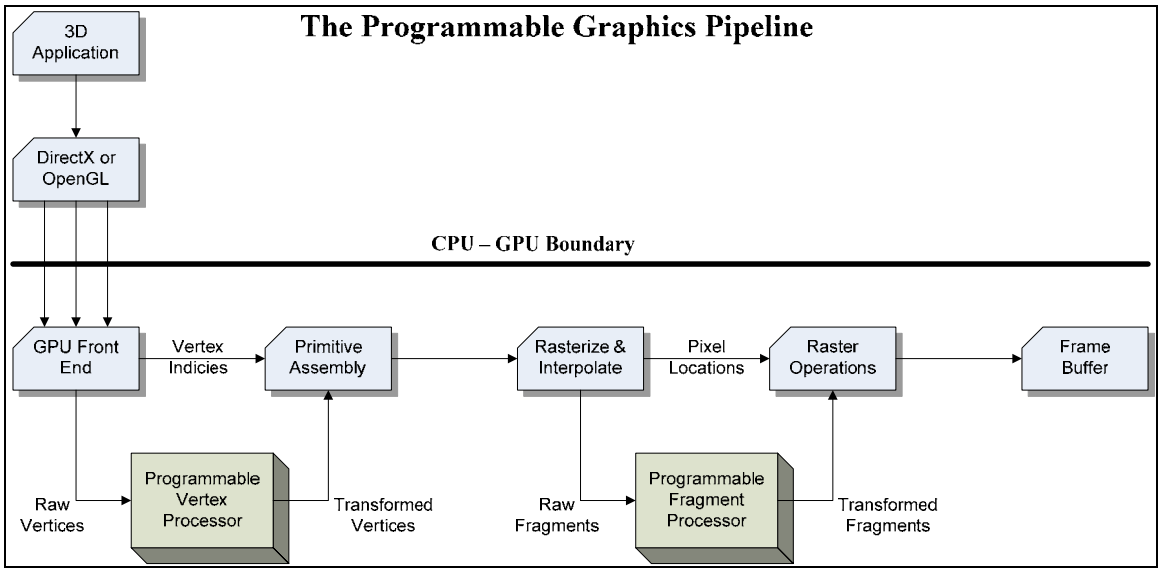


Figure A.1: The programmable graphics pipeline for a graphic card with Shader 3 capabilities. Two units in this pipeline execute user-defined programs (shaders). The programmable vertex processor transforms the real world coordinate vertices into screen coordinates. The programmable fragment processor defines the final color or value of a pixel.

Once you have fed the data to the gpu it goes through a series of stages. In the first stage, the gpu separates the vertex and index streams. The vertex stream(s) is then passed on to the programmable vertex processor (pvp). This is the gpu unit that is responsible for converting the untransformed real work vertex coordinates into screen coordinates and, as its name suggests, it is capable executing user defined programs. The output from the pvp is recombined with the index stream and the gpu then assembles the 3D surface using the now transformed vertices and the original index stream. The output from this stage is a series of transformed triangles.

The next stage in the pipeline is the rasterization and interpolation stage. At this stage, the gpu evaluates the individual triangles to see if it should cull them from

subsequent processing. Triangles that it does not cull are rasterized, which is the process of determining which pixels the triangle covers. There are two output streams for this stage. The first is a series of pixel locations, which are passed directly to the final raster operation stage. The second is a series of corresponding fragments, which are branched to the programmable fragment processor (pfp) which, like the pvp, is capable of executing user-defined programs.

There is often confusion between the terms pixel and fragment. The term pixel refers to a single location on an output surface, be it a frame buffer or render target. The term fragment refers to the data that the gpu needs to generate a pixel. That data usually includes the pixel location, depth, and various colors and texture coordinates. Fragments are fed directly into the pfp for processing. The gpu then recombines the output from the pfp with the pixel stream, feeds it into the final stages of the pipeline, and then on to either the gpu's frame buffer for display or an off-screen render target for access by the cpu.

It is important to note that this is an evolving pipeline and Figure A.1 only applies to fourth generation gpu's (NVidia GeForce FX series and ATI Radeon 9700) which were state of the art when this thesis began. Modern state of the art gpu's are massively parallel and may contain a third processing unit, the programmable geometry shader. In addition, modern gpu's are capable of passing more information between the vertex and pixel shaders and also of accessing textures directly in the vertex shaders as well as in the pixel shaders.

A.3 Programmable Shaders

In the early days of digital signal processing, the majority of the mathematical operations were performed on highly vectorized array processors, while the cpu functioned primarily as an I/O device. With the increased power of workstations and pc's, this programming model died out and the cpu functioned as both an I/O and a mathematical engine.

The programmable-shader programming model is very similar to the early array processor model. The cpu once again functions primarily as an I/O and data assembly system whereas the graphic card performs the majority of the mathematical calculations. There is one major difference, array processors were vector based whereas graphic cards are linear. Array processors were ideal for performing mathematical tasks on large arrays of data, i.e. performing Fourier Transforms, convolutions and filtering. In contrast, shaders operate on a single vertex or a single fragment at a time. Because of its inherently parallel nature, a modern gpu is capable of operating on multiple vertices and fragments at the same time but each instance of a shader program only operates on a one at a time.

This is not to imply that a shader program only has access to the data from a single vertex or fragment. Typically, information is passed to the shader as either 2D or 3D textures and a shader program may access any part of those textures. This makes it possible to perform complex tasks such as interpolation, filtering and smoothing in the shaders themselves. However, the output from a shader only pertains to a single vertex of fragment.

Because there are two shaders on the graphic card, the vertex shader and the pixel shader, it is important to understand which of the two to use for any given task. It is not as important when you use the gpu as a mathematical processing engine but it is vitally important when you use it as a rendering engine. This is because performing an operation in an inappropriate shader can result in a dramatic loss of performance. Vertex shaders operate on every vertex in the model and pixel shaders operate on every pixel. For models with a low triangle count it is best to do as many operations in the vertex shader as possible and let the rasterizer interpolate any results. On the other hand, seismic displays have more triangles than there are pixels on the screen. Consequently, they are best processed by pixel shaders because the rasterizer culls most of the triangles.

SeisScape displays produce two triangles per data sample and consequently have very large triangle counts, sometimes on the order of millions of triangles per section. As a result, the only operations that the vertex shader does are ones that affect the physical location of the vertex (i.e. displacement mapping, morphing between versions, clipping

and scaling of sample values). Operations such as determining the final lighting and coloring of the pixel are usually left until the pixel shader.

A.3.1 GPU Programming Languages

There are two separate graphics languages that must be considered when writing a graphics application. The first is the cpu based graphic language and the second is the gpu-based language. There are two predominant cpu based graphic languages available, OpenGL and Direct3D and three gpu based graphic languages, GLSL, Cg and HLSL.

OpenGL (Open Graphics Library) is a cross-platform toolkit for writing 3D graphic applications. It was initially developed by Silicon Graphics in 1992 and designed for what were then their state of the art workstations. From its inception, OpenGL was designed as a feature rich API to render a complete three-dimensional scene from simple graphic primitives. Despite its richness, OpenGL is considered a low-level graphic language requiring the programmer to specify the exact steps required to render a scene. There are higher-level languages based upon OpenGL which makes the job of building and rendering a complex scene easier. One such language is OpenInventor, which is the high-level graphic language used to produce the original SeisScape displays in 1999.

Direct3D is the Windows toolkit for writing predominantly gaming applications and is one component of Microsoft's DirectX API. It was originally developed in 1992 by RenderMorphics, a company that was subsequently purchased by Microsoft in 1995. Although it competes directly against it, Direct3D is fundamentally different than OpenGL. Whereas OpenGL is essentially a low-level rendering engine, Direct3D is essentially a low-level interface to the registers on the graphic card. As such it has from its inception been in more direct communication with the hardware but has also been harder to work with.

Historically, OpenGL has been the predominant language for scientific computing whereas Direct3D has dominated gaming. Currently this division is starting to blur. With the advent of programmable gpu's the use of both languages is being deprecated to defining the geometry of the scenes and configuring the graphic card. Almost all of the

rendering code is now written in one of the gpu languages. As a result, the decision as to which language is better for a given purpose is becoming moot. Up until 2003 SeisScape displays were created using the OpenInventor software, which is based upon OpenGL. After that time, I switched to using Direct3D, the rational being that since I was developing exclusively for the Windows platform and was anticipating using technology developed for the gaming industry, that using Microsoft's gaming api would be the better choice.

On the gpu side, all of the original shader programs were written in assembler language and loaded directly onto the graphic card. This made programming very difficult and often dependant upon the hardware itself. Modern shader programs are now generally written in one of three high-level shader languages; GLSL, Cg and HLSL, all of which have similar capabilities. GLSL is the OpenGL shading language and is based upon the C programming language but since it can only be used with OpenGL which I do not use in this thesis, I will not discuss it further.

The other two languages, Cg and HLSL, are essentially the same language. Cg stands for "C for graphics" and was developed by NVIDIA to abstract the developer from the nuances of gpu assembly language. As its name suggests, it uses a "C" type of syntax although some features of C were modified to make them more suitable for graphics programming. In addition, Cg contains several data types such as the half[4] data type which are needed for high-resolution imaging. Cg is available for both the OpenGL and the Direct3D platforms.

HLSL (High Level Shader Language) is Microsoft's implementation of Cg under its Direct3D platform and is virtually identical to it. The primary advantage of using HLSL over Cg in a Windows application is that Direct3D contains functions to compile HLSL

---

[4] A half data type is a 16 bit floating point value. It has a dynamic range of ±32767 with three decimal places of precision.

programs on the fly. Given that commercial applications generally release pre-compiled shaders and that HLSL can load compiled Cg shaders, this benefit is dubious.

Unless otherwise noted, all of the software used in this thesis was produced using C# as the high-level programming language, Direct3D as the low-level graphics language and HLSL as the shader language.

A.3.2 The Anatomy of an HLSL Shader

As previously mentioned, there are two variants of shader programs that can be written for the gpu, vertex shaders and pixel shaders. DirectX 10 has a third – the geometry shader which, because of its newness, will not be considered here. In this section I provide a very brief introduction to shaders and discuss some of the considerations that must be kept in mind when designing them. For a full explanation of the Cg/HLSL language see the excellent reference "The Cg Tutorial by NVidia" – better reference here.

Under HLSL, vertex and pixel shaders are combined together to form "techniques". Code Fragment A-1 shows a simple technique, circa late 2003, that contains both a vertex and pixel shader. Although it is usual to have both shaders in a technique, it is not strictly necessary and if one or the other is missing, the technique will use the native functionality on the card. Techniques can be multi-pass and can use different shaders for each pass. They can also change the configuration of the gpu between passes and set various rendering states.

```
Effect.Technique = "DynamicTexture";
```

```
technique DynamicTexture
{
    pass P0
    {
        // shaders
        VertexShader = compile vs_1_1 DynamicTexture_V();
        PixelShader  = compile ps_1_1 TextureColor();
    }
}
```

Code Fragment A-1: A single pass technique that compiles both a vertex and a pixel (fragment) shader using Shader 1.1 functionality. Techniques can have multiple passes and can use different vertex and pixel shaders in each.

```
void DynamicTexture_V(
    in   float2   inputPosition       : POSITION,
    in   float2   inputTexCoord       : TEXCOORD0,
    in   float    inputzValue         : BLENDWEIGHT1,
    out  float4   outputPostion       : POSITION,
    out  float2   outputTexCoord0     : TEXCOORD0)
{

    if( TimeVariantScaler > 0 )
        TimeVariantScaling( inputPosition[1], inputzValue );

    Clip( inputzValue );

    TextureScale( inputzValue, outputTexCoord0 );

    YAxisRotate( inputPosition.x, inputzValue );

    WorldTransform( inputPosition, inputzValue, outputPostion );

    return ;
}
```

Code Fragment A-2: SeisScape vertex shader circa late 2003. The primary function of a vertex shader is to transform the vertex coordinates into screen coordinates. In the above code, this is accomplished in the function WorldTransform, which applies a transformation matrix to the input, coordinates. The terms POSITION, TEXCOORD0 and BLENDWEIGHT1 are semantics. They tell the shader which of the input streams of data to use for the corresponding input variable.

```
void TextureColor (
    in float2 textureCoords   : TEXCOORD0,
    out float4 diffuseColor   : COLOR0 )
{
    // get the diffuse texture color
    diffuseColor = tex2D(alphaSampler, textureCoords);
}
```

Code Fragment A-3: SeisScape pixel shader circa late 2003. The primary function of the pixel shader is to determine the final color of the pixel. In the above code, the pixel shader receives an interpolated texture coordinate, which is then used to extract the pixel color from a texture. The function tex2D is an intrinsic HLSL function. The output color has four floating-point components, one each for red, green, blue and alpha (transparency), which vary between 0 and 1.

Both vertex and pixel shaders have the same basic syntax. Code Fragment A-2 and Code Fragment A-3 illustrates the relationship between Cg programs and C itself. Cg's syntax is very similar to C's but there is a difference in how information is passed in and out of a shader as opposed to a C function. The cpu code sends the data to the gpu in the

streams and instructs the shader which of the various streams corresponds to the input variables. There are five variables in the header of Code Fragment A-2, three are marked with the “in” prefix and two with the “out” prefix. The “in” prefix means that variable is passed into the program from one of the streams in the graphics pipeline; the “out” prefix means that it is passed out of the shader and back into the pipeline.

Another difference between C and HLSL are the float2 and float4 data types, these data types do not exist in C. These data types (and others with a similar syntax, i.e. int2, int4 etc) are arrays. HLSL also contains traditional arrays but these are arranged differently internally so that they can be manipulated more efficiently by the GPU. Using float4 as opposed to float[4] is more efficient.

Each variable is followed by an identifier (POSITION, TEXCOORD0 etc.). These identifiers are called semantics and they tell the gpu where to get the information from, in the case of “in” variables, and where to put them in the pipeline in the case of “out” variables. In the DynamicTexture_V function, for each vertex to be processed, the GPU will pass in the x, y position of the vertex (inputPosition), a set of texture coordinates (inputTexCoord) that in this case are not used and a further float value (inputzValue) which it will obtain from what are called the BlendWeights.

When writing shaders it is important to keep in mind that the model for their operation is fundamentally different than for conventional “C” code. In an environment where the shaders are being used in rendering (as opposed to processing), vertex shaders are run on every vertex for every rendering pass; likewise, pixel shaders are run on every pixel for every rendering pass. Consequently, the shaders may be executed hundreds of millions of times per second. This is different from most conventional C functions, which usually operate a few dozen times during the execution cycle of the program. Because shaders operate so frequently it is imperative that shader code be highly optimized and the developer must always keep in mind that whereas each mathematical operation may not by itself appear to be onerous that it is adding hundreds of millions of calculations to the gpu’s load.

A.3.3 Floating Point Textures

Throughout this thesis, I make extensive use of the gpu as both a processing and a display engine. Both cases require that seismic data be passed to the graphic card in a form that is readily accessible. There are two methods for passing this information; (1) as data streams or (2) as floating point textures. The original SeisScape displays used data streams when passing seismic to the card but this technique proved limited in what it could accomplish. This is because vertex shaders only have access to the information from a single vertex at a time and so using data streams restricted my early vertex programs to processing a single sample value at a time. It is possible to pass multiple streams of data to the shader and so such techniques as morphing between different seismic versions is possible. However, more complex techniques such as smoothing and filtering, which require values from surrounding samples, are not.

A superior technique is to pass the seismic data to the card in the form of a floating-point texture. A floating-point texture is, as its name suggests, simply a texture where the pixel values contain floating-point values rather than RGB color values. Both vertex and fragment shaders can access any part of these textures using texture coordinates. In this way, even though the output from a vertex shader represents one vertex, the information used to assemble the output can come from many. Floating-point textures are an efficient means of passing multiple arrays of seismic data to the graphic card. Whereas they are conceptually arrays of values, they are in practice something else entirely and extracting values from them is fundamentally different from reading values out of an array. You extract values from a two-dimensional array by specifying the two indices of the desired value. However, this is not how you obtain values from a texture. To get a value from a texture the programmer uses one of several texture lookup functions that use texture coordinates to specify the location of the value in the texture.

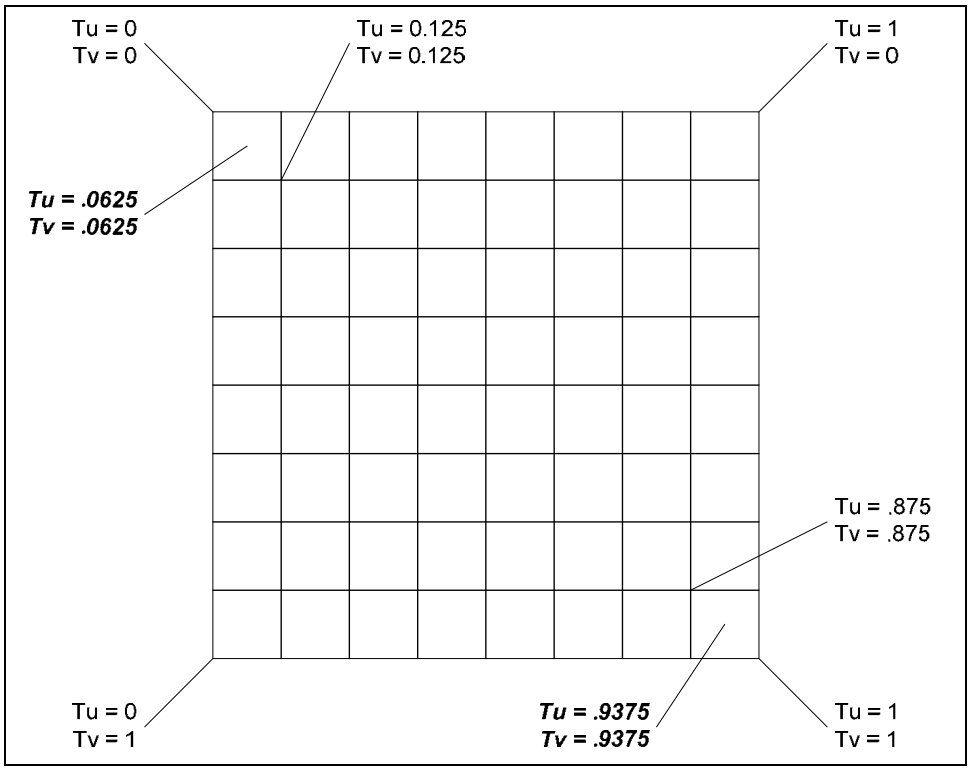


Figure A.2: The texture coordinates for an 8 x 8 texture. Unlike bitmaps, which you access by pixel location, textures are accessed using coordinates, which vary between 0 and 1. Pixels occupy a rectangle in texture coordinate space, which in the case of the 8 x 8 texture is an area of 0.125 x 0.125. There are several ways of sampling a texture, the most common being POINT mode and LINEAR mode. In POINT mode, specifying any coordinate within a pixels area will return the value of the pixel. In LINEAR mode, the value returned is linearly interpolated between the adjacent pixels. In LINEAR mode the true value of a pixel is obtained by specifying a coordinate exactly in the middle of the rectangle, in the case above this would be Tu = 0.0625, Tv = 0.0625 for the first pixel and Tu = 0.9375, Tv = 0.9375 for the last.

All of the texture retrieval functions use texture coordinates that vary between zero and one. Textures may be one-dimensional, two-dimensional or three-dimensional. Consequently, accessing them requires one, two or three texture coordinates designated Tu, Tv and Tz. Figure A.2 is an illustration of the texture coordinates required to access values in an 8 x 8 two-dimensional texture. Pixels are mapped to rectangles in texture coordinate space, that rectangle being a square .125 x .125 in Figure A.2.

On the surface, this appears to be an overly complex way of getting values from a texture but in practice, it has two major advantages over specifying pixel coordinates directly. The first is that textures, unlike arrays, often have multiple resolutions. By using a texture coordinate rather than a pixel location, the shader does not have to know which resolution it is using. The second advantage is that the value returned by a texture

sampling call changes depending upon the mode of access for the texture. There are several modes for accessing a texture, the most common being POINT mode and LINEAR mode. In POINT mode, the texture functions as an array and any texture coordinate within a pixels mapped rectangle returns the same value, that being the value of the pixel.

In contrast, in LINEAR mode, the texture represents a field of continuously changing values. The value returned by a texture lookup call is linearly interpolated between the surrounding pixels. The interpolation is built into the gpu itself and consequently, linear interpolation of floating point values is computationally free. As a tradeoff, it is more difficult to obtain the exact value of a sample when using LINEAR mode and to obtain it the texture coordinate must correspond to the exact center of the pixels rectangle (2D texture) or cell (3D texture). In the case of the 8 x 8 texture shown in Figure A.2, this corresponds to texture coordinates of Tu = 0.0625, Tv = 0.0625 for the first pixel and Tu = 0.9375, Tv = 0.9375 for the last.

A.3.4 Dynamic Coloring: an Early Pixel Shader

Shaders are rapidly becoming one of the simplest and most powerful forms of coding but in 2003, when I began work on this thesis, they were in their infancy and their benefits were not immediately obvious. In this section, I discuss the first fragment shader that I developed and use it to illustrate both the simplicity and the power of shader technology.

Figure A.3 is an example of an early SeisScape display that predates my use of shader technology. In this display, peaks represent positive seismic amplitudes and troughs represent negative amplitudes. The central focus of this section is the color of the display, which in this pre-shader example, I calculated in the cpu. To determine the color, I converted the seismic amplitude to color with a color lookup table created from a yellow-red-white-blue-cyan palette. In this example, yellow represents extreme positive amplitudes, white represents zero amplitude and cyan represents extreme negative amplitudes. After the color of each sample was calculated, it was attached to its

appropriate vertex and passed to the graphic card. This was, in theory, a very simple and obvious way of coloring the display but in practice, it suffered from several problems.

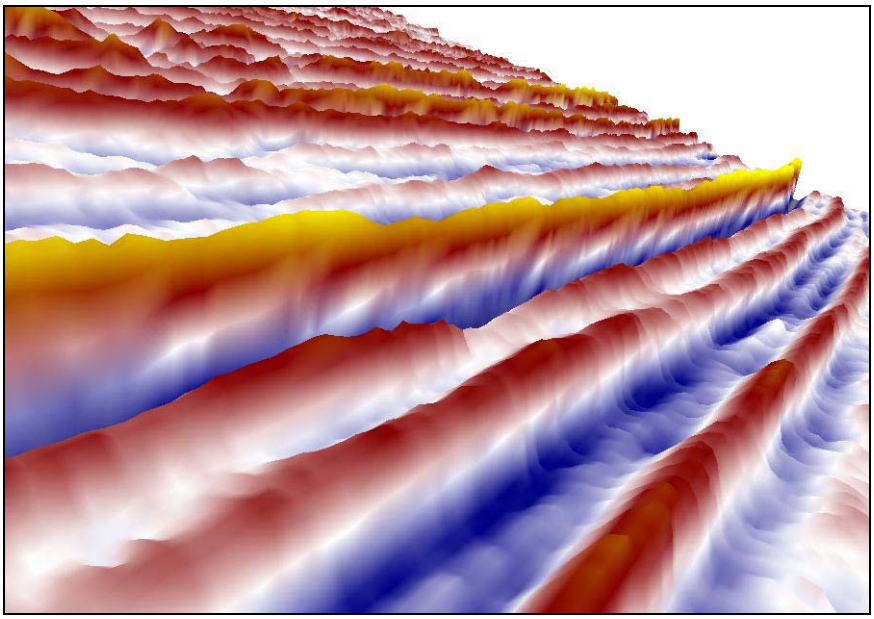


Figure A.3: A SeisScape display of a small section of a seismic line. Peaks represent positive seismic amplitudes whereas troughs represent negative seismic amplitudes. The seismic amplitude is the Z coordinate of the vertex. The color value for each vertex was calculated in the cpu and passed to the card as a data stream. The final display colors are interpolated during rasterization a process that causes the zero amplitude line, which should be a thin white streak, to become jagged and ill defined.

Looking at Figure A.3 it is obvious that the first problem is that the colors do not seem to match the seismic amplitudes. In fact, they do match, but only at the vertex locations themselves, everywhere else they are wrong. You can see this by looking along any one of the seismic events. Since I used white for the zero amplitude color there should be a thin white streak along each event. Instead, the zero amplitude line is jagged and ill defined. The same is true of other amplitudes because they should also be smooth from trace to trace but in practice, they appear jagged or “stepped” along the events.

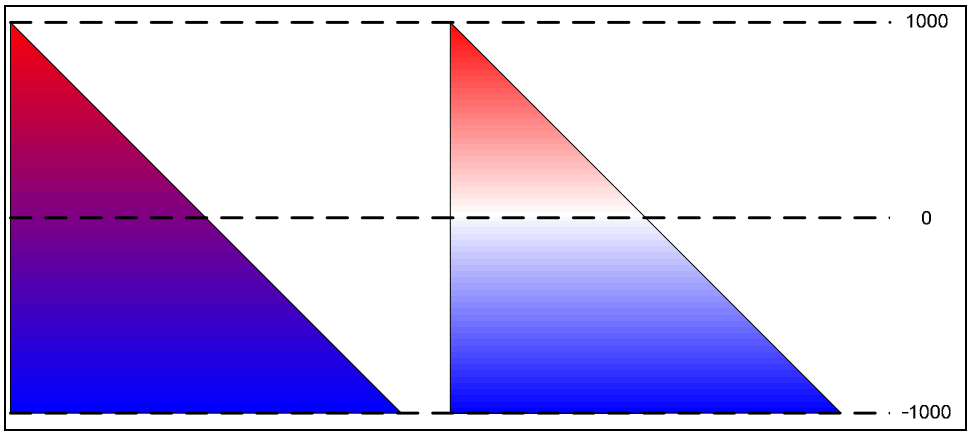


Figure A.4: Conventional vs. Fragment based color generation. Both images represent one triangle from a SeisScape display formed from two adjacent samples on one trace and one sample on the adjacent trace. The triangle on the left illustrates the coloring technique used in the pre-shader SeisScape displays. In this case, the colors (red and blue) are interpolated in the rasterizer. Consequently, the zero amplitude line which passes through the triangle and which should be white cannot be produced. The triangle on the right illustrates coloring done using a fragment shader. In this case, the seismic amplitude is passed to the rasterizer and interpolated. This interpolated value is then converted to color in the fragment shader. Since the fragment shader now "sees" the zero amplitude line it can correctly color it.

I illustrate the reason for the "stepped" nature of the coloring in the left triangle of Figure A.4. Both triangles in Figure A.4 represent one hypothetical triangle from a SeisScape display. I form each of them from two adjacent samples on one trace and one sample on the adjacent trace. I color the left triangle using the same technique that I used to color Figure A.3. The color of each pixel is determined by the rasterizer, which takes the value of the color at the three vertices (red, blue, blue) and interpolates them. Here is the problem, since the rasterizer interpolates color rather than seismic amplitude, it can only produce the correct amplitude/color match at the vertex itself. Everywhere else, it is wrong.

This illustrates an admittedly simple but crucial point that you must consider when writing shaders; vertices define the scenes but the viewer sees pixels and what lies between the two is the rasterizer. The rasterizer is responsible for interpolating the vertex values down to the pixel level and consequently it is incumbent upon the developer to make sure it is interpolating the correct things.

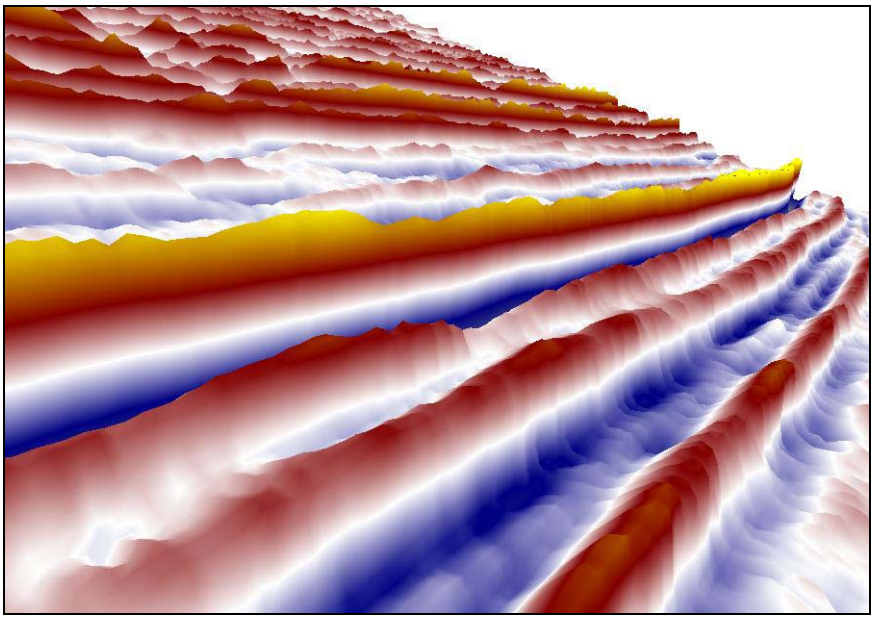


Figure A.5: A SeisScape display of a small section of a seismic line, coloring produced in the fragment processor. Peaks represent positive seismic amplitudes whereas troughs represent negative seismic amplitudes. The seismic amplitude is the Z coordinate of the vertex. The seismic amplitude is converted to a texture coordinate in the vertex shader and then passed out of the vertex shader and onto the rasterization engine where it is interpolated and passed into the fragment processor. The interpolated texture coordinate is then used to extract the color from a 1D-palette texture. Note that the zero amplitude line, which should be a white streak, is now sharp and well defined.

In Figure A.3 the rasterizer is interpolating color whereas it should be interpolating seismic amplitude and the result is a display that lacks precision. By contrast, the colors and amplitudes shown in Figure A.5 matches perfectly. This display was created using my first ever vertex and fragment shader. To color the display, the vertex shader converted the seismic amplitude into a texture coordinate. This texture coordinate was then passed out of the shader and onto the rasterizer, which subsequently passed the interpolated coordinates onto the fragment shader. Once there, the interpolated texture coordinates were used to extract a color from a one-dimensional texture. Passing the texture coordinates to the rasterizer is analogous to passing the seismic amplitudes themselves and then converting the interpolated amplitudes into a texture coordinate in the fragment shader. This latter technique is the one that I use today but due to the limitations of the early shaders, it was not possible in 2003. Both approaches, however, produce the same results that being to produce colors that correctly match the true seismic amplitudes.

A comparison of Figure A.3 and Figure A.5 graphically shows the benefits of what you can achieve with the appropriate use of shaders. What is more dramatic is the simplicity of the shader code that I used to produce the coloring in Figure A.5. Code Fragment A-4 shows the details of the function that I called from within the vertex shader to convert the seismic amplitude into a texture coordinate. Code Fragment A-5 is the fragment shader code that uses the interpolated texture coordinate to extract a color from a palette texture.

```
void TextureScale( in float inputzValue, out float2 outputTexCoord )
{
    inputzValue = (0.5 + (polarity * 0.5 * inputzValue / ColorPeakValue));
    outputTexCoord[0]   = 0;
    outputTexCoord[1]   = inputzValue;
}
```

Code Fragment A-4: Function called from within a vertex shader to convert an amplitude value into a texture coordinate. Texture coordinates vary between 0 and 1 whereas the seismic amplitude is assumed to vary between ± ColorPeakValue. The ColorPeakValue is a variable that is set by the calling code to change the color saturation of the display.

```
void BumpTextureColor (    in      float2 textureCoords0      : TEXCOORD0,
                           in      float4 lightColor          : COLOR0,
                           out     float4 diffuseColor        : COLOR0 )
{
    // get the diffuse texture color,
    diffuseColor = tex2D(alphaSampler, textureCoords0);
    diffuseColor = diffuseColor * lightColor;
}
```

Code Fragment A-5: Fragment shader code that converts the interpolated texture coordinates into a color value. The texture coordinates calculated in Code Fragment A-4 are interpolated by the rasterizer and passed into the shader as textureCoords0. These coordinates are then used to extract a color value from a "palette" texture. The lightColor variable passed into the shader is the diffuse lighting which is calcualted in the vertex shader.

In both functions, the operative code is contained in a single line of shader code. This may not appear significant until you consider the steps necessary to change the color saturation of the display. The color saturation, i.e. the amplitude range that the colors in the palette are spread over, is set in Code Fragment A-5 by the variable ColorPeakValue. The calling cpu code sets this value for each rendering pass. Consequently, all the user has to do is change the value and re-render the display, the colors of the display change automatically. By contrast, passing the color itself to the gpu, as is done in Figure A.3, requires the cpu code to load the seismic samples, convert them to color and then attach

them to the vertices. In my original SeisScape code this took several hundred lines of code.

A.3.5 Conclusions about Shaders

In conclusion, gpu shaders are a new form of programming. They were originally developed strictly for rendering applications but with the introduction of floating point textures and off-screen render targets they can now be used as high-performance mathematical engines as well. Because they may be executed hundreds of millions of times per second shaders must be kept short and simple and the developer must always keep in mind the effect and importance of the rasterizer which converts vertex values into pixel values.

Shaders are more than simply another way to program the same thing. Because shaders operate every time a scene is rendered they dramatically improve the communication between the user and the data. Operations which previously took seconds or minutes to complete can now be performed in real-time and, as I will show in later chapters, shaders are capable of performing actions that simply would not be possible in the cpu. The appropriate use of shaders can also result in a dramatic reduction of cpu code volumes. Typically, in my own work, I have seen reductions of about 100 lines of cpu code removed for every line of shader code added.

APPENDIX B: PIXEL SHADER TECHNIQUES

B.1 Down-Dip Normals Calculation

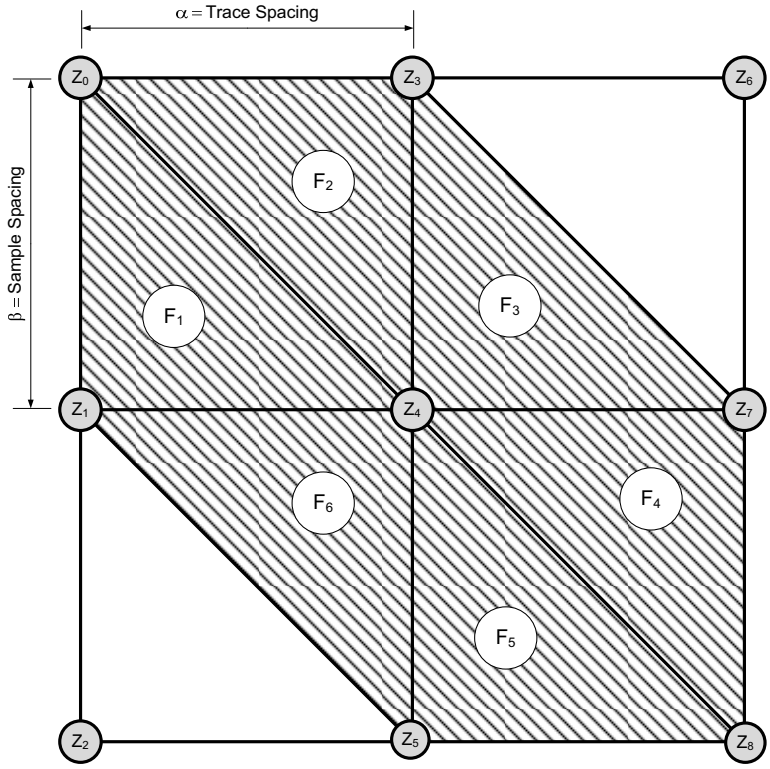


Figure B.1: The six faces (F_1-F_6) that contribute to the normal at sample Z_4. The direction of tessellation favors down-dipping seismic events.

Given the nine samples ($Z_1 - Z_9$) shown in Figure B.1, the objective is to develop a pixel shader technique which calculates the surface normal at sample Z_4. The surface normal for any point on a surface is calculated as the average of the normals for each face (triangle) to which it contributes. Sample Z_4 contributes to faces $F_1 - F_6$, consequently its normal can be written as:

$$N_0 = Normalize(\overrightarrow{F1} + \overrightarrow{F2} + \overrightarrow{F3} + \overrightarrow{F4} + \overrightarrow{F5} + \overrightarrow{F6}) \tag{B.1}$$

The normal vector to any two vectors is given by the cross-product:

$$Vn = V1 \otimes V2 \tag{B.2}$$

Where the cross-product is defined as:

$$
\begin{aligned}
V_n x &= V_1 y \times V_2 z - V_1 z \times V_2 y \\
V_n y &= V_1 z \times V_2 x - V_1 x \times V_2 z \\
V_n z &= V_1 x \times V_2 y - V_1 y \times V_2 x
\end{aligned}
\tag{B.3}
$$

If we define:

α = trace spacing, β = sample spacing

Then the face normals can be written as:

F_1	**F_2**	**F_3**
$V1 = (0, -\beta, (z0 - z1))$	$V1 = (0, \beta, (z4 - z3))$	$V1 = (0, -\beta, (z3 - z4))$
$V2 = (\alpha, 0, (z4 - z1))$	$V2 = (-\alpha, 0, (z0 - z3))$	$V2 = (\alpha, 0, (z7 - z4))$
$F_1 x = -\beta(z4 - z1)$	$F_2 x = \beta(z0 - z3)$	$F_3 x = -\beta(z7 - z4)$
$F_1 y = \alpha(z0 - z1)$	$F_2 y = -\alpha(z4 - z3)$	$F_3 y = \alpha(z3 - z4)$
$F_1 z = - - \beta \times \alpha = \alpha\beta$	$F_2 z = -\beta \times -\alpha = \alpha\beta$	$F_3 z = - - \beta \times \alpha = \alpha\beta$
F_4	**F_5**	**F_6**
$V1 = (0, \beta, (z8 - z7))$	$V1 = (0, -\beta, (z4 - z5))$	$V1 = (0, \beta, (z5 - z4))$
$V2 = (-\alpha, 0, (z4 - z7))$	$V2 = (\alpha, 0, (z8 - z5))$	$V2 = (-\alpha, 0, (z1 - z4))$
$F_4 x = \beta(z4 - z7)$	$F_5 x = -\beta(z8 - z5)$	$F_6 x = \beta(z1 - z4)$
$F_4 y = -\alpha(z8 - z7)$	$F_5 y = \alpha(z4 - z5)$	$F_6 y = -\alpha(z5 - z4)$
$F_4 z = - - \alpha\beta = \alpha\beta$	$F_5 z = - - \beta\alpha = \alpha\beta$	$F_6 z = - - \alpha\beta = \alpha\beta$

Table B-1: The face normals for each triangle that contributes to the normal at position Z_4 in Figure B.1

We now sum the normals noting that if:

$$V_1(i, j, k) = X_1 i + Y_1 j + Z_1 k \tag{B.4}$$

$$V_2(i, j, k) = X_2 i + Y_2 j + Z_2 k \tag{B.5}$$

then

$$V_1 + V_2 = \left(X_1 + X_2\right) i + \left(Y_1 + Y_2\right) j + \left(Z_1 + Z_2\right) k \tag{B.6}$$

$$N_0x = -\beta(z4 - z1) + \beta(z0 - z3) + -\beta(z7 - z4) + \beta(z4 - z7) + -\beta(z8 - z5) + \beta(z1 - z4)$$
$$N_0x = \beta(z1 - z4 + z0 - z3 + z4 - z7 + z4 - z7 + z5 - z8 + z1 - z4) \quad \text{(B.7)}$$
$$N_0x = \beta(2z1 + z0 + z5 - (z3 + 2z7 + z8))$$

$$N_0y = \alpha(z0 - z1) - \alpha(z4 - z3) + \alpha(z3 - z4) - \alpha(z8 - z7) + \alpha(z4 - z5) - \alpha(z5 - z4)$$
$$N_0y = \alpha(z0 - z1 + z3 - z4 + z3 - z4 + z7 - z8 + z4 - z5 + z4 - z5) \quad \text{(B.8)}$$
$$N_0y = \alpha(2z3 + z0 + z7 - (z1 + 2z5 + z8))$$

$$N_0z = 6\alpha\beta \quad \text{(B.9)}$$

The equation for the unnormalized down-dip favoring average normal for sample Z_4 can then be written as:

$$\begin{aligned} N_0x &= \beta(2z1 + z0 + z5 - (z3 + 2z7 + z8)) \\ N_0y &= \alpha(2z3 + z0 + z7 - (z1 + 2z5 + z8)) \\ N_0z &= 6\alpha\beta \end{aligned} \quad \text{(B.10)}$$

Noting that both α and β are constants, and that the x-component of the normal depends entirely upon β and the y-component on α, calculating the normals in a pixel shader becomes reduced to calculating the following:

$$\begin{aligned} N_0x &= 2z1 + z0 + z5 - (z3 + 2z7 + z8) \\ N_0y &= 2z3 + z0 + z7 - (z1 + 2z5 + z8) \end{aligned} \quad \text{(B.11)}$$

Scaling by the trace spacing and sample interval and then normalizing can be done when the normals are used in rendering operations.

B.1.1 Down-Dip Normals Pixel Shader

The following pixel shader can be used to program (B.11).

Notes:

- To save texture memory, the output from this shader is stored in a G16R16F format which is why the calculation are done as half floats.
- The seismic data is stored in 128*128 textures which are accessed via the surfaceSampler variable. There is one sample per pixel in the input texture.
- smapInc is the spacing (in texture coordinates) between adjacent pixels.
- The input data is stored in a multi-sample format, i.e. there may be more than one sample version per pixel. The lightIndex variable is used to select which of the samples is used for the lighting calculation.

```
void DownDipNormalPS(
        in      float2  texco: TEXCOORD0,
        out     half4   color: COLOR0)
{
        // define the input values
        half4   sourcevals0;
        half4   sourcevals1;
        half4   sourcevals2;
        half4   sourcevals3;
        half4   sourcevals4;
        half4   sourcevals5;
        half4   sourcevals6;
        half4   sourcevals7;
        half4   sourcevals8;

        // The normalScaler is used to balance the lighting effect between seismic lines with
        // different amplitude ranges. Because of the limited dynamic range of the half float
        // format, the normalization must be done for each input value to avoid clipping
        float   nrm = normalScaler;

        // read in the input data from the input texture
        sourcevals0 = tex2D(surfaceSampler, texco + float2(-smapInc, -smapInc)) * nrm;
        sourcevals1 = tex2D(surfaceSampler, texco + float2(-smapInc, 0)) * nrm;
        sourcevals3 = tex2D(surfaceSampler, texco + float2(0, -smapInc)) * nrm;
        sourcevals4 = tex2D(surfaceSampler, texco + float2(0, 0)) * nrm;
        sourcevals5 = tex2D(surfaceSampler, texco + float2(0, smapInc)) * nrm;
        sourcevals7 = tex2D(surfaceSampler, texco + float2(smapInc, 0)) * nrm;
        sourcevals8 = tex2D(surfaceSampler, texco + float2(smapInc, smapInc)) * nrm;

        color.x = 2 * sourcevals1[lightIndex] + sourcevals0[lightIndex] + sourcevals5[lightIndex] -
                (sourcevals3[lightIndex] + 2 * sourcevals7[lightIndex] + sourcevals8[lightIndex]);

        color.y = 2 * sourcevals3[lightIndex] + sourcevals0[lightIndex] + sourcevals7[lightIndex] -
                (sourcevals1[lightIndex] + 2 * sourcevals5[lightIndex] + sourcevals8[lightIndex]);

        // Apply the normalization scaler
        color.x = color.x / 6;
        color.y = color.y / 6;

        // the output from a pixel shader is always a four component vector regardless of the
        // format of the texture being rendered to. The z and w components of the vector are
        // not used but must be defined before the shader will compile.
        color.z = 1;
        color.w = 1;
};
```

B.2 Up-Dip Normals Calculation

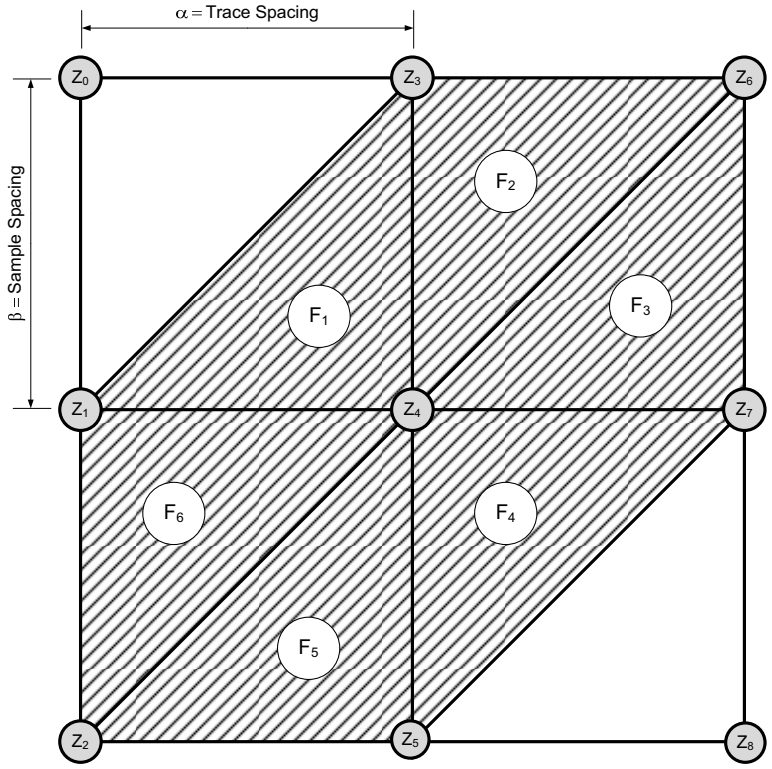


Figure B.2: The six faces (F_1-F_6) that contribute to the normal at sample Z_4. The direction of tessellation favors up-dipping seismic events.

F_1

$V1 = (0, -\beta, (z3 - z4))$

$V2 = (-\alpha, 0, (z1 - z4))$

$F_1x = -\beta(z1 - z4)$

$F_1y = -\alpha(z3 - z4)$

$F_1z = - - \beta \times -\alpha = -\alpha\beta$

F_2

$V1 = (0, \beta, (z4 - z3))$

$V2 = (\alpha, 0, (z6 - z3))$

$F_2x = \beta(z6 - z3)$

$F_2y = \alpha(z4 - z3)$

$F_2z = -\beta \times \alpha = -\alpha\beta$

F_3

$V1 = (0, -\beta, (z6 - z7))$

$V2 = (-\alpha, 0, (z4 - z7))$

$F_3x = -\beta(z4 - z7)$

$F_3y = -\alpha(z6 - z7)$

$F_3z = - - \beta \times -\alpha = -\alpha\beta$

F_4

$V1 = (0, \beta, (z5 - z4))$

$V2 = (\alpha, 0, (z7 - z4))$

$F_4x = \beta(z7 - z4)$

$F_4y = \alpha(z5 - z4)$

$F_4z = -\alpha\beta$

F_5

$V1 = (0, -\beta, (z4 - z5))$

$V2 = (-\alpha, 0, (z2 - z5))$

$F_5x = -\beta(z2 - z5)$

$F_5y = -\alpha(z4 - z5)$

$F_5z = - - \beta \times -\alpha = -\alpha\beta$

F_6

$V1 = (0, \beta, (z2 - z1))$

$V2 = (\alpha, 0, (z4 - z1))$

$F_6x = \beta(z4 - z1)$

$F_6y = \alpha(z2 - z1)$

$F_6z = -\alpha\beta$

Table B-2: The face normals for each triangle that contributes to the normal at position Z_4 in Figure B.2

The shader development for the up-dip tessellation lighting follows that laid out for the down-dip lighting shown in B.1. Using the same model, the six faces that contribute to the normal at Z_4 are shown in Table B-1:

Given this, the x, y & z components of the normal can be described by:

$$N_0x = -\beta(z1 - z4) + \beta(z6 - z3) - \beta(z4 - z7) + \beta(z7 - z4) - \beta(z2 - z5) + \beta(z4 - z1)$$
$$N_0x = -\beta(z1 - z4 - z6 + z3 + z4 - z7 + z4 - z7 + z2 - z5 + z1 - z4) \quad \text{(B.12)}$$
$$N_0x = -\beta(2z1 + z3 + z2 - (z6 + 2z7 + z5))$$

$$N_0y = -\alpha(z3 - z4) + \alpha(z4 - z3) - \alpha(z6 - z7) + \alpha(z5 - z4) - \alpha(z4 - z5) + \alpha(z2 - z1)$$
$$N_0y = -\alpha(z3 - z4 - z4 + z3 + z6 - z7 - z5 + z4 + z4 - z5 - z2 + z1) \quad \text{(B.13)}$$
$$N_0y = -\alpha(2z3 + z6 + z1 - (z2 + 2z5 + z7))$$

$$N_0z = -6\alpha\beta \quad \text{(B.14)}$$

The equation for the unnormalized up-dip favoring average normal for sample Z_4 can then be written as:

$$N_0x = \beta(2z1 + z3 + z2 - (z5 + 2z7 + z6))$$
$$N_0y = \alpha(2z3 + z6 + z1 - (z2 + 2z5 + z7)) \quad \text{(B.15)}$$
$$N_0z = 6\alpha\beta$$

Again, as with (B.10), we note that both α and β are constants, and that the x-component of the normal depends entirely upon β and the y-component on α, calculating the normals in a pixel shader becomes reduced to calculating the following:

$$N_0x = 2z1 + z3 + z2 - (z5 + 2z7 + z6)$$
$$N_0y = 2z3 + z6 + z1 - (z2 + 2z5 + z7) \quad \text{(B.16)}$$

B.2.1 Up-Dip Normals Pixel Shader

The following pixel shader can be used to program (B.16).

```
void UpDipNormalPS(
        in      float2  texco: TEXCOORD0,
        out     half4   color: COLOR0)
{
        // define the input values
        half4   sourcevals0;
        half4   sourcevals1;
        half4   sourcevals2;
        half4   sourcevals3;
        half4   sourcevals4;
        half4   sourcevals5;
        half4   sourcevals6;
        half4   sourcevals7;
        half4   sourcevals8;

        // The normalScaler is used to balance the lighting effect between seismic lines with
        // different amplitude ranges. Because of the limited dynamic range of the half float
        // format, the normalization must be done for each input value to avoid clipping
        float   nrm = normalScaler;

        // read in the input data from the input texture
        sourcevals1 = tex2D(surfaceSampler, texco + float2(-smapInc, 0)) * nrm;
        sourcevals2 = tex2D(surfaceSampler, texco + float2(-smapInc, smapInc)) * nrm;
        sourcevals3 = tex2D(surfaceSampler, texco + float2(0, -smapInc)) * nrm;
        sourcevals4 = tex2D(surfaceSampler, texco + float2(0, 0)) * nrm;
        sourcevals5 = tex2D(surfaceSampler, texco + float2(0, smapInc)) * nrm;
        sourcevals6 = tex2D(surfaceSampler, texco + float2(smapInc, -smapInc)) * nrm;
        sourcevals7 = tex2D(surfaceSampler, texco + float2(smapInc, 0)) * nrm;

        color.x = 2 * sourcevals1[lightIndex] + sourcevals3[lightIndex] + sourcevals2[lightIndex] -
                (sourcevals5[lightIndex] + 2 * sourcevals7[lightIndex] + sourcevals6[lightIndex]);

        color.y = 2 * sourcevals3[lightIndex] + sourcevals6[lightIndex] + sourcevals1[lightIndex] -
                (sourcevals2[lightIndex] + 2 * sourcevals5[lightIndex] + sourcevals7[lightIndex]);

        // Apply the normalization scaler
        color.x = color.x / 6;
        color.y = color.y / 6;

        // the output from a pixel shader is always a four component vector regardless of the
        // format of the texture being rendered to. The z and w components of the vector are
        // not used but must be defined before the shader will compile.
        color.z = 1;
        color.w = 1;
};
```

B.3 Forward Loop Subdivision Normals

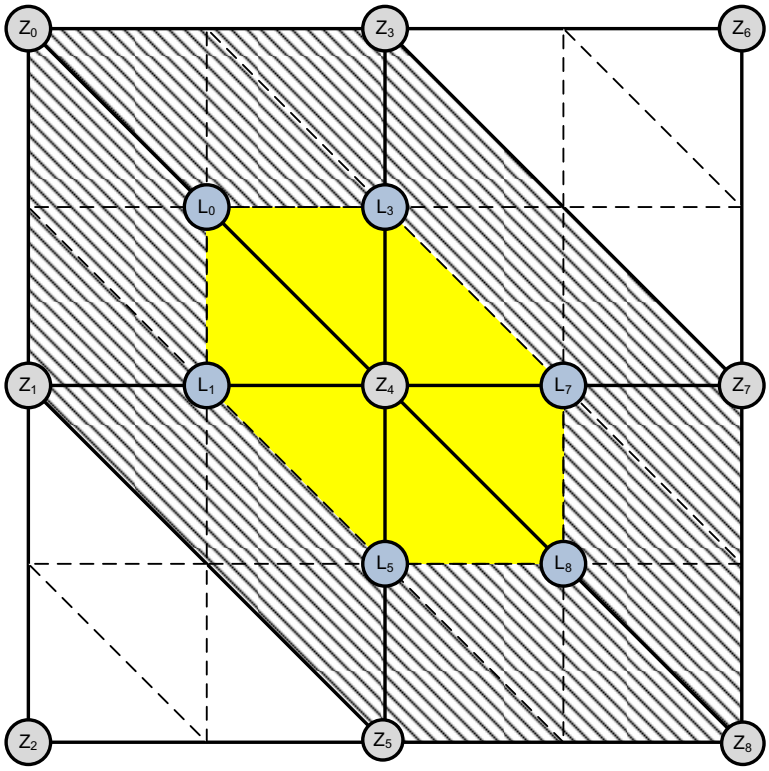


Figure B.3: Modified down-dip tessellation scheme for producing a normal at position Z_4. The vertices are calculated by Forward Loop subdivision. The face normals for the subdivided triangles in yellow contribute to the Z_4 normal.

In this section I develop a pixel shader technique to calculate normals based upon Loop Subdivision (Loop 1987). Forward Loop subdivision is a technique for splitting each triangle in a course mesh into four smaller triangles. It requires as input a regular mesh of triangles and produces as output a refined regular mesh of triangles. As such it is ideal for use in subdividing seismic data. Its drawback is that it quadruples the number of samples needed to render a given set of samples, increasing the number of triangles from two per sample to eight per sample. This places an undue burden on the resources of the graphic card and as a result forward Loop subdivision is not practical on most seismic lines. The Loop scheme is based upon the three-dimensional box spline, which produces C^2 continuous surfaces over regular meshes. Whereas it is not practical to produce the refined meshes it is still possible to make use of the scheme by calculating lighting for a course mesh based upon it.

Figure B.3 shows how a down-dip (see section B.1) tessellation scheme can be modified to produce normals based upon the Loop scheme. There are two types of vertices in this subdivided mesh. The Z_0 - Z_8 vertices are at the locations of the original input samples and are called even vertices. The L_0 - L_8 vertices are new vertices and are called odd vertices.

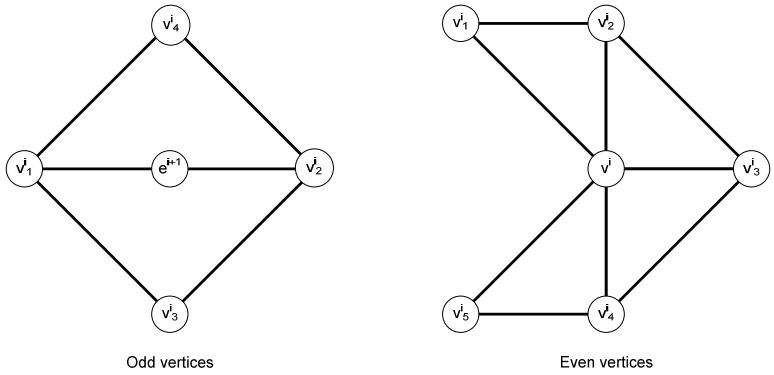


Figure B.4: Schematic of the even and odd vertices of Forward Loop subdivision.

The values of both the even and the odd vertices are calculated by a mask. For the case of the even vertices this mask is given by:

$$v^{i+1} = (1 - n\alpha)v^i + \alpha \sum_{j=1}^{n} v_j^i \tag{B.17}$$

Where:

i is the level of subdivision.

n is the number of vertices to which the given vertex connects.

α is calculated from:

$$\alpha = \frac{1}{n}\left(\frac{5}{8} - \left(\frac{3}{8} + \frac{1}{4}\cos 2\pi/n\right)^2\right) \tag{B.18}$$

For a seismic mesh, each sample is connected to six other samples and therefore:

$$\alpha = \frac{1}{6}\left(\frac{5}{8} - \left(\frac{3}{8} + \frac{1}{4}\cos 2\pi/6\right)^2\right) = \frac{1}{16} \tag{B.19}$$

The mask for the even vertex at Z_4 is then:

$$z_4^{i+1} = \frac{5}{8}z_4^i + \frac{1}{16}\left(z_0^i + z_1^i + z_3^i + z_5^i + z_7^i + z_8^i\right) \tag{B.20}$$

The mask for each odd vertex is given from:

$$e^{i+1} = \frac{3}{8}v_1^i + \frac{3}{8}v_2^i + \frac{1}{8}v_3^i + \frac{1}{8}v_4^i \tag{B.21}$$

B.3.1 Forward Loop Subdivision Normals Pixel Shader

Both equations (B.20) and (B.21) are readily programmable in a pixel shader. The following shader calculates the average unnormalized normal based upon the down-dip tessellation schema described in section B.1.

```
void GenerateLoopNormals(
        in      float2  texco: TEXCOORD0,
        out     float4  color: COLOR0)
{
        float z[9];
        float lp[9];

        // obtain the original sample values from the texture
        z[0] = tex2D(surfaceSampler, texco + float2(-smapInc, -smapInc))[0];
        z[1] = tex2D(surfaceSampler, texco + float2(-smapInc, 0))[0];
        z[3] = tex2D(surfaceSampler, texco + float2(0, -smapInc))[0];
        z[4] = tex2D(surfaceSampler, texco + float2(0, 0))[0];
        z[5] = tex2D(surfaceSampler, texco + float2(0, smapInc))[0];
        z[7] = tex2D(surfaceSampler, texco + float2(smapInc, 0))[0];
        z[8] = tex2D(surfaceSampler, texco + float2(smapInc, smapInc))[0];

        // calculate the odd vertices
        lp[0] = 0.375 * (z[0] + z[4]) + 0.125 * (z[1] + z[3]);
        lp[1] = 0.375 * (z[1] + z[4]) + 0.125 * (z[0] + z[5]);
        lp[3] = 0.375 * (z[3] + z[4]) + 0.125 * (z[0] + z[7]);
        lp[5] = 0.375 * (z[4] + z[5]) + 0.125 * (z[1] + z[8]);
        lp[7] = 0.375 * (z[4] + z[7]) + 0.125 * (z[3] + z[8]);
        lp[8] = 0.375 * (z[4] + z[8]) + 0.125 * (z[5] + z[7]);

        // calculate the even vertices
        lp[4] = 0.625 * z[4] + 0.0625 * (z[0] + z[1] + z[3] + z[5] + z[7] + z[8]);

        // generate the x and y components of the normal
        color.x = 2 * lp[1] + lp[0] + lp[5] - (lp[3] + 2 * lp[7] + lp[8]);
        color.y = 2 * lp[3] + lp[0] + lp[7] - (lp[1] + 2 * lp[5] + lp[8]);

        // The normalScaler is used to balance the lighting effect between seismic lines with
        // different amplitude ranges. Because of the limited dynamic range of the half float
        // format, the normalization must be done for each input value to avoid clipping
        color.x *= normalScaler;
        color.y *= normalScaler;

        // the output from a pixel shader is always a four component vector regardless of the
        // format of the texture being rendered to. The z and w components of the vector are
        // not used but must be defined before the shader will compile.
        color.z = 1;
        color.w = 1;
};
```

B.4 Low-Dip Correlative Dip Adaptive Tessellation

The following pixel shader uses normalized cross-correlation to determine the direction of local dip for a given sample. Where the local dip is determined to be upwards the shader returns 32767 and where it is downwards it return -32767. You must compile the shader using ps_3_0. There are a limited number of registers available to a pixel shader, consequently you can not load all of the samples from both traces in the cross-correlation. I load the samples from the primary trace and subsequently read the sample values for the secondary trace each time I need to use them. This is inefficient because texture reads are slow. However, reading the secondary values on the fly is the only way that this shader will compile. Once I upgrade this routine to Shader 4 (which is only available in DirectX 10) this limitation will go away.

The purpose of this shader is to determine if the seismic data at a sample is up-dip or down-dip. The shader determines this by cross-correlating a small window around the sample with the values from the next trace. The window that I use is ± 4 samples around the target sample and I only allow a ± 2 sample shift in the cross-correlation. The shader sub-samples both the input trace and the cross-correlation trace to one half the sample interval which is why I use halfInc in the texture coordinate calculations, smapInc is input to the shader as the sample interval in texture coordinate space.

```
void GenerateLowDipTexture_PS(
        in      float2  texco: TEXCOORD0,
        out     float4  color: COLOR0)
{
        // Attempts to calculate local dip using a small correlation window.
        // To keep within the limits of the texture I use a 9 point operator and allow +- 2 samples in
        // the correlation. The input is sub-sampled to one half the sample interval.
        // All we want to determine is whether the maximum cross correlation
        // occurs up dip or downdip
        int maxSamps = 8;
        int maxShift = 4;
        int totalSamps = 17;
        int totalShift = 9;
        float sourceData[17];
        float corrData;
        float halfInc = smapInc / 2;
        float2 coords = float2(texco);

        // the halfInc is the width of half a sample in texture coordinates.
        coords[1] -= maxSamps * halfInc;
        for( int i = 0; i < totalSamps; i++ )
                sourceData[i] = tex2D(surfaceSampler, coords + float2(0, i * halfInc))[0];

        coords.x += smapInc;

        float acS0, acC0;
        float corrVal=0;
        float mx = -1000;
        int index = 0;
        float results[9];
        int i, j;

        for( i = 0; i < totalShift; i++ )
        {
                coords.y = texco.y - (maxShift + maxSamps - i) * halfInc;
                results[i] = 0;
                acS0 = 0;
                acC0 = 0;
                for( j = 0; j < 17; j++ )
                {
                        corrVal = tex2D(surfaceSampler, coords)[0];
                        coords.y += halfInc;
                        results[i] += (sourceData[j] * corrVal);
                        acS0 += pow(sourceData[j], 2);
                        acC0 += pow(corrVal, 2 );
                }

                if( results[i] != 0 )
                {
                        results[i] /= sqrt( acS0 * acC0 );

                        if( results[i] > mx )
                        {
                                mx = results[i];
                                index = i;
```

```
                        }
                }
        }

        if( index < maxShift )
                color.x = 32767; // updip
        else
        {
                if( index > maxShift )
                        color.x = -32767; // downdip
                else
                {
                        if( results[index-1] > results[index+1] )
                                color.x = 32767; // updip
                        else
                                color.x = -32767; // downdip
                }
        }

        // these three are not needed since they are never written to the output
        // however you can't just comment them out otherwise it won't compile
        color.y = 1;
        color.z = 1;
        color.w = 1;
};
```

B.5 High-Dip Correlative Dip Adaptive Tessellation

In the previous section I developed a shader that calculated the direction of the local dip for a given sample. In this section I extend that shader to return the magnitude of the local dip as well as the direction. The returned magnitude is in samples per trace. I use the same ± 4 samples around the target sample for the data window but in this shadier I allow a ± 4 sample shift in the cross-correlation as opposed to the ± 2 sample shift that I used in the previous one. To faciliate that change and still maintain the same efficiency I also drop the subsampling. That is why the halfInc, used throughout the previous example, is replaced by smapInc, which is defined as the sample interval in texture coordinate space.

```
void GenerateHighDipTexture_PS(
	in	float2	texco: TEXCOORD0,
	out float4	color: COLOR0)
{
	// Attempts to calculate local dip using a small correlation window.
	// Uses a 9 point operator and allowing +- 4 samples in the correlation
	// so there may be problems as you cross texture boundaries.
	//The difference between this shader and the low dip version
	// is that the low dip version steps by half a sample.
	// The amount of work done here is the same but we step by an entire sample
	int maxSamps = 8;
	int maxShift = 4;
	int totalSamps = 17;
	int totalShift = 9;

	float sourceData[17];
	float corrData;

	float2 coords = float2(texco);
	coords[1] -= maxSamps * smapInc;
	for( int i = 0; i < totalSamps; i++ )
		sourceData[i] = tex2D(surfaceSampler, coords + float2(0, i * smapInc))[0];

	coords.x += smapInc;

	float acS0, acC0;
	float corrVal=0;
	float mx = 0.25;
	int index = maxShift;
	float results[9];
	int i, j;
	for( i = 0; i < totalShift; i++ )
	{
		coords.y = texco.y - (maxShift + maxSamps - i) * smapInc;
		results[i] = 0;
		acS0 = 0;
		acC0 = 0;
		for( j = 0; j < 17; j++ )
		{
			corrVal = tex2D(surfaceSampler, coords)[0];
			coords.y += smapInc;
			results[i] += (sourceData[j] * corrVal);
			acS0 += pow(sourceData[j], 2);
			acC0 += pow(corrVal, 2 );
		}

		if( results[i] != 0 )
		{
			results[i] /= sqrt( acS0 * acC0 );

			if( results[i] > mx )
			{
				mx = results[i];
				index = i;
			}
```

```
                }
        }

        color.x = maxShift - index;

        if( color.x == 0 )
        {
                // handle small shifts
                if( results[index-1] > results[index+1] )
                        color.x = 0; // updip
                else
                        color.x = -1; // downdip
        }

        // these three are not needed since they are never written to the output
        // however you can't just comment them out otherwise it won't compile
        color.y = 1;
        color.z = 1;
        color.w = 1;
};
```

White hands cling to the tightened rein
Slipping the spur from the booted heel,
Tenderest voices cry "Turn again!"
Red lips tarnish the scabbarded steel.
High hopes faint on a warm hearth-stone –
~~*He travels the fastest who travels alone*~~.
"The Winners"

Turn now to Simoorie where, lapped in his ease,
The Captain is petting the Bride on his knees,

Where the whit of the bullet, the wounded man's scream
Are mixed as the mist of some devilish dream –

Forgotten, forgotten the sweat of the shambles
Where the hill-daisy blooms and the gray monkey gambols,

From the sword-belt set free and released from the steel,
The Peace of the Lord is on Captain O'Neil.
"The Ballad of Boh Da Thone"

"Audentis Fortunas Iuvat"

Fortune Favors the Bold

Allegedly, the final known words of the great Roman scholar, historian and general, Pliny the Elder. Spoken as he boarded his ship to mount a very bold rescue operation for the citizens of Pompeii where, very unfortunately, he died; roasted alive in a pyroclastic flow.

Ah what the hell!

Sanity is very much overrated.

Even if we had known … we would have done it anyway.