COMPUTER GRAPHICS, PERIPHERAL VISION AND NON EUCLIDIAN GEOMETRY

By

Jayanth Rama Varma

W P No. 852
February 1990

The main objective of the working paper series of the IIMA is to help faculty members to test out their research findings at the pre-publication stage.

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD-380 056
INDIA

COMPUTER GRAPHICS, PERIPHERAL VISION AND NON EUCLIDIAN GEOMETRY

Jayanth Rama Varma

ABSTRACT

Computer graphics in Decision Support Systems is often confronted with the task of providing the decision maker with a visual picture of some object which is too large to fit on a computer screen unless the image is scaled down so drastically that much of the detail is lost. The viewer is then asked to work with a partial view of the object, and use a keyboard or a mouse to (a) scroll this image horizontally or vertically, or (b) zoom in or out, or (c) rotate the object. These techniques are strikingly similar to those that the human visual system uses to deal with a similar problem. One crucial difference is that of peripheral vision — the human eye while concentrating on a small part of the field of vision still retains a hazy view of the peripheral region preventing it from losing sight of the total picture. This paper argues that the lack of a similar peripheral vision is perhaps the single gravest deficiency in computer graphics today. It then goes on to develop a mapping technique which simulates this peripheral vision, and thereby makes computer graphics truly powerful and versatile. The paper analyses the distortions induced by such a mapping, and argues at length why these do not pose serious problems. The suggested mapping is closely related to non Euclidian geometry; this ties in with the fact known to psychologists for over fifty years that the perceptual geometry of human vision is strongly non Euclidian. Thus, if one were to adapt the Turing test for artificial intelligence to computer

vision, then non Euclidian geometry can be expected to play a key role in any attempt to satisfy that test. Building on these ideas, the paper demonstrates that computer graphics has a great deal to learn from non Euclidian geometry, and that in turn computer graphics can contribute significantly to the intelligent application of non Euclidian geometries to real life problems. What is needed is the willingness to set aside the shackles and shibboleths of Euclidian geometry.

COMPUTER GRAPHICS, PERIPHERAL VISION AND NON EUCLIDIAN GEOMETRY

Jayanth Rama Varma

Computer graphics is often used in Decision Support Systems to provide the decision maker with a visual picture of some object which I shall hereafter call the "world". Usually, the "world" is too large to fit on a computer screen unless the image is scaled down so drastically that much of the detail is lost. The viewer is then presented with a partial view of the "world" at which he can see the relevant detail, and is allowed to zoom in or zoom out so that he can either see a smaller portion of the "world" with a great deal of detail (high resolution), or he can see a larger portion of the "world" at low resolution. (I shall use resolution to denote the relationship between the size of an element of the "world" and its image on the screen. In other words, if a centimetre on the screen represents 10 metres of the "world", the resolution is higher than if that centimetre represents 100 metres of the "world". Hardware vendors, on the other hand, define resolution in terms of either the total number of pixels or the pixel density.) The user is also given the option to scroll the image up or down or sideways so that he can see different portions of the "world" at the same level of resolution. He can also rotate the "world" to see it from different angles and perspectives.

All these techniques have a striking resemblance to the techniques of human vision itself. For example, Wiener(1961) gives a description of vision in cybernetic terms :

> "The human eye has economically confined its best form and color vision to a relatively small fovea, while its perception of motion is better on the periphery. When the peripheral vision has picked up some object conspicuous by brilliancy or light contrast or color or above all by motion, there is a reflex feedback to bring it into the fovea. This feedback is accompanied by a complicated system of interlinked subordinate feedbacks, which tend to converge the two eyes so that the object attracting attention is in the same part of the visual field of each, and to focus the lens so that its outlines are as sharp as possible. These actions are supplemented by motions of the head and body, by which we bring the object into the centre of vision if this cannot be done readily by a motion of the eyes alone, or by which we bring an object outside the visual field picked up by some other sense into that field. In the case of objects with which we are more familiar in one angular orientation than another — writing, human faces, landscapes and the like — there is a mechanism by which we tend to pull them into the proper orientation."
> (p 134)

When we compare this mechanism of human vision with that of computer graphics, we see that when the viewer manipulates the keyboard or the mouse to bring an object into the centre of the screen, the techniques that he uses are essentially no different from that adopted by the eye. The graphics facilities implement most of the techniques of human vision with two notable exceptions :

1. The stereoscopic (or binocular) vision provided by our two eyes is not available in computer graphics. This is a very important topic by itself, but I shall not deal with it in this paper.
2. Computer graphics does not provide the peripheral vision which is so central in Wiener's description. When the viewer brings an object to the centre of the screen, the object in question has been picked up not by the peripheral vision, but in Wiener's words "by some other sense". Usually, the viewer goes by memory or some prior knowledge of the "world".

This concept of peripheral vision is the key idea of this paper. In the human eye, we find a gradual loss of visual acuity as we move away from the fovea to the periphery; we do not find an abrupt loss of vision at some point. The computer screen on the other hand uses line clipping to implement what one might call a "cookie cutter" vision — a small portion of the "cookie" is neatly cut out and given to us. The screen is treated as a window to the "world" — everything visible from this window is displayed at the same resolution, and what is outside is simply cut out. (In the jargon of computer graphics, this is known as the pyramid of vision.)

It is as if, while I sit in front of my table and write, I can see only the paper, and cannot see somebody walking into the room until he calls me by name. (Of course, I could change the resolution so that I can see the whole room, but then I cannot read what I have written on the paper.) I think it is quite obvious that an organism whose vision was similar to our computer graphics displays would not survive for long in this harsh world. Even in the case of humans, the peripheral vision and the associated arousal mechanism can sometimes fail. I am reminded of Archimedes staring at his diagram oblivious of the soldier approaching him with drawn sword — Archimedes did not survive. Of course, even the Archimedes staring at the central portion of the diagram is not oblivious of the rest of the diagram; if he did, he would be not only a poor survivor but also a poor mathematician. Thus, even in the case of Archimedes, the failure is not so much that of the peripheral vision of the eye as that of the arousal mechanism of the brain (the reticular formation) under extreme cortical control (cf. Kilmer et al (1968) p 290).

The point that I wish to make is that the lack of peripheral vision in our graphics displays is a very serious deficiency; I am inclined to regard it as the single greatest deficiency in these systems today. Neither as system users nor as system designers should we accept this as the natural state of affairs and be satisfied with it.

The question then is what can we do about it. One alternative is to follow the lead of the cinema which has used such large screens (cinemascope etc.) that they almost cover the field of vision of the human eye (foveal and peripheral). By contrast, the typical computer screen has a width of only about 15° out of the 150° field of vision — even if we wore blinkers like horses, we would have a far wider field of vision than that. It is interesting to note that in this respect, television faces an even more acute problem as the TV screen is usually only about 10° wide at normal viewing distance. The principal obstacle in using the "cinemascope" solution is the high cost, and one hopes that, in the long run, development in flat screen technology will make this solution feasible. But let me clarify that it would not do to simply project the image from a conventional graphics display onto a large screen. For, then the viewing distance would have to be increased to prevent the image from becoming too coarse and grainy; at the enhanced viewing distance, the image would still subtend the same old narrow angle of vision. What we need is a large cinemascope-like screen with a pixel density comparable to that of present-day graphic terminals. This obviously means a very much larger number of pixels, and a proportional increase in computational load on the graphic processor. Even if all this were achieved, it is arguable that other means of simulating peripheral vision could be profitably used in addition to this.

Another possibility is to simulate the technique of the human eye itself to create peripheral vision on the screen. Unfortunately, this is likely to be computationally infeasible. The massive parallelism of the human nervous system can perform a complex averaging process which our von-Neumann machines would find quite time consuming. Advances in supercomputer technology could change this situation, but one may be pardoned for suggesting that supercomputers ought to be used to replicate (and transcend!) the human brain rather than the human eye.

That leaves us with the alternative of using mathematical (analytical) techniques to achieve a substitute for peripheral vision. I propose to regard it as a mapping problem — how do we map the "world" onto the screen so that the central portion is seen in high resolution and the peripheral region in low resolution. By contrast, the conventional display system uses a constant resolution within the screen and a zero resolution for the rest of the "world". The moment the problem is posed this way, the solution seems obvious : use a variable scaling, that is to say, let the scale of the map vary from point to point decreasing sharply towards the periphery.

At this point some analytic geometry is inevitable. Typically, the "world" is three dimensional; this 3D "world" is projected onto the plane of the screen using the laws of geometric optics; and it is this planar image that is clipped onto the screen by using line clipping. Since I have no quarrel with the laws of geometric optics, I shall start with the intermediate planar image which I shall call the "world plane". The problem is to map this plane onto the screen. For mathematical simplicity, I shall initially assume that the screen is circular with a radius of unity (I shall call this the screen circle). In reality, computer screens are rectangular; if we use only the largest inscribed circle in it, we would have to waste a lot of screen space. I shall subsequently discuss how this problem can be alleviated. For the moment, the problem is to map the "world plane" to the screen circle. To this end, I introduce polar coordinates (r, θ) in the "world plane"; the origin is at the centre of vision (whichever point of the "world plane" the viewer is focussing on at a given time), and r tells us how far a given point is from this centre while θ tells us in what direction it lies (see Fig. 1a). Similarly, I introduce polar coordinates (r', θ') in the screen circle; here, the origin is at the physical centre of the screen, and r' tells us how far a given point is from this centre while θ' tells us in what direction it lies. While r ranges from 0 to ∞, r' ranges from 0 to 1; both θ and θ' range from 0 to 2π.

The essential idea of the proposed mapping is that the center of vision should be mapped into the centre of the screen and points close to the centre of vision should be mapped unchanged into corresponding points on the screen. For points far from the centre of vision, the image should lie in the same direction (θ') from the screen centre as the point lies from the centre of vision (i.e. θ), but the distance from the screen centre (r') should be shrunk so that a large peripheral area can be fitted onto the screen. I suggest the following class of mappings (see

Fig 1) :

$$\theta' = \theta$$

$$r' = f(r) \qquad r \leq R,\ R \gg 1.$$

where $f(0) = 0$, $f(R) = 1$, $f'(0) = 1$, and f is strictly increasing though at a decreasing rate i.e. $f'(r) > 0$, $f''(r) < 0$. The condition $\theta' = \theta$ means that the direction is always left unchanged in the mapping; $f(0) = 0$ means that the centre of vision is mapped into the screen centre; $f(R) = 1$ means that points at distance R from the centre of vision are mapped to the edge of the screen and points beyond this are not shown on the screen; $f'(0) = 1$ means that for points close to the centre of vision $f(r)$ is close to r itself. $f'(r)$ can be regarded as the local scale factor (more precisely it is what I shall call the radial scale factor); $f'(r) > 0$ states the requirement that the scale should be strictly positive; $f''(r) < 0$ means that this scale falls off as we move away from the centre of vision so that the peripheral region is shrunk. I call $f'(r)$ the radial scale because when a point at distance r from the centre of vision moves a small distance d radially outward (or inward), its image on the screen moves a distance $d' = f'(r)d$ radially outward (or inward). On the other hand, if the same point moves a small distance δ tangentially, its image moves a distance $\delta' = \delta f(r)/r$ tangentially; the tangential scale is, therefore, $f(r)/r$ (see Fig 2). These two scales will not in general be the same; in fact, for no choice of f satisfying the conditions imposed above will the two scales be equal throughout the screen[1*].

A good candidate for the mapping function f is the hyperbolic tangent :

$$f(r) = \tanh(r) = \frac{\exp(r) - \exp(-r)}{\exp(r) + \exp(-r)}$$

whose inverse is given by :

$$f^{-1}(r') = \tanh^{-1}(r') = \frac{1}{2} \ln \frac{1 + r'}{1 - r'}$$

Interestingly, in this case, $R = \infty$ so that the entire "world plane" is actually mapped onto the screen circle. This means that the peripheral vision includes the entire "world", but distant parts of the "world" will be very hazy indeed. The reader may readily verify that this choice of f satisfies all the conditions that were imposed earlier. The radial scale is given by :

$f'(r) = 1 - \tanh^2(r) > 0$

and we have

$f''(r) = -2\tanh(r)(1-\tanh^2(r)) < 0.$

The tangential scale is :

$f(r)/r = \tanh(r)/r.$

---

* Most of the mathematical detail has been banished to the notes which appear at the end.

The tabulation below gives an idea of how these scales behave :

r	0	0.10	0.20	0.30	0.40	0.50	1.00	2.00	3.00	4.00
f(r)	0	0.10	0.20	0.29	0.38	0.46	0.76	0.96	1.00	1.00
f'(r)	1	0.99	0.96	0.92	0.86	0.79	0.42	0.07	0.01	0.00
f'(r)/r	1	1.00	1.00	0.97	0.95	0.92	0.76	0.48	0.33	0.25

It will be observed that a circle of radius 0.50 in the "world plane" is mapped without much distortion into the circle of radius 0.46 in the screen. A circle of radius 1 fills 3/4ths of the screen, and at this point the radial resolution has dropped to less than half. A circle of radius 2 almost fills the screen, and the radial resolution at this point is about 7%.

It will also be observed that the radial and tangential scales are not equal; in fact towards the periphery, the scales are markedly different[2]. As argued earlier, this is not peculiar to our choice of f = tanh(r); the same phenomenon is inevitable for any reasonable choice of f. This divergence does however have several very serious implications :

1. Angles are not preserved in the periphery — any angle between a radial line and any curve is distorted towards a right angle, i.e. acute angles are increased and obtuse angles are diminished[3]. A curve intersecting the radial line at right angles in the world plane continues to do so in the screen also.
2. Straight lines are not always preserved in the periphery. Any straight line in the "world plane" passing through the centre of vision is mapped into a diameter of the screen

circle and, therefore, remains straight. Lines in the "world plane" parallel to this line are mapped into curves; near the screen centre, these curves are approximately straight lines parallel to the diameter of the screen circle mentioned above, but they gradually bend to meet the diameter at the edge of the screen[4].

Thus, the mapping gives a substantially accurate picture of the central field of vision, but gives an increasingly distorted picture of the peripheral region (see Fig 3). This in my opinion is quite reasonable as the purpose of the peripheral vision is only to situate the central field of vision in the larger "world". If the viewer finds a peripheral object interesting, he would move the centre of vision to that point; that now becomes the new central field, and is presented to him without distortion. This is exactly what happens with human vision also; when I detect a moving object in the periphery, I do not know whether it is a cat or a piece of paper or what have you; I then move my head or eyes bringing the object into the centre of vision to find out what it is; having satisfied my curiosity, I might turn back to whatever I was looking at earlier or I might spend the next half an hour observing the cat until something else catches my attention.

It remains to show how the circle into which the "world plane" has been mapped can be transformed into a geometrical shape that fits into a rectangular screen without much loss of space. In my view, the simplest and most practical solution is to dilate one of the axes to deform the circle into an ellipse whose major and minor axes are equal to the width and height of the screen. Further discussion of this matter is relegated to the notes which also highlight the possible use of the conformal mappings of complex analysis[5].

I would like to emphasize that none of the features currently available in graphics systems is being given up to obtain peripheral vision. Peripheral vision is in addition to all these things.

1. Scrolling : As pointed out in the last paragraph, the viewer can scroll the picture vertically or horizontally by moving the centre of vision in the appropriate direction; the only difference is that while scrolling, objects do not disappear from the edge of the screen as with conventional line clipping; instead they just shrink and fade away into the periphery.
2. Zooming : It might appear that the scaling function condemns us to seeing the "world" at a fixed level of resolution. This is not so. The hyperbolic scaling,

   $r' = \tanh(r)$

   is applied only to the "world plane" which is the end result of all the projective transformations of conventional graphics. The "world plane" is the plane which conventional graphics "clips" onto the screen. Zooming changes the scale relationship between the "world" and the "world plane"; hyperbolic scaling then rescales this "world plane" to the screen circle. For example, the tabulation given earlier showed that the circle of radius half in the "world plane" is mapped into the central portion of the screen without much of distortion, and the rest is peripheralized. Zooming determines how much of the "world" was contained in that circle of radius half in the "world plane". When the viewer zooms in, only a small portion of the world occupies this circle; hyperbolic scaling than focuses on this bit of the world and peripheralizes the rest. Put differently, hyperbolic scaling is a substitute for line clipping; it does not supplant, usurp or obstruct any other function. This viewpoint makes for conceptual clarity and software modularity; otherwise, one could view zooming as introducing another adjustable parameter μ ($\mu > 0$) into the hyperbolic scaling function to yield the mapping :

$$r' = \tanh(\mu r)$$

Raising μ above one would zoom into the picture; reducing it below one would zoom out. But, conceptually, it is better not to do this; the cleaner solution is to perform hyperbolic scaling only after the zooming has been done. Strictly, one would view even the scrolling function discussed above as being done before the hyperbolic scaling is carried out; at the software level that is how it would be done. In existing graphics packages, one has to replace only the line clipping algorithm by the hyperbolic scaling algorithm; ideally, this algorithm must be implemented directly in hardware for maximum speed.

At the end of it all, the distortion of the peripheral region is what many readers would find distasteful in the whole proposal. Can the viewer interpret pictures in which all these distortions exist? My answer is an unequivocal yes.

First of all, distorted maps are very common : every map of the world is highly distorted. In world maps, the only choice one has to make is what to distort and what to leave unchanged : the Mercator projections map the rhumb lines into straight lines, the equal area projections leave areas invariant and so on. We do interpret these maps, and in fact derive a great deal of our knowledge of geography from them.

Secondly, just as we do misinterpret maps of the world, we may also misinterpret the hyperbolically scaled screen. But this is not dangerous because we would never misinterpret the central area of the screen (the object on which we are focussing), and by scrolling and zooming, we would rapidly correct any misinterpretations of the peripheral regions that may exist.

Thirdly, let us consider the principal misinterpretations that the viewer would make when viewing the hyperbolically scaled image. I think there are two principal misinterpretations possible.

1. The viewer may wrongly assume that straight lines on the screen are straight in the "world plane" also. In case of

world maps also, we tend to make a similar assumption which is valid in the gnomonic projection but not most other projections.

2. The viewer though realizing that there is a great shrinkage of scale towards peripheral regions, may wrongly assume that the radial and tangential scales are the same, and, therefore, angles are correctly depicted. In the technical jargon, he may assume that the image is conformal, i.e. small shapes are not distorted.

It turns out that both of these misinterpretations of the map have one major common implication : the geometry in question becomes non Euclidian (specifically hyperbolic)[6]. This to my mind, is a very important insight because psychologists have in fact known for over fifty years now that perceptual geometry of human vision is in fact strongly non Euclidian – specifically hyperbolic (see, for example, Blank (1959)). This experimental evidence is at first quite surprising and inexplicable. Living as we do in a Euclidian world (the relativistic non Euclidian nature of the world is negligible for our purposes), why do we have non Euclidian vision and how do find our way about in the world? Peripheral vision suggests an answer to both questions. We find our way about because for that we rely one our foveal vision which is Euclidian (hyperbolic geometry is locally Euclidian); we never trust our peripheral (non Euclidian) geometry for that. We have hyperbolic vision in a Euclidian world because that is the way to accommodate peripheral vision which is more important for human survival than the niceties of

Euclidian geometry. This suggests that our hyperbolic scaling is on the right track after all. In its attempt to replicate human peripheral vision, it seems to be producing as an unintended side effect the same non Euclidian characteristics to which the human eye is also subject for, possibly, the same reason. Thus what appears at first sight to be a disadvantage of hyperbolic scaling turns out in the end to be an argument in its favour. Long ago, Turing (1950) proposed a test for artificial intelligence in terms of the ability to simulate a human being across a teletype line. Turing's paper contained a hypothetical conversation in which the interrogator asks the computer to add 34957 to 70764. After a pause of 30 seconds, the computer replies 105621. The correct answer, of course, is 105721. In other words, to satisfy the Turing test, the computer must, in simulating human thought, be able to simulate human mistakes also. If we extend the idea to the case of computer vision, we would have to conclude that the "mistake" of hyperbolic geometry that human vision makes in interpreting the Euclidian physical world is one of the important things that a successful computer vision system must be able to replicate. On a deeper analysis, of course, hyperbolic geometry is not so much a mistake as an inevitable side effect of achieving other more important objectives — peripheral vision.

Fourthly, often the "world" is a graph or other similar object whose topological properties are extremely important but whose metrical properties are totally irrelevant. Since the hyperbolic scaling preserves all incidence relationships (and other topological properties), it faithfully reproduces all important properties of a graph. For example, a hyperbolically scaled PERT chart might have some curved edges; but, in interpreting a PERT chart, it matters little whether the edges are curved or straight.

Fifthly, often the "world" which is being displayed on the screen is not a physical space at all but an imaginary "attribute space". This happens, for example, in Multi Dimensional Scaling (MDS), where statistical techniques are applied to infer a set of unobservable attribute dimensions from data on human perceptions or choices; e.g., data on consumer preferences among various brands of toothpaste may be used to infer a set of perceptual attributes on which the consumers are evaluating the toothpastes, and the various brands of toothpastes can be plotted in this attribute space which the viewer can explore. In these cases, there is no a priori reason for assuming that this perceptual space is Euclidian; in fact, the evidence that even physical space which is Euclidian is not so perceived should lead to a presumption that these perceptual spaces will also be non Euclidian — specifically hyperbolic. It is only sheer prejudice, ignorance or mathematical naivete which leads to the continued prevalence of Euclidian models in these fields (thankfully, the situation is changing (Carroll and Arabie (1980))). If the "world plane" is in fact hyperbolic, then a slight reinterpretation of the hyperbolic scaling produces a faithful, undistorted (conformal) map of the hyperbolic plane.

If non Euclidian models are being used in MDS and similar applications then computer graphics will also have to prepare to map non Euclidian spaces onto the screen. In fact, computer graphics has an even greater role to play in the graphical representation of non Euclidian spaces than it has in the Euclidian case. This is because the non Euclidian space does not have a ready-made physical model as the Euclidian space has. We are, therefore, forced to rely on mathematical models of these spaces. Computer graphics provides the best mechanism for the visual representation and manipulation of these models. Computer graphics will thus be the principal educational tool by which viewers can obtain an intuitive feel for non Euclidian geometry, and learn to apply it intelligently and creatively to solve real life problems. I, for one, am prepared to claim unhesitatingly that, given some exposure and practice, non Euclidian geometries are as intuitively appealing as the Euclidian ones. All that needs to be done is to demystify these geometries and make them accessible to the non specialist. This is a function that computer graphics is best suited to perform.

The conclusion is thus that computer graphics has a great deal to learn from non Euclidian geometry, and that in turn computer graphics has a very significant educational contribution to make to the intelligent application of non Euclidian geometries to real life problems. It is indeed high time that we as users and designers of computer graphics systems began to discover and utilize the richness and power of non Euclidian geometry. What is needed for that to happen is the willingness to set aside the shackles and shibboleths of Euclidian geometry.

REFERENCES

Blank, A.A. (1959), "The Luneberg Theory of Binocular Space Perception", in Koch. S., (ed), Psychology : A Study of a Science, Vol 1, New York, McGraw Hill.

Carroll, J.D. and Arabie, P. (1980), "Multidimensional Scaling", Annual Review of Psychology, 31, 607-649.

Turing, A.M. (1950), "Computing Machines and Intelligence", Mind, 59, 433-460, reprinted in Feigenbaum, E.A. and Feldman, ed, (1963)Computers and Thought, New York, McGraw Hill.

Wiener, N. (1961), Cyberneertics, 2d edn., New York, MIT Press.

MATHEMATICAL NOTES

(1) Radial and Tangential Scales

When the point (r,θ) in the world plane is displaced by an infinitesimal amount to the point $(r+dr, \theta+d\theta)$, the distance moved in the world plane is given by :

$$ds^2 = dr^2 + r^2 d\theta^2$$

which is the equivalent in polar coordinates of the more familiar formula :

$$ds^2 = dx^2 + dy^2$$

in rectangular coordinates.

At the same time, the image on the screen moves to the point $(r'+dr', \theta'+d\theta')$ where

$$dr' = f'(r)dr$$
$$d\theta' = d\theta$$

and the distance moved in the screen is

$$ds'^2 = dr'^2 + r'^2 d\theta'^2$$

For radial movements, $d\theta = 0$ and the preceding formulas give

$$ds^2 = dr^2 \Rightarrow ds = dr$$
$$ds'^2 = dr'^2 \Rightarrow ds' = dr' = f'(r)dr$$

so that $ds'/ds = f'(r)$. This is the radial scale.

For tangential movements, $dr = 0$, and we have

$$ds^2 = r^2 d\theta^2 \Rightarrow ds = r d\theta$$
$$ds'^2 = r'^2 d\theta'^2 \Rightarrow ds' = r' d\theta' = f(r) d\theta$$

so that $ds'/ds = f(r)/r$ (for $r > 0$). This is the tangential scale. As r tends to 0, the tangential scale tends to $f'(0)$ by L'Hospital's rule. By assumption, $f'(0) = 1$ implying that both the radial and tangential scales are equal to unity at the centre of vision.

For the two scales to be the same everywhere (i.e. for all $r > 0$), we must have

$$f(r)/r = f'(r)$$
$$\Rightarrow f'(r)/f(r) = 1/r$$

$$\Rightarrow \frac{d}{dr} \ln f(r) = 1/r$$

$\Rightarrow \ln f(r) = \ln r + c$ where c is a constant of integration.
$\Rightarrow f(r) = ar$ where $a = \exp(c)$.

Further, $f'(0) = 1$ implies that $a = 1$; thus the map must be the identity. The map has $f''(r) = 0$ violating the requirement that $f''(r) < 0$.

(2) Limiting Behaviour of the Radial and Tangential Scales

In the limit, as r tends to infinity, the radial scale diminishes relative to the tangential scale, and the ratio of the scales tends to zero. This is valid for all scaling functions f(r) which map the whole world plane into the screen circle, i.e., for which $R = \infty$. I shall, however, prove this first for the hyperbolic map $f(r) = \tanh(r)$.

Tangential scale = $1/r$
Radial scale = $1 - \tanh^2(r)$
Radial/Tangential = $(1 - \tanh^2(r))r$

Substituting $t = \tanh(r)$ we get

$$\lim_{r \to \infty} (1 - \tanh^2(r))r = \lim_{t \to 1} (1-t^2)\tanh^{-1}(t)$$

$$= \lim_{t \to 1} (1-t^2)\frac{1}{2}\ln\frac{1+t}{1-t}$$

$$= \lim_{t \to 1} (1/2)(1-t^2)\ln(1+t) - (1/2)(1-t^2)\ln(1-t)$$

As t tends to 1, the first term tends to zero. It remains to evaluate the second term which is still of the form $0\cdot\infty$. To this end, I substitute $s = 1-t$ and evaluate

$$\lim_{s \to 0} -(1/2)[1 - (1-s)^2]\ln s$$

$$= \lim_{s \to 0} -(1/2)s(2-s)\ln s$$

The term $-(1/2)(2-s)$ remains bounded as s tends to zero, while, as is well known (use L'Hospital's rule), $s \ln s$ tends to 0. The limit in question is thus zero.

To prove the result for the general case, first define the function g(r) to denote the ratio

$$\frac{f'(r)}{f(r)/r}$$

whose limit (as r tends to infinity) is to be evaluated). The assumptions on f are :

$f(0) = 0$, $f'(r) > 0$, $f''(r) < 0$, and

$$\lim_{r \to \infty} f(r) = 1$$

Suppose g(r) were greater than ϵ for all r sufficiently large. Then for such r we have

$$f'(r)/f(r) > \epsilon/r$$
$$\Rightarrow (d/dr) \ln f(r) > \epsilon/r$$
$$\Rightarrow \ln f(r) > \epsilon \ln r + c$$

As r tends to infinity, ln f(r) would have to increase without limit contradicting the fact that f(r) tends to 1 and ln f(r) tends to 0. Therefore, given an s howsoever large, g(r) cannot remain above any positive ϵ howsoever small for all r > s. This and the fact that g(r) is non negative preclude the possibility of g(r) tending to any non zero limit; but this is not the same thing as saying that it tends to zero. It is necessary to prove that the limit in question exists at all; as usual, it takes more effort to prove that a limit exists than to evaluate the limit assuming its existence.

Since ln f(r) tends to zero monotonically, we can choose s so large that ln f(r) > − ϵ for all r > s. By the preceding paragraph, g(r) cannot remain above ϵ for all such r. It is, however, possible that g(r) may drop below ϵ and then rise above it, and keep oscillating in this manner. I shall now prove that having once dropped below ϵ for some r > s, it cannot rise above eϵ where e is the base of the natural logarithms. Assume to the contrary that this possible. Then we can select t and t* as follows :

$$s < t < t^*$$
$$g(t) = \epsilon$$
$$g(t^*) = e\epsilon$$
$$\epsilon \leq g(r) \leq e\epsilon \text{ for } t \leq r \leq t^*.$$

I shall show that this leads to a contradiction. First of all, I derive an upper bound on g′(r).

$$\begin{aligned} g &= rf'/f \\ \Rightarrow g' &= [f(f'+rf'') - rf'^2]/f^2 \\ &< [ff' - rf'^2]/f^2 \\ &= [rf'/f - (rf'/f)^2]/r \\ &= (g - g^2)/r \\ &\leq g/r \end{aligned}$$

The first inequality follows from the fact that f′′ < 0. Further,

$$g' < g/r$$
$$\Rightarrow g'/g < 1/r$$
$$\Rightarrow (d/dr) \ln g < 1/r$$
$$\Rightarrow \ln g(t^*) - \ln g(t) < \ln(t^*) - \ln(t)$$
$$\Rightarrow \ln (t^*/t) > \ln [g(t^*)/g(t)] = 1$$

On the other hand,

$$\ln f(t^*) - \ln f(t) = \int_t^{t^*} \frac{g(r)}{r}\, dr \geq \int_t^{t^*} \frac{\epsilon}{r}\, dr = \epsilon \ln(t^*/t)$$

The preceding two inequalities together imply that

$\ln f(t^*) - \ln f(t) \geq \epsilon$.

This is a contradiction because $\ln f(t^*) \leq 0$ and $\ln f(t) > -\epsilon$, and, therefore, $\ln f(t^*) - \ln f(t) < \epsilon$.

This contradiction establishes that for r sufficiently large, g(r) remains below $e\epsilon$. Since ϵ was arbitrary, g(r) does tend to zero as claimed.

(3) Distortion of Angles

It is convenient to work in rectangular coordinates converting to polar coordinates and back where necessary. The transformation is given by :

$x = r \cos \theta$
$y = r \sin \theta$

Now consider the angle formed by a curve intersecting a radial line in the world plane; it is required to determine the distortion of this angle when it is mapped into the screen. Since the mapping clearly possesses circular symmetry, it is sufficient to consider the case where the radial line in question coincides with the positive X-axis, i.e., $\theta = 0$. I, therefore, consider a curve intersecting the X-axis at the point $(x = r, y = 0)$. The angle between two curves is the angle between their tangents. In other words, the angle Φ is given by

$dy/dx = \tan \Phi$

where dy/dx is the slope of the curve at the point of intersection.

Conversion into polar coordinates yields :

$dr = dx$
$d\theta = dy/r$

In the screen, we have

$dr' = f'(r)dr = f'(r)dx$
$d\theta' = d\theta = dy/r.$

Converting back into rectangular coordinates in the screen

$dx' = dr' = f'(r)dx$
$dy' = r'd\theta' = r'dy/r$
$dy'/dx' = [(r'/r)/f'(r)]dy/dx$

The expression in square brackets is the ratio of tangential and radial scales (it is $1/g(r)$ in the notation of the preceding note). Denote this ratio by μ. Then,

$dy'/dx' = \mu\ dy/dx$
$\tan \xi' = \mu \tan \xi$

$\xi' = \tan^{-1}(\mu u)$ where $u = \tan \xi$

For $\mu = 1$ (as in the screen centre), $\xi' = \xi$.

Differentiating w.r.t. μ, we get

$$\frac{d}{d\mu}\xi' = u/(1+\mu^2 u^2)$$

Consequently

$$\xi' = \xi + \int_1^{\mu} \frac{u}{1 + \mu^2 u^2}\, d\mu$$

The integrand is positive or negative according as u is positive or negative, i.e., according as ξ is acute or obtuse. If ξ is obtuse, then ξ' rises steadily as μ is increased from unity. On the other hand, if ξ is obtuse, the integrand is negative and the obtuse angle is reduced as we raise μ above unity. In other words, all angles are distorted towards a right angle; acute angles are made less acute, and obtuse angles less obtuse. (If the angle is a right angle the integrand is zero and the angle is undistorted.) It is also interesting to examine the limiting case as the same angle is moved towards the periphery in the world plane. By the preceding note μ (which is $1/g(r)$) tends to infinity as r tends to infinity. Consequently $\xi' = \tan^{-1}(\mu u)$ becomes a right angle. Near the boundary of the screen, all curves intersect a radial line at right angles.

(4) Images of Straight Lines

The equation of a straight line passing through the origin can be written in rectangular coordinates as $y = mx$ or in polar coordinates as $\theta = \theta^*$ where $m = \tan(\theta^*)$. Its image is therefore given by $\theta' = \theta^*$ which is a diameter of the screen circle.

Now consider a straight line $y = mx + c$ which is parallel to the line $y = mx$ (or $\theta = \theta^*$). In polar coordinates, the point $(x, mx+c)$ on this line becomes

$$r^2 = x^2 + (mx+c)^2$$
$$\theta = \tan^{-1}([mx+c]/x).$$

As x tends to $\pm\infty$, r tends to ∞ and θ tends to θ^*. Hence, r' tends to 1 and θ' tends to θ^*.

Thus the image of the line $y = mx + c$ meets the image of the line $y = mx$ at the edge of the screen.

In the world plane, drop a perpendicular from the origin to the line $y = mx + c$. The foot of this perpendicular is the closest point on this line to the origin. This perpendicular is a straight line passing through the origin; it is, therefore, mapped into a diameter of the screen circle. In the world plane, this line cuts the lines $y = mx$ and the line $y = mx + c$ at right angles. By the preceding note, these right angles remain right angles in the screen circle also. Consequently, at the points of intersection with this perpendicular diameter, the images of the line $y = mx + c$ and the line $y = mx$ are parallel. This point of intersection is also clearly the point of closest approach of the image of $y = mx + c$ to the origin. From this point of parallelism near the origin, the image of $y = mx + c$ must then bend to meet the image of $y = mx$ at the edge of the screen.

(5) Fitting into a Rectangular Screen

The simplest solution, as suggested in the text, is to transform the circle into an ellipse whose major and minor axes are equal to the width and height of the screen respectively. This is a very simple linear transformation. If we introduce rectangular coordinated with the X-axis along the longer side of the rectangular screen, we need only dilate the X-axis by the aspect ratio (typically 1.33).

Those who are not satisfied with this solution and insist on a better fit would perhaps find some solace in the theory of conformal mappings in complex analysis. The Riemann mapping theorem asserts that if D and D' are two simply connected domains other than the entire plane (or the entire plane with one point omitted), it is possible to map D into D' conformally by using an analytic function. In other words, if we treat the x and y coordinates as a single complex variable $z = x + iy$, then we can find an analytic function $f(z)$ such that f maps points z belonging to the domain D into points $f(z)$ belonging to the domain D′; moreover, angles (and, therefore, shapes of infinitesimal objects) are preserved by this mapping.

I am quite content with using an ellipse and wasting some pixels at the end of the screen, and do not intend to search for the optimal conformal mapping of the circle into a rectangular screen. I leave this to the ingenuity of the interested reader. Perhaps, somebody should build elliptical screens instead!

Since, we are on the subject of conformal mappings, I might as well state what was implicit in the statement of the Riemann mapping theorem above, viz., it is not possible to map the entire "world plane" conformally onto a rectangular screen or any part thereof. The reason is very simple : if it were so possible, the function $f(z)$ inducing that map would be everywhere bounded contradicting Liouville′s theorem. This is the principal reason why complex analysis and conformal mapping are not very useful in solving the original problem of peripheral vision. If conformal mappings are to be used at all as suggested by the preceding paragraph, it will be after the use of hyperbolic scaling to solve the main problem.

(6)Non Euclidian Geometry

Euclidian geometry is characterized by the parallel postulate : given a line L and a point P not lying on it, one and only one line parallel to L can be drawn through P. Denial of this postulate leads to non Euclidian geometry. Elliptic geometry is characterized by the impossibility of drawing any parallel, while hyperbolic geometry is characterized by the existence of more than one parallel. Two straight lines are said to be parallel if they do not meet however far they are produced, or alternatively if they meet only at "infinity". Straight lines are defined as the paths of shortest distance (technically known as geodesics); hence changing the measure of distance can change the notion of straightness, and can also change the nature of the geometry.

If we take a circle and choose the measure of distance in such a way that the boundary of the circle is infinitely remote from the centre, and yet the "straight lines" (geodesics) are the chords of the circle, then we obtain the Beltrami-Klein model of hyperbolic geometry. (Beltrami and Klein were able to prove that a distance measure could be constructed with the above mentioned properties; their measure of distance agrees with that induced by the hyperbolic mapping $r' = \tanh(r)$ for distances along a radial line but not for other distances.). In Fig. 4, it will be seen that, the lines CB and DA passing through P are both parallel to AB because they meet AB only at the boundary of the circle which is infinitely far away. The line EF does not intersect AB even at an infinite distance; in hyperbolic geometry such lines are said to be ultraparallel. Through the point P, exactly two parallels and an infinite number of ultraparallels can be drawn to a line AB.

It is interesting to observe that in Fig. 4, while BC approaches AB as we move towards B, it moves away from B as we move in the opposite direction. In hyperbolic geometry, it is NOT true that parallel lines are at the same distance from each other. The locus of all points equidistant from a given straight line is not a straight lines at all but a so called equidistant curve. This is one of the predictions of hyperbolic geometry which has found strong experimental confirmation in the case of human vision.

The other route to hyperbolic geometry mentioned in the text is to assume that the tangential scale is also equal to the radial scale. The radial scale $1-\tanh^2(r)$ can also be expressed as $1-r'^2$. If the scale is the same in all direction, distances can be defined in the screen circle by the formula

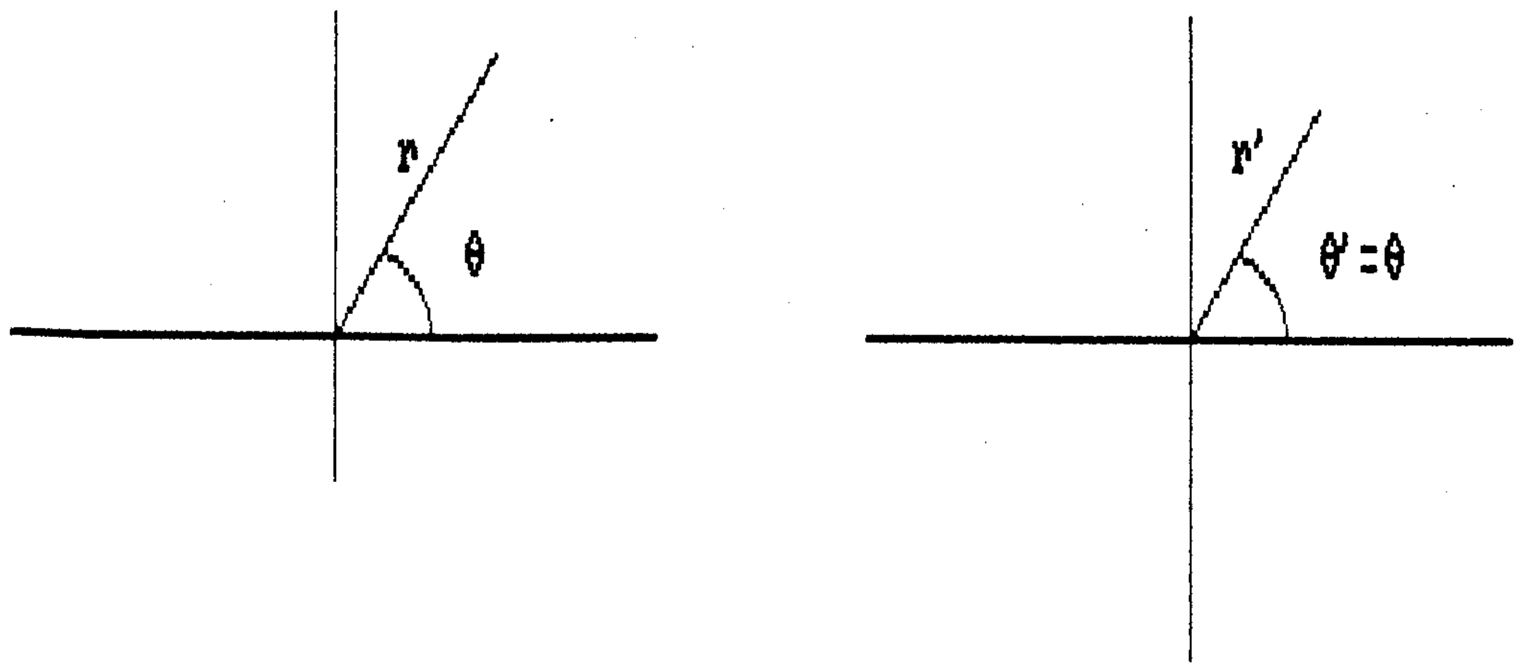
$$ds'^2 = (dr'^2 + r'^2 d\theta'^2)/(1-r'^2)^2$$

or in rectangular coordinates as

$$ds'^2 = (dx'^2 + dy'^2)/(1 - (x'^2+y'^2))^2$$

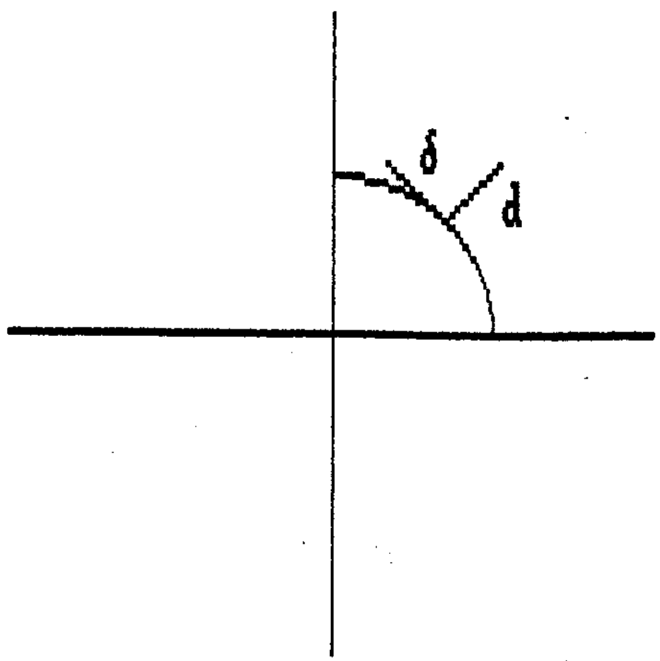
This is the Poincare model of non Euclidian geometry. With this measure of distance it can be proved that the "straight lines" are arcs of circles intersecting the screen circle at right angles; a limiting case of this is a diameter of the circle. Again, the geometry turns out to be non Euclidian.

Thus the two misinterpretations mentioned in the text – assuming straight lines in the screen to be straight, and assuming that the radial and tangential scales are equal – both lead to hyperbolic geometry. The former leads to the Beltrami-Klein model and the latter to the Poincare model.

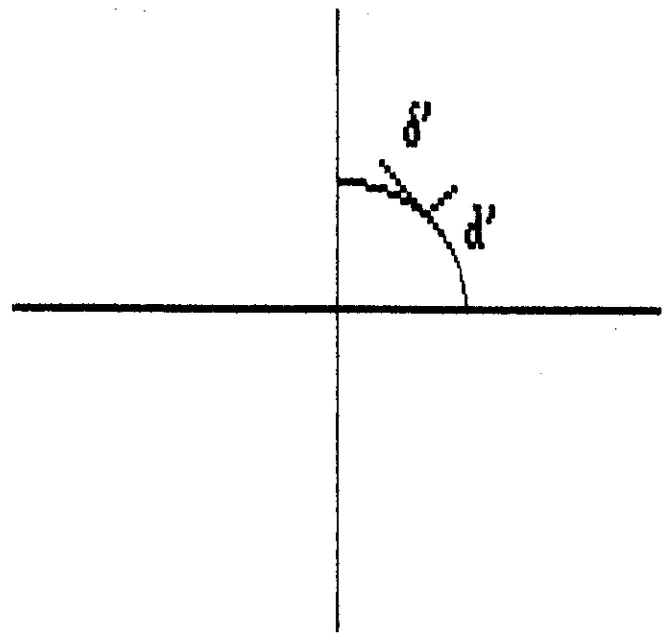


(a) The World (b) The Image

Fig. 1 Polar Coordinates in The World Plane and its Image in the Screen



(a) The World



(b) The Image

Fig. 2 Radial Scale (d'/d) and Tangential Scale (δ'/δ)

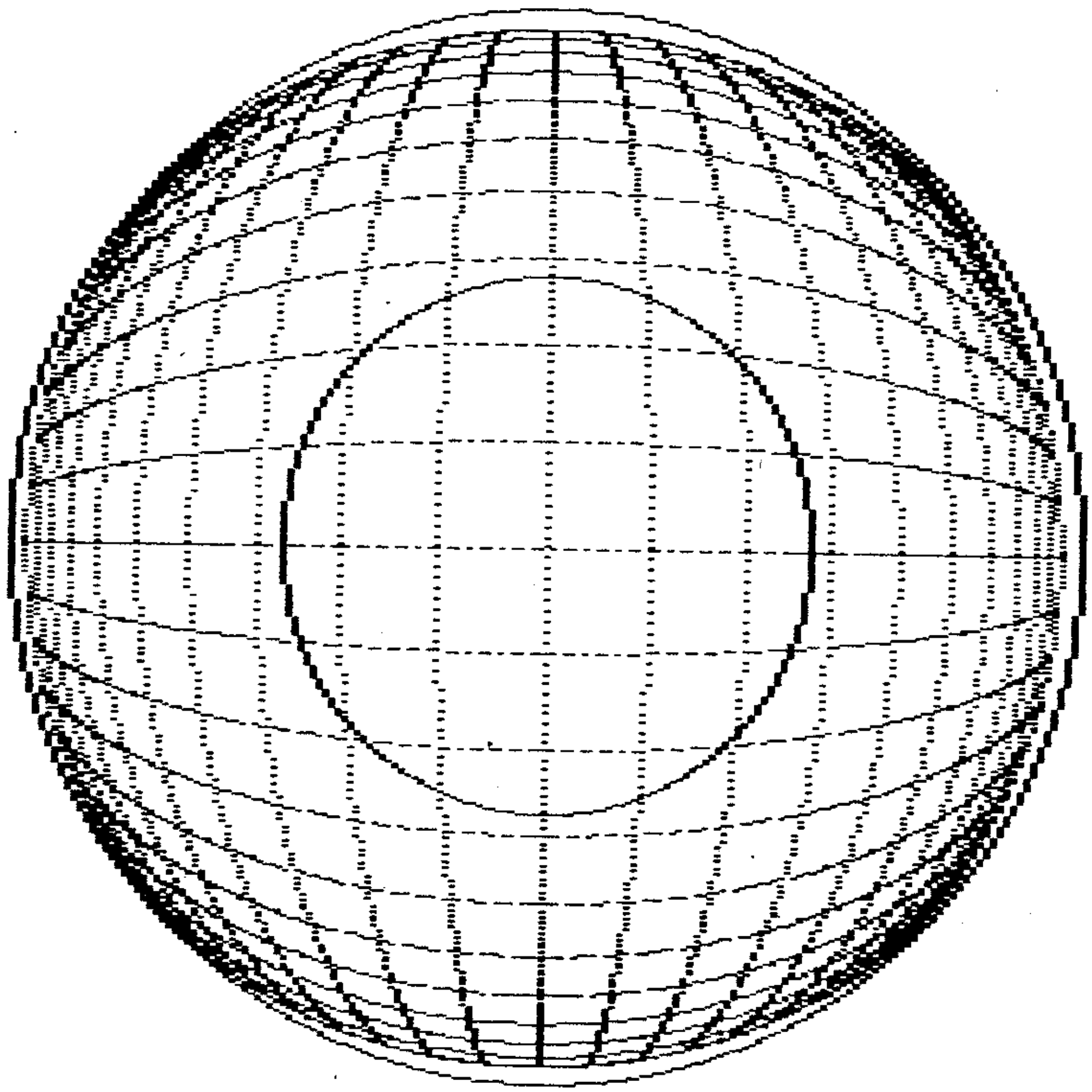


Fig. 3 Image in the Screen of a Uniform Square Grid in the World Plane.

1- Outer Circle is the Screen.

2- Inner Circle of Radius 1/2 is not distorted.

3- In the Periphery :

a) Straight lines are Curved

b) Angles are Distorted.

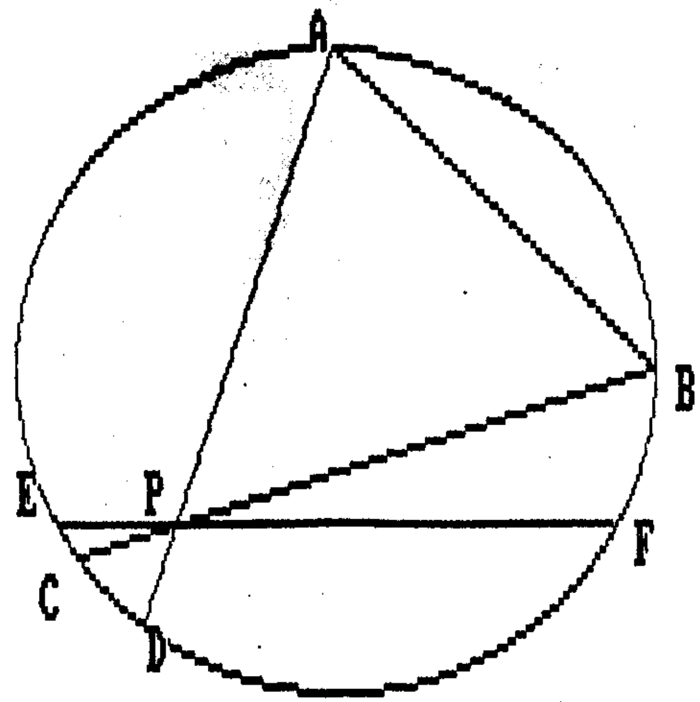


Fig 4   The Beltrami Klein Model of Hyperbolic Geometry
AD and BC are parallel and EF is ultraparallel to AB