THE ORIGIN AND PROPAGATION OF ACOUSTIC-GRAVITY WAVES DUCTED IN THE THERMOSPHERE

By A. F. WICKERSHAM*

[Manuscript received March 13, 1968]

Summary

The theory of acoustic-gravity waves in a thermodinal atmosphere shows that damping decreases as the thermal gradient increases. If the thermal gradient is large enough it offsets the increase in amplitude caused by upward propagation into a region of decreasing ambient density and the motion is completely undamped, neglecting viscosity. Thus it appears that long distance propagation can occur in the principal duct at 150 km altitude, where the thermal gradient is almost large enough to eliminate nonviscous damping.

It is shown that thermal perturbation of the principal duct by the long winter-solstice sunrise in the polar regions is a source of ducted gravity waves consistent with the seasonal variations in directions of travelling ionospheric disturbances (TID's) observed in Australia.

The observed tilts of TID's are compared with theory and can be used to determine the period of a disturbance. The effect of ionospheric charge equilibration is discussed briefly, and a leaky-duct gravity wave model is proposed which reconciles the observations of tilted wave fronts and long distance propagation.

I. INTRODUCTION

Naturally occurring travelling ionospheric disturbances (TID's) have been detected for many decades, and several writers have suggested that they are propagating atmospheric acoustic-gravity waves (e.g. Martyn 1950; Hines 1960). For more than 10 years it has been known that TID's have fronts that are inclined in the direction of travel. Munro and Heisler (1956b) measured the tilt in 430 naturally occurring TID's and found it to be 65°, measured from the horizontal, for disturbances travelling in the direction of the geomagnetic meridian; however, they found that disturbances travelling at right angles to such direction also have tilted fronts, the tilt being about 54°. This probably means that the tilt is in the neutral disturbance itself and is not principally a result of charged particle interaction with the geomagnetic field. Other work by Munro and Heisler (1956a) shows that such tilted-front TID's propagate horizontally over large distances, and Munro (1958) has measured the seasonal variation in number and direction of such naturally occurring disturbances.

In the past few years it has been shown that there is quantitative agreement between the observed speed distributions of TID's and those expected theoretically for ducted gravity waves; also, there is agreement between the theoretical speeds expected for various ducted modes and the observed speeds of disturbances caused

* Radio Physics Laboratory, Stanford Research Institute, Menlo Park, California 94025, U.S.A.

by a nuclear detonation at Novaya Zemla (Wickersham 1964, 1965b, 1966). The theoretical speeds used for comparison with observations are the numerical results of Pfeffer and Zarichny (1963) and are based on a realistic atmosphere.

It is the present purpose to further examine the theory of acoustic-gravity waves in a thermoclinal atmosphere in order to attempt additional comparisons with observations. We shall see, from an approximate form of the equations of motion, that a principal duct exists in the thermosphere where minimal damping occurs because of the large thermal gradient. By analysing the consequences of a thermal perturbation of the duct in the thermosphere we find that winter sunrise in the polar regions is a source of ducted acoustic-gravity waves consistent with the seasonal changes in direction observed by Munro.

From the equations of motion appropriate for a thermoclinal atmosphere we can find an expression for the tilt of a gravity wave disturbance. If we assume that the duct in the thermosphere is leaky and that some energy flows up into the ionosphere, we can compare theoretical and observed tilts to determine the period of the portion of the disturbance that appears in the ionosphere. The leaky-duct hypothesis reconciles tilted fronts and horizontal ducting.

II. Dispersion Equations for an Atmosphere with a Linear Thermocline

For an atmosphere with a linear gradient in the scale height \( H = H_0(1+ax) \), the equation satisfied by the hydrodynamical divergence \( \chi \) is

\[
\chi'' + (H'-1)H^{-1}\chi' + (\omega^2/c^2 - K_z^2 + K_x^2 \omega_B/\omega)\chi = 0
\]  

(1) 

(Lamb 1945; Martyn 1950). In this notation \( \omega \) is the frequency, \( c \) the local speed of sound, \( K_x \) the horizontal wave number, which may be complex \( (K_x = k_x + i\beta_x, K_z \approx k_z^2 + i2k_z\beta_z) \), \( \omega_B \) the angular Brunt frequency, and primes denote differentiation with respect to \( z \), the vertical coordinate. The reciprocal of the Brunt frequency is sometimes called the thermobaric period,

\[
\tau_B = 2\pi(H/g)\{(\gamma-1)/\gamma+H'\}^{-1}.
\]  

(2) 

A graph of thermobaric period based on the 1961 CIRA standard atmosphere is shown in Figure 1.

In the present discussion we shall use an approximate form of equation (1) obtained by replacing \( H^{-1} \) by \( H_0^{-1} \) in the second term. With the understanding that the equation of motion now is approximately valid only in a region \( z_0 \pm \Delta z \), such that \( \Delta z \delta H/\delta z \ll H_0 \), we can seek solutions of the form \( \chi_0 \exp(-iK_zz) \), where \( K_z \) is complex, \( K_z = k_z + i\beta_z \). With the assumption that \( \beta_z \) is small, so that \( K_z^2 \approx k_z^2 + i2k_z\beta_z \), we obtain the dispersion equations

\[
\omega^4/c^2 - \omega^2[k_x^2 + \beta_x(x-H_0^{-1})] + k_x^2 \omega_B^2 = 0,
\]  

(3) 

\[
\beta_x = \frac{\tau_B}{\tau^2 - \tau_B^2} k_x \left( \beta_x + \frac{1}{2}x - \frac{1}{2}H_0^{-1} \right),
\]  

(4)
where \( k^2 = k_x^2 + k_z^2 \). Introducing \( \cot \phi = k_z/k_x \) into (3), we find a useful expression for the angle \( \phi \) that a wave front makes with the horizontal, namely,

\[
\cot^2 \phi = \frac{\tau^2}{\tau_B^2} + \frac{\lambda_x^2}{c^2 T^2} \left( 1 + \frac{\beta_z^2 c^2 T^2}{4\pi^2} (\alpha - H_0^{-1}) \right),
\]

where \( \lambda_x \) is the horizontal wavelength. We also can use equation (3) to compute the direction of energy flow, \(-\delta k_x/\delta k_z\), and find that it is given by the same angle \( \phi \). In other words, the direction of energy flow is in the plane of the phase front, a result pointed out by Hines (1960) for an isothermal model atmosphere.

We see from equation (4) that if the conditions \( \alpha = H_0^{-1} \) and \( \beta_z = 0 \) prevail then \( \beta_x \) vanishes and the motion is undamped. From the definition of linear variation of scale height, \( H = H_0 (1 + \alpha z) \), the first condition can be expressed as \( \partial H/\partial z = 1 \). On examining these conditions, we shall find that the first implies the second; further, the first is nearly satisfied in the normal atmosphere at about 150 km altitude.

The energy in a gravity wave is partitioned equally into a kinetic form, \( \frac{1}{2} \rho_0 u^2 \), and a potential form, the latter being further partitioned into elastic energy, \( \frac{1}{2} \rho_0 c_0^2 \xi \), where \( \xi \) is the condensation \( \delta \rho_\epsilon/\rho \), and into thermobaric energy (e.g. Eckart 1960). The thermobaric energy is the product of the average thermobaric force during a quarter-cycle of the motion and the vertical displacement \( W_t = \frac{1}{2} (-\rho_\epsilon^2 \partial z \partial z) \partial z \) (Eckart 1960). From the definition of scale height we can define a thermobaric condensation

\[
\dot{\xi}_t = -\frac{\delta z}{H} \left( \frac{\gamma - 1}{\gamma} + \frac{\partial H}{\partial z} \right),
\]

so that the thermobaric energy can be written as

\[
\frac{1}{2} \rho_0 \omega_B^2 \left( \frac{\gamma - 1}{\gamma} + \frac{\partial H}{\partial z} \right)^{-2} H^2 \xi_t^2.
\]

With the help of equation (2) and the relation \( c^2 = \gamma g H \), we obtain

\[
W_t = \frac{1}{2} \rho_0 \frac{1}{\gamma} \left( \frac{\gamma - 1}{\gamma} + \frac{\partial H}{\partial z} \right)^{-1} c^2 \xi_t^2,
\]
and the total energy in a gravity wave is

\[ E = \frac{1}{2} \rho_0 u^2 + \frac{1}{2} \rho_0 c^2 \xi^2 + \frac{1}{2} \rho_0 \left( \frac{\gamma - 1}{\gamma} H' \right)^{-1} c^2 \xi^2. \]  

(7)

Since \( E = \rho_0 u^2 \), we find from equation (7)

\[ u = c \xi, \]  

(8)

where

\[ \xi^2 = \frac{1}{\gamma} \left( \frac{\gamma - 1}{\gamma} + \frac{\partial H}{\partial z} \right)^{-1} \xi_l^2 + \xi_e^2. \]

The energy carried across a unit area must be conserved; that is, \( \rho u^2 c \) is constant or, by virtue of equation (8), \( \rho c^2 |u_z| \xi \) is constant. If now we ask that the product of the kinetic and elastic amplitudes be constant or independent of altitude, then we must have \( \partial(\rho c^2)/\partial z = 0 \), or

\[ \frac{1}{c^2} \frac{\partial c^2}{\partial z} + \frac{1}{\rho} \frac{\partial \rho}{\partial z} = 0. \]  

(9)

From equation (9), the perfect gas law, and the definition of scale height, it follows that \( H^{-1} + H^{-1} \partial H/\partial z = 0 \), or \( \partial H/\partial z = 1 \). Thus we see that the motion is undamped when the thermal gradient is large enough to offset the growth in amplitude that results from decrease in ambient density with altitude (equation (9)). If amplitude is independent of altitude, then \( \beta_z = 0 \) and it follows that \( \beta_x = 0 \) and the motion is undamped, neglecting viscosity.

The condition \( \partial H/\partial z = 1 \) is almost satisfied (\( \partial H/\partial z \approx 0.7 \)) from about 150 to 160 km altitude in the 1961 CIRA standard atmosphere. This altitude range is part of a gravity wave duct that has boundaries at about 100 and 165 km (Wickersham 1965a) and, because it is almost free of nonviscous damping, we conclude that this is a major duct for the long distance propagation of gravity waves that perturb the lower ionosphere.

III. A Source of Ducted Acoustic-gravity Waves

From Figure 1 we see that in the principal thermospheric duct the thermobaric period varies from 270 sec at the lower boundary near 100 km to 650 sec at the upper boundary near 160 km. Since gravity waves with short periods, say 270–430 sec, would be confined to the lower half of the duct and suffer nonviscous damping, we shall confine our attention to the upper portion of the duct, 150–160 km, where the period varies from \( \tau_{\text{min}} = 430 \) sec to \( \tau_{\text{max}} = 650 \) sec and has a mean or central value of about \( \tau_0 = 540 \) sec or 9 min. We shall regard this altitude region as an ensemble of thermobaric oscillators having an approximately uniform density in frequencies and subject to thermal perturbations varying in time. Beyond the altitude interval of interest we shall consider the thermobaric period to be constant, as indicated in the idealized distribution shown in Figure 2.

A major thermal perturbation must occur diurnally when dawn, first at 160 km and then down to 150 km, increases the temperature and local scale height, causing
an atmospheric heave or vertical displacement. If \( z \) is the vertical displacement within the region, the corresponding thermobaric force is

\[-\rho \omega_B^2 z \sim -\rho \omega_B (\partial z / \partial H) (\partial H / \partial t) \delta t,\]

the \( \partial H / \partial t \) factor being proportional to the time rate of change in temperature. We shall neglect equilibration time lag and represent this perturbing force in time by a simple square pulse, starting at dawn at 160 km and ending at dawn at 150 km (Fig. 3). The width of the pulse varies from about 1·6 min in the equatorial regions to 1 hr 20 min for the winter solstice dawn nearest the pole.

The square pulse shown in Figure 3 can be regarded as a superposition of up and down Heaviside step functions. The second step function will excite an oscillator with a phase opposite to that of the first; therefore, maximum excitation occurs when the pulse width is equal to half the period of the oscillator. For the central oscillator this would be a pulse width of 4·5 min. A time difference of 4·5 min corresponds to 1·125° of rotation of the Earth. The segment between dawn lines at 160 and 150 km altitude that includes 1·125° of polar arc is shown in Figure 4, a projection of the dawn lines onto a plane tangent to the South Pole. The segment is about 6° north of the Antarctic Circle. A line drawn from the segment and normal
to the dawn line indicates the most probable directions of propagation of disturbances excited in the duct. Extended on a great circle to Sydney, Australia, the line indicates a direction of arrival there of about 30° east of north.

We see that during the Australian winter, when the dawn line near the Antarctic Circle is at its closest approach to Australia, we should expect ducted gravity waves to arrive in Australia from the south-east. Conversely, during the Australian summer, disturbances should arrive from the north-west from the winter dawn line near the Arctic Circle. These indeed are the seasonal characteristics observed by Munro (1958, Fig. 11).

To obtain a quantitative prediction of the observed angular spectrum, we note that the amplitude, or response, of a pulse-excited oscillator is proportional to its bandwidth. The bandwidth of the central oscillator is the largest and is approximately $B_0 \approx \tau_0(\tau_{\text{max}} - \tau_{\text{min}})/\tau_{\text{max}} \tau_{\text{min}}$. For an oscillator displaced $\pm \Delta \tau$ sec in period from the central period, the relative bandwidth is approximately $B \approx B_0 - 2\Delta \tau/\tau_0$, from which we have

$$\Delta B \approx 2\Delta \tau/\tau_0.$$  \hspace{1cm} (10)

At points along the dawn line to the left or right of the 4·5 min dawn interval, there are dawn intervals shorter or longer than 4·5 min. These pulses will excite oscillators that have bandwidths less than that of the central oscillator; thus, their response will be less than that of the central oscillator. According to equation (10), the response decreases linearly as the time interval becomes less than or greater than $\tau_0$.

If we assume that probability of occurrence is proportional to response, we can associate a relative probability with each directional normal on the dawn line and obtain the relative angular spectrum shown in Figure 5. In the same figure we show a histogram of Munro's 5-year average of observations in the month of June. The two spectra are in agreement. A mirror image of Figure 4 serves for the northern hemisphere and can be used to predict the spectrum observed during the Australian summer; however, the source in this case is now more than twice as far removed from Australia, and, as might be expected, the observed spectrum has somewhat more angular spread than predicted. During the equinoxes both spectra appear to be observed.

IV. THE TILT AND PERIOD OF TID's

The morphology of TID's as deduced from ionosonde data by Munro and Heisler appears to be confirmed by the incoherent scattering measurements recently presented by Thome (1964). The electron density perturbation caused by a disturbance as
measured by Thome is reproduced here in Figure 6. Unfortunately, Thome did not report the speed of the disturbance but gave only a range of 50-100 m/sec as typical of most disturbances. If we assume a horizontal speed of 57 m/sec we obtain the isodensity contours in spatial coordinates, as shown in Figure 7. We shall consider this to be a representative TID since the 45° slope near the centre is almost as large as the mean values of 51° to 65° observed by Munro and Heisler.

![Figure 6](image)

**Fig. 6.—Charge density contours associated with a TID detected by the incoherent scattering technique (from Thome 1964).**

![Figure 7](image)

**Fig. 7.—Charge density contours (C), associated with the TID observed by Thome (1964), transformed to spatial coordinates by assuming a horizontal propagation speed of 57 m/sec. Ambient electron density (D), from Thome, is shown at left; an energy density profile (E) for a ducted gravity wave, from Pfeffer and Zarichny (1963), is shown at right. At bottom centre there is an assumed cosine-squared antenna radiation pattern (A).**

From equation (5), since the last term is small, we see that when the wave front is vertical the period of the disturbance must be \( \tau \approx \tau_B \). From either Figure 6 or Figure 7 we see that the leading edge of the disturbance is vertical just above 300 km altitude, and thus from Figure 1 we conclude that the period of the disturbance is about 705 sec or 11·7 min. This conclusion is independent of the speed assumed for the disturbance.

At the centre of the disturbance, near 225 km, the slope of an 11·7 min gravity wave should be, by equation (5), about 74°. This is somewhat greater than the average upper limit of 65° measured by Munro and Heisler and considerably greater than the slope of the TID shown in Figure 7. The slope near the centre of the TID shown in Figure 7 is, of course, dependent on the speed assumed for the disturbance. We can obtain agreement with Munro and Heisler’s upper value of 65° if we assume \( \beta_z \) small and \( \lambda_x = 200 \) km. To obtain agreement with the centre slope of the TID in Figure 3 we should have to assume \( \lambda_x \geq 445 \) km. So large a value of horizontal wavelength would be difficult to reconcile with our first determination of an 11·7 min
period and the theoretical and observed propagation speeds of TID's. The difficulty increases at still lower altitudes. Even with a horizontal wavelength as large as 445 km we cannot reconcile theory and observation at 160 km altitude. From equation (5) we should predict a slope of 38° where the observed slope is zero. It is possible that the observed slope is zero at 160 km because the disturbance does not extend to lower altitudes. However, if this were so, then the disturbance would consist of an oblique rising of energy from no apparent source at 160 km altitude to an ultimate dissipation at 300 km altitude, an interpretation difficult to reconcile with long-range horizontal propagation. We find a more satisfactory model by assuming that the neutral disturbance extends to lower levels and that the disagreements between theoretical and observed slopes at the lower edges derive from the effects of ionospheric chemistry.

It is possible that the trailing edge of the disturbance is more a measure of ionospheric chemical equilibration time rather than the duration or width of the neutral wave. We note that above 180 km altitude the ionospheric electron reservoir is controlled by the charge-exchange reaction $O^{+} + N_2 \rightarrow NO^+ + N$. Using the atmospheric model and reaction-rate compilation of Ferguson et al. (1965), and assuming a linear, positive temperature dependence for the reaction rate, we find that the effective time constant for the above reaction is about 500 sec at 270 km altitude. This is also the time interval between the 6\% or half-amplitude levels at 270 km altitude for the TID shown in Figures 6 and 7. Just below 170 km altitude the reaction $NO^+ + e^- \rightarrow N + O$ controls the electron reservoir; however, at this altitude both reactions have effective time constants less than 20 sec. Because of the rapid equilibration, the charge density no longer reflects the neutral particle perturbation and the TID is almost unobservable by radio means. This expectation appears to be confirmed by the data shown in Figure 6 as well as by other data shown by Thome. If chemical equilibration time determines the duration of the charge perturbation caused by a TID then the slope of the electron density perturbation will be less than that of the neutral density perturbation, especially at the lower altitudes, and at 170 km the almost zero slope may be a result of fast equilibration.

From the above brief consideration we see a difficulty in determining the period from the slope; a difficulty, however, that may ultimately prove useful in studying ionospheric chemistry. We shall see also that it is difficult to determine the period on the basis of the time during which the charge perturbation interacts with radio equipment.

Curve E of Figure 7 shows the relative kinetic energy profile for a 13 min period, fully ducted, third-order gravity mode computed by Pfeffer and Zarichny. This profile is typical of many of the gravity modes in that some energy appears in the $F$ region but the centroid is between 100 and 165 km altitude. Curve D of this figure shows the ambient electron density measured by Thome at about the time of the TID. We see that the charge density perturbation can be expected to be considerably higher than the maximum disturbance in the neutral particles; thus, it is possible that an observed $F$ region charge perturbation is only the wake of a stronger disturbance in a lower region.
In Figure 7 we also show a cosine-squared radiation pattern (A) which is typical of many sweep-frequency ionosondes. The half-power points of such radiation patterns are at 45°, and these are shown as broken lines in the figure. With such sounders one can detect disturbances up to the maximum in the electron density profile (\( \sim 250 \text{ km altitude} \)). The sensitivity of most vertical incidence sounders is greatest in the shaded region shown in the figure. The average charged particle perturbation caused by TID's, as measured by Munro and Heisler, is 20% of equilibrium, with maximum perturbations as large as 30%. A TID that causes a 20% charge perturbation possibly could affect the edge of the antenna radiation pattern and continue to do so as it moves through the shaded region. A disturbance similar to that shown in Figure 7, moving at a speed of 57 m/sec, would interact with the sounder for 130 min. It is possible that the first half of such an interaction could be observed and a "period" ascribed to the disturbance of about 65 min. Some observers, however, would take the time interval between this disturbance and the next one to be the period. We see that interaction times might be longer than the periods we have discussed earlier and should not necessarily be taken as a measure of the period of a TID.

V. A Leaky-duct Model of TID's

We now have a large number of theoretical and experimental restrictions and can attempt to construct a gravity wave model of TID's. A model should be consistent with all of the following information.

1. TID's have fronts inclined in the direction of travel.
2. The slopes of TID's are nearly independent of the direction of the geomagnetic field.
3. TID's propagate large distances with little attenuation.
4. The speed distributions of TID's agree with those predicted theoretically for ducted acoustic-gravity waves.
5. The effective time constant of the reaction 
   \[ \text{O}^+ + \text{N}_2 \rightarrow \text{NO}^+ + \text{N} \]
   corresponds to the observed half-width of a TID observed by Thome.
6. Most of the centroids of the energy profiles of ionospheric, long-period, ducted gravity waves, as computed numerically by Pfeffer and Zarichny, lie in the strong thermoclineal channel between 100 and 165 km altitude.
7. Fast chemical equilibration in the 100–165 km altitude region would prevent observation of long-period neutral particle disturbances by means of radio detection of charged particle perturbations.
8. Fast equilibration in the lower ionosphere appears to be consistent with the fact that observed slopes of TID's are less than can be predicted by gravity wave theory.

To obtain a model consistent with the above information we turn to a leaky-duct model somewhat similar to one suggested originally by Hines (1960). We assume a model in which a wave packet, or pulse, of gravity waves is ducted in the thermoclineal channel between 100 and 165 km altitude. The longer period components of
the wave packet leak through the upper boundary of the channel—the knee in the thermobaric period profile of Figure 1—to provide detectable perturbations of the upper ionospheric charge density.

This model is analogous to the bow wave caused by a ship. To an observer in a nearby rowboat, the bow wave appears to move normal to its wave front. To an observer in an aeroplane, the wake of the bow wave has an apparent velocity equal to and in the direction of that of the ship. Thus the horizontal speed of an unseen disturbance, ducted in the channel at 100–165 km altitude, can be determined from measurements of the wake detectable in the upper ionosphere; however, the wake itself has neither the morphology nor direction of motion of a ducted disturbance. A principal difference between the model and the analogy is the forward slope of the energy leaking through the top of the duct, as contrasted with the backward slope of a ship’s bow wave. This effect is peculiar to gravity waves. Energy propagates upward in the plane of the downward-sloping wave front.

Note added in proof. The conditions under which a plane wave solution is appropriate to the equation of motion are more general than stated in the text. If the nonviscous damping is small as a result of the condition \( \partial H / \partial z \approx 1 \) being satisfied, then the coefficient of the second term of the equation of motion, equation (1), vanishes and plane wave solutions are valid for the remaining equation.

If \( \partial H / \partial z < 1 \), there is nonviscous damping as a result of amplitude increase with increasing altitude. In this case the amplitude increase is due to decreasing ambient density. If \( \partial H / \partial z > 1 \), there is nonviscous damping as a result of amplitude increase with decreasing altitude. In this case the amplitude increase is due to decreasing sound speed, or temperature, in the ambient medium. The condition \( \partial H / \partial z > 1 \) is not an instability condition for a horizontally growing wave.

VI. References


