Chapter 2.

Signal Processing and Modulation

2.1. The Nature of Electronic Signals

2.1.1. Static and Quasi-Static Signals

Static signals are by definition unchanging over a long period of time. Such signals are essentially DC levels, while quasi-static signals are those that change very slowly such as the drift on a sensor.

2.1.2. Periodic and Repetitive Signals

Periodic signals are those that repeat themselves on a regular basis. These include sine, square and sawtooth waves. Their nature is such that each waveform is identical.

Repetitive signals are periodic in nature, but the exact shape may change slightly with time. ECG signals are an example of this type.

2.1.3. Transient and Quasi Transient Signals

Transient signals are either one time only signals while quasi-transient signals are those which are periodic but with a duration which is very short compared to the period of the waveform. Pulsed radar signals are good examples of these.

2.2. Sinusoidal Signals

Most acoustic and electromagnetic sensors exploit the properties of sinusoidal signals

In the time domain, such signals are constructed of sinusoidally varying voltages or currents constrained within wires,

$$v_c(t) = A_c \cos\omega_c t = A_c \cos 2\pi f_c t , \qquad (2.1)$$

where $v_c(t)$ – Signal,
  A_c – Signal amplitude (V),
  ω_c – Frequency (rad/s),
  f_c – Frequency (Hz),
  t – Time (s).

Sinusoidal electrical signals can be generated by the appropriate frequency selective feedback (shown below) or by feedback across an inductive-capacitive tank circuit.
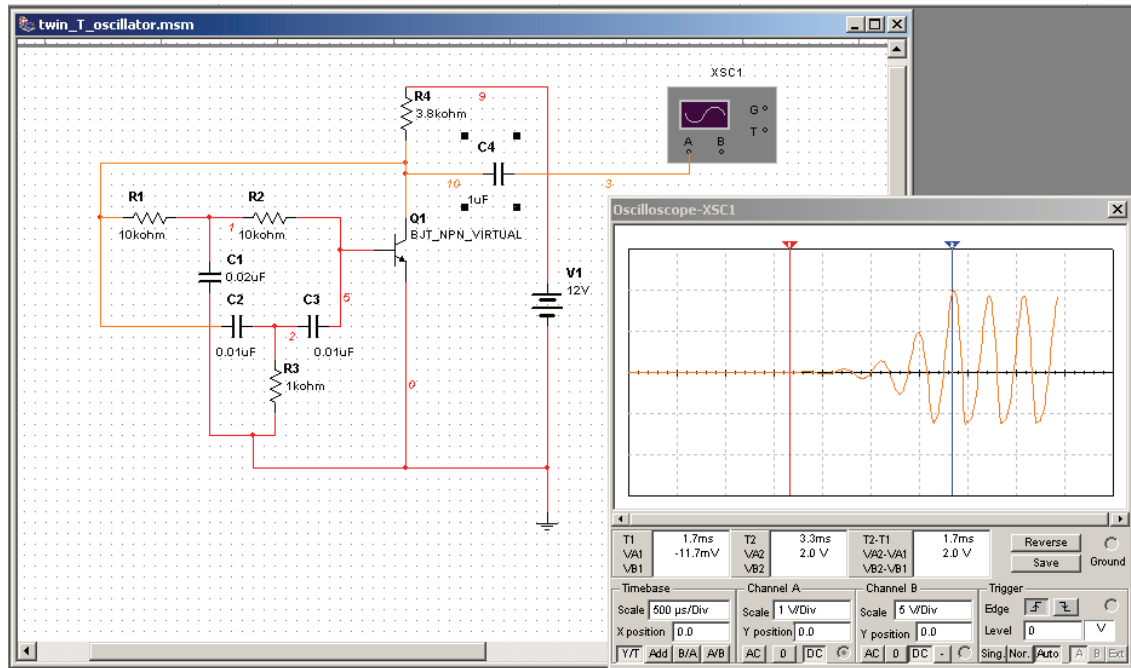

Figure 2.1: Sinusoidal voltage signal generated by an oscillator

In the frequency domain, a continuous sinusoidal signal of infinite duration can be represented in terms of its position on the frequency continuum and its amplitude only.

Most practical signals are not of infinite duration and so there is some uncertainty in the measured frequency which is represented in the frequency domain by a finite spectral width.

From a mathematical perspective, this is equivalent to windowing the continuous sinusoidal signal using a rectangular pulse of duration τ. Because windowing, or multiplication, in the time domain becomes convolution in the frequency domain, the continuous signal spectrum must be convolved by the spectrum, or Fourier transform, of a rectangular pulse to obtain the spectrum of the windowed signal.

As this Fourier transform is the Sync function

$$F(\omega) = \tau \frac{\sin(\omega\tau/2)}{\omega\tau/2} \tag{2.2}$$

and the spectrum of a continuous sinusoidal signal is an impulse $\delta(\omega)$, the resultant convolution is just the Sync function.

It can be seen from the equation for the Sync function that as the duration of the signal decreases, $\tau \to 0$, its spectral width increases until, in the limit, when the signal can be represented by an impulse $\delta(t)$, the spectral width is infinite. This relationship is shown graphically in the following figure.
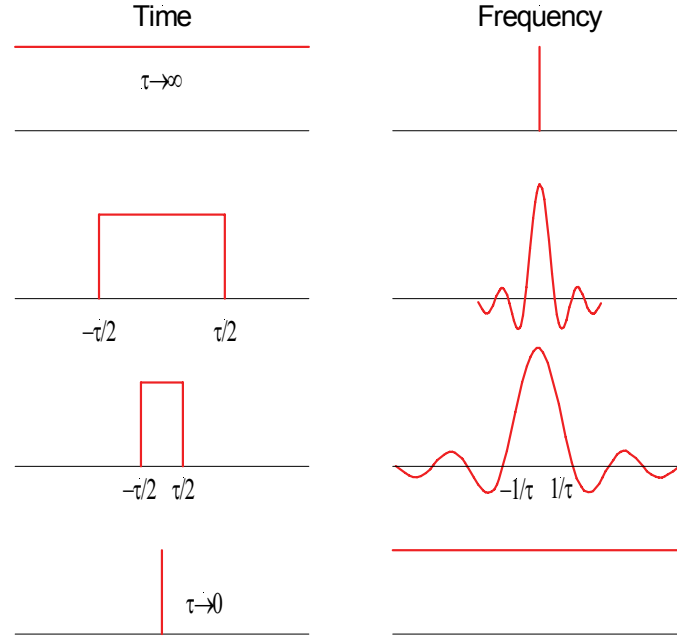
Figure 2.2: Mapping the relationship between the duration of a pulse and its spectrum

2.3. The Fourier Series

All continuous periodic signals can be represented by a fundamental frequency sine wave and a collection of sine and/or cosine harmonics of that fundamental sine wave.

The Fourier series for any waveform can be expressed as

$$v(t) = \frac{a_0}{2} + \int_{n=1}^{\infty} a_n \cos(n\omega t) + \int_{n=1}^{\infty} b_n \sin(n\omega t), \qquad (2.3)$$

where: $a_n \ b_n$ - Amplitudes of the harmonics (can be zero),
$\qquad n$ - Integer.

The amplitude coefficients can be calculated as follows,

$$a_n = \frac{2}{T}\int_0^t v(t)\cos(n\omega t)dt$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2.4)$$
$$b_n = \frac{2}{T}\int_0^t v(t)\sin(n\omega t)dt$$

The amplitude terms are non-zero at specific frequencies determined by the Fourier series. Because only certain frequencies, determined by integer n, are allowed, the spectrum is discrete.

The term $a_0/2$ is the average value of $v(t)$ over a complete cycle.

In general, though the harmonic series is infinite, the coefficients become so small that their contribution is considered to be negligible.

An ECG trace, for example with a fundamental frequency of about 1.2Hz can be reproduced with 70 to 80 harmonics (a bandwidth of about 100Hz)
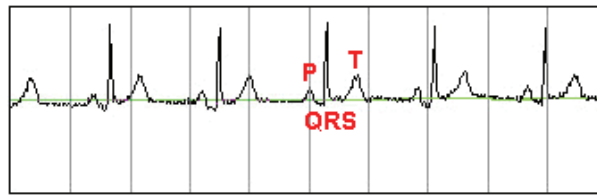

Figure 2.3: Typical ECG trace

A square wave on the other hand may require up to 1000 harmonics to reproduce the sharp transitions as shown in the figure below, depending on the application.
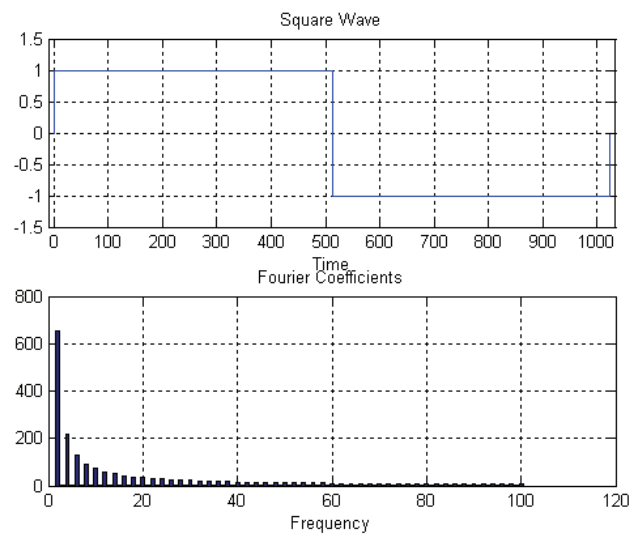

Figure 2.4: Amplitudes of Fourier coefficients to produce a square wave

If insufficient Fourier coefficients are used in the reconstruction of the signal it will be distorted as shown in the following example.
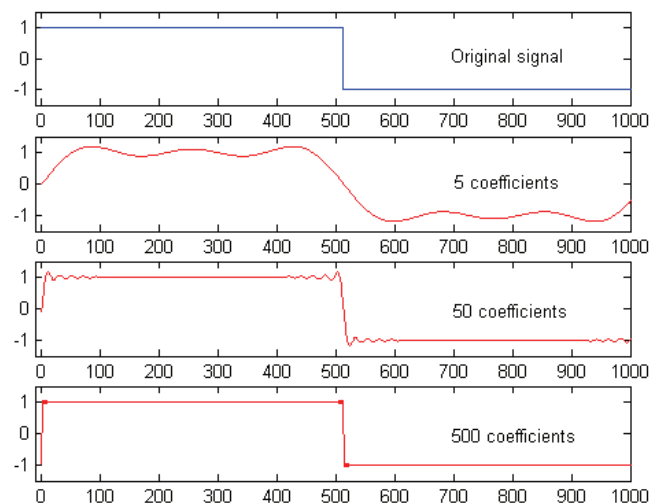

Figure 2.5: Effect on the reconstructed signal of limiting the number of Fourier coefficients

2.4. Sampled Signals

To process signals within a computer (Digital Signal Processing) requires that they be sampled periodically and then converted to a digital representation using an Analog to Digital Converter (ADC).
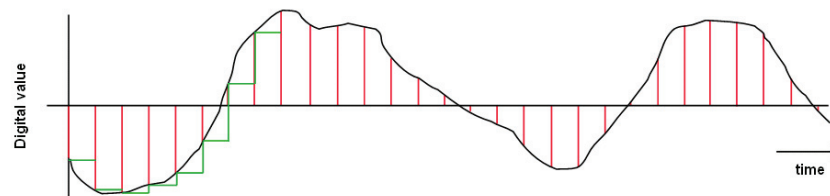

Figure 2.6: Digitising a signal

To ensure accurate representation the signal must be sampled at a rate which is at least double the highest significant frequency component of the signal. This is known as the Nyquist rate.

In addition, the number of discrete levels to which the signal is quantised must also be sufficient to represent variations in the amplitude to the required accuracy. Most ADCs quantise to 12 or 16 bits which represent $2^{12} = 4096$ or $2^{16} = 65536$ discrete levels.

After the signal has been processed, it is often necessary to generate an analog output. This function is performed by a Digital to Analog Converter (DAC)
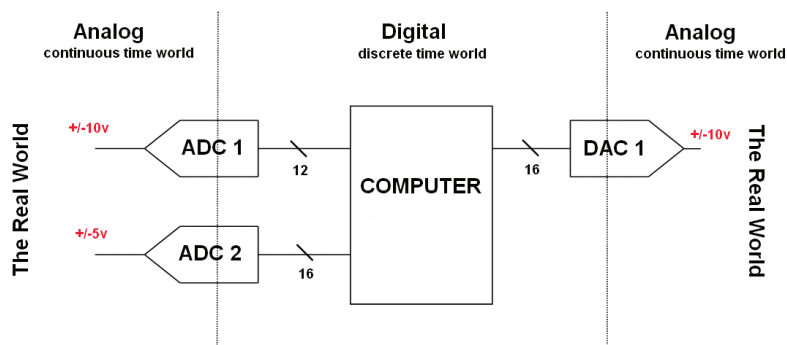

Figure 2.7: Typical Configuration for a digital signal processing application

The reconstruction process generally involves holding the signal constant (zero order hold) during the period between samples as shown in the following figure. This signal is then cleaned up by passing it through low-pass filter to remove high frequency components generated by the sampling process.
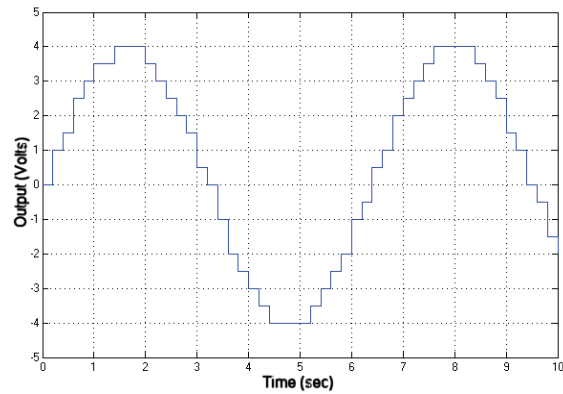
Figure 2.8: Analog Reconstruction of a sampled signal using a zero order hold

2.4.1. Generating Signals in MATLAB

MATLAB includes a number of built-in functions that make it easy to generate both periodic and aperiodic signals.
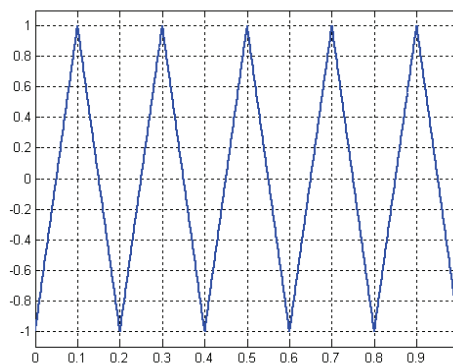
Generation of a square wave with amplitude A, fundamental frequency w0 (rad/s) and duty cycle rho as a percentage.

```
% generate a square wave
A = 1;
w0 = 10*pi;
rho = 50;
t = 0:0.001:1;
sq = A*square(w0*t, rho);
plot(t,sq)
axis([0,1,-1.1,1.1]);
```
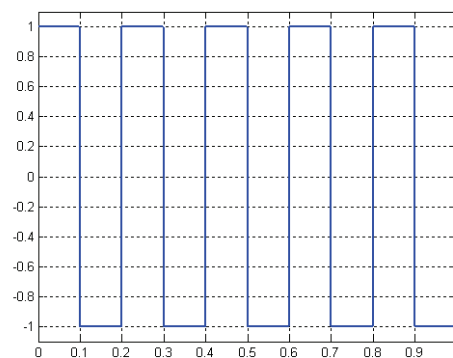

Generation of a triangular wave with amplitude A, fundamental frequency w0 (rad/s) and width W.

```
% generate a triangular wave
A = 1;
w0 = 10*pi;
W = 0.5;
t = 0:0.001:1;
tri = A*sawtooth(w0*t, W);
plot(t,tri)
grid
axis([0,1,-1.1,1.1]);
```


Generation of a sine wave with amplitude A, fundamental frequency w0 (rad/s) and start phase angle phi (rad).
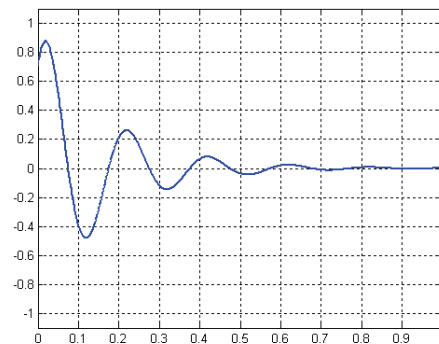
```
% generate a sine wave
A = 1;
w0 = 10*pi;
phi = pi/4;
t = 0:0.001:1;
sine = A*sin(w0*t + phi);
plot(t,sine)
grid
axis([0,1,-1.1,1.1]);
```
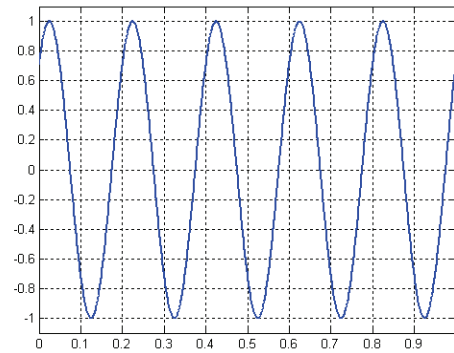
An exponentially damped sine wave is easily generated by taking the product of an exponential function and a sine wave.

$$x(t) = A\sin(\omega_o t + \phi)e^{-at} \tag{2.5}$$

```
% generate a exponentially
% decaying sine wave
A = 1;
w0 = 10*pi;
phi = pi/4;
a = 6;
t = 0:0.001:1;
expsine = A*sin(w0*t + phi).*exp(-a*t);
plot(t,expsine)
grid
axis([0,1,-1.1,1.1]);
```

Other useful MATLAB functions include the following:, COS, CHIRP, DIRAC, GAUSPULS, PULSTRAN, RECTPULS, SINC and TRIPULS.

2.4.2. Aliasing

If the analog signal is not sampled at at least twice the frequency of the highest frequency component, then these high frequency signals are "aliased" down to a lower frequency as shown in the following figure.
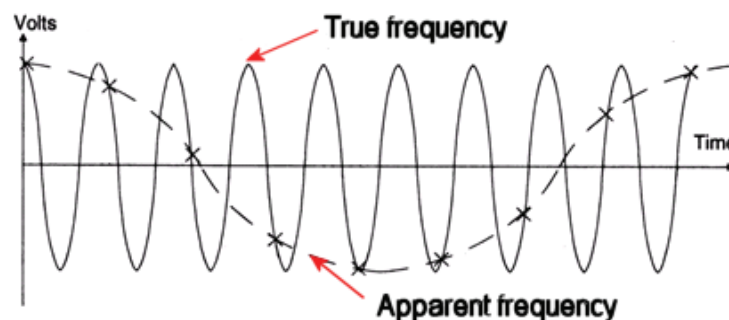

Figure 2.9: Interpretation of aliasing effects in the time domain

In the frequency domain, a generic analog signal may be represented in terms of its amplitude and total bandwidth as shown in the figure below.

A sampled version of the same signal can be represented by a repeated sequence spaced at the sample frequency (generally denoted f_s).

If the sample rate is not sufficiently high, then the sequences will overlap, and high frequency components will appear at a lower frequency (albeit with reduced amplitude) as shown in the figure.
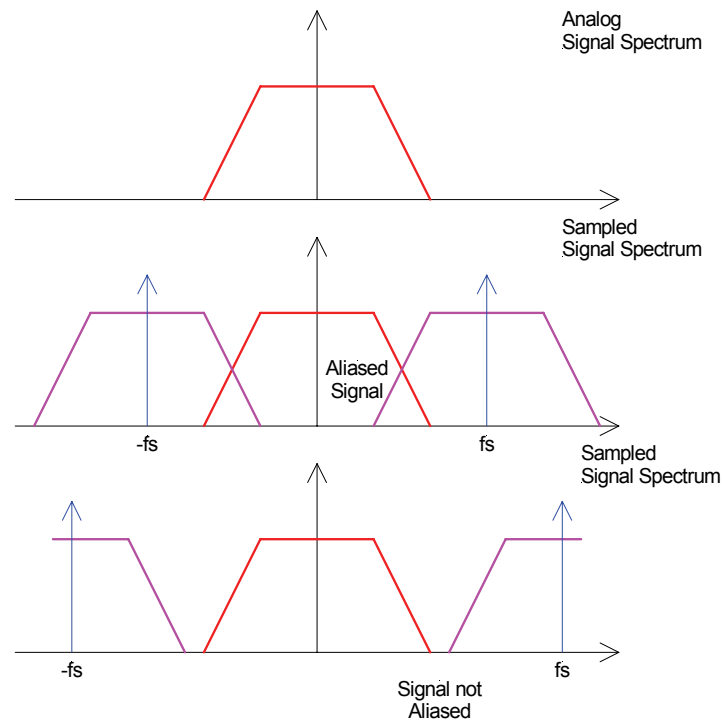

Figure 2.10: Interpretation of aliasing effects in the frequency domain

In most applications, an anti aliasing (low pass) filter ensures that the high frequency signals are sufficiently attenuated prior to sampling. A typical filter will attenuate these unwanted signals by between 40 and 60dB (1/100 to 1/1000) in voltage.

2.5. Filtering

A filter is a frequency selective network that passes certain frequencies of an input signal and attenuates others

The three common types of filter are:
- High Pass
- Low Pass
- Band Pass

High pass filter blocks signals below its cutoff frequency and passes those above.

Low pass filter passes signals below its cutoff frequency and attenuates those above.

Band pass filter passes a range of frequencies while attenuating those both above and below that range.

A fourth, less common configuration, is a band-stop or notch filter that attenuates signals at a specific frequency or over a narrow range of frequencies and passes all other frequencies.

Filters can be characterised by their impulse responses in the time domain, but are usually characterised by their frequency domain amplitude response $|H(\omega)|$ or Gain which presents the ratio of the output to input voltage over the frequency range of interest.

In this course we will only use sampled data filters that can be synthesized in MATLAB as shown in the figure.

```
% Banspass filter
fs = 200e+03;
ts = 1/fs;
fmat = 40.0e+03;
bmat = 10.0e+03;
wl=2*ts*(fmat-bmat/2);% lower band
wh=2*ts*(fmat+bmat/2);% upper band
wn=[wl,wh];
% 6th order Butterworth filter
[B,A]=butter(3,wn);
[h,w]=freqz(B,A,1024);
freq=(0:1023)/(2000*ts*1024);
%semilogx(freq,20*log10(abs(h)));
plot(freq,abs(h));
grid
title('Bandpass Filter Transfer Fn')
xlabel('Frequency (kHz)');
ylabel('Gain')
```
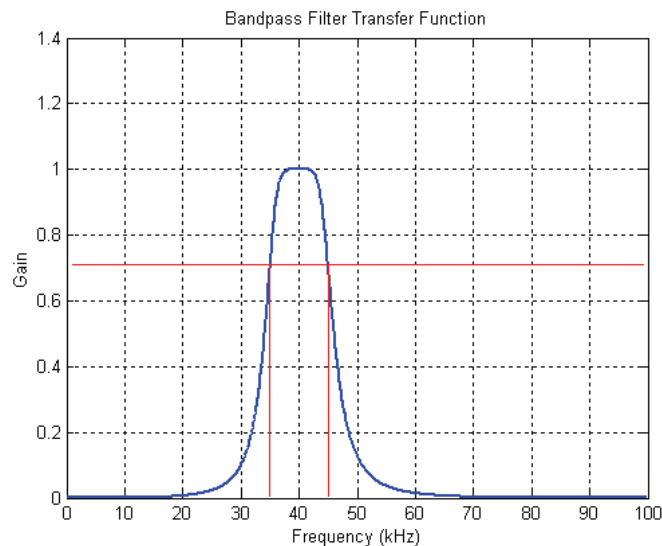

Figure 2.11: Butterworth bandpass filter transfer function generated by MATLAB

Note that the upper and lower limits of the pass band represent the half power points (0.707 of the peak voltage gain)

If the gain was plotted in dB then it would be calculated using $20*\log_{10}(\text{Gain})$ to convert to power.

MATLAB implements filters as the "Direct Form II Transposed" of the standard difference equation

$$a(1)*y(n) = b(1)*x(n) + b(2)*x(n-1) + ... + b(n_b+1)*x(n-n_b)$$
$$-a(2)*y(n-1) - ... - a(n_a+1)*y(n-n_a)$$

where the coefficients $A = [a(1), a(2)...a(n_a+1)]$ and $B = [b(1), b(2) ...b(n_b+1)]$ are generated when the filter is synthesised

2.5.1. Filter Categories

The major filter categories are as follows:
- Butterworth (maximally flat).
- Chebyshev (equi ripple).
- Bessel (linear phase).
- Elliptical

Butterworth

This approximation to an ideal low pass filter is based on the assumption that a flat response at zero-frequency is most important. The transfer function is an all-pole type with roots that fall on the unit circle.

It exhibits fairly good amplitude and transient characteristics.

Chebyshev

The transfer function is also all-pole, but with roots that fall on an ellipse. This results in a series of equi amplitude ripples in the pass band and a sharper cutoff than the Butterworth.

It exhibits good selectivity but poor transient behaviour.

Bessel

Optimised to obtain a linear phase response which results in a step response with no overshoot or ringing and an impulse response with no oscillatory behaviour.

Poor frequency selectivity compared to the other response types.

Elliptic

These filters have zeros as well as poles which create equi-ripple behaviour in the pass band similar to Chebyshev filters.

Zeros in the stop band reduce the transition region so that extremely sharp roll-off characteristics can be achieved, for that reason they are commonly used in anti-aliasing filters.
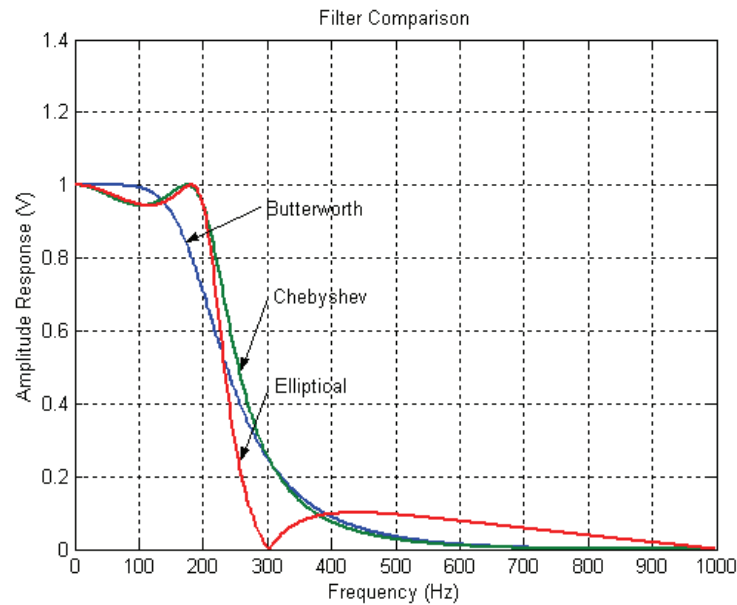
Figure 2.12: Lowpass filter comparison

Note that the cutoff frequency (200Hz in this case) specified in MATLAB is equal to the 3dB point for the Butterworth filter and to the passband ripple for the Chebyshev and Elliptic filters

The rate at which the signal is attenuated as a function of frequency is proportional to the order of the filter. The following figure shows the roll-off for Butterworth filters. The theoretical slope is 6n dB/octave.
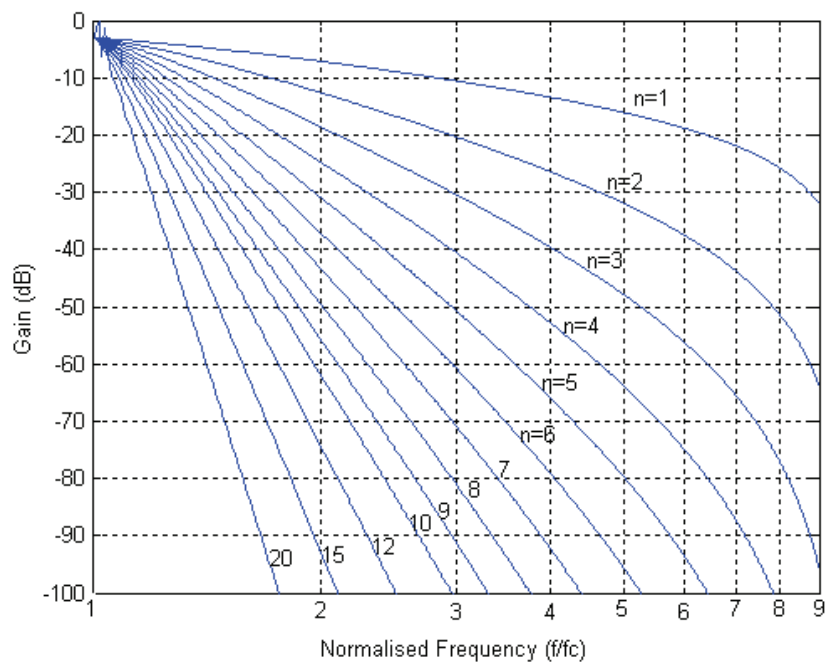

Figure 2.13: Roll off for Butterworth lowpass filters

2.5.2. The Ear as a Filter Bank

In the ear, sound waves are transmitted into the cochlea which tapers in size like a cone. Through the middle stretches the basilar membrane which gets wider as the cochlea gets narrower.

The vibratory movement is transmitted as a standing wave in the basilar membrane (much like a snapping rope) with the amplitude reaching a peak at a location dependent on frequency due to the varying resonant characteristics of the membrane as shown schematically in the following figure.

High frequency peaks occur toward the base (where the membrane is stiffest and narrowest) while the low frequency peaks occur towards the apex. Hair cells rest on the basilar membrane and convert these vibrations to electro-chemical signals which are transmitted to the brain for interpretation.

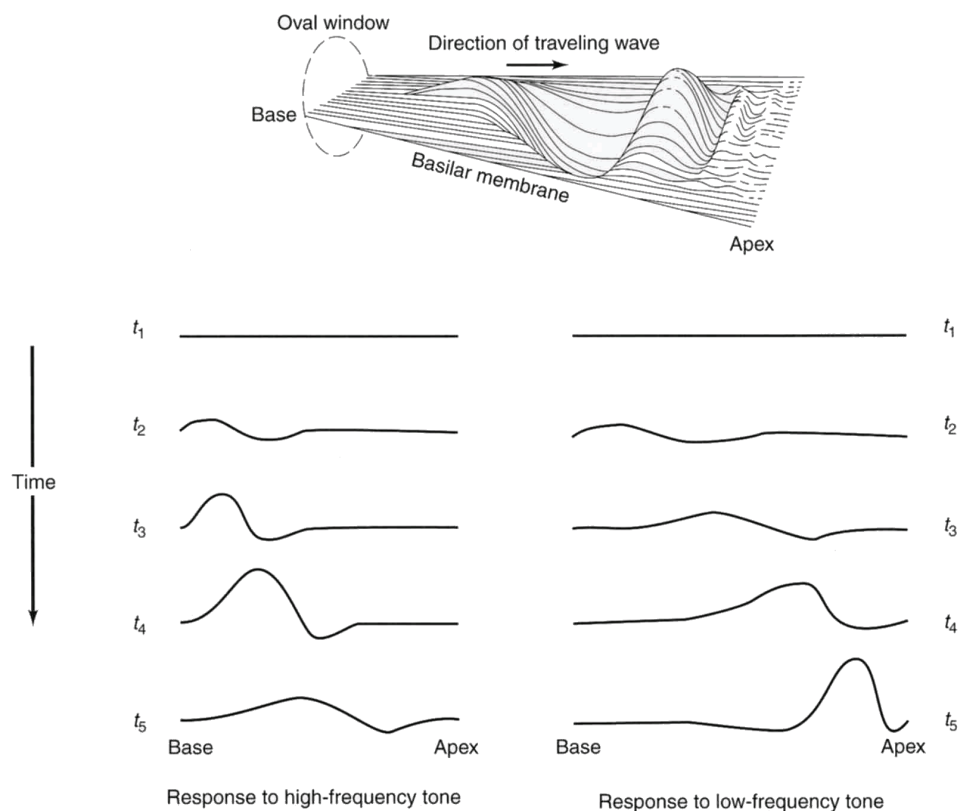The base resonates at about 20kHz while the apex resonates at about 20Hz

Figure 2.14: The Basilar Membrane of the Cochlea depicted uncoiled and flattened showing the resonance for travelling waves of different frequencies

The cochlear implant consists of a string of electrodes, each excited by a narrow band of frequencies, which stimulate the hair cell nerves directly. This allows some hearing to be restored in the case when either the cilia or basilar membrane has been damaged.

2.6. Analog Modulation and Demodulation

A continuous unmodulated signal cannot be used to measure range as there is no way of determining when the signal was transmitted.

For most sensor applications, the transmitted signal is "marked" in some way either by altering its amplitude or frequency.

The round trip time from the moment that the "mark" is transmitted to when it is received can then be used to determine the range to the target if the speed of propagation is known.

Marking the transmitted signal is known as modulation.

2.6.1. Amplitude Modulation (AM)

AM is defined as a modulation technique in which the amplitude of the carrier is varied in accordance with some characteristic of the baseband modulating signal.

It is the most common form of modulation because of the ease with which the baseband signal can be recovered from the transmitted signal.
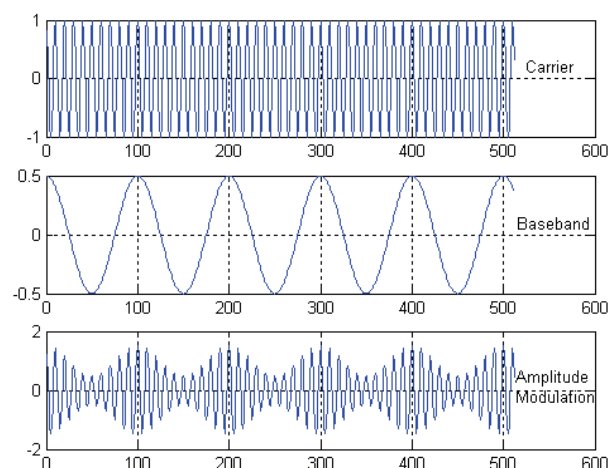

Figure 2.15:Time domain representation of amplitude modulation

For an unmodulated carrier described earlier and for a baseband signal $v_b(t)$, the AM signal $v_{am}(t)$ is described by the following equation

$$v_{am}(t) = A_c \big[1 + v_b(t) \big] \cos 2\pi f_c t . \tag{2.6}$$

For the example shown above, where the modulating signal is a sinusoid with a frequency f_a and amplitude A_{am} then the revised formula,

$$v_{am}(t) = A_c \big[1 + A_{am} \cos 2\pi f_a t \big] \cos 2\pi f_c t . \tag{2.7}$$

In general $A_{am}<1$ otherwise a phase reversal occurs and demodulation becomes more difficult.

The extent to which the carrier has been amplitude modulated is expressed in terms of a *percentage modulation* which is just calculated by multiplying A_{am} by 100.

To determine the characteristics of the signal in the frequency domain, it can be rewritten in the following form (using a trig identity),

$$v_{am}(t) = A_c \cos 2\pi f_c t + \frac{A_c A_{am}}{2} \left[\cos 2\pi (f_c - f_a)t + \cos 2\pi (f_c + f_a)t \right]. \qquad (2.8)$$

It can be seen that this is made up from three independent frequencies:
- The original carrier at a frequency of f_c.
- A frequency at the difference between the carrier and the baseband signal
- A frequency at the sum of the carrier and the baseband signal

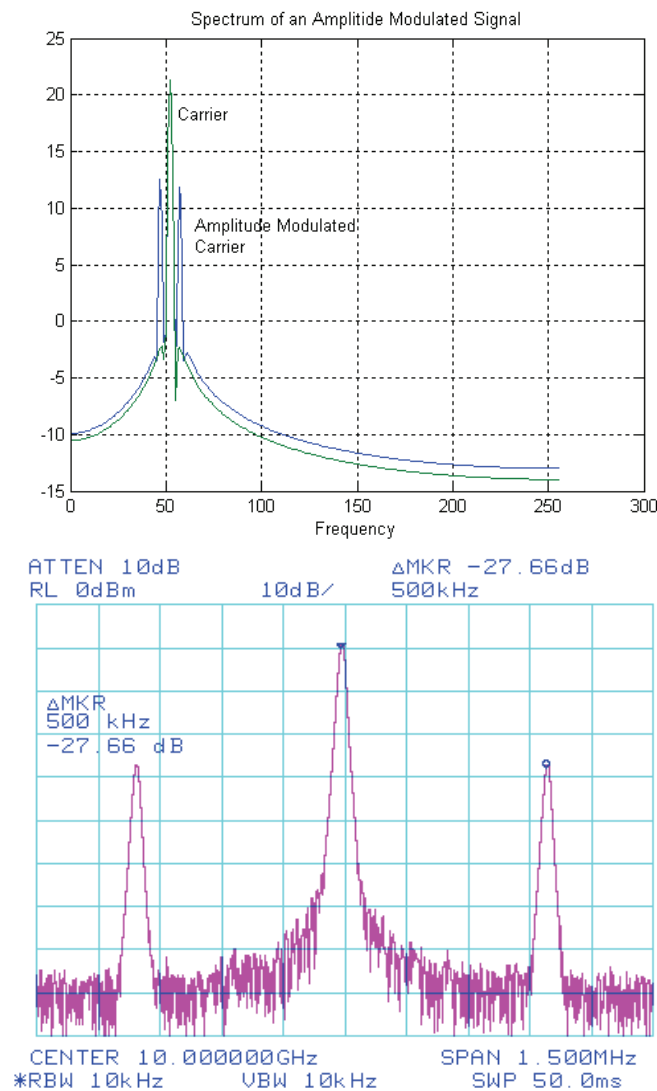The spectrum is shown in the figure.

Figure 2.16: Simulated and measured frequency domain representation of amplitude modulation

The "tank" circuit in the crystal radio shown in the figure below is tuned to resonate at the carrier frequency and so operates to select only a single radio program which passes into the detection stage.

Demodulation is then achieved with the use of a simple diode rectifier (crystal) and lowpass filter (capacitor) and a pair of high-impedance headphones converts the resultant envelope current to an audio signal that accurately reproduces the original modulation.

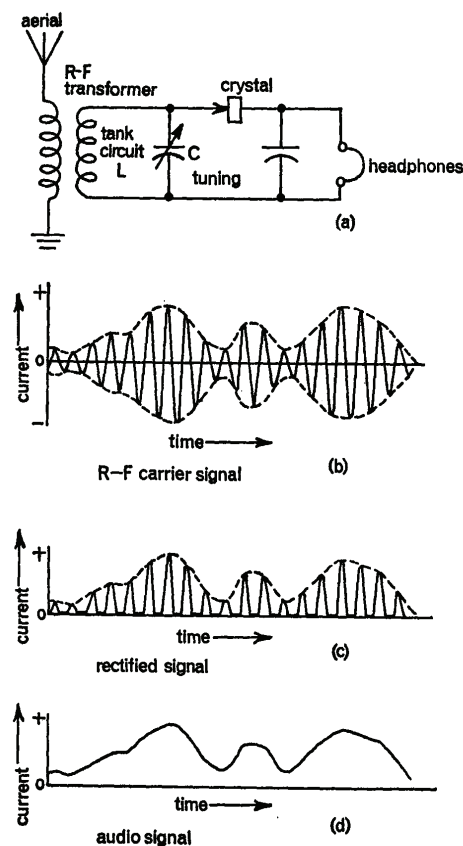From the figure below it can be seen how simple the demodulation of an AM signal can be.

Figure 2.17: Demodulation of an AM signal by a crystal radio

Note that there is sufficient energy in the RF signal (with the appropriate long wire antenna) to drive a set of high-impedance headphones directly without any amplification (hence no battery).

2.7. Frequency Modulation (FM)

FM is a modulation technique in which the frequency of the carrier is varied in accordance with some characteristic of the baseband modulating signal.

$$v_{fm}(t) = A_c \cos\left[\omega_c t + k \int_{-\infty}^{t} v_b(t)dt \right] \qquad (2.9)$$

The reason that the modulating signal is integrated is because variations in the modulating term equate to variations in the carrier phase.

The instantaneous angular frequency can be obtained by differentiating the instantaneous phase,

$$\omega = \frac{d}{dt}\left[\omega_c t + k \int_{-\infty}^{t} v_b(t)dt\right] = \omega_c + kv_b(t) .$$

(2.10)

The deviation of the instantaneous frequency from the carrier frequency, $\omega_c/2\pi$, is

$$\delta f = f - f_c = \frac{k}{2\pi} v_b(t)$$

(2.11)

This shows that the deviation of the instantaneous frequency is directly proportional to the amplitude of the modulating signal. Hence the combination of an integrator and phase modulator produces frequency modulation.
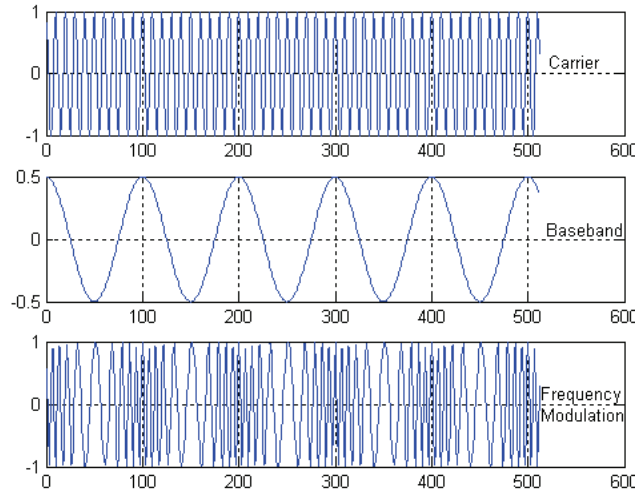

Figure 2.18: Time domain representation of frequency modulation

For sinusoidal modulation, the formula for FM is

$$v_{fm}(t) = A_c \cos[\omega_c t + \beta \sin \omega_a t],$$

(2.12)

where β, which is the maximum phase deviation, is usually referred to as the modulation index.

The instantaneous frequency in this case is

$$f = \frac{\omega_c}{2\pi} + \frac{\beta\omega_a}{2\pi}\cos\omega_a t$$

$$f = f_c + \beta f_a \cos\omega_a t$$

(2.13)

So the maximum frequency deviation defined as Δf, when $\cos\omega_a t = 0$, is

$$\Delta f = \beta f_a .$$
(2.14)

Even though the instantaneous frequency lies within the range $f_c +/- \Delta f$, the spectral components of the signal don't lie within this range.

Some manipulation of the formula for $v_{fm}(t)$ shows that the spectrum comprises a carrier with amplitude $J_o(\beta)$ with sidebands spaced symmetrically on either side of the carrier at offsets of $\omega_a, 2\omega_a, 3\omega_a, \dots$ as shown in the figure below.

Theoretically, the bandwidth is infinite, however, for any β, most of the power is confined within a finite bandwidth.

As a rule of thumb (Carson's Rule), the bandwidth is twice the sum of the maximum frequency deviation plus the modulating frequency.
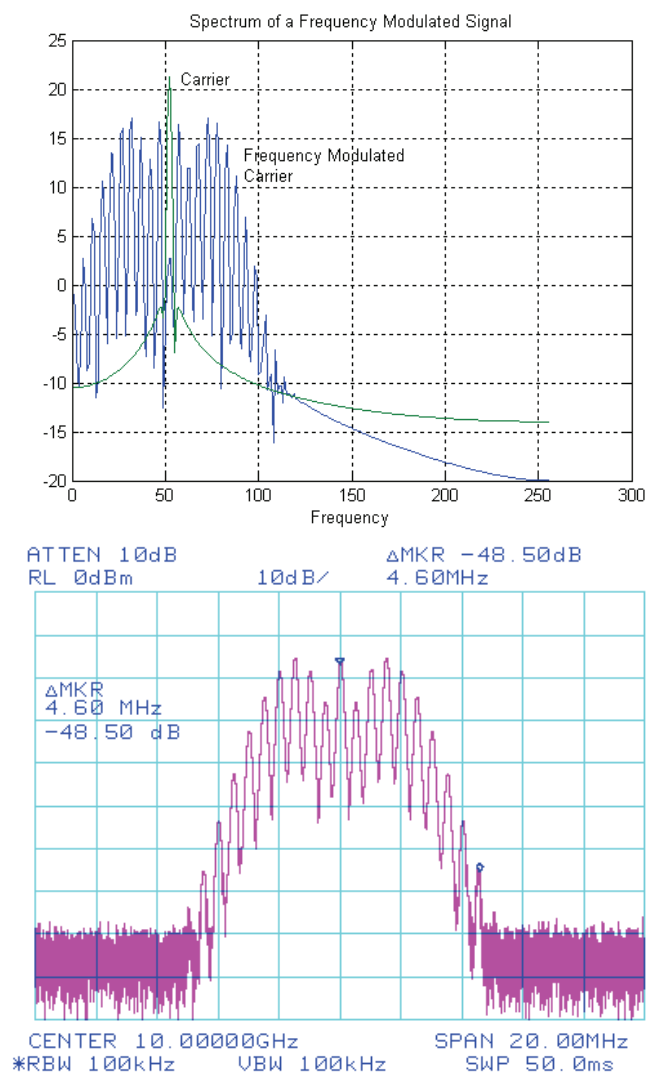

Figure 2.19: Simulated and measured frequency domain representation of frequency modulation

Demodulation of an FM signal is commonly achieved by converting it into AM and then envelope detecting it.

The simplest method to perform the conversion to AM is to pass the signal through a frequency sensitive circuit such as a low-pass or bandpass filter. The circuitry to perform this function is known as a discriminator
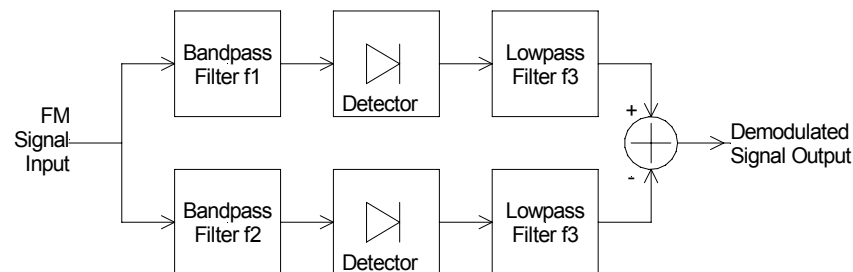

Figure 2.20: A discriminator converts the FM signal into an amplitude variation and envelope detects the resulting AM signal

In this example the FM signal is split and passed through two bandpass filters with centre frequencies just below, and just above the carrier frequency. The two signals are then detected and filtered to remove the residual carrier before the difference is taken. The transfer function for this process is shown in the figure below.
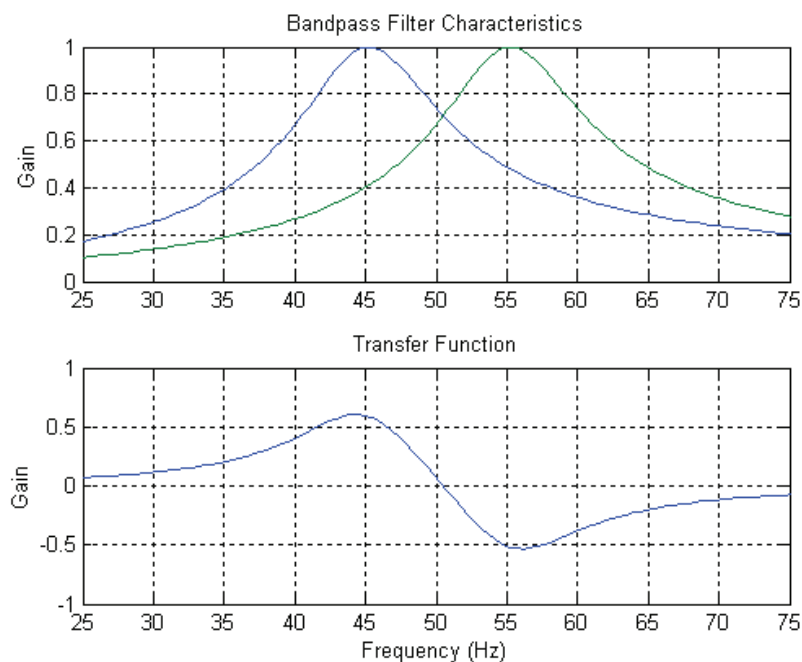

Figure 2.21: The difference signal from a pair of offset bandpass filters produces a symmetrical transfer function to convert variations in frequency to variations in amplitude

In the time domain, the FM signal after detection and filtering produces two symmetrical demodulated signals with DC offsets. The difference between these reproduces the original baseband signal as shown in the figure below.
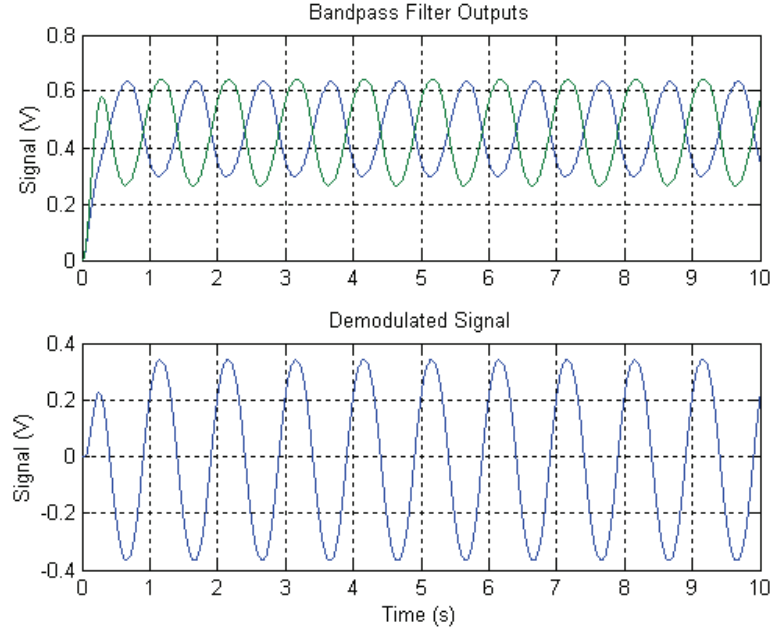
Figure 2.22: Stages of FM demodulation showing (a) the detected and filtered outputs from the bandpass filters and (b) the final demodulated output obtained by taking the difference between these signals

Alternative techniques used in most modern radio receivers use either phase-locked loop or quadrature detection techniques to perform this function.

2.8. Linear Frequency Modulation

In most active sensors that operate using the Frequency Modulated Continuous Wave (FMCW) principle, the frequency is not modulated sinusoidally, but in a linear manner with time.

$$\omega_b = A_b t . \tag{2.15}$$

Substituting into the standard equation for FM, we obtain the following result

$$v_{fm}(t) = A_c \cos\left[\omega_c t + A_b \int_{-\infty}^{t} t dt\right]$$

$$v_{fm}(t) = A_c \cos\left[\omega_c t + \frac{A_b}{2} t^2\right] \tag{2.16}$$

Note that the phase modulation follows a quadratic function.

In general it is not possible to continue to increase the frequency indefinitely, so the modulation often follows a sawtooth or triangular function.
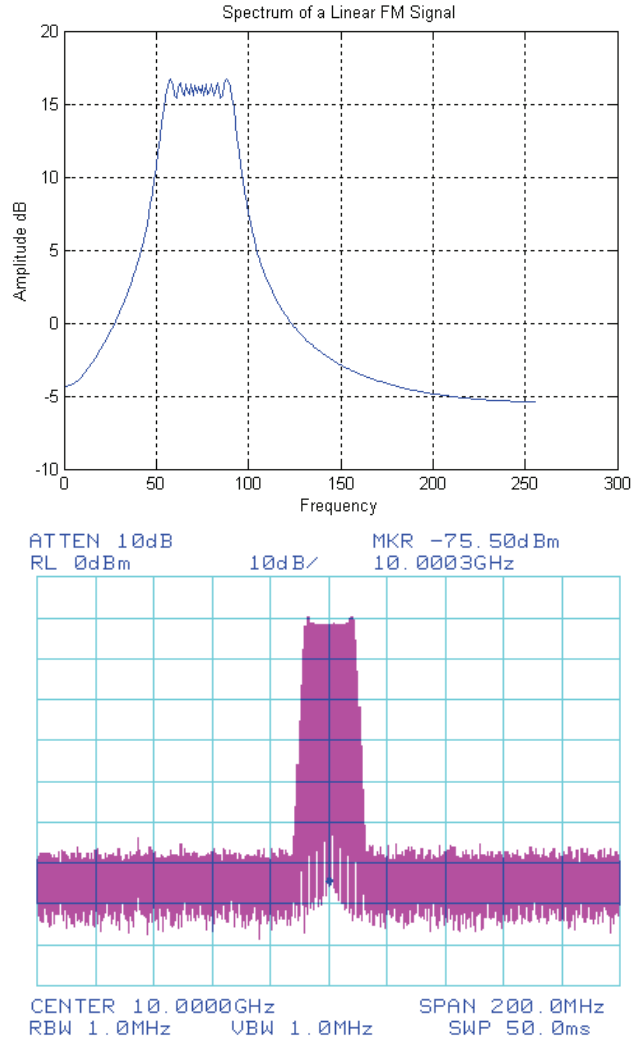
Figure 2.23: Simulated and measured frequency domain representation of linear FM

In FMCW systems, a portion of the transmitted signal is mixed with (multiplied by) the returned echo as discussed in Chapter 11.

The transmit signal will be shifted from that of the received signal because of the round trip time, τ,

$$v_{fm}(t - \tau) = A_c \cos\left[\omega_c(t - \tau) + \frac{A_b}{2}(t - \tau)^2\right].$$
(2.17)

Calculating the product of the transmitted signal and the delayed echo

$$v_{fm}(t - \tau)v_{fm}(t) = A_c^2 \cos\left[\omega_c t + \frac{A_b}{2}t^2\right]\cos\left[\omega_c(t - \tau) + \frac{A_b}{2}(t - \tau)^2\right]$$
(2.18)

Equating using the trig identity $\cos A \cos B = 0.5[\cos(A+B)+\cos(A-B)]$

$$v_{out}(t) = \frac{1}{2} \left[\begin{array}{l} \cos\left\{ \left(2\omega_c - A_b\tau\right)t + A_b t^2 + \left(\frac{A_b}{2}\tau^2 - \omega_c\tau\right)\right\} \\ + \cos\left\{ A_b\tau.t + \left(\omega_c\tau - \frac{A_b}{2}\tau^2\right)\right\} \end{array} \right] \tag{2.19}$$

The first cosine term describes a linearly increasing FM signal (chirp) at about twice the carrier frequency with a phase shift that is proportional to the delay time τ. This term is generally filtered out.

The second cosine term describes a beat signal at a fixed frequency,

$$f_{beat} = \frac{A_b}{2\pi}\tau . \tag{2.20}$$

It can be seen that the signal frequency is directly proportional to the delay time τ, and hence is directly proportional to the round trip time to the target.

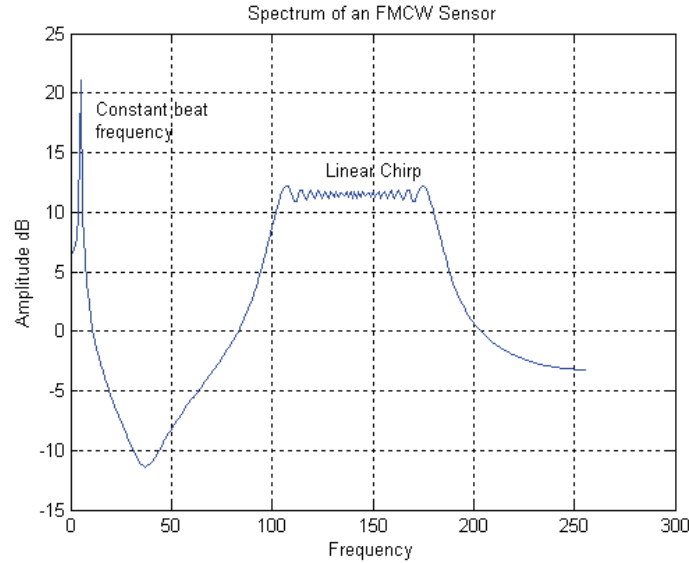The spectrum of this output signal is shown in the following figure.

Figure 2.24: Frequency domain representation of FMCW sensor output

2.9. Pulse Coded Modulation Techniques

2.9.1. Pulse Amplitude Modulation (PAM)

Modulation in which the amplitude of individual, regularly spaced pulses in a pulse train is varied in accordance with some characteristic of the modulating signal. For time-of-flight sensors, the amplitude is generally constant.

The ability of a pulsed sensor to resolve two closely spaced targets is determined by the pulse width as shown in more detail in Chapter 11.

The following figure shows the relationship between the width of a single pulse and its spectral content.
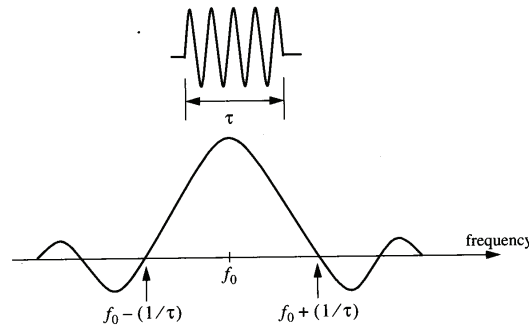

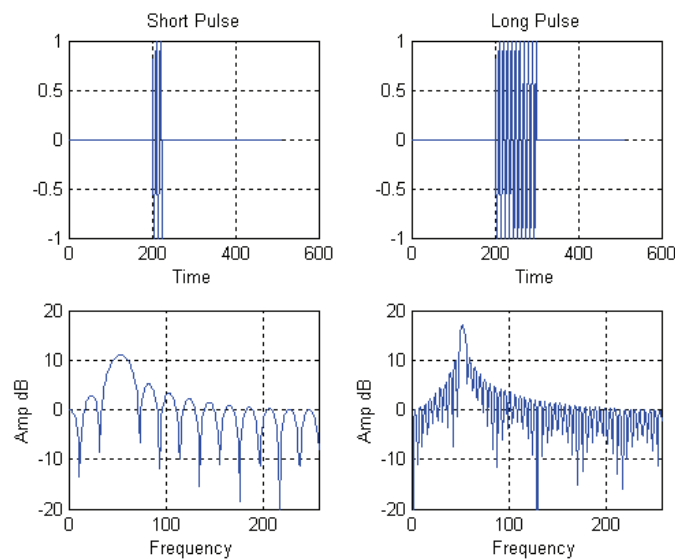**Figure 2.25: Relationship between pulsewidth and frequency for a single pulse**

Figure 2.26: Relationship between the duration of a pulse and its spectrum for pulsed amplitude modulation

Note that the width of the pulse spectrum in the frequency domain increases as the pulse width decreases with the following relationship,

$$\beta \approx \frac{1}{\tau}, \tag{2.21}$$

where β is the spectral width (Hz) between the half power (3dB) points and τ is the pulsewidth (sec). In a time-of-flight sensor, this relationship determines the bandwidth required by a receiver to receive pulses of a specific duration. A filter that conforms to this relationship is known as a "matched" filter. It is formally defined in Chapter 11.

In a repeated sequence of pulses, the fine structure of the spectrum is determined by the total length of the observed sequence as shown in the figure below.
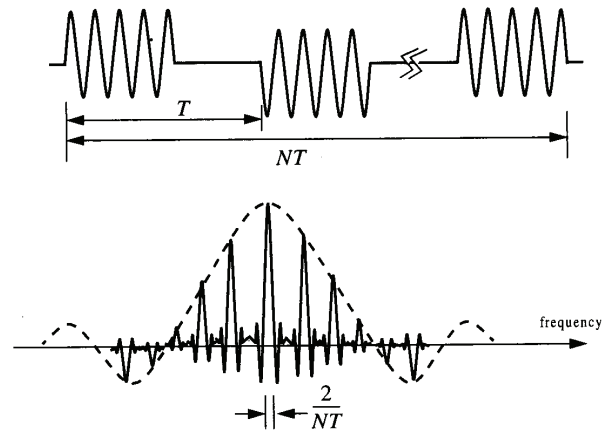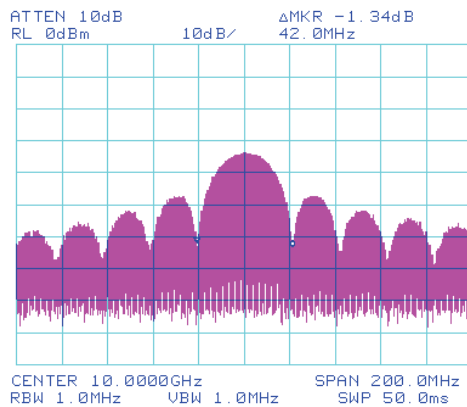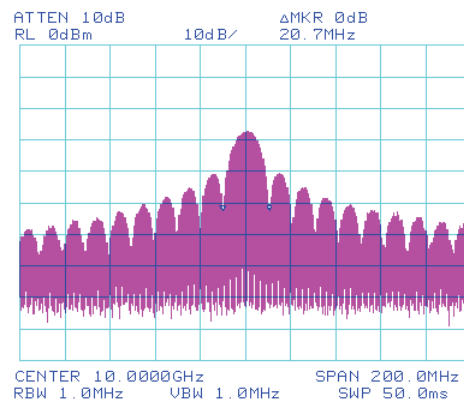
Figure 2.27: Relationship between pulsewidth and frequency for a finite length sequence of pulses
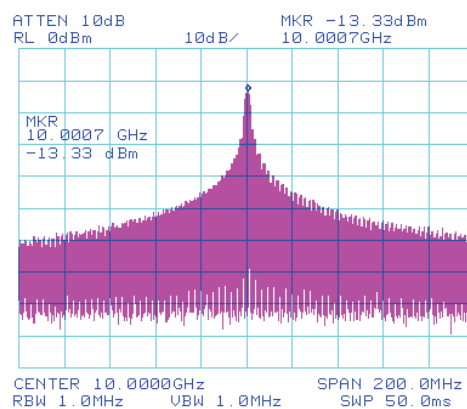
The following figures show the measured spectra of a number of continuous and pulsed signals with different periods
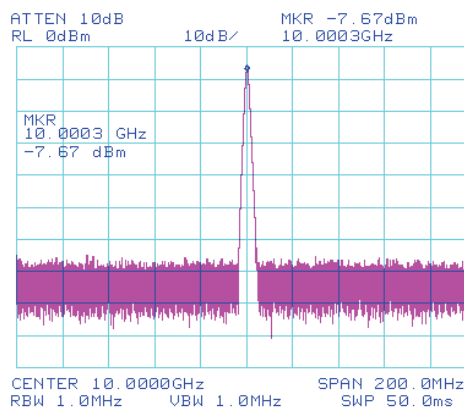

(a)

(b)

(c)

(d)

Figure 2.28: Spectra of pulsed signals with durations of (a) 50ns, (b) 100ns, (c) 500ns, (d) continuous

2.10. Frequency Shift Keying (FSK)

This modulation technique is the digital equivalent of linear FM where only two different frequencies are utilised. A single bit can be represented by a single cycle of the carrier, or if the data rate is not critical, then multiple cycles can be used.
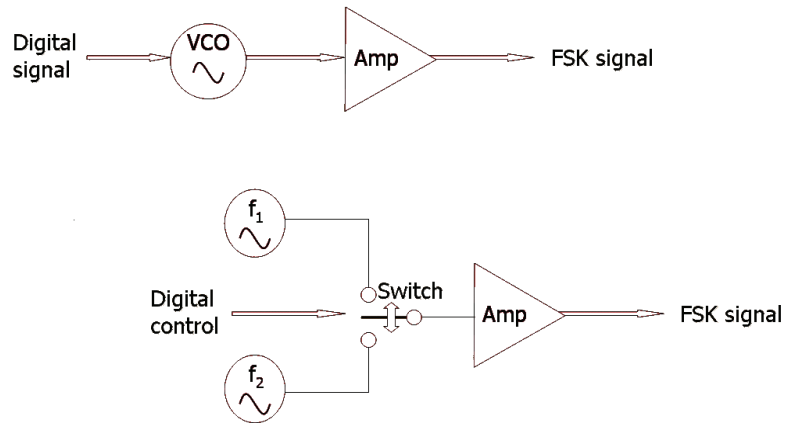

Figure 2.29: Two common methods of generating FSK

Demodulation can be achieved by detecting the outputs of a pair of filters centred at the two modulation frequencies, f_1 and f_2 as shown, or by using a phase locked loop.
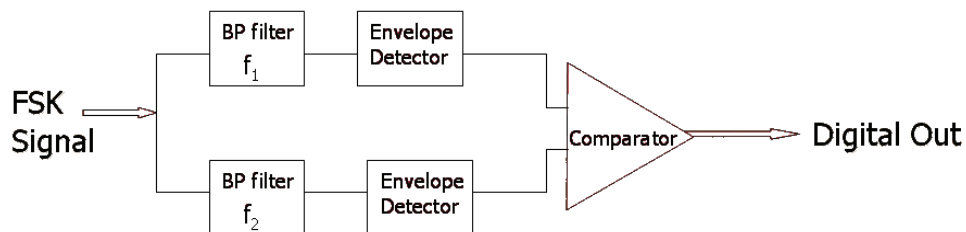

Figure 2.30: Band pass filter method of demodulating FSK signals

The following figure shows an example of FSK modulation with multiple cycles per bit.
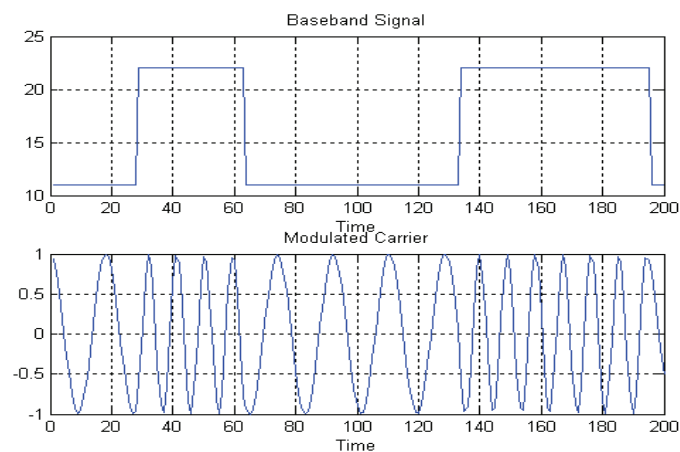

Figure 2.31: Example of frequency shift keying (FSK)

In coherent FSK, the bit transition is synchronised with the carrier frequency so that there are no phase discontinuities. In the non-coherent option there is no synchronisation and large phase discontinuities can occur

As the number of cycles per bit decreases, the spectrum shifts from being two distinct peaks centred t the two frequencies to a flatter, broad almost uniform peak as shown in the figures below.
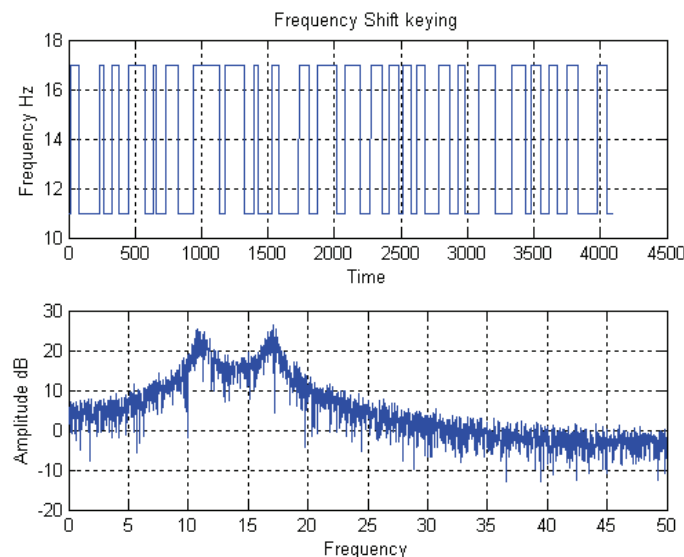
Figure 2.32: FSK spectrum, 5 cycles per bit
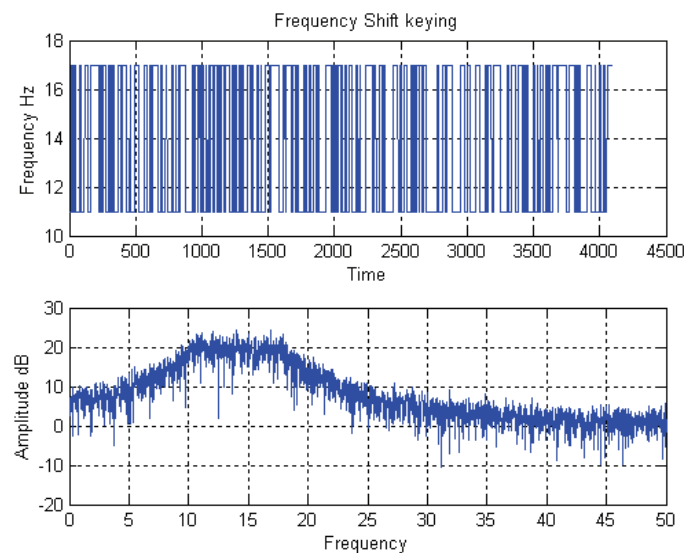
Figure 2.33: FSK spectrum, 1 cycle per Bit

FSK is a very simple modulation technique and is still extremely popular. It was originally used by teleprinters which operated at about 45bps, and was introduced in 1962 for a Bell modem which operated at up to 300bps. Early PC's used a Kansas City Interface which used FSK to store software on audio cassettes at up to 1200bps. It is now used in touch-tone phones and a myriad of other communications systems, operating at speeds in excess of 1Mbps.

2.11. Phase Shift Keying (PSK)

The most common form of PSK is binary phase coding. The carrier, f_o is switched between +/-180° according to a digital base band, f_m sequence.

This modulation technique can be implemented quite easily using a balanced mixer as shown in the figure, or with a dedicated BPSK modulator.
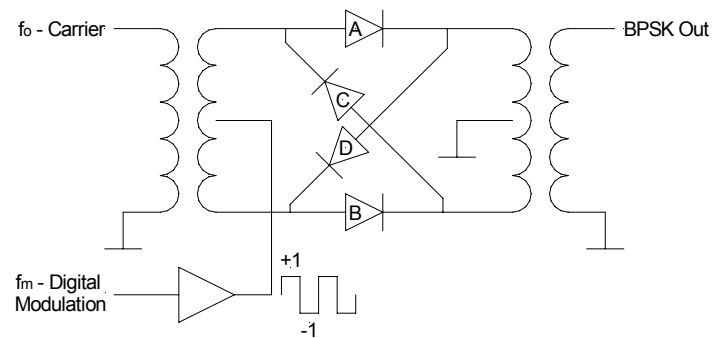
Figure 2.34: Implementing BPSK using a balanced mixer

When the modulation signal f_m is high (+1), diodes A and B are forward biased and the carrier f_o is coupled to the directly to the output transformer. However, when f_m is low (-1), then diodes C and D are forward biased and f_o is coupled to the to the opposite terminals of the output transformer which results in a reversal of the phase.

The modulation and output waveforms are shown in the following figure.

Figure 2.35: Example of binary phase shift keying with one cycle per bit

Demodulation is achieved by multiplying the modulated signal by a coherent carrier (a carrier that is identical in frequency and phase to the carrier that originally modulated the BPSK signal).

This produces the original BPSK signal plus a signal at twice the carrier which can be filtered out.
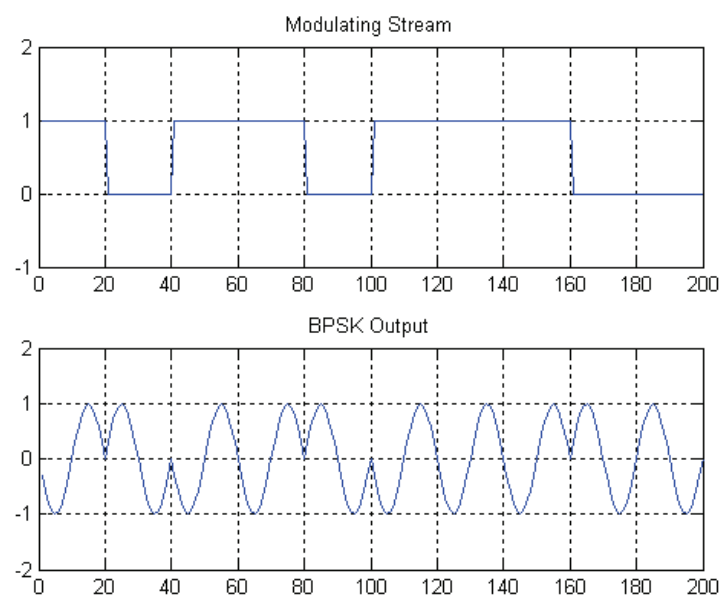
The spectral width of the carrier is widened by the modulation process as shown in the figure below. This is one of the modulation techniques that is used for "broadband" communications for obvious reasons.

Note in this example that the amplitude of the transmitted signal has been reduced by 30dB because the power has been spread, even though the total power transmitted (and hence the range performance) remains unaltered (see Chapter 11).
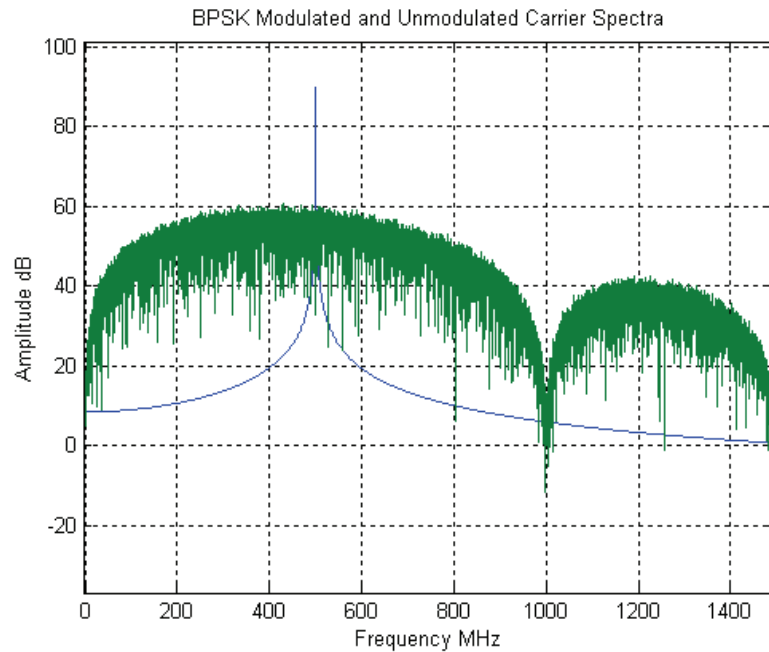
Figure 2.36: PSK Spectrum for a 500MHz carrier and 1 cycle per bit modulation

2.12. Stepped Frequency

In this modulation technique, a sequence of pulses are transmitted each at a slightly different frequency.

The pulse width (in the radar case) is made sufficiently wide to span the region of interest, but because it is so wide, it cannot resolve individual targets within that region. However, if all of the pulses are processed together, the effective resolution is improved because the total bandwidth is widened by the total frequency deviation of the sequence of pulses as discussed below.

If a signal is transmitted at a frequency ω_c and it reflects off a target at a range R, then the round trip delay time is

$$\tau = \frac{2R}{c}.$$
(2.22)

The phase difference between the transmitted signal and the echo can be determined by mixing a portion of the transmitted signal with the echo as follows and filtering out the high frequency components,

$$v_{out}(t) = \cos \omega_c t . \cos \omega_c (t - \tau)$$

$$v_{out}(t) = \frac{1}{2}\left[\cos \omega_c(2t - \tau) + \cos \omega_c \tau\right] \tag{2.23}$$

Substituting for τ and filtering the high frequency components leaves the phase term only,

$$v_{out}(t) = \frac{1}{2}\cos \frac{2\omega_c R}{c}. \tag{2.24}$$

The phase shift, ϕ, can be obtained by replacing ω_c by $2\pi f_c$

$$\phi = \frac{4\pi f_c R}{c}. \tag{2.25}$$

It can be seen that the phase shift will change with each new frequency f_c

This phase change is sinusoidal in nature and can be considered to be a synthetic Doppler frequency and can be processed in the frequency domain to determine the target range more accurately.

If a number of closely spaced reflectors are present they will each have their own unique Doppler frequency and so can each be resolved as a separate target.

2.13. Convolution

2.13.1. Linear Time Invariant Systems

It is convenient to describe the relationship between the input $x(t)$ and output $y(t)$ signals of Linear Time Invariant (LTI) systems in terms of the impulse response $h(t)$ of the system. It can be shown that an LTI system is completely characterised by its impulse response $x(t) = \delta(t)$ applied at time $t = 0$ (or $n = 0$).

Applying an impulse to the input of an unknown LTI system is therefore a good method of determining its characteristics. This is easily achieved in the discrete time case where the input is set equal to an impulse $\delta(n)$. However, in the continuous time case, it is not possible to produce a true impulse with zero width and infinite amplitude and an approximation is used.

As an arbitrary input signal can expressed as the weighted superposition of time shifted impulses $x(t) = x(\tau)\delta(t-\tau)$, the system output is just the weighted superposition of time-shifted impulse responses.

$$y(t) = \int_{\tau=-\infty}^{t} h(t - \tau)x(\tau)d\tau = \int_0^{\infty} h(\tau)x(t - \tau)d\tau \tag{2.26}$$

This weighted superposition is termed the *convolution integral* for continuous time systems and the *convolution sum* for discrete time systems.

Given a system defined by its impulse response $h(t)$, the output, $y(t)$ is found by convolving the input $x(t)$ with this function.

If the Laplace transform is taken of this convolution integral, it can be shown that

$$Y(s) = H(s)X(s) \qquad\qquad (2.27)$$

That is, convolution in the time domain is equivalent to multiplication in the Laplace domain. Additionally, if s is replaced by $j\omega$, then the equation can be replaced by

$$Y(\omega) = H(\omega)X(\omega) \qquad\qquad (2.28)$$

which describes the frequency response of the system. It is the magnitude of this transfer function, $|H(\omega)|$, that is considered when the characteristics of various filters are considered in the following section.

In this case, the mapping between the time domain and the frequency domain is through the Fourier Transform. In the following example, the impulse response of a 3rd order Butterworth bandpass filter (normalised passband 0.2 to 0.3) is generated. As this characterises the filter completely, its Fourier Transform should reproduce the frequency response accurately.
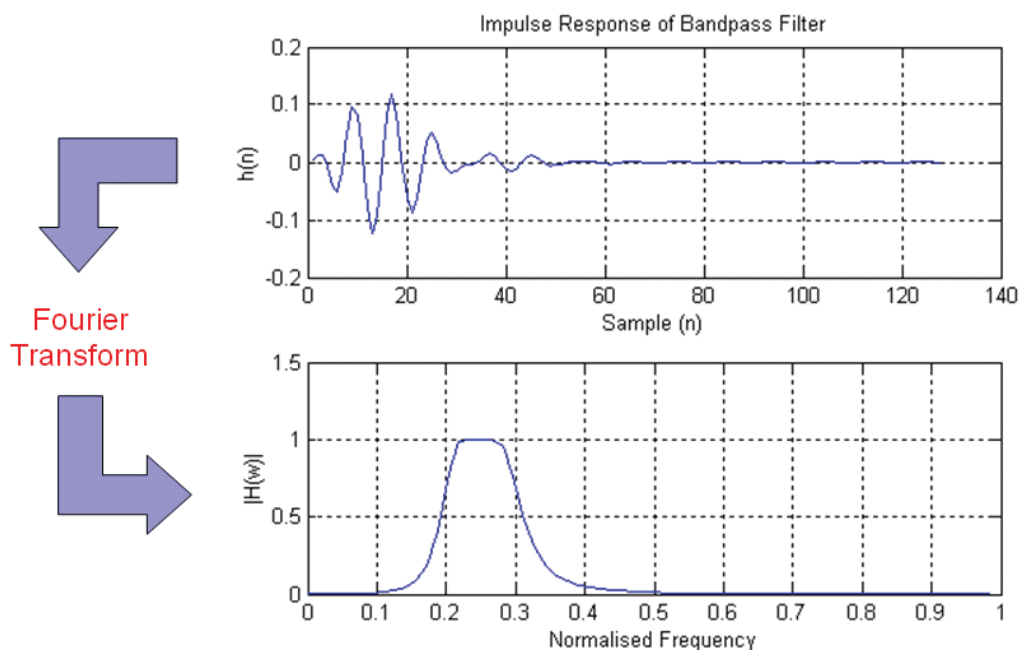

Figure: MATLAB example showing the relationship between the impulse response of a bandpass filter and its Fourier Transform

```
% relationship between filters in the time and frequency domain
%
% generate a Butterworth bandpass filter
wn=[0.2,0.3];
[b,a]=butter(3,wn)

% generate the impulse response of the filter
x=zeros(1,128);
x(1)=1;
y=filter(b,a,x);
subplot(211), plot(y),grid, xlabel('Sample (n)'), ylabel('h(n)')
title('Impulse Response of Bandpass Filter')

% take the Fourier transform of the impulse response
f=fft(y);
freq=(0:63)*1/64;
subplot(212), plot(freq, abs(f(1:64))), grid, xlabel('Normalised
Frequency'),ylabel('|H(w)|')
%title('Frequency Response')
```
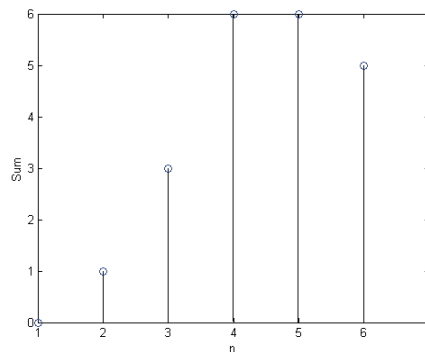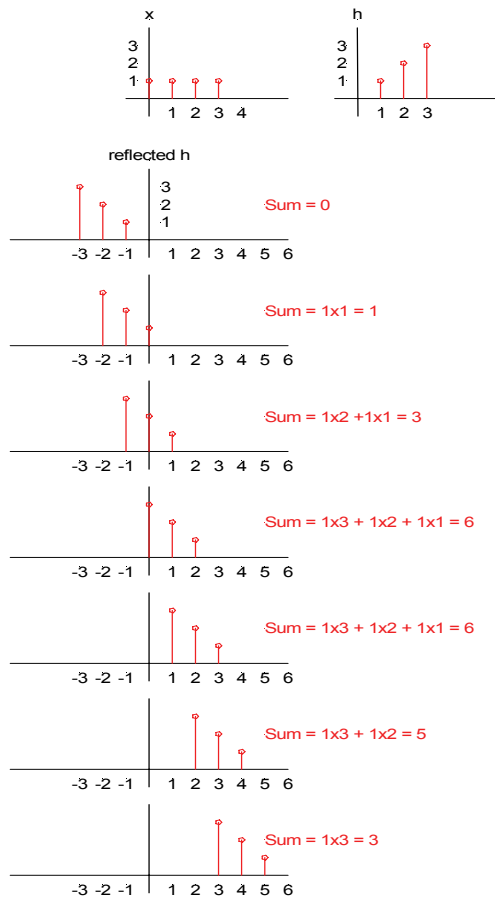
2.13.2. The Convolution Sum

Most modern systems are digital which requires that the input and output data be sampled. In this case the *convolution sum* replaces the convolution integral and is defined by the following equation where $y(n)$ is the output, $x(n)$ the input and $h(n)$ the system impulse response

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k) \tag{2.29}$$

However, to be tractable $x(n)$ and $h(n)$ are nonzero only over a finite interval. In the example below $N_x = 4$ and $N_h = 3$ which makes the total duration of the convolved sequence $N_x + N_h - 1 = 6$

To perform this function manually, the order of $h(n)$ is first reversed before being shifted across $x(n)$ one sample at a time. The output $y(n)$ is equal to the sum of the products of each of the aligned terms

```
% convolution example
% convsum.m
x = ([1,1,1,1]);
h = ([0,1,2,3]);
y = conv(x,h);
plot(y,'o')
```

2.13.3. Convolution Example

An unusual application for the convolution process is to use it to model the propagation of a pulsed time-of-flight sensor.

In this application the impulse response of the round-trip propagation to the point target is a signal that is delayed in time and attenuated in amplitude $h(t) = a\delta(T-\tau)$ where a represents the attenuation and τ the round trip time to the target and back.

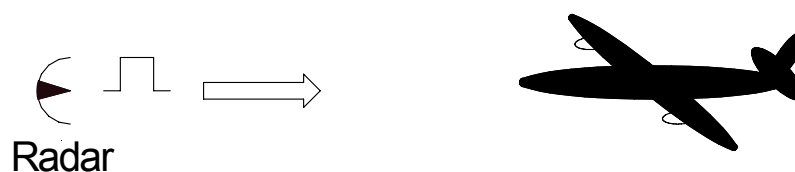Consider an aircraft flying towards a radar system as shown in the following figure:

Figure 2.37: Radar illuminating an aircraft target

Assume that the main points of reflection from the aircraft are listed in the following table:

Table 2.1: Reflection contributions from various parts of the aircraft

Contribution	Distance from Datum (m)	Magnitude
Nose	2	1
Wing	6	1
Engine	8	1
Tail	14	1

Determine the magnitude of the echo signal if a pulse with a length of 3m is transmitted.

- The transmit signal is treated as a sequence of impulses separated by the sample interval of 10cm covering a total of 3m in range
- The target is treated as a sequence of four impulses with separations as listed in the table
- The reflected signal received back at the radar is well described by the convolution of the transmitted pulse and the array of point reflectors as shown in the figure.

```
% convolution demo
%
% generate the transmit pulse as a sequence of impulses 10cm apart
a=([zeros(1,70),ones(1,30),zeros(1,70)]);

% generate the point target reflectors
b=zeros(1,170);
b(20)=1;                    % Nose
b(60)=1;                    % Wings
b(80)=1;                    % Engine
b(140)=1;                   % Tail

% take the convolution to determine the return from all of the reflectors
c=conv(a,b);

% plot the results
x=(1:170)/10;
subplot(211) ,plot(x,a,x,a,'o'), grid;
title('Transmit Pulse'),ylabel('amplitude');
subplot(212), plot(x,b,x,b,'o'),grid;
title('Target Reflectors'),xlabel('range(m)'),ylabel('amplitude');
figure(2)
xx=(1:339)/10;
plot(xx,c,xx,c,'+');
grid
title('Echo Pulse')
xlabel('range(m)')
ylabel('amplitude')
```
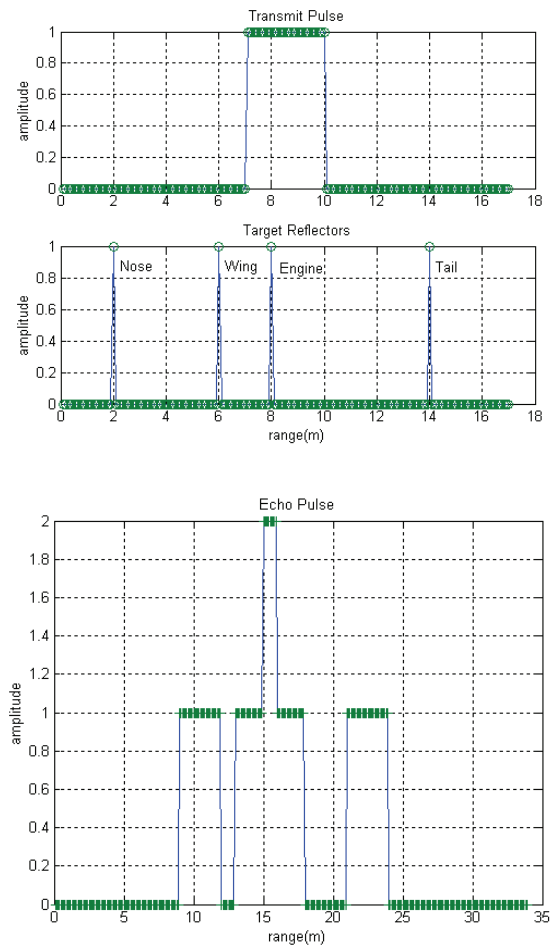
Figure 2.38: Matlab simulation output showing transmit signal, target structure and echo signal

This example does not consider that the transmitted signal is in fact sinusoidal in nature and hence the convolution should include both amplitude and phase effects. Nor does not consider the effects of the round-trip distance which doubles the effective spacing between the reflectors as explained in Chapter 5.

2.14. References

[1]   N.Taub & D.Schilling, *Principles of Communication Systems.* McGraw Hill. 1971
[2]   B.Mahafza, *Radar Systems Analysis and Design using MATLAB.* Chapman & Hall/CRC. 2000
[3]   N.Currie, *Radar Reflectivity Measurement: Techniques and Applications.* Artech House, 1989
[4]   M.Skolnik, *Radar Handbook 2nd Ed.* McGraw Hill. 1990
[5]   N.Currie & C.Brown. *Principles and Applications of Millimeter-Wave Radar.* Artech House. 1987
[6]   T.Duncan, *Electronics and Nuclear Physics.* John Murray. 1973
[7]   J.Carr, *Electronic Circuit Guidebook: Sensors.* Prompt. 1997
[8]   H.Jacobovitz, *Electronics Made Simple.* W.H.Allen. 1972
[9]   *The ARRL Handbook for Radio Amateurs*: 2000
[10]   S.Haykin & B.van Veen, *Signals and Systems.* John Wiley. 2003
[11]   H. Durrant-Whyte, *Signal Analysis and Transform Methods,* Course Notes