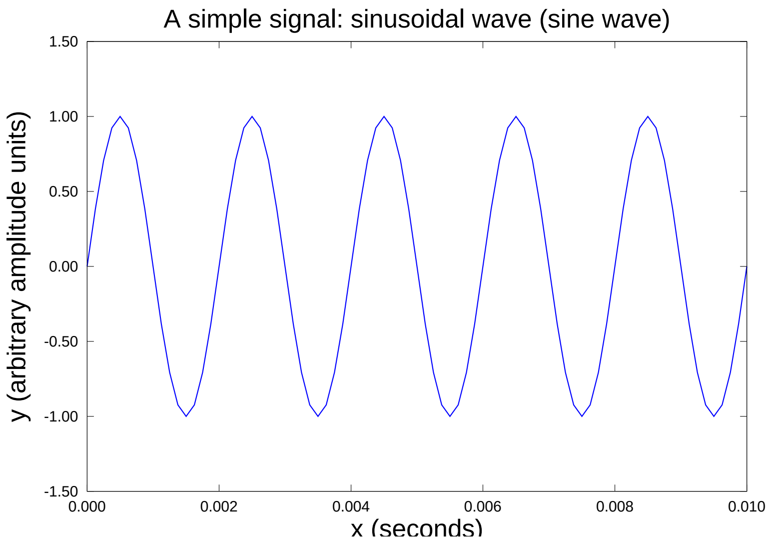
Introduction to Digital Signals

Contents

1 What is a Signal?

DSP is a about processing of signals that are typically in digital form. Signals are therefore very important to DSP. A simple example of a signal is a sine wave, e.g.



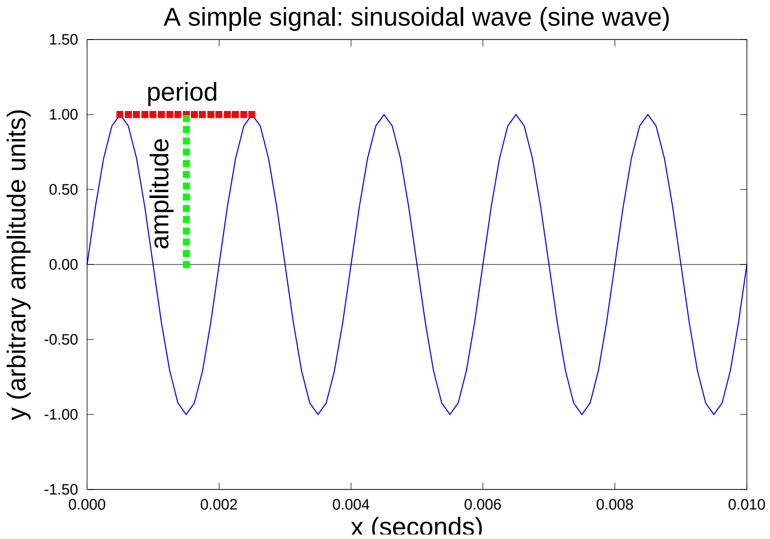
A signal can contain information for many reasons including:

- *Communication*
- *Storage*

- *Calculation*

Information is carried in the signal typically by varying one or more of the following parameters:

- *amplitude, "a"*;
- *period, "T"*;
- *frequency, "$f = 1/T$"*;
- and *phase, "ϕ"*.



We should all be familiar with the equation for a sine wave:

$$y(x) = a \sin(2\pi f x + \phi)$$

where "x" is time in seconds for this example. Amplitude "$a = 1$" controls the height of the wave.

Frequency and Period are very important. In particular, the topic of frequency is studied in great detail in DSP:

- f is the frequency
- Measured in Hertz or Hz
- Here period, $T = 0.002$s
- $f = 1/T$ Hz, therefore $f = 1/0.002 = 500$Hz.

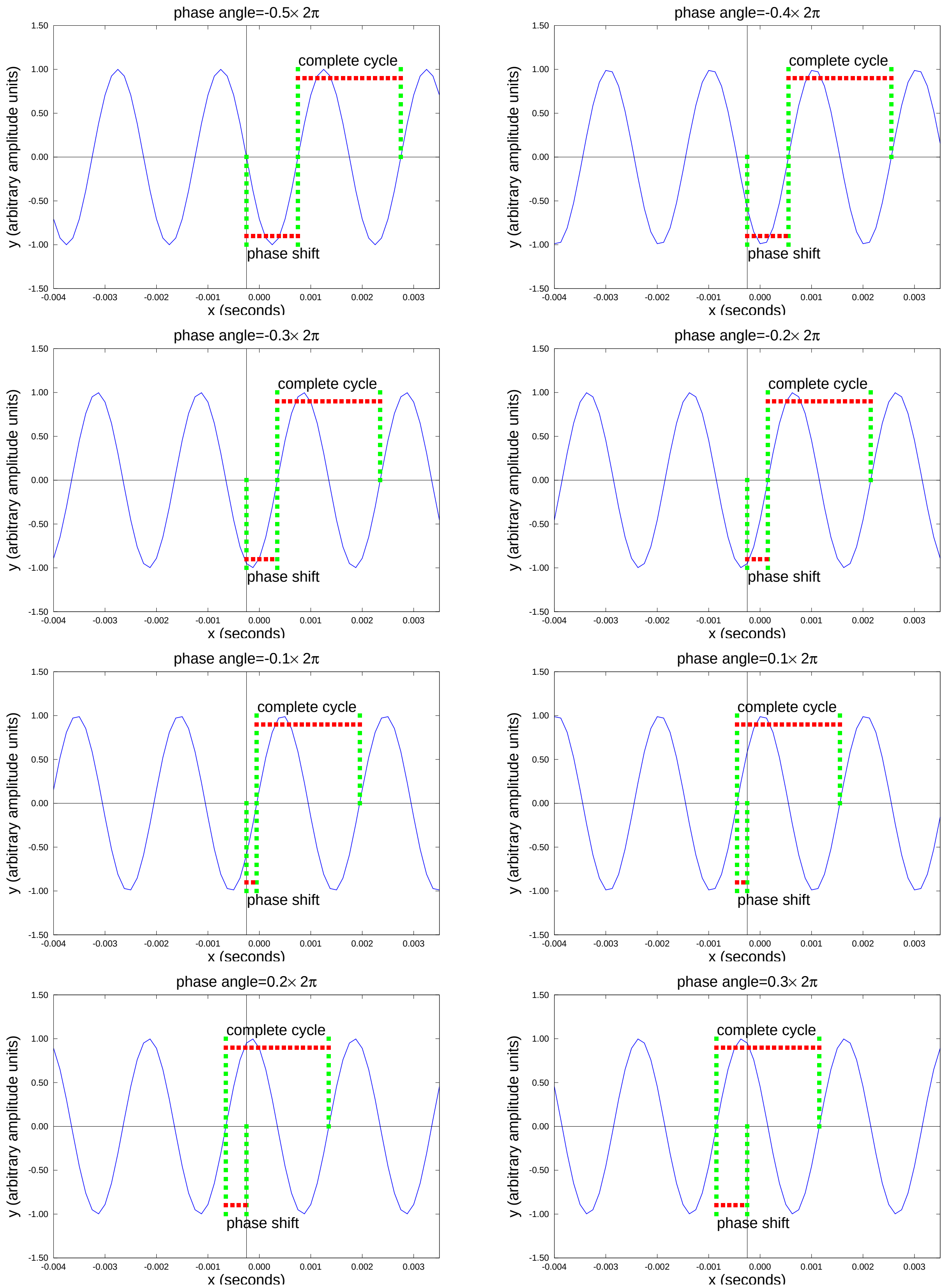
Phase Phase is important as it indicates the signal has been delayed for some reason. This may be used to encode important information, e.g. in Phase Shift Keying.

- ϕ is the phase
- Here $\phi = 0$

For the above example, $\phi = 0$, so

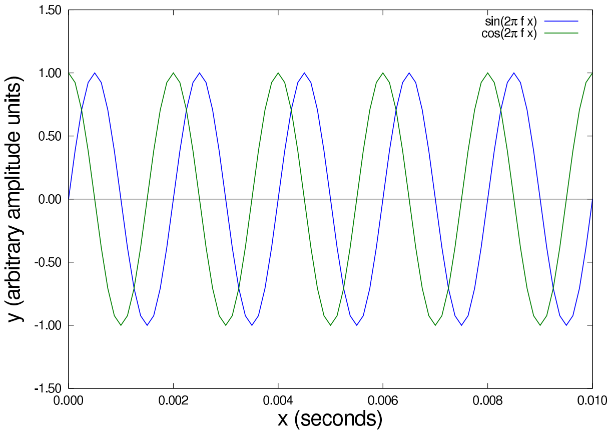
$$y(x) = y(x, \phi = 0) = a \sin(2\pi f x).$$

Examples of non-zero phase can be seen in the following:



1.1 Cosine Vs Sine

Cosine and Sine functions are equivalent except for a phase shift ($1/4\times$period).

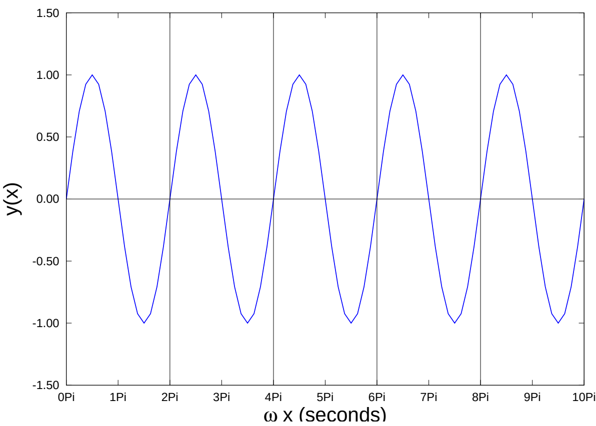


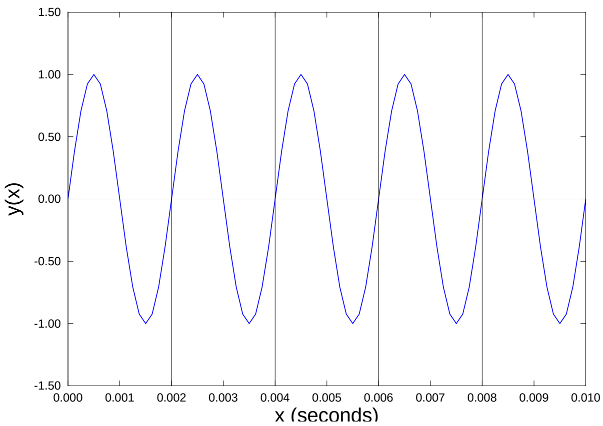
- $\cos(2\pi f x) = \sin(2\pi f x + \phi)$ where $\phi = \pi/2$.
- $\sin(2\pi f x) = \cos(2\pi f x + \phi)$ where $\phi = -\pi/2$.

1.2 Angular Frequency

- Frequency, $f = 1/T$
- Angular frequency, $\omega = 2\pi f$
- 1 period or cycle $= 2\pi$ radians

$$\begin{aligned} y(x) &= \sin(2\pi f x + \phi) \\ &= \sin(\omega x + \phi) \end{aligned}$$





1.3 Phasor Representation

As already mentioned, signals such as sine waves have frequency associated with them as an inherent property, especially for signals that *repeat* or which are periodic. In any case, the concept of frequency is something that we will be looking at in great detail in this course using techniques such as the Fourier transform. A Fourier transform is a mathematical operation that converts the signal to a representation that helps us to understand the frequency content of a signal. We will not just yet look at the Fourier transform. However we do need to know about complex numbers as the Fourier transform is defined in terms of complex numbers. So let us remind ourselves about complex numbers. Recall that a cosine (or sine) wave:

$$y(x) = a\cos(\omega x + \phi)$$

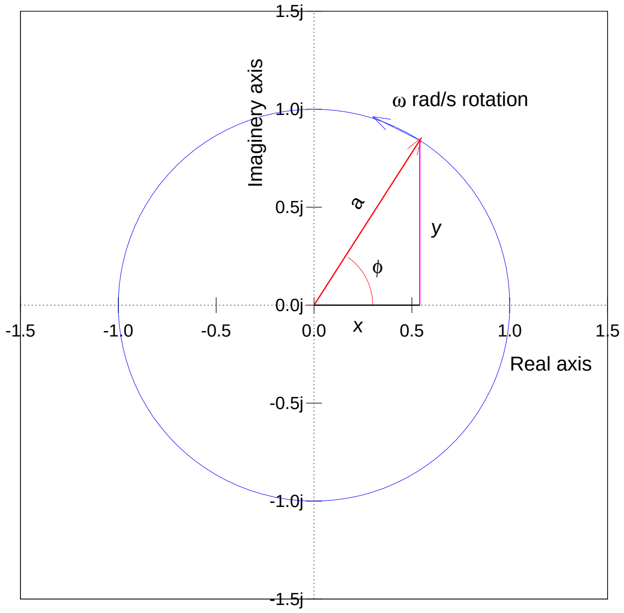
can be represented as a phasor. A phasor is a complex number:

$$z = x + jy = a(\cos(\phi) + j\sin(\phi))$$

where x is known as the real part or $\mathrm{Re}(z) = x$ and y is known as the imaginery part or $\mathrm{Im}(z) = y$. x and y can be calculated with $x = a\cos(\phi)$ and $y = a\sin(\phi)$.

Also remember $j = \sqrt{-1}$.

Argand *or* Phasor Diagram:



1.4 Complex Numbers

The square root of minus one is not defined so a symbol, j is used (sometimes i):

$$j = \sqrt{-1}.$$

Powers of imaginary numbers have the following properties:

- $j^2 = -1$
- $j^3 = -j$
- $j^{-1} = \frac{1}{j} = -j$

If $z = x + jy$ (rectangular form) then alternative representations are:

- Polar form: $z = a\angle\phi$
- Exponential form: $z = a\exp(j\phi)$

where $a = \sqrt{x^2 + y^2}$ and $\phi = \tan^{-1}(y/x)$.

1.5 Properties of Complex Numbers

If $z = x + jy$, $z_1 = x_1 + jy_1$ and $z_2 = x_2 + jy_2$ then

- Addition: $z_1 + z_2 = x_1 + x_2 + j(y_1 + y_2)$
- Subtraction: $z_1 - z_2 = x_1 - x_2 + j(y_1 - y_2)$
- Multiplication: $z_1 z_2 = a_1 a_2\angle(\phi_1 + \phi_2)$
- Division: $z_1/z_2 = a_1/a_2\angle(\phi_1 - \phi_2)$
- Reciprocal: $1/z = 1/a\angle(-\phi)$

- Square root: $\sqrt{z} = \sqrt{a}\angle(\phi/2)$
- Complex conjugate: $z^* = x - jy = a\angle - \phi$

The polar form simplifies some operations such as multiplication and division of complex numbers.

Euler's identity is an important identity in DSP because it tells us that a complex exponential is actually equivalent to a real cosine signal combined with an imaginary sine part.

$$\exp(j\phi) = \cos(\phi) + j\sin(\phi)$$

The complex exponential is used in the Fourier transform and in many other aspects of DSP. Using Euler's identity we can state the following:

- $\cos(\phi) = \mathrm{Re}(\exp(j\phi)) \longrightarrow$ or the real part, x
- $\sin(\phi) = \mathrm{Im}(\exp(j\phi)) \longrightarrow$ or the imaginary part, y

Recall the cosine wave:

$$y(x) = \cos(\omega x + \phi)$$

which can be written as:

$$y(x) = \mathrm{Re}(a\exp(j(\omega x + \phi))) = \mathrm{Re}(a\exp(j\omega x)\exp(j\phi))$$

$$= \mathrm{Re}(A\exp(j\omega x))$$

where A is the phasor representation of $y(x)$ given by

$$A = a\exp(j\phi) = a\angle(\phi).$$

1.6 Complex Exponentials, Sines and Cosines

The complex exponential has the following properties:

- $y_1(x) = b\exp(j\omega x) = b\cos(\omega x) + jb\sin(\omega x)$
- $y_2(x) = b\exp(-j\omega x) = b\cos(\omega x) + jb\sin(-\omega x)$
  as
  - ◇ $\cos(-\omega x) = \cos(\omega x)$ (even function)
  - ◇ $\sin(-\omega x) = -\sin(\omega x)$ (odd function)

This means that

$$y_1(x) + y_2(x) = 2b\cos(\omega x).$$

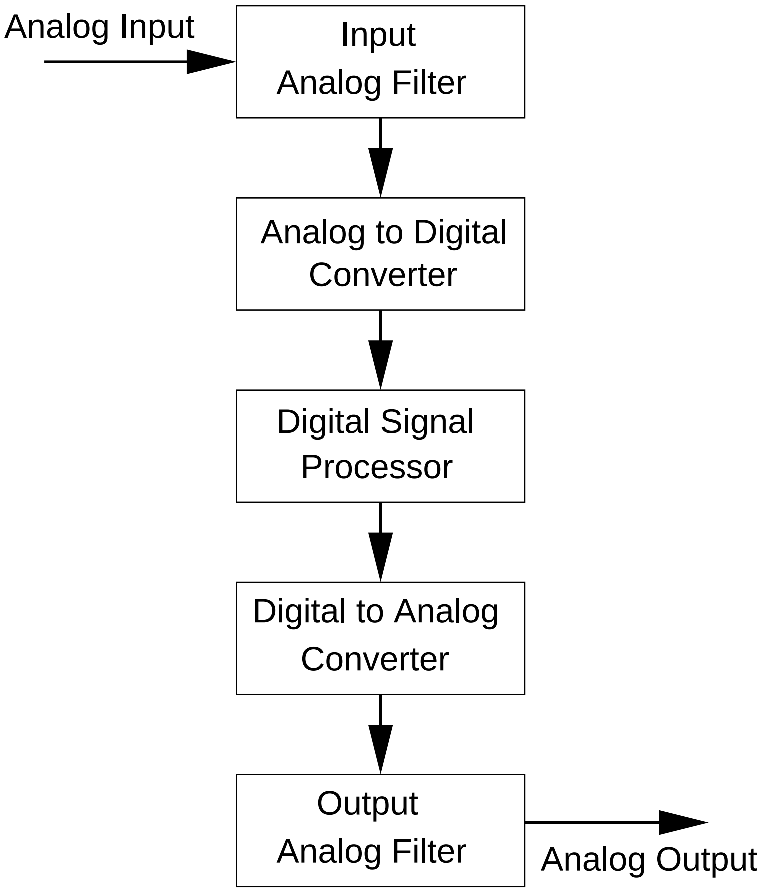
So that

$$b\cos(\omega x) = \frac{a}{2}\exp(j\omega x) + \frac{a}{2}\exp(-j\omega x).$$

A similar approach can be used to derive a sine function.

2 A Digial Signal Processing System

A typical DSP system consists of a number of stages, often including an input i n the form of an analogue signal and perhaps even an output in the form of processed analog signal. In between the steps to represent the signal in the digital form will include a Digital Signal Processor that performs some sort of DSP function:



In summary the steps illustrated above perform the following operations:

- Input Analog Filter (antialiasing): *Limits frequency range*;
- Analog to Digital Converter: *Converts signal to digital samples*;
- Digital Signal Processor: *Storage, Communication and or Calculations*;
- Digital to Analog Converter: *Convert to continuous signal*;
- Output Analog Filter: *Removes sharp transitions*.

2.1 Analog to Digital Converter

The analog to digital converter is an important component in many DSP systems and even many other digital systems that may not be immediately considered to be a DSP type system.

- Real world is typically *analog* (continuous) however computers are digital and if we want to perform some sort of computer operation on a signal we therefore need to convert it to digital form.
- Digital signal approximates analog signal with discrete quantised samples
- ADC converts an analog signal to a digital signal
- Signal is digitised in two ways:
  - ◇ Signal is sampled at a sampling rate or frequency: Information is collected about the signal at regular intervals.
  - ◇ The continuous or analog signal is then quantised: *i.e.* put into digital form, where only a finite set of numbers are represented.

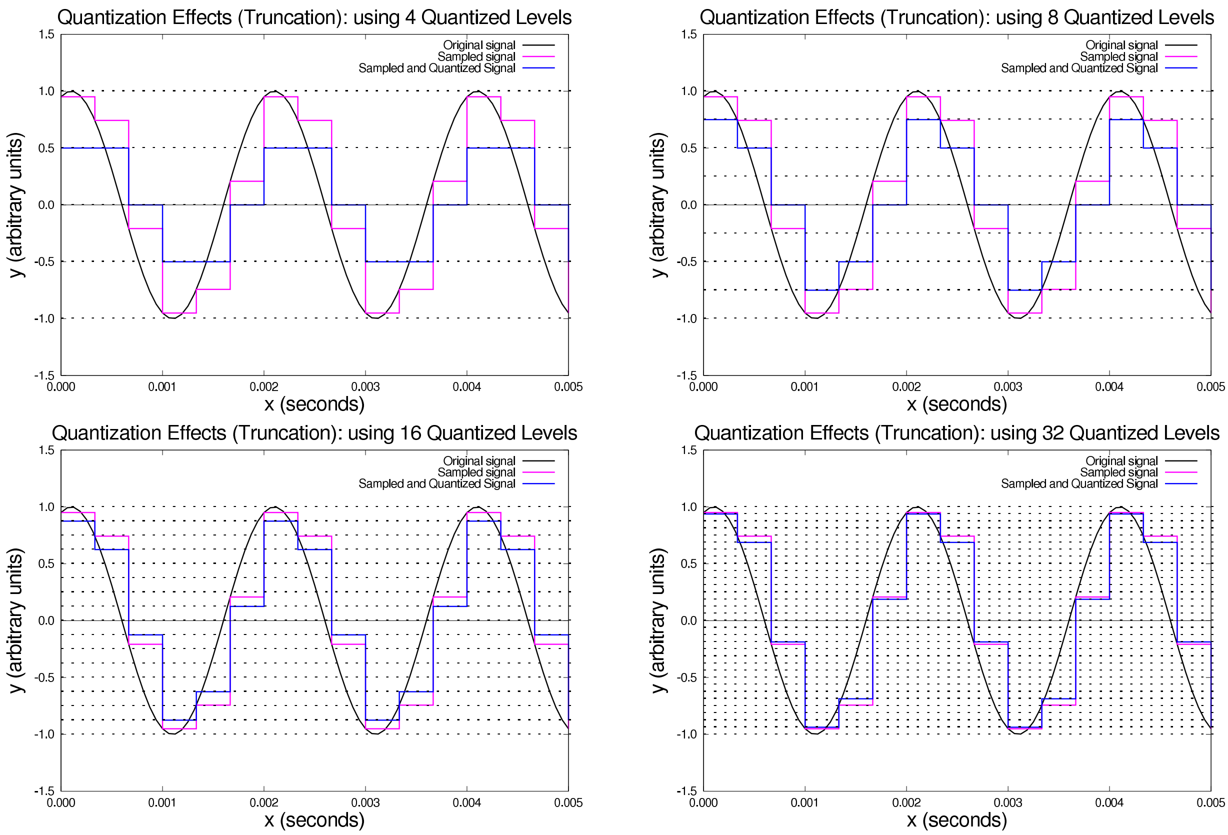
2.2 Quantisation using Truncation

Unfortunately the digitisation or ADC does remove some important information. Part of the digitisation process is known as quantisation which means:

- Signal can be quantised using *e.g.* truncation where numbers following specified position are removed.
- Examples:
  - ◇ 5.7 truncated to integer is 5
  - ◇ 5.11 truncated to 1 decimal place is 5.1
- Negative numbers are truncated in the same way (note different to the common *floor* function in matlab), *e.g.*
  - ◇ -5.78 truncated to integer is -5
  - ◇ -5.135 truncated to 2 decimal places is -5.13

2.3 Truncation Quantisation ***examples***

It is important to understand the implications of quantisation. Illustrations of various levels of quantisation can be seen in the following:



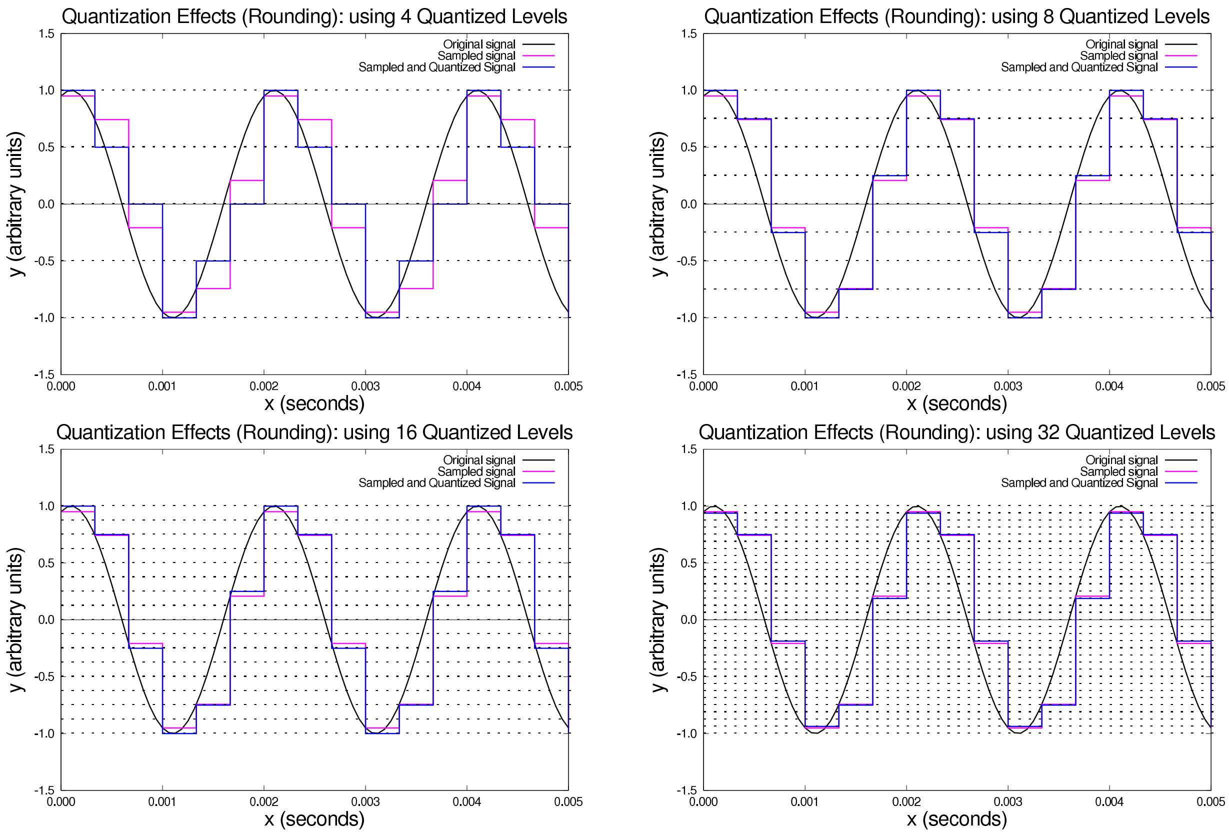
- Errors can be seen between the sampled and the sampled and quantized signals.

These errors are typically referred to as quantisation noise.

2.4 Quantisation using Rounding

- Rounding can be a quantization method associated with smaller errors, *e.g.*
  - ◇ 5.7 rounded to nearest integer is 6
  - ◇ 5.11 rounded to 1 decimal place is 5.1
  - ◇ -5.78 rounded to nearest integer is -6
  - ◇ -5.135 rounded to 2 decimal places is -5.14

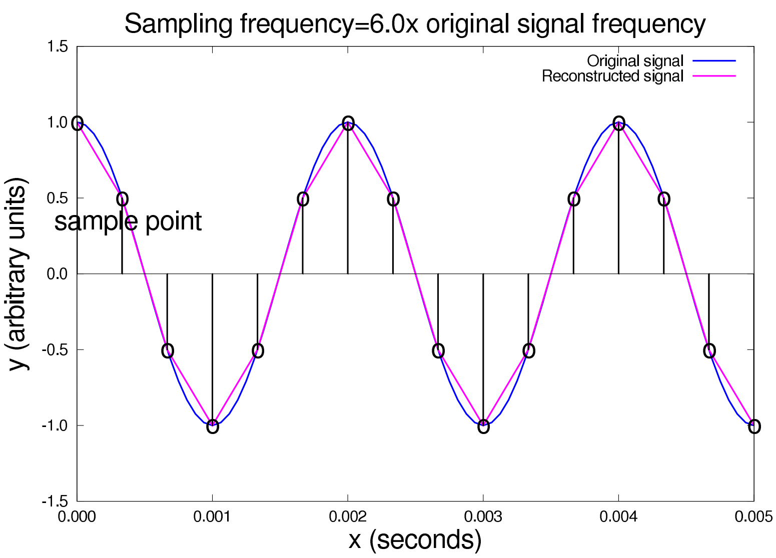
2.5 Rounding Quantisation *examples*



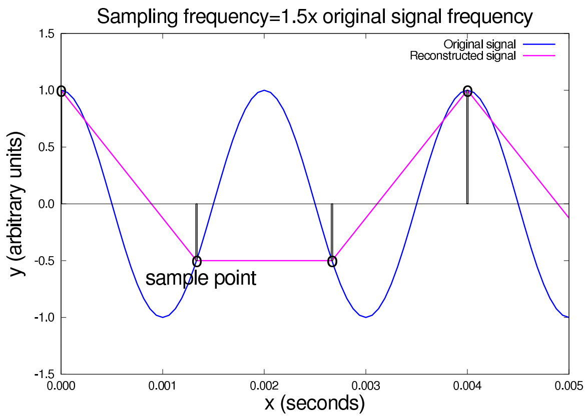
- Errors can be seen between the sampled and the sampled and quantized signals.

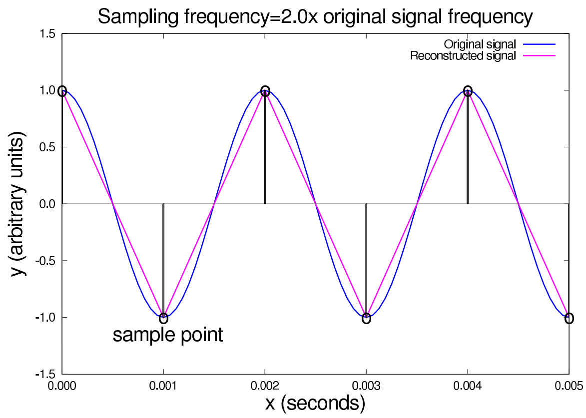
2.6 Sampling

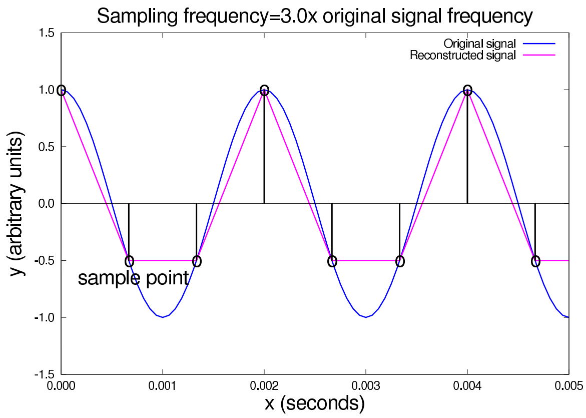
- Sampling also affects the quality of the digitised signal.
- Higher sampling rate reduces error and enables better representation of the original analog signal in digital form.

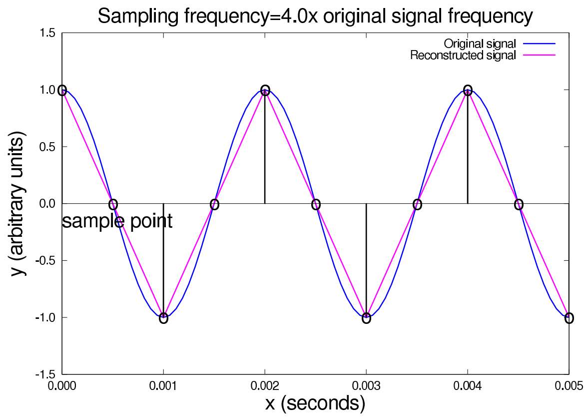


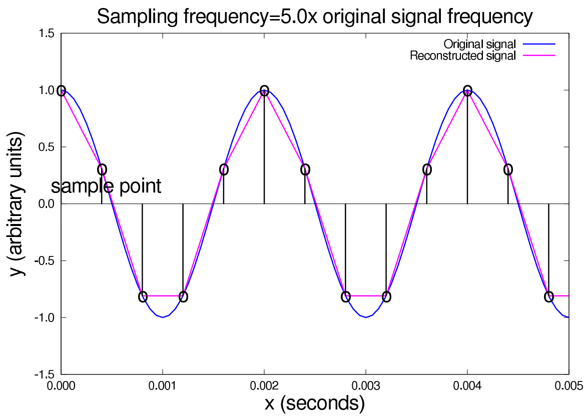
2.7 Sampling *examples*

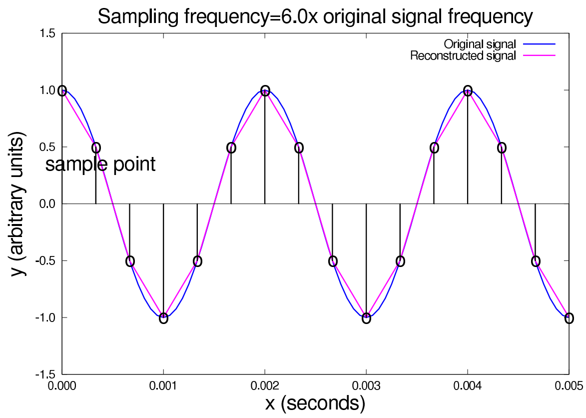


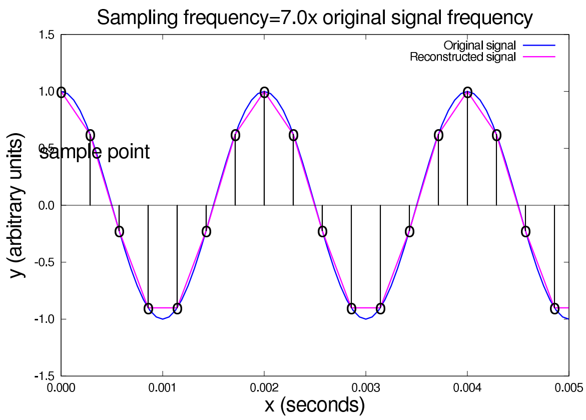


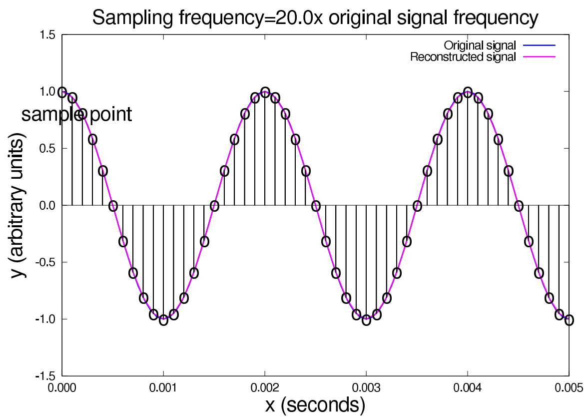






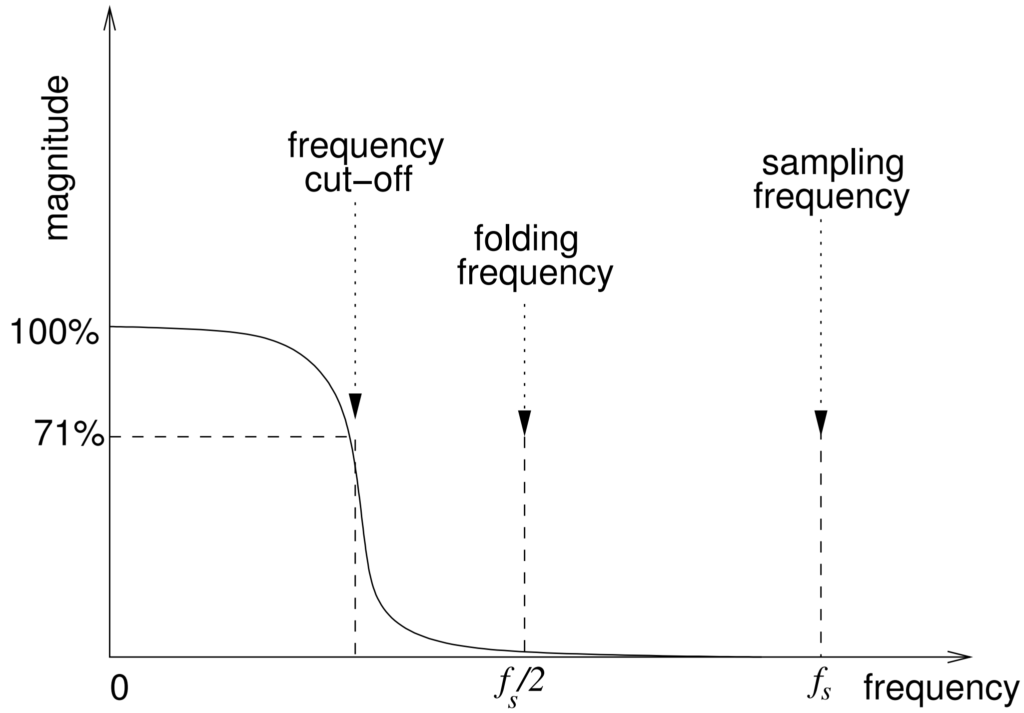






2.8 Input Analog Filter: Antialiasing Filter

- Analog to Digital Converter (ADC) requires signal below a particular frequency known as the folding or Nyquist Frequency[1]
- $\therefore$ Limit frequency range to below Nyquist frequency ($f_s/2$) before Analog to Digital Conversion. The antialiasing filter needs to remove all signal above the Nyquist frequency otherwise aliasing will occur where the inadvertently high frequency signal content will not be sampled at a sufficiently high enough frequency and aliasing frequencies will occur in the digitised signal.



maximum frequency < folding frequency

- Otherwise next stage produces frequency errors (*i.e.* aliasing)
- Sampling produces copies of signal at multiples of sampling frequency
- Aliasing occurs when copies of signal overlap each other

Aliasing frequencies will also be present in the signal even after conversion back to analog form. So all frequencies have to be removed above the folding frequency. This means that the cut-off frequency of the antialiasing filter needs to be below the folding frequency because there will still be signal content above the cut-off point. The cut-off frequency is the frequency at which the filter is removing the signal content by half-power (or 71% magnitude). This means there will still be signal content above the cut-off frequency. So the cut-off frequency has typically to be much less than half the sampling frequency.

2.9 Digital to Analogue Conversion

(DAC) is the process of converting a digital signal back to analogue form. The initial stage of a DAC will often take the individual samples and produce a waveform composed of those digital samples. This waveform will however often resemble a staircase because it is not continuous and it will still contain the discrete changes in the quantisation. This is known as quantisation noise. Often a low pass filter will then be applied, after the DAC to remove this quantisation noise by smoothing the staircase like waveform. This is possible because the sudden changes in the staircase waveform are composed of many high frequencies. Low pass filtering the signal removes these high frequencies and allows the lower frequency part of the signal to be retained which should represent the signal that was original converted into digital form.

[1] Nyquist was a scientist working at Bell Labs in the 20th century. His first name was Harry.

3 Basic Communication System

A communication system is a system in which a signal is communicated from an (information) source to an (information) sink. A communication system can be analogue or digital. A communication system will typically involve some sort of coding. Coding in communication systems can mean something quite specific. But for the purposes of this introduction it means preparing a signal so that it can arrive at the information sink without too much degradation. Degradation may occur possibly because of problems in the channel.

A signal will often be modulated by the transmission side of a communication system and then demodulated by the receiving side of a communication system. This modulation / demodulation process is where the information signal is combined with another usually much higher frequency type of signal to enable it to be effectively carried across, usually a radiofrequency channel.

Other intentional changes to a signal prior to transmission can include: encryption to hide the signal from anyone other than the intended recipient; compression to reduce the signal bandwidth by removing unecessary parts of the signal that are not needed by the receiver; and also channel coding to help improve the possibility of the signal being irreversibly damaged because of some degrading effect that might occur as part of the transmission process.

4 Frequencies

A high frequency signal is a signal that varies very quickly; a low frequency signal varies much more slowly. A simple example of a high frequency signal in comparison to a low frequency signal could include a sine wave $\sin(2\pi f t)$ with different values for the frequency variable f:

$$x_1(t) = \underbrace{\sin(2\pi 0.1 t)}_{f=0.1} \quad \text{(A)} \qquad\qquad x_2(t) = \underbrace{\sin(2\pi 1000 t)}_{f=1000} \quad \text{(B)} \tag{1}$$

For equation (1.A) the frequency variable is much smaller $f = 0.1$ than the frequency variable for equation (1.B) $f = 1000$. This means that the signal represented by $x_1(t)$ completes one cycle every 10 seconds but the signal represented by $x_1(t)$ completes 1000 cycles every second! This is quite a big difference in frequency.

These signals are represented in the time domain. However think about an alternative representation where we have frequency along the horizontal axis. We can then think that $x_1(t)$ transformed to this alternative representation would show a single point at $f = 0.1$ and then the signal $x_2(t)$ transformed to this alternative representation would show a single point at $f = 1000$. This frequency representation does exist and is known as the Fourier domain. The Fourier transform can transform a function from the time domain to the frequency domain. We will meet this transform again in a few weeks time and look at it in more detail.

An important characterisation of an engineering system is the way the system responds to different frequencies. Some frequencies might be amplified (made stronger), other frequencies might be attenuated (reduced in strength). Consider a system that completely attenuates high frequencies (e.g. above $f = 10$). Now imagine a signal composed of two sinusoidal frequencies like so:

$$x(t) = \sin(2\pi 2t) + \sin(2\pi 20t). \tag{2}$$

If this signal was input to the system that attenuates high frequencies then the output of the system could be a new signal:

$$y(t) = \sin(2\pi 2t) \tag{3}$$

where the higher frequency component $\sin(2\pi 20t)$ has been completely removed from the signal. A system such as this has a frequency response that can be considered as a low pass filter.

In communication systems a signal composed of frequencies that range from 0 Hertz up to some other frequency is known as a **baseband system**. Examples of baseband systems might be a communication system that communicates over an actual cable not over the air where the original signal is baseband and the communication channel does not require any form of modulation. In contrast to this transmission via radiofrequencies in the air means that the signal has to be carried by a carrier frequency signal and the resulting modulated signal will have a lowest frequency well above 0 Hertz. This type of system is known as a **bandpass system**.

A communication system's range of frequencies is an important property of a communication system. This range of frequencies is known as the *Bandwidth* of a system or signal. The bandwidth of these types of systems are typically defined in terms of the frequency range of a typical signal. For example, by definition a baseband system will operate correctly across a range of frequencies starting from 0 Hertz; such a system may only operate correctly for this range of frequencies and may in fact not actually attenuate frequencies above or below its specified limit.

4.1 Fractional bandwidth

ΔB is a useful measure of a communication system's frequency content defined by

$$\Delta B = \frac{B}{f_c} \tag{4}$$

which relates the bandwidth B with the centre frequency f_c. If f_c is much higher than half the bandwidth then the system or signal can most definitely be classed as bandpass. However if $f_c \leq B/2$ then the system is baseband. This leads to the following classification of signals:

- If $0 \leq \Delta B \leq 1$ then the signal or system is **baseband**;
- If $1 \leq \Delta B \leq 2$ then the signal or system is **practically baseband**.
- If $\Delta B > 2$ then the signal or system is **bandpass**.

5 Example Type of Signals: Audio Signals

Audio is a very common signal type. However there are different types of audio signal depending on the intended application. Human hearing has a relatively wide *bandwidth* ranging between 20Hz upto around 20kHz with a dynamic range of around 140 decibels. Dynamic range quantifies the range of intensities (on the vertical axis) that a signal can carry. Analogue dynamic range is dependent on the minimum and maximum measurable signal values, e.g. for $V_{\min}$ and $V_{\max}$ then:

$$\text{DNR}_{\text{analogue}} = 20 \times \log_{10}\left(\frac{V_{\max}}{V_{\min}}\right). \tag{5}$$

Digital dynamic range is quantified slightly differently because the minimum and maximum measurable signal values are dependent on the number of steps that the signal has been divided on the vertical axis, m. So that:

$$\text{DNR}_{\text{digital}} = 20 \times \log_{10}(m). \tag{6}$$

The range of frequencies for which the typical human ear can hear is far greater than what is needed to communicate via speech. For example, the old fashioned telephone is typically specified to have a frequency range from 300 Hertz up to 3.4kilo Hertz with a dynamic range of only 48 decibels. This frequency range is much less than the range of frequencies that the human ear can actually hear. This is one of the reasons why music played down the telephone whilst we wait to speak to a person may often sound strange or of poor quality. However when we actually speak to someone, we can usually easily understand what they are saying. So the most useful parts of human speach actually do not occupy the majority of the *bandwidth* of the human hearing system. *Why is the frequency range of a telephone system exclude the higher frequencies?* As discussed, the higher frequencies are somewhat redundant because the speech can be understood without retaining and reproduction of the higher frequencies. Bandwidth is an important commodity. If a signal has a large bandwidth then it is more expensive to communicate because more wires are needed, higher computing power is needed (faster CPUs, more computers) and fewer telephone calls could be carried at the same time. Reducing the bandwidth of the speech signal to the minimum necessary to enable most speech to be understood removes some *redundancy* and makes it more efficient to communicate.