Proc. Indian Acad. Sci. **A49** 251–261 (1959)

The optics of mirages

C V RAMAN
and
S PANCHARATNAM
(Memoir No. 115 from the Raman Research Institute, Bangalore-6)

Received May 6, 1959

1. Introduction

Chapter IV of the celebrated *Traite de la Lumiere* by Christiaan Huyghens published in the year 1690 deals with "the Refraction of the Air". Huyghens discussed the effects which arise by reason of the atmosphere having a refractive index varying with the height above the surface of the earth and also with the physical conditions. The problem was considered by him from two distinct points of view, viz., that light consists of rays travelling out from the source, alternatively of wave-fronts propagated through the medium; he showed that the two approaches give completely identical results. Each small piece of the wave-front in an inhomogeneous but isotropic medium moves forward normally to itself with the velocity appropriate to the refractive index at its position; the pieces thus moving forwards join up to form a continuous wave-front. As the result, we find that the rays of light traverse curved paths which are everywhere normal to the wave-fronts.

Mirages constitute one of the most remarkable effects arising from a variation of refractive index of the atmosphere. A very familiar type of mirage is that seen when the light rays are nearly parallel to the heated surface of the earth; the sky and objects on the horizon then appear to be reflected by the surface or rather by the cushion of hot air in contact with it. The phenomenon is frequently observed over asphalted or concrete roads on sunny days and over firm smooth sands in warm weather. Very striking pictures of such mirages are reproduced in a book by Minnaert.[1] Mirages can also be artificially reproduced in the laboratory. R W Wood[2] in his *Physical Optics* gives photographs of the phenomenon as thus obtained.

On an examination of the explanations usually given for the formation of mirages, it becomes apparent that they are inadequate and indeed unsatisfactory. The authors felt that the subject should be investigated more fully both from the theoretical and the experimental standpoints. The present memoir records the results of such study and throws new light on the nature and origin of mirages.

2. The nature of the problem

Consider two media of refractive indices μ_1 and μ_2 which we suppose in the first instance to have a sharp plane boundary of separation that is normal to the z-direction. If a plane wave from the first medium be incident on the boundary at a glancing angle ϕ_1 then, in general, there would exist both a reflected and a refracted wave; the former follows the usual law of reflection, while the direction ϕ_2 of the latter is given by Snell's law of refraction $\mu_1 \cos\phi_1 = \mu_2 \cos\phi_2$, the intensities of the reflected and refracted beams being given by Fresnel's formulae. If however $\mu_1 \cos\phi_1 > \mu_2$, there exists no refracted beam, so that the incident energy is totally reflected.

In the present paper, we are interested in the case in which the change from the refractive index μ_1 to the refractive index μ_2 does not occur abruptly at a sharp boundary, but takes place continuously and progressively over a certain transition layer—the refractive index being constant along planes normal to the z-direction. In this case, general considerations based on wave-optics show that the refracted wave in the second medium should still satisfy Snell's law of refraction $\mu_1 \cos\phi_1 = \mu_2 \cos\phi_2$ while, if the refractive index does not change appreciably within distances of the order of the wavelength, there would normally be no reflected wave in the first medium, the entire energy going into the refracted wave. However, an exception to this statement must clearly arise when $\mu_1 \cos\phi_1 > \mu_2$. Since in these circumstances, there can be no refracted wave in the second medium, a reflected wave must be formed, the reflection being in fact total. Such a situation occurs in the phenomenon of the mirage. Though it is clear that the reflected wave in the first medium must be a plane wave following the usual laws of reflection, the nature of the disturbance within the transition layer can only be ascertained by a rigorous analytical treatment from wave-theoretical considerations. This has been done by Epstein[3] choosing a particular profile for the refractive index variation and the solution obtained is expressed in terms of hypergeometric functions; however, the only application of Epstein's work which he has made and is relevant to the present problem is to remark that there is total reflection in the conditions mentioned, the angle of incidence being equal to the angle of reflection.

3. Some elementary considerations

The theory of the mirage which is usually accepted purports to base itself on geometrical optics. Any incident ray initially making an angle ϕ_1 with the plane of the stratifications would be progressively deviated from its course so that when it reaches a limiting layer of refractive index $\mu_l = \mu_1 \cos\phi_1$, it would be tangential to that layer. It is then usually stated that the ray continues to curve round, but it must be noted that on the basis of geometrical optics the ray should really continue parallel to the stratifications along the limiting layer. This would happen to every one of the rays in the incident beam. As a consequence, the entire energy of the incident radiation would be concentrated in an infinitely thin layer having the refractive index μ_l. If we take two adjacent rays a finite distance apart, the part of the wave-front between them would on entry into the transition layer swing round and at the same time contract in its extension and ultimately become reduced to a point travelling along the limiting layer of refractive index μ_l.

It is evident that the foregoing situation deduced from elementary considerations presents some fundamental difficulties. In the first place, as we have already noted, considerations of a general nature indicate that the energy reaching the limiting layer which cannot penetrate further must of necessity be returned back through the transitional region into the first medium as a reflected wave. Then, again, as has been stressed by Planck[4] in his *Theory of Light*, the equivalence of the results of geometrical optics with the exact results of wave theory is subject to certain limitations. One of them is that the amplitude of the wave-function must be only slowly variable in space. This condition is clearly violated in the present case, for according to what has been said above, the disturbance would be infinitely great at the limiting layer of refractive index μ_l and zero immediately below it. It follows that while the actual facts may bear some resemblance in their general features to this elementary picture, the real situation would nevertheless be rather different; firstly, a reflected wave must emerge from the limiting layer; secondly, there should exist a finite intensity below the limiting layer, while above that layer, the intensity though large would exhibit successive maxima and minima arising from the interference of the incident and reflected waves and progressively diminish as we move from the limiting layer toward the first medium.

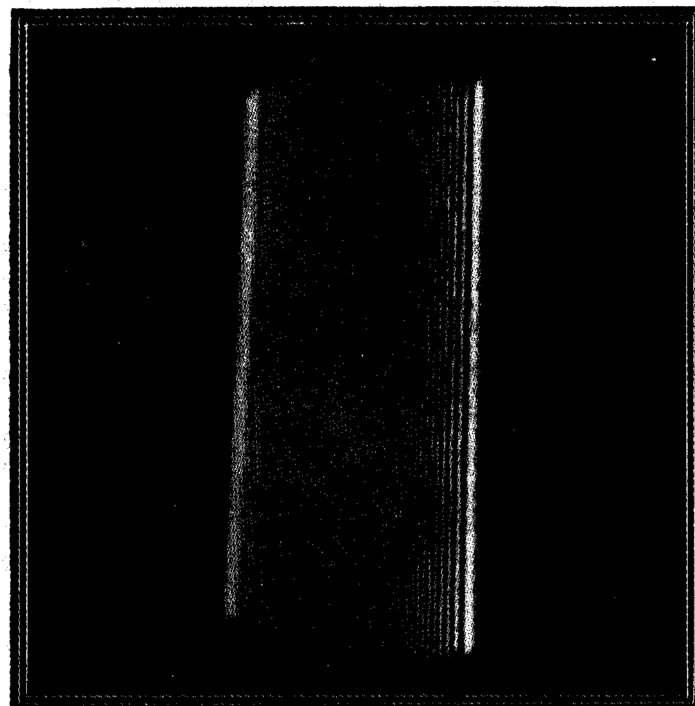
It is clear from the foregoing remarks that the situation in the vicinity of the limiting layer would resemble that well known in physical optics to appear in the vicinity of caustic surfaces formed by reflection or refraction of light. The effects observed in such cases are usually described as due to the propogation of a cusped wave-front along the caustic surface. Such a result is clearly capable of experimental test in the present case and has been confirmed by the observations made by us presently to be described.

4. Observation of the caustic and accompanying interferences

The arrangement first tried for producing the mirage under laboratory conditions was similar to that described by R W Wood. A long steel plate, approximately $1\frac{1}{2}$ metres long, 10 cm in width and 2 cm in thickness was supported horizontally and heated from below by four gas burners, the upper surface being covered with soot. The arrangement did not prove very satisfactory since the rising of the hot air from the heated surface prevented a sufficiently sharp temperature gradient from being established, while at the same time such phenomena as could be observed were of an exceedingly labile nature owing to the turbulent convection of air near the hot surface. However, the conditions were greatly improved when an electric fan was used to blow away the hot air from above the surface. When this was done, it was possible to see clearly the reflected image of the 'mountain peaks' formed by holding the serrated edge of a cardboard against an illumined ground glass screen at the far end of the plate. The action of the fan in sharpening the temperature gradient was picturesquely revealed by suitably lowering the eye, when a luminous thin cushion of air was seen to form over the heated surface. In order to study the phenomena critically, the object viewed was replaced by an illuminated slit kept parallel to the heated surface and the light diverging from it was rendered parallel by a collimating lens and allowed to fall obliquely on the hot plate. The beam was allowed to cover the whole length of the plate, the angle of incidence being adjusted merely by moving the slit. Further, the plate was heated electrically to ensure a fairly uniform temperature. When the temperature was sufficiently high and the glancing angle sufficiently small, the field at the nearer end of the plate, when examined through a pocket lens, displayed a bright strip of light separated by a dark region from the plane of the hot surface. It was clear, however, that the use of the fan was causing a smearing out of the caustic and preventing phenomena of the nature of interferences from being clearly seen.

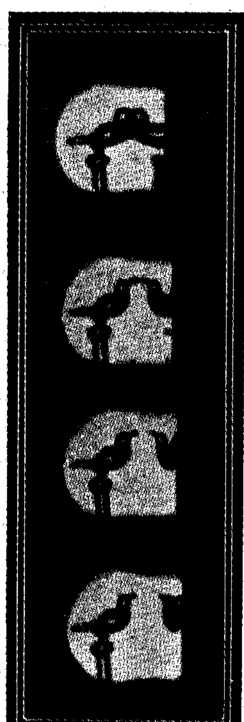
The above difficulty was removed by the artifice of turning the plate edgewise so that the hot surface was now vertical, though its length remained horizontal. The hot air now flowed up in streamlines parallel to the surface of the plate, thus rendering the use of the fan unnecessary. With the slit now vertical, it was possible to observe clearly with the aid of a pocket lens a bright caustic bordered by a large number of interferences of the type referred to in the last section. A photograph of this phenomenon is given in plate I, figure 1. In order to completely arrest the oscillations of the caustic and its fringes, a short exposure (1/100th of a second) was necessary so that the slit had to be brilliantly illuminated. Accordingly, sunlight was used for this purpose and this, in fact, enabled also the introduction of a red filter for increasing the clarity of the fringes.

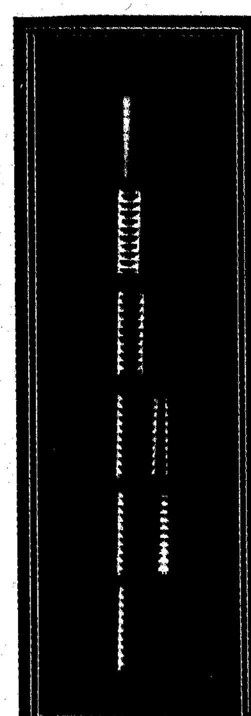
It will be seen that the field in plate I, figure 1 consists of three parts. To the right of the bright caustic (i.e., towards the heated surface) the field is dark, while to its left lies an illumined strip containing a large number of interference fringes whose separation narrows down to a constant value as we move away from the caustic; this is of course due to the increasing inclination between the two branches of the cusped wave-front. Since the heated plate is necessarily of finite extension, the reflected part of the cusped wave-front does not extend to infinity but is terminated. This manifests itself in the field of view by the occurrence of a second edge to the left of which the intensity is considerably less (though not zero), the edge being bordered by some broad fringes.



(1)

(2)





(3)

Plate I

5. Relation of the mirage to the cusped wave-front

The actual mirage is observed when the eye is kept at any point which lies on the bright strip of light lying to the left of the caustic, the eye being focussed on the plane containing the object, i.e., at infinity. It is to be expected that two images would then be seen whose positions lie respectively along the directions of the normals drawn from the eye to the two branches of the cusped wave-front leaving the nearer edge of the plate. In fact, the fringes observed to the left of the caustic in plate I, figure 1 may be regarded as due to the interference between the light from two such virtual sources, the progressive narrowing of the spacing of the fringes to the left of the caustic corresponding to the increasing separation of the sources. That the separation of the two images observed depends on the position of the eye or aperture through which the phenomenon is viewed is illustrated in plate I, figure 3. In order to make the nature of the image evident, the serrated edge of a hacksaw blade has been used to form one of the edges of the slit. An aperture was kept in front of the lens and the succession of photographs exhibit the alteration in the phenomena as the aperture is gradually moved to the left. When the aperture is on the bright caustic, a single image is seen, while as it is moved to the left this separates into two images, one of which is a direct (or more properly, a refracted) image, the second being an inverted reflected image, the separation between the two gradually increasing. A remarkable feature of the sequence of phenomena illustrated in plate I, figure 3 is the occurrence of a third erect image close to the reflected image in the fourth and fifth photographs of the sequence; this image starts developing when the aperture has been moved towards the outer edge of the central illumined strip (where broad fringes start appearing in plate I, figure 1), becoming coincident with the reflected image when the aperture is exactly at the edge mentioned. As the aperture is moved further left only the 'direct' image continues to be visible, as is shown in the last photograph of the sequence; this is to be expected since the reflected part of the cusped wave-front is no longer received through the aperture.

The third image mentioned above can be cut off by inserting an opaque screen near the farther end of the heated plate and adjusting it so that its edge protrudes a little beyond the surface of the plate. This shows that the image is due to the ordinary refraction of rays directly entering the edge of the heated layer at the farther end of the plate. The main features of the path of such rays may be deduced from the experimental observations described in the previous paragraph. The terminus of the reflected part of the cusped wave-front corresponds to certain limiting rays entering the region at the farther end of the plate. Rays which are able to enter the edge of the heated layer at a closer distance to the plate than these limiting rays proceed a longer distance before emerging from the heated stratum and also suffer a larger deviation. These rays give rise to the erect third image; in fact, the second refracted wave-front, obtained by drawing the surfaces orthogonal to these rays, meets the termination of the reflected part of the wave-front so as to form a second cusp.

It may be mentioned that an erect third image has been observed and photographed by Hiller[5] in his study of the mirage produced by a long vertical wall. He has attributed the phenomenon to multiple reflection, but it is more probably due to ordinary refraction as in the present case.

Till now we have dealt mainly with the case when the distant object is of negligible angular dimensions. When an object of finite angular dimensions is used, the point on the image which corresponds to any particular point on the object is to be determined as before for each setting of the eye. In this case there will be a distortion of the images because the position of the limiting layer as well as the inclinations of the cusped wave-fronts corresponding to any particular point on the object varies with the position of the object-point. Plate I, figure 2 shows the photographs taken using as the object a small model of a bird made of glass. This was placed near the focal point of the collimating lens. The sequence of photographs show the variation in the appearance of the phenomena as the eye is moved away from the plate. The appearance of a third erect image in addition to the usual reflected image may be discerned in the last two photographs of the sequence.

6. Mathematical treatment

Let the z-axis be taken normal to the direction of the stratifications of refractive index, the plane of incidence being taken as the xz plane. The disturbance $\psi \exp(iwt)$ at every point must satisfy the wave equation

$$\nabla^2\psi + k^2\mu^2\psi = 0.$$

Since the refractive index is a function only of z, we may put $\psi = \exp(ikpx)u(z)$ where u satisfies the one-dimensional wave equation

$$\frac{d^2u}{dz^2} + Q^2(z)u = 0, \tag{1}$$

where

$$Q(z) = k\sqrt{\mu^2 - p^2}. \tag{2}$$

We shall take the layer at which the refractive index becomes equal to p to be the plane $z = 0$, and the law of variation of refractive index to be given by

$$\left.\begin{array}{ll} \mu^2 = \mu_1^2 & \text{for} \quad z \geqslant z_1 \\ \mu^2 = \alpha^2 z + p^2 & \text{for} \quad z \leqslant z_1 \end{array}\right\}. \tag{3}$$

The general solution of the wave equation (1) for the region $z \geqslant z_1$, can be written

$$\begin{aligned} u = A\{&\exp - ik(\sqrt{\mu_1^2 - p^2 z} + \theta) \\ &+ R\exp ik(\sqrt{\mu_1^2 - p^2 z} + \theta)\}. \end{aligned} \tag{4}$$

If the incident wave be given by $A\exp - ik(\mu_1 \sin\phi_1 \cdot z + \theta)$, the first term will represent the incident wave with $p = \mu_1 \cos \phi_1$ so that from (3) the plane $z = 0$ is the limiting layer; the second term represents a reflected wave with the angle of reflection equal to the angle of incidence, R being the coefficient of reflection. We shall take R to be a real quantity, since this can be ensured by appropriately choosing the phase factor θ, the complex reflection coefficient being $R \exp(2i\theta)$.

Writing $k\alpha = 3\gamma/2$ the wave equation for the region $z \leqslant z_1$ can be written

$$\frac{d^2u}{dz^2} + \left(\frac{3}{2}\gamma\right)^2 zu = 0. \tag{5}$$

This is one of the transformations of Bessel's equation, the general solution being expressed in terms of the cylinder functions of order one-third:

$$u = z^{1/2} C_{1/3}(\gamma z^{3/2})$$

i.e., with the usual notation for the Bessel functions

$$u = \xi^{1/3}\{A_1 J_{1/3}(\xi) + B_1 J_{-1/3}(\xi)\} \tag{6}$$

where

$$\xi = \gamma z^{3/2}.$$

For negative values of z the above may be rewritten to avoid the use of imaginary arguments, by putting

$$\begin{aligned}
&\xi = i|\xi| = i\gamma|z|^{3/2} \\
&u = \xi^{1/3}\{A_1 I_{-1/3}(|\xi|) - B_1 I_{1/3}(|\xi|)\} \\
&\quad = \xi^{1/3}\left\{A_1 \frac{2}{\pi}\sin\frac{\pi}{3} K_{1/3}(|\xi|) + (A_1 - B_1) I_{1/3}(|\xi|)\right\}.
\end{aligned}$$

Since $I_{1/3}(|\xi|)$ tends to infinity as $z \to -\infty$, the condition that u remain finite below the limiting layer gives $A_1 = B_1$. Hence below the limiting layer $z = 0$, we have

$$u = \frac{\sqrt{3}}{\pi}|\xi|^{1/3} A_1 K_{1/3}(|\xi|) \tag{7}$$

while above the limiting layer

$$u = A_1 \xi^{1/3}\{J_{1/3}(\xi) + J_{-1/3}(\xi)\}. \tag{8}$$

For layers not close to the limiting layer, the asymptotic expansion of (8) for large z may be used:

$$\begin{aligned}
u &\sim \left(\frac{3}{2\pi}\right)^{1/2} \xi^{-1/6} A_1 \cdot 2\cos\left(\xi - \frac{\pi}{4}\right) \\
&\sim \left(\frac{3}{2\pi}\right)^{1/2} \xi^{-1/6} A_1 \left\{\exp - i\left(\xi - \frac{\pi}{4}\right) + \exp i\left(\xi - \frac{\pi}{4}\right)\right\}.
\end{aligned} \tag{9}$$

The constants A_1 and θ may be determined by the condition that the solutions (4) and (8) must join smoothly, i.e., by the condition that u and du/dz must be continuous at $z = z_1$. Since the solution u as given by (8) is a real quantity below the plane z_1, it is clear from (4) that the reflection coefficient R must be unity, as is to be expected. The constants A_1 and θ may be determined with sufficient accuracy by equating separately the amplitudes and arguments of the periodic functions in (4) and (9) since it is the periodic functions which contribute mainly to du/dz. Thus

$$A_1 \sim \left(\frac{2\pi}{3}\right)^{1/2} \xi_1^{1/6} A$$

$$\theta \sim \left(\xi_1 - \frac{\pi}{4}\right) - \mu_1 \sin\phi_1 z_1 \tag{10}$$

where $\xi_1 = \gamma z_1^{3/2}$.

Before proceeding to discuss the solution obtained, we may remark that the function ξ in terms of which the solution is expressed can be given a physical interpretation. Using the fact that $\xi = \int_0^z Q dz$, it may be easily verified that

$$[\xi(z) + kpy] = \int_0^P \mu ds\ .$$

where the right-hand side represents the optical distance from the point $P(x, z)$ to the limiting layer, measured along the limiting ray of geometrical optics. Thus in (9) the solution has been represented as the sum of two disturbances, the surface of constant phase of one of the wave-fronts being taken to be that given by geometrical optics; the other reflected wave-front is symmetrical to the first but has suffered a phase change of $\frac{1}{2}\pi$. The asymptotic solution (9) does not apply near the limiting layer as may also be seen by the fact that it gives an infinite intensity at that layer. The correct solution (8) applicable near the limiting layer may be written in a form similar to (9)

$$u = A_1 \xi^{1/3} \frac{J_{1/3}(\xi) + J_{-1/3}(\xi)}{2\cos(\xi - \frac{1}{4}\pi)} \{\exp - i(\xi - \tfrac{1}{4}\pi) + \exp i(\xi - \tfrac{1}{4}\pi)\}. \tag{11}$$

Thus, the foregoing mathematical treatment provides a justification for the qualitative description given earlier in the paper of the optical field as a cusped wave which moves along the limiting layer.

Using equations (7) and (8) and a table of Bessel functions, the function $(u/A_1)^2$ which is proportional to the intensity has been plotted against $|\xi|^{2/3}$ which is proportional to the distance z from the limiting layer. The graph is shown in figure 1 in the text. It is seen that slightly above the limiting layer there is a large concentration of intensity. Above this again we have a series of interferences which progressively diminish in spacing and intensity, while below the maximum, the intensity rapidly falls to zero.

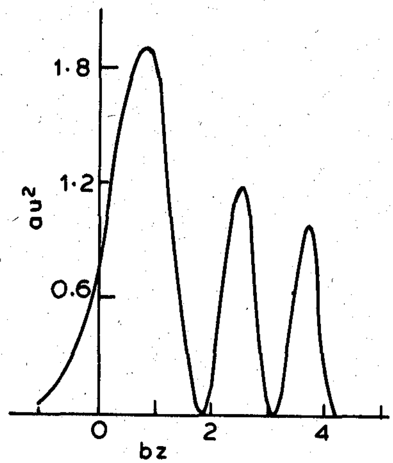


Figure 1

It will be seen that the method of solution adopted is similar to that adopted in quantum mechanics where similar mathematical problems arise. In fact, though we have for simplicity considered only the case when μ^2 varies linearly, a similar method[6] can be adopted even in the general case, the solution being again expressible as the sum of two Bessel functions if we neglect the square of the wavelength.

7. Summary

Experimental studies reveal that when a collimated pencil of light is incident obliquely on a heated plate in contact with air, the field of observation exhibits a dark region adjacent to the plate into which the incident radiation does not penetrate, followed by a layer in which there is an intense concentration of light and then again by a series of dark and bright bands of progressively diminishing intensity. Photographs of these features have been obtained and are reproduced in the paper. The observed facts find a satisfactory explanation when the problem is considered from the standpoint of wave-theory and more completely on the basis of a formal analytical treatment. The optical characters of the mirage are observed to stand in the closest relation to the wave-optical phenomena referred to, changing with the position of the observer's eye in the field of interference. The explanations of the mirage usually given are thus seen to be inadequate and unsatisfactory.

8. References

1. Minnaert M *Light and Colour in the Open Air*, G Bell and Sons, London, 1940, Plate V.
2. Wood R W *Physical Optics*, The Macmillan Co., New York, 1911, Plate III.
3. Epstein S *Proc. Natl. Acad. Sci.*, U.S.A., 1930, **16**, 628.
4. Planck M *Theory of Light*, Macmillan and Co., London, 1932, 204.
5. Hiller W *Phys. Z.*, 1913, **14**, 718; 1914, **15**, 303.
6. Morse and Feshbach *Methods of Theoretical Physics*, McGraw-Hill, 1953, Part II, 1096.