Radio based navigation techniques

When navigating, we can also use a certain type of electromagnetic radiation, known as radio signals. In this chapter, we'll examine how this works. First we examine some basic principles. Then we examine certain types of beacons. And after that, we look at how we can use space based signals for navigation.

1 Radio navigation basics

1.1 Information carrying waves

Radio navigation makes use of **electro-magnetic waves**. These waves travel at the speed of light $c = 2.998 \cdot 10^8 m/s$. Their wavelength λ and frequency f satisfy $c = \lambda f$. To send a signal, we usually take a **wire/antenna** with length $L = \lambda/2$ and put an alternating current on it with the corresponding frequency.

But by just emitting a wave, we don't send any information. We need to be coding information on the radio waves. This principle is called **modulation**. When applying modulation, we modify a (high-frequency sinusoidal) **carrier wave**. There are three important types of modulation.

- **On-off modulation** – When the data signal is 1, we emit the signal. And when the data signal is 0, we emit nothing. The signal thus consists of pulses.
- **Amplitude modulation** (AM) – The amplitude of the carrier wave is proportional to the data signal. We thus have

$$V(t) = V_M(t)\cos(\omega_c t + \phi), \qquad \text{where} \qquad V_M(t) = V_0(1 + m\cos\omega_s t). \tag{1.1}$$

  In this equation, $V_M(t)$ is the **data signal**, ω_s is the **modulating signal frequency**, ω_c is the **carrier frequency** and m is the **depth in modulation**, satisfying $0 \leq m \leq 1$.
- **Frequency modulation** (FM) – The frequency of the carrier wave is proportional to the data signal. We now have

$$V(t) = V_0\cos(\omega_c t + \Delta\omega\sin(\omega_s t)\, t). \tag{1.2}$$

Radio signals can be categorized, based on their carrier frequencies. We have **very low frequency** (VLF), **low frequency** (LF), **medium frequency** (MF), **high frequency** (HF), **very high frequency** (VHF), **ultra high frequency** (UHF), **super high frequency** (SHF) and **extremely high frequency** (EHF). The frequency for VLF ranges from 3 to $30kHz$. The frequency of LF ranges from 30 to $300kHz$. This continues to increase with a factor 10 every step until we reach EHF, ranging from 30 to $300GHz$.

1.2 Propagation of radio waves

Normally, electro-magnetic waves simply travel in a straight line. Therefore, they only allow so-called **line-of-sight navigation**: you can only navigate using beacons you can actually see. It is now important to know over what distance we can see beacons. This **range** R can be found, using the equation

$$R = 1.2\sqrt{h_T} + 1.2\sqrt{h_R}. \tag{1.3}$$

In this equation, R is the range in nautical miles, h_T is the transmitter height in feet and h_R is t he receiver height in feet. (These units are important. The above equation won't work for other units.) For flying aircraft, the range is often roughly 200 nautical miles.

The downside of line-of-sight navigation is its limited range. However, there are waves that don't travel in a straight line. For frequencies up to $3MHz$ (HF), radio waves tend to follow the curvature of the Earth. These **ground waves** thus enable long-range radio-navigation systems, like the **Long Range Navigation** (LORAN) system. The downside of ground waves is that they quickly lose power with increasing distances.

For frequencies up to $30MHz$ (VHF), there are also **sky waves**. This time, the signals are 'reflected' by the ionosphere of the Earth. Since sky waves need to travel through the atmosphere first, they won't work on short distances. Only on relatively long distances are sky waves useful.

1.3 Multipath effects

Let's suppose we send out a low-frequency signal. This signal can now reach its destination along various paths: as a ground wave and as a sky wave. This means that **multipath effects** may occur. The two signals have travelled a different distance, and thus may have a phase difference. This can cause the signal to be **amplified**, or it may be **fading** away.

But this isn't the only situation where multipath effects occur. Let's suppose we have a transmitter at a height h_T. When sending out a signal, this signal can reach a nearby receiver directly. However, it can also reach the receiver after being reflected by the ground. The result of this is that a sort of **lobing** pattern occurs around the transmitter. In some lobes, the signal is amplified by the multipath effects. But in other lobes, the signal fades away.

Of course, it's inconvenient if we can't receive a signal at certain places. How can we solve the lobing problem? One solution is to raise the height h_T of the transmitter. This will result in a lobing pattern with a lot of thin lobes. So, we will never be without a signal for long. Another possibility is to apply **counterpoising**. This is a fancy word for placing a plate under the transmitter, to make it seem that the transmitter is placed on the ground. This will result in one very big lobe in which the signal can be received.

1.4 Navigation system types

Now let's try to use radio signals to navigate our airplane. How can we find our position? One way is to just look in which direction there are beacons. This will give us the **radial** θ of the corresponding beacons. (Beacons that give you the radial are called **θ-systems**. An example is the **VHF omnidirectional range** (VOR) beacon.) To find our position in 2D in this way, we need at least two such beacons.

Another type of beacons are **ρ-systems**. These beacons give you the distance ρ with respect to that beacon. An example is the **Distance Measuring Equipment** (DME) beacon. This time, two beacons can't unambiguously say where you are. You thus need at least three beacons to find your position.

We can also combine a θ-system with a ρ-system. This **ρ-θ-system** then tells us the radial and the distance with respect to the beacon. Examples now include the co-located VOR/DME beacon or the **TACAN beacon**. Now, only one beacon is sufficient to tell us our position.

Finally, there are **hyperbolic systems**, like the LORAN system. With hyperbolic systems, you measure the time difference between two signals arriving from two beacons. This data then results in a hyperbolic line on which your position must lie. You now need at least two beacons (and sometimes even three) to find your position.

1.5 Geometric dilution of precision

All the systems treated above are called **line of position** systems. A beacon alone can't tell you your position, but only a 'line' on which your position must be. In reality, the system can't even tell you that.

And this is because accuracy needs to be taken into account.

Let's suppose that we're trying to find our position using two θ-systems. These systems give us the radials θ_1 and θ_2. However, these radials aren't exactly accurate. They can be off by (for example) $0.5°$ on either side. We now don't get a point anymore where our position must be. Instead, we get a region. This is displayed in figure 1.

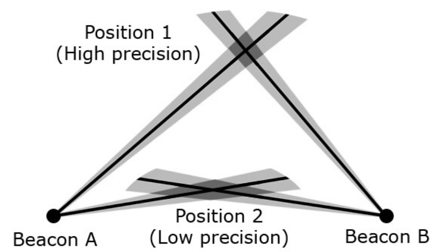

Figure 1: Geometric dilution of precision (GDOP), shown in action.

This region, however, has a special property. Its size depends on our position. If we are placed inconveniently with respect to the beacons, then the region is quite big. This 'dependence' of the precision, based on our position, is called **geometric dilution of precision** (GDOP). The GDOP is smallest when the aircraft is positioned on two perpendicular lines of position.

2 Navigation systems

It's time examine some navigation systems, and take a look at how they work.

2.1 Non-directional beacons and automatic direction finders

A **non-directional beacon** (NDB) is a beacon on the ground, sending a signal in all directions. To use it, we need to equip our aircraft with an **automatic direction finder** (ADF). This ADF has two types of antennas.

- The **loop antenna** has the shape of a loop with two arms AB and CD. When the loop is aligned with the direction of the signal, the arms AB and CD will get the same excitation: nothing happens. This is the **null position** of the loop antenna. On the other hand, when the loop is placed facing the signal, then AB gets a (for example) positive current, while CD gets a negative. Thus, a voltage is created in the loop. This voltage has maximal value V_{max} when the loop exactly faces the direction of the signal.
- The **sense antenna** is a lot simpler. It is omni-directional. It always measures the maximum voltage V_{max}.

The loop antenna alone can't detect the direction: it has a $180°$ ambiguity. But, by adding up the loop antenna voltage V_1 and the sense antenna voltage V_2, we get a **cardioid** pattern. Now, the ADF can find the direction of the beacon. The ADF thus yields a bearing to an NDB ground station.

2.2 VHF omnidirectional range

A **VHF omnidirectional range** (VOR) beacon gives three types of informations: a voice (for example, to carry wheather information), the identity of the VOR beacon and the radial on which the aircraft is positioned, relative to the beacon. This means that a lot of data is put on one wave. In fact, we first take a sub-carrier wave of $30Hz$. This wave is frequency-modulated with the data signal. Then, we use this wave to amplitude-modulate the actual carrier wave of $9960Hz$.

Think this is complex? The working of the VOR itself isn't very easy either. To explain the working principle of the VOR, we examine a 'lighthouse' with two 'lights'. Light 1 is always on, but the direction in which it shines varies. In fact, the light beam rotates at 30 RPM. Light 2 is only on when light 1 shines North. At this moment, light 2 shines in all directions. At every other time, light 2 is off. The result of this is a repeating pattern with a period of $T = 1/30s$. To find the radial, we can now use the equation

$$\frac{t_{\text{light 1 on}} - t_{\text{light 2 on}}}{T} \cdot 360°. \tag{2.1}$$

In reality, we don't have two lights. Instead, we use a special **limacon** antenna. This antenna sends signals in a limacon pattern: some sides get a strong signal, while other sides get a weak signal. And the limacon antenna is also rotating at 30 RPS. This results in another modulation: a $30Hz$ amplitude modulation. The signal that is sent thus contains two $30Hz$ signals. By measuring the phase difference between them, we can find our radial θ with respect to the beacon.

VORs have been around for quite a while. In fact, they were used when defining airways. When following an airway, you usually keep on flying from one VOR to the next, until you have reached your destination.

2.3 Distance measuring equipment

A **distance measuring equipment** (DME) system is based on two-way communcication. First, an airborne **DME interrogator** sends out a UHF pulse. (This is done about 150 times each second.) The ground-based **DME transponder** receives the signal, waits exactly $50\mu s$, and then sends a signal back. The difference between the frequency of the incoming signal and the frequency of the replied signal is always $63MHz$. The airborne equipment now computes the **slant range** d, being the line-of-sight distance. (So, if you're flying directly above a DME transponder, the slant range is not zero, but h.) This slant range is given by

$$d = \frac{1}{2} c \left(\Delta T - 50\mu s \right). \tag{2.2}$$

One downside of the DME system is its relatively low accuracy. The accuracy is only $\frac{1}{4}NM + 0.0125R$. Another downside is that it requires active communication. A DME beacon can only send out a limited amount of signals. This means that at most 50 to 100 aircraft can make use of a beacon simultaneously.

But, with 50 aircraft using a beacon, how do we know which signal is meant for us? A first suggestion to solve this problem might be to just look at multiple intervals. Let's suppose that, every time we send the signal, we get a response a fixed amount of time Δt later. Then this will probably be the reply to our signal! Sadly, if all aircraft send their signals at regular intervals, this trick doesn't work. But the solution to this problem is **jitter**. Instead of sending out our signal at constant intervals, we apply very small variations in our interval length. (Of up to $10\mu s$ only.) Now, the trick that we just described does work: there's only one signal with a constant time difference Δt.

2.4 Long range navigation system

The **long range navigation** (LORAN) system is, surprisingly, used for long range navigation. It's especially useful for oceanic regions or other remote places where no other beacons are present. LORAN makes use of ground waves for navigation. It thus uses LF signals.

The LORAN system consists of transmitting stations (beacons) put together in groups known as **chains**. Every group has one **master station** and at least two (but often more) **secondary (slave) stations**. This is because at least three beacons are needed for a position fix. All beacons in a LORAN chain send out pulses in all directions. This is done at a regular pattern. (This pattern can even be used to identify the LORAN chain.)

A receiver, listening to the LORAN stations, measures the time differences between receiving the pulses. The receiver also knows the time difference between the sending of the pulses. Based on this, he can calculate the difference in distance with respect to the beacons. If he does this for multiple pairs of beacons, he can derive his own position.

The accuracy of the LORAN system varies with position (GDOP). The accuracy is best when the receiver is on the **baseline**: the line between two beacons. Furthermore, the accuracy decreases as the distance from the beacons increases. Up to $350NM$, the accuracy is roughly $130m$. Up to $1000NM$, it is about $550m$. Starting from $1500NM$, multipath effects due to sky waves start to play a role. The accuracy then drops to over $10NM$.

3 Satellite radio navigation

Radio signals don't always need to come from ground-based beacons. We might as well use satellites. But how does this work?

3.1 Satellite navigation equations

Satellite navigation is all about timing. The signal is sent by the satellite and received by the receiver. The **pseudo range** PR_i to satellite i is now given by

$$PR_i = c\tau = R_i + c\Delta t_u + c\Delta t_{s_i} + \epsilon_{PR_i}, \quad \text{with} \quad R_i = \sqrt{(x_{s_i} - x_u)^2 + (y_{s_i} - y_u)^2 + (z_{s_i} - z_u)^2}. \tag{3.1}$$

τ is the **time delay**, Δt_u is the **user clock error**, Δt_{s_i} is the **satellite clock error** and ϵ_{PR_i} is the sum of **other measurement errors**. (Think of atmospheric effects, multipath effects, etcetera.) Now let's assume that Δt_{s_i} and ε_{PR_i} are either known (due to models) or ignored; we can eliminate them. Also, the position of the satellites is known. We then remain with four unknowns: x_u, y_u, z_u and Δt_u. We thus also need at least four satellites to find these unknowns. Note that the four equations that we can find in this way are nonlinear. Solving it manually isn't easy. Instead, we usually use some sort of least-squares algorithm to solve it.

GPS systems have an accuracy of about 5 to 10 meters. The main cause for the inaccuracies are atmospheric effects. However, there is a nice way to increase the accuracy of GPS, called **differential GPS**. How does this work? First, we need a reference station. This station has an accurately known position. It compares this position to the position measured by the GPS system. Of course, there will be some very small differences. These **differential pseudorange corrections** are then sent to the user. The user can then also apply these corrections to make his own measurements more accurate. In this way, all **common errors** are cancelled. How well this method works depends mainly on the distance between the user and the reference station.

3.2 The global positioning system

It is time to examine a satellite system. Let's just start with the most well-known example: the **global positioning system** (GPS). This system was intended to have 24 satellites, although it often has a bit more. These satellites fly in 6 different orbital planes, each having an inclination of 55° and a height of approximately $20.000km$. This ensures that there are always 5 to 10 satellites in sight.

With the GPS system, every satellite has a **pseudo random noise** (PRN) code $C(t)$. This so-called **coarse/acquisition code** (C/A code), having a frequency of $1.023 MHz$, consists of a series of -1s and 1s. The code is then multiplied by the $1575.42 MHz$ carrier signal to find the **spread spectrum signal** $S(t)$. The reason for this is that energy is spread out over different frequencies. So, the signal can't be jammed anymore by any enemies we might have. Also, since every satellite has its own PRN code, we can use the PRN to identify the satellite.

Next to the PRN, we also send a $50 Hz$ **data message** $D(t)$ along with the signal. This data signal contains information like the position/velocity of the satellite, its clock error, and so on. The full signal that is sent is thus

$$S(t) = C(t)D(t)\cos(2\pi f t)\,. \tag{3.2}$$

However, the receiver receives the signal

$$S_{rec}(t) = S(t-\tau) = C(t-\tau)D(t-\tau)\cos(2\pi(ft-\tau))\,, \tag{3.3}$$

where τ still is the time delay. The most important problem is that we don't know τ. However, we know the PRN codes of the satellites. So, we just try to match these codes with the signal as well as possible. The best match will give us the time delay τ. And once we know the time delay, we can also find the message $D(t)$ that was sent along with the signal.