Chapter 1: Definitions and Applications of DSP

1.1 What is digital signal processing?

Digital systems are all the rage these days. Mention that a consumer product is digital and it's sure to sell. For home computers, MP3 players, mobile telephones, DVD players and surround sound systems, sales have never been better. Why is this? What does digital mean anyway? Is it true that digital implies high quality, better than we could get from traditional analogue techniques? The answer to the last question is yes, but why this is so may not always be obvious. And where does DSP sit in all of this? If DSP involves the processing of digital data, surely everything that we do on computers can be said to be DSP – including, for example, word processing or web browsing? Well not really. Most practitioners would agree that, strictly speaking, DSP involves manipulation of signals that have their origins in the analogue world. Such signals may be produced for example, by video, audio, radio telemetry, radar, thermal, magnetic or ultrasonic sensor systems, to name but a few from a truly enormous range of devices and instruments. The point here is that the signals are originally *analogue* and *continuous* in nature to start with (don't worry too much about these terms just yet – just think about a microphone. It produces a continuous electrical signal whose magnitude is proportional to the intensity of sound it detects).

The problem with artificial definitions of this kind is that there are always grey areas that do not lie within their boundaries, yet should still be included within the definition, in this case because of the nature of the operations applied. Take for example, computer networking and data transferral. In many cases, digital information is passed between computer and computer, information that owes nothing to the analogue world (for example a file representing a word processed document). The information may be encoded in a variety of lossless compression formats, to minimise the time and bandwidth constraints on the network. Undoubtedly, compression is an important aspect of DSP, and so the original limited definition given above is, as we see, not entirely accurate. Despite this, in the great majority of cases, DSP is used to enhance or change (sometimes even to degrade) signals obtained from "real world" sensors and instrumentation systems.

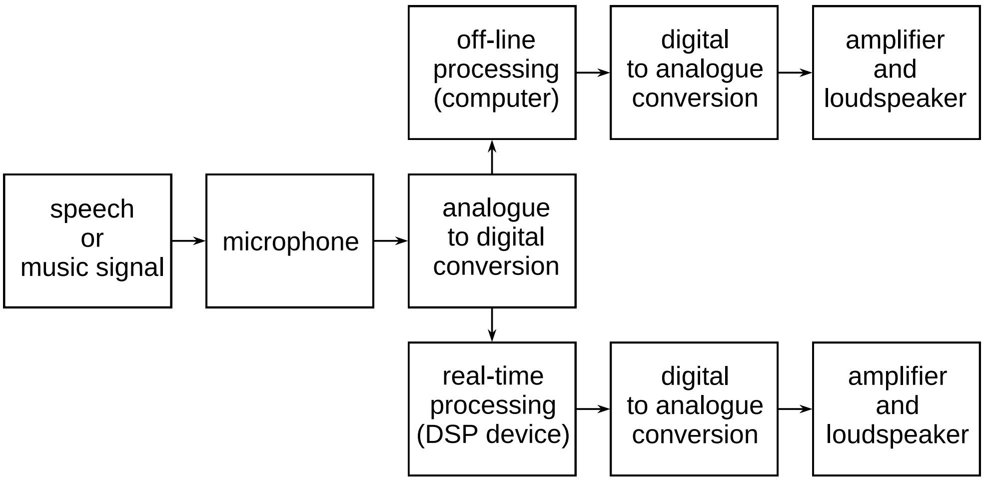

Figure 1.1 Sequence of events in the DSP chain. The processing may be performed in real-time or off-line.

In brief, the simplified chain of events goes something like this: the transducer converts some form of energy to which it is designed to respond into an electrical signal. This signal is converted into digital form, processed by a computer or DSP system and sometimes, but not always, re-converted back into the analogue domain. This process is illustrated in Figure 1.1. It is stressed that the word *simplified* is used above, because we have omitted some links in this chain that are essential from a practical perspective (such as anti-aliasing and reconstruction filters, which we will cover), but which do not impinge on the essential principle of the process. In this illustrated example, audio is being recorded and converted into a digital data stream. That may not seem particularly interesting or revolutionary, until you consider that while the audio is still in its “number form”, we can do some really rather amazing things with it. Manipulating this kind of “numerical audio”, if you like, is called Digital Signal Processing (DSP), and is one of the most important technological achievements of the age – in many ways it is transforming our world beyond recognition. This is no overstatement – without DSP, the lives that we now live in the modern world would simply not be possible. What is startling is that many experts working in this field consistently fail to predict events and developments of global significance. This is perhaps because we have not yet fully grasped that the DSP revolution is fundamentally and qualitatively different from all technological revolutions that have gone before. With digital signal processing, we can effectively re-write reality, because, just as the currency of our brains is thought, so too the currency of digital signal processing equipment is number. The DSP revolution is not just about fast hardware, although it is undeniably important. DSP is also about ideas and, more importantly, ideas about ideas.

If something is in number form, it is very easy to manipulate, because transformations are strictly a matter of software. Hence a digital filter system is inherently flexible, since changing the characteristics of the filter merely involves changing the program code or filter coefficients; with an analogue filter, physical reconstruction is required. Furthermore, it is immune to the effects of ageing and environmental conditions, since the filtering process is dependent on numerical calculations, not mechanical characteristics of the components. This makes it particularly suited for very low frequency signals. For the same reason, the performance of a digital filter can be specified with extreme precision, in contrast to analogue filters where a 3% figure is considered excellent.

And what about the distinction between off-line and real-time DSP? If you’ve already recorded and stored your data on a PC and want to process it, then speed is not critical. As long as the processing algorithm takes a reasonable time to produce the desired result, then it doesn’t matter that there is no synchronicity between the signal input and output. Digital recording studios, for example, invariably resort to recording the music, digitally enhancing it at leisure, and producing the final version many days after the final guitar note has faded away. Such off-line luxury, is however, not always available or possible. What about live performances, video and audio broadcasting, mobile phone telephony, radio telemetry, and a host of other circumstances where the data are being generated and consumed in real-time? In this case, any DSP that is performed must, by definition, be applied in real-time. And so we reach an important conclusion: real-time DSP must produce one new output value for every input value. Invariably, there will be a constant delay within the DSP system (which represents the processing operation), but as long as this delay is constant and small, no data logjam will build up. How small is small? Well, it all depends on the nature of the consumer. If it is a live audio performance, the constant delay in the system should not really exceed 50 ms, otherwise the movement of the performer’s lips will not correspond to the perceived sound. Even here, though, we have some flexibility, because if the performance is taking place in a large auditorium or an open air stadium, the delay resulting from the sound travelling through the air will exceed the delay introduced by the DSP system (it takes sound about 50ms to travel 16.5m, i.e. the length of a small lecture theatre).

In order to perform real-time DSP that is effective, we need, above all, a fast processor. Why? Because data are streaming in and out at kHz or MHz speeds, and we need to multiply, add and shift many times per sample point for our algorithms to work. As we will learn, multiplication, addition and shifting are the three operations that lie at the heart of all DSP algorithms, and the big semiconductor corporations such as Texas and Motorola invest billions of dollars in developing chips that do these three things as fast as possible. How fast is fast? Well, let’s take a typical example. Say you have a mono audio signal sampled at 48 kHz. You design a high-pass finite impulse response (FIR) filter with 256 coefficients, to remove mains hum. These filters operate through convolution – every time a new signal point is acquired, we multiply the most recent 256 signal values with the coefficients of our filter and sum them all to produce one new (output) signal value. Hence we need to perform 48000 × 256 =

12.288 million multiplications *and* accumulations per second, or MMACS, as they are known in the business. Is this possible? Yes indeed – modern DSP chips would still be in first gear! The Motorola DS56309, which costs around $20, operates at 100 MMACS. Another member of this family, the DSP56321 goes up to 400 MMACS. It doesn't stop there. The Motorola Starcore MSC8102 operates at 48000 MMACS – all this in a package the size of a 50p piece. In comparison, when the world's first digital computer, ENIAC, was first completed in 1945, it contained 18,000 valves, consumed 150 kilowatts of power and performed 5,000 additions or 357 multiplications per second.

Given all this power, DSP can achieve truly wonderful things, as evidenced by Figure 1.2a. Buried in this seemingly random data is an electrocardiogram (ECG) signal. In this case, the noise exists in a band between 350 Hz and 1 kHz. Using a digital brick-wall band pass filter with very steep transition zones, we can recover the ECG trace, shown in Figure 1.2 (b). Although the recovery looks impressive, digital filters do this job very easily. Since the bandwidth of the noise does not encroach on the signal bandwidth, which does not extend beyond about 100 Hz, total signal restoration is possible, as long as the filter is sharp enough with pure phase linearity. Incidentally, digital filters can separate signals whose band gap margins can be arbitrarily small - this kind of recovery would be very difficult with analogue filters – the sharper the filter, the more the risk of instability and phase distortion. These problems can be completely avoided in digital designs – we'll see how later.

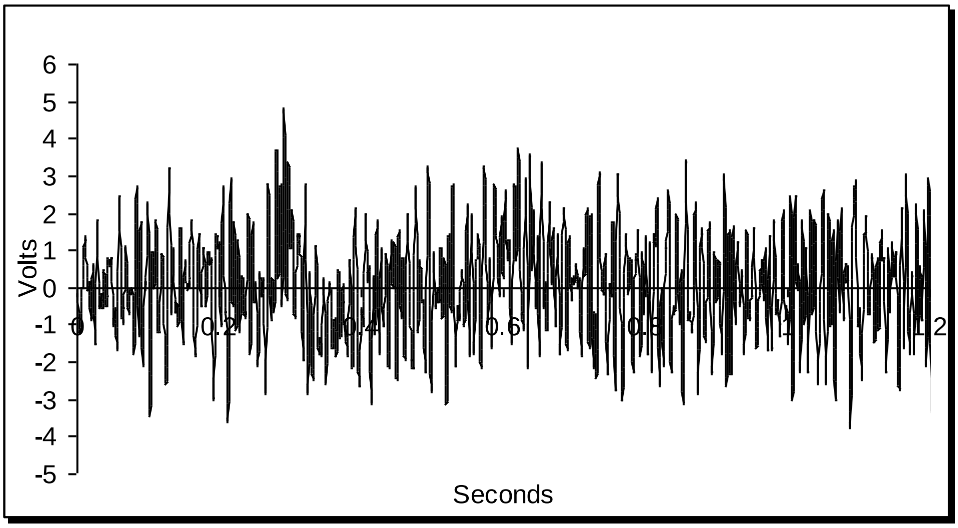

Figure 1.2 (a). ECG signal severely contaminated by out-of-band noise.

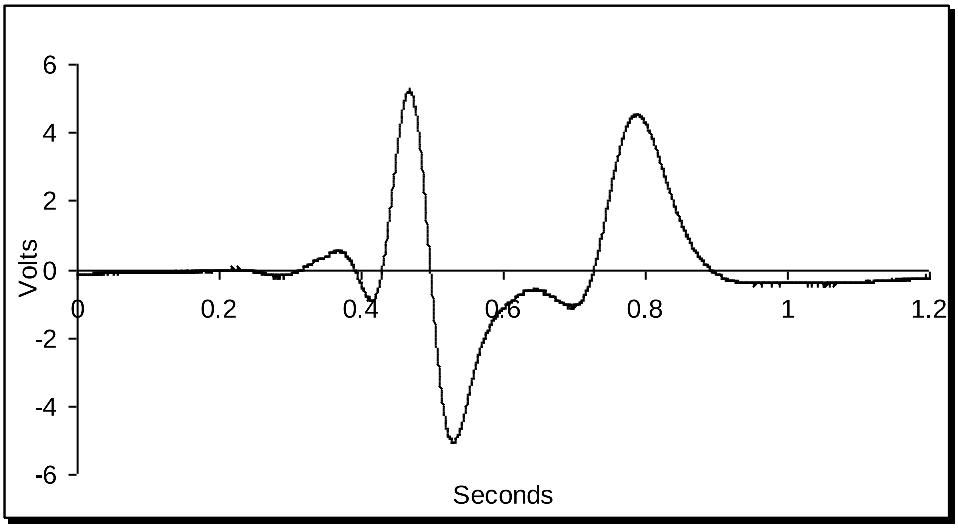

Figure 1.2 (b). ECG Signal shown above, recovered after processing with a linear band pass filter whose, bandwidth extends over the range of the ECG signal only. In addition to removing the noise, the filter has preserved the shape of the signal by virtue of its linear-phase characteristics. Signal shape preservation is an essential feature of filters intended for use with biomedical signals.

1.2 What is DSP used for?

Although there are almost as many DSP algorithms as there are stars in the heavens, in essence their functions fall into a small number of categories. These are:

- Noise removal
- Enhancement
- Special effects
- Compression/decompression
- Encryption/decryption
- (Feature analysis)
- (Feature extraction/recognition)

Once again, these categories are slightly artificial and often the boundaries between them are blurred. Noise removal is one of the most common applications of DSP; in a simple case, the noise might be band limited or lie outside the bandwidth of the signal. Thus a simple, fixed band stop filter will suffice. In another situation, the noise distribution may be more complex, demanding arbitrarily shaped or adaptive filters. In both circumstances however, the objective is to maximise the signal-to-noise (SNR) ratio. Signal enhancement is an allied but distinct subject area. The signal may have a good SNR, but certain signal frequency components may be too weak (or too strong). For example, a loudspeaker should ideally have a flat frequency response, but never does in practice. DSP can be used with great facility here to remedy an imperfect system – see Figure 1.3 for example. Special effects naturally follow on from signal enhancement – if you extend the loudspeaker correction example, you can make it sound truly bizarre. Never discount special effects as an area unworthy of your attention. They are a multi-billion dollar business, universally applied in the music and entertainments industry. What is intriguing in all of this is that the algorithms for noise removal, signal enhancement and special effects share significant commonality of the mathematical principles upon which they are based.

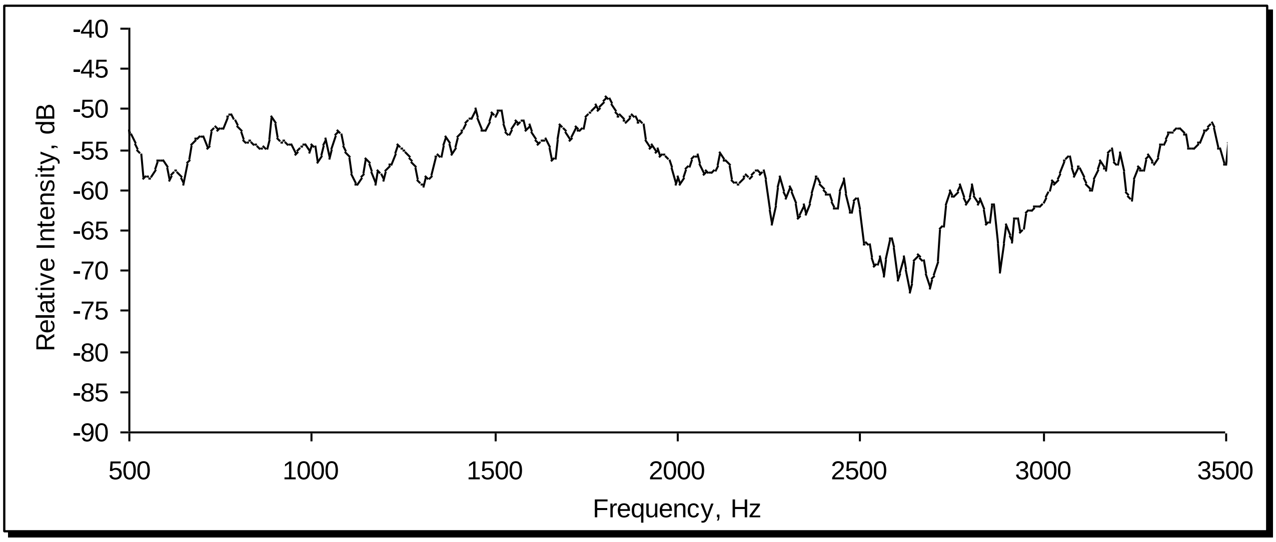

Figure 1.3(a). Midrange frequency response of a cheap computer loudspeaker. The response was obtained by stimulating the speaker with a swept sine wave, and performing a high-resolution Fourier analysis on the signal received from a microphone. Ideally, the response should be a flat line.

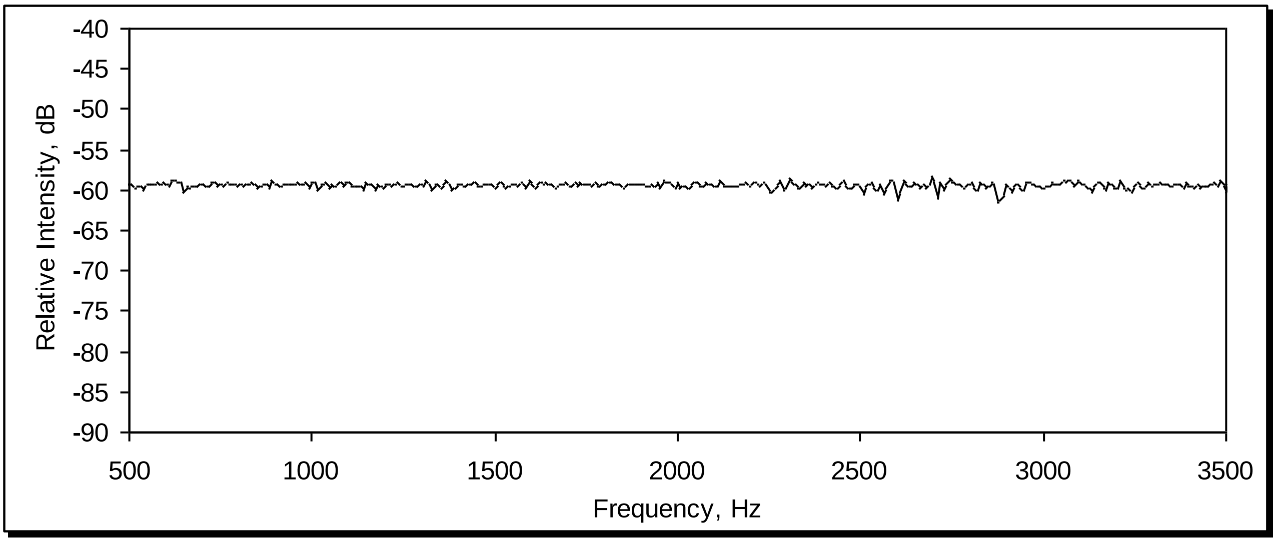

Figure 1.3(b). The frequency response after digital equalisation. Equalisation was performed by generating a time-domain filter from the inverse Fourier transform of the inverted frequency response.

Modern computer and digital systems are associated with large amounts of information, information that is often transmitted over a network or a mobile telephony system. If you want to transmit a lot of information quickly, you need a high bandwidth, and this costs money. Similarly, the longer you use the network channel, the more it costs. So there is great incentive to compress data before transmission, expanding it upon reception. Compression/decompression algorithms generally come in two varieties, loss-free and lossy, and both are widely used. Loss-free are clearly required if, for example, you are downloading a program file from the web. Upon decompression, every byte must be identical with the original; otherwise it will not work correctly, if at all. Lossy are used when some information loss can be tolerated in the decompressed version, and are frequently applied to video and audio data. By definition, lossy algorithms can achieve far greater compression ratios than loss-free. Both are the subjects of much research in DSP labs all over the globe.

To emphasise the point made earlier, progress in DSP algorithms follows in the wake of advances in hardware systems. Nowhere is this truer than in the field of data encryption/decryption. The establishment of the Web as *the* global communication network has led to a veritable explosion in developing secure means of transmitting data. Whole journals are devoted to this subject, but it is not something that we will cover in any detail here. The reasons are more pragmatic than anything else; to do the subject justice requires a lot of space, and on balance it is preferable to devote this to dealing with the more commonly encountered subjects in DSP.

The final two categories that are mentioned – feature analysis and extraction/recognition, do not, strictly speaking, belong to DSP, although they are often taught as part of this discipline. The reason why they are interlopers is fairly simple. Signal processing necessarily involves *changing* the signal in some way, to make it better, different or sometimes, deliberately worse (think of the sound-effects system of a professional theatre, for example, that is used to make an actor sound as if he or she is speaking through a telephone). Signal analysis, feature extraction and pattern recognition essentially involve the *observation* of a signal – information is extracted from the data, but the data remain unchanged. Once more however, our definition is a bit leaky. The Fourier transform is an analytical tool – although it maps the data into a different domain, the data remain the same – they are just presented in a different way. Yet the Fourier transform is also ubiquitously applied to change or process signals, and in fact, lies at the heart of nearly all DSP.

1.3 Application areas

How many of us are aware, for example, that the Pythagoras' Theorem we learned at school – the square of the hypotenuse of a right angled triangle is equal to the sum of the squares of the remaining two sides – is applied every second of the day in some form or another to maintain our digital world. But what on earth does a triangle have to do with, for example, digital mobile communication systems? Well, the *vector sum* calculation, as it's formally known, is used to obtain signal magnitudes (for one thing), and is applied to the Fourier transform to obtain magnitudes of frequency components. But it isn't just communication systems that exploit DSP; have a look at the list below, which just scratches the surface:

- Analogue emulation systems
- Audio/visual/multimedia
- Biomedical
- Control
- Control systems engineering
- Digital waveform synthesis
- Earth-based telecommunications
- Image processing
- *Image processing in all its representations*
- Industrial signal conditioning
- Mechatronics
- Military/surveillance
- Multiplexing
- Multirate processing and non-linear DSP
- Networking
- Noise cancellation systems
- Nondestructive testing
- Pattern recognition and matching
- Radar
- Remote sensing
- Robotics
- Satellite telemetry
- Seismology
- Speech recognition/synthesis
- Scientific Instrumentation; signal analysis

Table 1.1 Some application areas of DSP.

There are two points to make about Table 1.1. First, each of these areas is a world in itself, and to do any of them justice requires at least an entire volume. To emphasise this point, think about digital audio for a moment; both amateurs and professionals have made extensive use of DSP for recording and signal manipulation over the last thirty years, and a vast industry has been built up around this. Funnily enough, many of the algorithms developed for audio have found their way into what some people might call "serious" scientific research. If you look at Table 1.2 for example, DSP is applied to audio in many different ways.

- Chorusing
- Digital composition
- Distortion
- Echo
- Flanging
- Mono to pseudo stereo conversion
- Noise removal, hiss removal etc
- Phasing
- Pitch scaling
- Restoration of old recordings
- Reverberation
- Signal delay
- Surround sound
- Time scaling
- Virtual surround sound
- Waveform synthesis

Table 1.2 Just a few areas in audio DSP

The second point about Table 1.1 concerns the topic of image processing, which appears in italics. This is more than an application area, but an extension of DSP into at least one more dimension and a whole new scientific discipline. It should be borne in mind however, that anything we do in one dimension extends without (conceptual) difficulty into 2 or even *N* dimensions.