

Chapter 1

What Is Digital Signal Processing?

A *signal*, technically yet generally speaking, is a formal description of a phenomenon evolving over time or space; by *signal processing* we denote any manual or “mechanical” operation which modifies, analyzes or otherwise manipulates the information contained in a signal. Consider the simple example of ambient temperature: once we have agreed upon a formal model for this physical variable – Celsius degrees, for instance – we can record the evolution of temperature over time in a variety of ways and the resulting data set represents a temperature “signal”. Simple processing operations can then be carried out even just by hand: for example, we can plot the signal on graph paper as in Figure 1.1, or we can compute derived parameters such as the average temperature in a month.

Conceptually, it is important to note that signal processing operates on *an abstract representation* of a physical quantity and not on the quantity itself. At the same time, the *type* of abstract representation we choose for the physical phenomenon of interest determines the nature of a signal processing unit. A temperature regulation device, for instance, is not a signal processing system as a whole. The device does however contain a signal processing core in the feedback control unit which converts the instantaneous *measure* of the temperature into an ON/OFF trigger for the heating element. The physical nature of this unit depends on the temperature model: a simple design is that of a mechanical device based on the dilation of a metal sensor; more likely, the temperature signal is a voltage generated by a thermocouple and in this case the matched signal processing unit is an operational amplifier.

Finally, the adjective “digital” derives from *digitus*, the Latin word for finger: it concisely describes a world view where everything can be ultimately represented as an integer number. Counting, first on one’s fingers and then

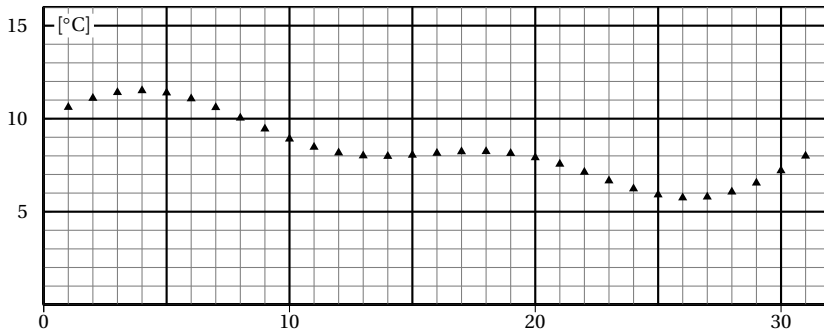


Figure 1.1 Temperature measurements over a month.

in one’s head, is the earliest and most fundamental form of abstraction; as children we quickly learn that counting does indeed bring disparate objects (the proverbial “apples and oranges”) into a common modeling paradigm, i.e. their cardinality. Digital signal processing is a flavor of signal processing in which everything *including time* is described in terms of integer numbers; in other words, the abstract representation of choice is a one-size-fit-all countability. Note that our earlier “thought experiment” about ambient temperature fits this paradigm very naturally: the measuring instants form a countable set (the days in a month) and so do the measures themselves (imagine a finite number of ticks on the thermometer’s scale). In digital signal processing the underlying abstract representation is always the set of natural numbers regardless of the signal’s origins; as a consequence, the physical nature of the processing device will also always remain the same, that is, a general digital (micro)processor. The extraordinary power and success of digital signal processing derives from the inherent universality of its associated “world view”.

1.1 Some History and Philosophy

1.1.1 Digital Signal Processing under the Pyramids

Probably the earliest recorded example of digital signal processing dates back to the 25th century BC. At the time, Egypt was a powerful kingdom reaching over a thousand kilometers south of the Nile’s delta. For all its latitude, the kingdom’s populated area did not extend for more than a few kilometers on either side of the Nile; indeed, the only inhabitable areas in an otherwise desert expanse were the river banks, which were made fertile

by the yearly flood of the river. After a flood, the banks would be left covered with a thin layer of nutrient-rich silt capable of supporting a full agricultural cycle. The floods of the Nile, however, were⁽¹⁾ a rather capricious meteorological phenomenon, with scant or absent floods resulting in little or no yield from the land. The pharaohs quickly understood that, in order to preserve stability, they would have to set up a grain buffer with which to compensate for the unreliability of the Nile's floods and prevent potential unrest in a famished population during "dry" years. As a consequence, studying and predicting the trend of the floods (and therefore the expected agricultural yield) was of paramount importance in order to determine the operating point of a very dynamic taxation and redistribution mechanism. The floods of the Nile were meticulously recorded by an array of measuring stations called "nilometers" and the resulting data set can indeed be considered a full-fledged digital signal defined on a time base of twelve months. The Palermo Stone, shown in the left panel of Figure 1.2, is a faithful record of the data in the form of a table listing the name of the current pharaoh alongside the yearly flood level; a more modern representation of an flood data set is shown on the right of the figure: bar the references to the pharaohs, the two representations are perfectly equivalent. The Nile's behavior is still an active area of hydrological research today and it would be surprising if the signal processing operated by the ancient Egyptians on their data had been of much help in anticipating for droughts. Yet, the Palermo Stone is arguably the first recorded digital signal which is still of relevance today.

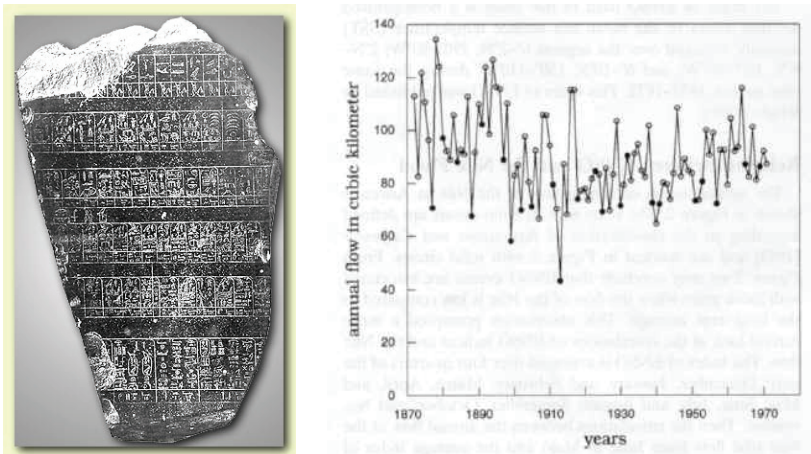


Figure 1.2 Representations of flood data for the river Nile: circa 2500 BC (left) and 2000 AD (right).

⁽¹⁾The Nile stopped flooding Egypt in 1964, when the Aswan dam was completed.

1.1.2 The Hellenic Shift to Analog Processing

“Digital” representations of the world such as those depicted by the Palermo Stone are adequate for an environment in which quantitative problems are simple: counting cattle, counting bushels of wheat, counting days and so on. As soon as the interaction with the world becomes more complex, so necessarily do the models used to interpret the world itself. Geometry, for instance, is born of the necessity of measuring and subdividing land property. In the act of splitting a certain quantity into parts we can already see the initial difficulties with an integer-based world view;⁽²⁾ yet, until the Hellenic period, western civilization considered natural numbers and their ratios all that was needed to describe nature in an operational fashion. In the 6th century BC, however, a devastated Pythagoras realized that the the side and the diagonal of a square are incommensurable, i.e. that $\sqrt{2}$ is *not* a simple fraction. The discovery of what we now call irrational numbers “sealed the deal” on an abstract model of the world that had already appeared in early geometric treatises and which today is called *the continuum*. Heavily steeped in its geometric roots (i.e. in the infinity of points in a segment), the continuum model postulates that time and space are an uninterrupted flow which can be divided arbitrarily many times into arbitrarily (and infinitely) small pieces. In signal processing parlance, this is known as the “analog” world model and, in this model, integer numbers are considered primitive entities, as rough and awkward as a set of sledgehammers in a watchmaker’s shop.

In the continuum, the infinitely big and the infinitely small dance together in complex patterns which often defy our intuition and which required almost two thousand years to be properly mastered. This is of course not the place to delve deeper into this extremely fascinating epistemological domain; suffice it to say that the apparent incompatibility between the digital and the analog world views appeared right from the start (i.e. from the 5th century BC) in Zeno’s works; we will appreciate later the immense import that this has on signal processing in the context of the sampling theorem.

Zeno’s paradoxes are well known and they underscore this unbridgeable gap between our intuitive, integer-based grasp of the world and a model of

⁽²⁾The layman’s aversion to “complicated” fractions is at the basis of many counting systems other than the decimal (which is just an accident tied to the number of human fingers). Base-12 for instance, which is still so persistent both in measuring units (hours in a day, inches in a foot) and in common language (“a dozen”) originates from the simple fact that 12 happens to be divisible by 2, 3 and 4, which are the most common number of parts an item is usually split into. Other bases, such as base-60 and base-360, have emerged from a similar abundance of simple factors.

the world based on the continuum. Consider for instance the dichotomy paradox; Zeno states that if you try to move along a line from point A to point B you will never in fact be able to reach your destination. The reasoning goes as follows: in order to reach B, you will have to first go through point C, which is located mid-way between A and B; but, even before you reach C, you will have to reach D, which is the midpoint between A and C; and so on *ad infinitum*. Since there is an infinity of such intermediate points, Zeno argues, moving from A to B requires you to complete an infinite number of tasks, which is humanly impossible. Zeno of course was well aware of the empirical evidence to the contrary but he was brilliantly pointing out the extreme trickery of a model of the world which had not yet formally defined the concept of infinity. The complexity of the intellectual machinery needed to solidly counter Zeno's argument is such that even today the paradox is food for thought. A first-year calculus student may be tempted to offhandedly dismiss the problem by stating

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = 1 \quad (1.1)$$

but this is just a void formalism begging the initial question if the underlying notion of the continuum is not explicitly worked out.⁽³⁾ In reality Zeno's paradoxes cannot be "solved", since they cease to be paradoxes once the continuum model is fully understood.

1.1.3 "Gentlemen: *calculemus!*"

The two competing models for the world, digital and analog, coexisted quite peacefully for quite a few centuries, one as the tool of the trade for farmers, merchants, bankers, the other as an intellectual pursuit for mathematicians and astronomers. Slowly but surely, however, the increasing complexity of an expanding world spurred the more practically-oriented minds to pursue science as a means to solve very tangible problems *besides* describing the motion of the planets. Calculus, brought to its full glory by Newton and Leibnitz in the 17th century, proved to be an incredibly powerful tool when applied to eminently practical concerns such as ballistics, ship routing, mechanical design and so on; such was the faith in the power of the new science that Leibnitz envisioned a not-too-distant future in which all human disputes, including problems of morals and politics, could be worked out with pen and paper: "gentlemen, *calculemus*". If only.

⁽³⁾An easy rebuttal of the bookish *reductio* above is asking to explain why $\sum 1/n$ diverges while $\sum 1/n^2 = \pi^2/6$ (Euler, 1740).

As Cauchy unsurpassably explained later, everything in calculus is a limit and therefore everything in calculus is a celebration of the power of the continuum. Still, in order to apply the calculus machinery to the real world, the real world has to be modeled as something calculus understands, namely a function of a real (i.e. *continuous*) variable. As mentioned before, there are vast domains of research well behaved enough to admit such an *analytical* representation; astronomy is the first one to come to mind, but so is ballistics, for instance. If we go back to our temperature measurement example, however, we run into the first difficulty of the analytical paradigm: we now need to model our measured temperature as a function of continuous time, which means that the value of the temperature should be available at *any* given instant and not just once per day. A “temperature function” as in Figure 1.3 is quite puzzling to define if all we have (and if all we *can* have, in fact) is just a set of empirical measurements reasonably spaced in time. Even in the rare cases in which an analytical model of the phenomenon is available, a second difficulty arises when the *practical* application of calculus involves the use of functions which are only available in tabulated form. The trigonometric and logarithmic tables are a typical example of how a continuous model needs to be made countable again in order to be put to real use. Algorithmic procedures such as series expansions and numerical integration methods are other ways to bring the analytic results within the realm of the practically computable. These parallel tracks of scientific development, the “Platonic” ideal of analytical results and the slide rule reality of practitioners, have coexisted for centuries and they have found their most durable mutual peace in digital signal processing, as will appear shortly.

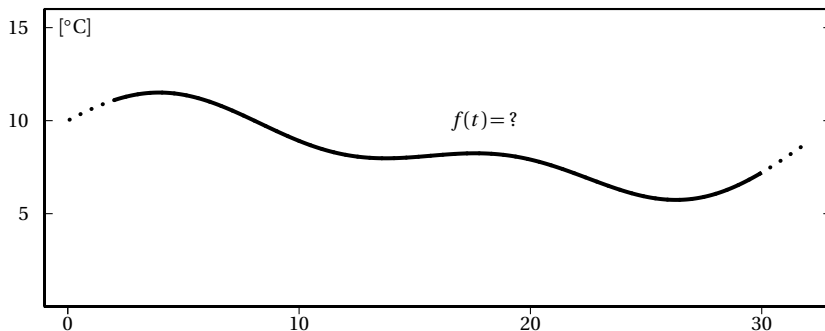


Figure 1.3 Temperature “function” in a continuous-time world model.

1.2 Discrete-Time

One of the fundamental problems in signal processing is to obtain a permanent record of the signal itself. Think back of the ambient temperature example, or of the floods of the Nile: in both cases a description of the phenomenon was gathered by a naive *sampling* operation, i.e. by measuring the quantity of interest at regular intervals. This is a very intuitive process and it reflects the very natural act of “looking up the current value and writing it down”. Manually this operation is clearly quite slow but it is conceivable to speed it up mechanically so as to obtain a much larger number of measurements per unit of time. Our measuring machine, however fast, still will never be able to take an *infinite* amount of samples in a finite time interval: we are back in the clutches of Zeno’s paradoxes and one would be tempted to conclude that a true analytical representation of the signal is impossible to obtain.

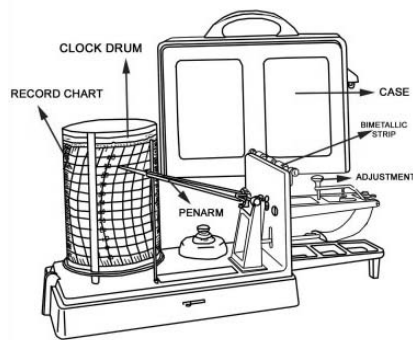


Figure 1.4 A thermograph.

At the same time, the history of applied science provides us with many examples of recording machines capable of providing an “analog” image of a physical phenomenon. Consider for instance a thermograph: this is a mechanical device in which temperature deflects an ink-tipped metal stylus in contact with a slowly rolling paper-covered cylinder. Thermographs like the one sketched in Figure 1.4 are still currently in use in some simple weather stations and they provide a chart in which a temperature function as in Figure 1.3 is duly plotted. Incidentally, the principle is the same in early sound recording devices: Edison’s phonograph used the deflection of a steel pin connected to a membrane to impress a “continuous-time” sound wave as a groove on a wax cylinder. The problem with these analog recordings is that they are not abstract signals but *a conversion of a physical phenomenon into another physical phenomenon*: the temperature, for instance, is con-

verted into the amount of ink on paper while the sound pressure wave is converted into the physical depth of the groove. The advent of electronics did not change the concept: an audio tape, for instance, is obtained by converting a pressure wave into an electrical current and then into a magnetic deflection. The fundamental consequence is that, for analog signals, a different signal processing system needs to be designed *explicitly* for each specific form of recording.

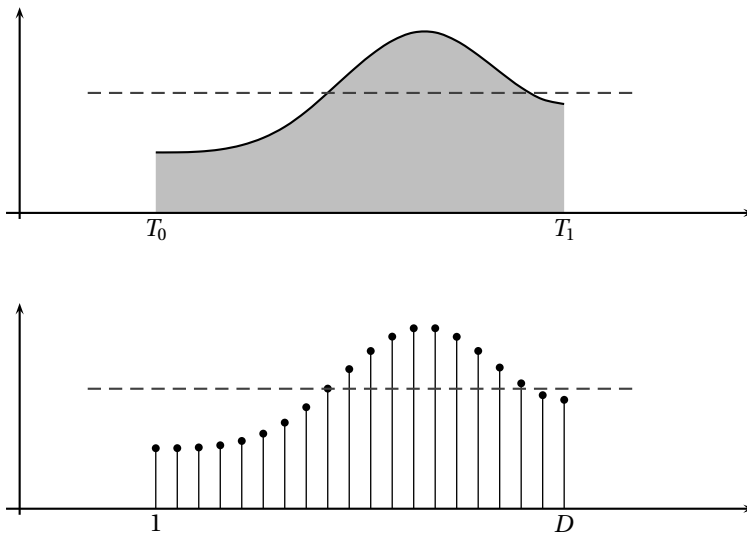


Figure 1.5 Analytical and empirical averages.

Consider for instance the problem of computing the average temperature over a certain time interval. Calculus provides us with the exact answer \bar{C} if we know the elusive “temperature function” $f(t)$ over an interval $[T_0, T_1]$ (see Figure 1.5, top panel):

$$\bar{C} = \frac{1}{T_1 - T_0} \int_{T_0}^{T_1} f(t) dt \quad (1.2)$$

We can try to reproduce the integration with a “machine” adapted to the particular representation of temperature we have at hand: in the case of the thermograph, for instance, we can use a planimeter as in Figure 1.6, a manual device which computes the area of a drawn surface; in a more modern incarnation in which the temperature signal is given by a thermocouple, we can integrate the voltage with the RC network in Figure 1.7. In both cases, in spite of the simplicity of the problem, we can instantly see the practical complications and the degree of specialization needed to achieve something as simple as an average for an analog signal.

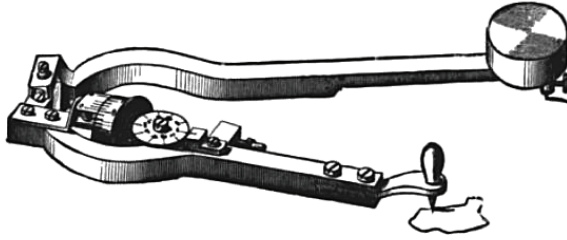


Figure 1.6 The planimeter: a mechanical integrator.

Now consider the case in which all we have is a set of daily measurements c_1, c_2, \dots, c_D as in Figure 1.1; the “average” temperature of our measurements over D days is simply:

$$\hat{C} = \frac{1}{D} \sum_{n=1}^D c_n \quad (1.3)$$

(as shown in the bottom panel of Figure 1.5) and this is an elementary sum of D terms which anyone can carry out by hand and which does not depend on how the measurements have been obtained: wickedly simple! So, obviously, the question is: “How different (if at all) is \hat{C} from \bar{C} ?” In order to find out we can remark that if we accept the existence of a temperature function $f(t)$ then the measured values c_n are *samples* of the function taken one day apart:

$$c_n = f(nT_s)$$

(where T_s is the duration of a day). In this light, the sum (1.3) is just the Riemann approximation to the integral in (1.2) and the question becomes an assessment on how good an approximation that is. Another way to look at the problem is to ask ourselves how much information we are discarding by only keeping samples of a continuous-time function.

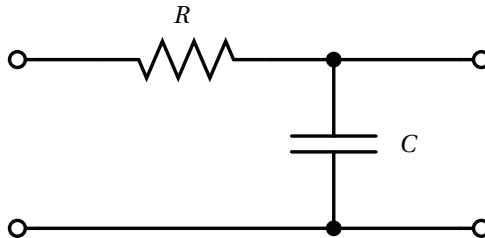


Figure 1.7 The RC network: an electrical integrator.

The answer, which we will study in detail in Chapter 9, is that in fact the continuous-time function and the set of samples *are perfectly equivalent representations* – provided that the underlying physical phenomenon “doesn’t change too fast”. Let us put the proviso aside for the time being and concentrate instead on the good news: first, the analog and the digital world can perfectly coexist; second, we actually possess a *constructive* way to move between worlds: the *sampling theorem*, discovered and rediscovered by many at the beginning of the 20th century⁽⁴⁾, tells us that the continuous-time function can be obtained from the samples as

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \frac{\sin(\pi(t - nT_s)/T_s)}{\pi(t - nT_s)/T_s} \quad (1.4)$$

So, in theory, once we have a set of measured values, we can build the continuous-time representation and use the tools of calculus. In reality things are even simpler: if we plug (1.4) into our analytic formula for the average (1.2) we can show that the result is a simple sum like (1.3). So we don’t need to explicitly go “through the looking glass” back to continuous-time: the tools of calculus have a discrete-time equivalent which we can use directly.

The equivalence between the discrete and continuous representations only holds for signals which are sufficiently “slow” with respect to how fast we sample them. This makes a lot of sense: we need to make sure that the signal does not do “crazy” things between successive samples; only if it is smooth and well behaved can we afford to have such sampling gaps. Quantitatively, the sampling theorem links the speed at which we need to repeatedly measure the signal to the maximum frequency contained in its spectrum. Spectra are calculated using the Fourier transform which, interestingly enough, was originally devised as a tool to break periodic functions into a *countable* set of building blocks. Everything comes together.

1.3 Discrete Amplitude

While it appears that the time continuum has been tamed by the sampling theorem, we are nevertheless left with another pesky problem: the precision of our measurements. If we model a phenomenon as an analytical function, not only is the argument (the time domain) a continuous variable but so is the function’s value (the codomain); a practical measurement, however, will never achieve an infinite precision and we have another paradox

⁽⁴⁾Amongst the credited personnel: Nyquist, Whittaker, Kotel’nikov, Raabe, Shannon and Someya.

on our hands. Consider our temperature example once more: we can use a mercury thermometer and decide to write down just the number of degrees; maybe we can be more precise and note the half-degrees as well; with a magnifying glass we could try to record the tenths of a degree – but we would most likely have to stop there. With a more sophisticated thermocouple we could reach a precision of one hundredth of a degree and possibly more but, still, we would have to settle on a maximum number of decimal places. Now, if we know that our measures have a fixed number of digits, the set of all possible measures is actually countable and we have effectively mapped the codomain of our temperature function onto the set of integer numbers. This process is called *quantization* and it is the method, together with sampling, to obtain a fully digital signal.

In a way, quantization deals with the problem of the continuum in a much “rougher” way than in the case of time: we simply accept a loss of precision with respect to the ideal model. There is a very good reason for that and it goes under the name of *noise*. The mechanical recording devices we just saw, such as the thermograph or the phonograph, give the illusion of analytical precision but are in practice subject to severe mechanical limitations. Any analog recording device suffers from the same fate and even if electronic circuits can achieve an excellent performance, in the limit the unavoidable thermal agitation in the components constitutes a noise floor which limits the “equivalent number of digits”. Noise is a fact of nature that cannot be eliminated, hence our acceptance of a finite (i.e. countable) precision.

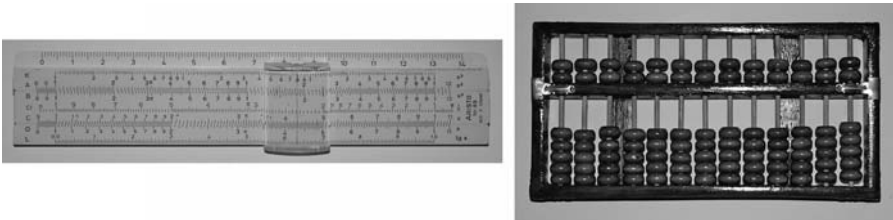


Figure 1.8 Analog and digital computers.

Noise is not just a problem in measurement but also in processing. Figure 1.8 shows the two archetypal types of analog and digital computing devices; while technological progress may have significantly improved the speed of each, the underlying principles remain the same for both. An analog signal processing system, much like the slide rule, uses the displacement of physical quantities (gears or electric charge) to perform its task; each element in the system, however, acts as a source of noise so that complex or,

more importantly, *cheap* designs introduce imprecisions in the final result (good slide rules used to be *very* expensive). On the other hand the abacus, working only with integer arithmetic, is a perfectly precise machine⁽⁵⁾ even if it's made with rocks and sticks. Digital signal processing works with countable sequences of integers so that in a digital architecture no processing noise is introduced. A classic example is the problem of *reproducing* a signal. Before mp3 existed and file sharing became the bootlegging method of choice, people would “make tapes”. When someone bought a vinyl record he would allow his friends to record it on a cassette; however, a “peer-to-peer” dissemination of illegally taped music never really took off because of the “second generation noise”, i.e. because of the ever increasing hiss that would appear in a tape made from another tape. Basically only first generation copies of the purchased vinyl were acceptable quality on home equipment. With digital formats, on the other hand, duplication is really equivalent to copying down a (very long) list of integers and even very cheap equipment can do that without error.

Finally, a short remark on terminology. The amplitude accuracy of a set of samples is entirely dependent on the processing hardware; in current parlance this is indicated by the number of *bits per sample* of a given representation: compact disks, for instance, use 16 bits per sample while DVDs use 24. Because of its “contingent” nature, quantization is almost always ignored in the core theory of signal processing and all derivations are carried out as if the samples were real numbers; therefore, in order to be precise, we will almost always use the term *discrete-time* signal processing and leave the label “digital signal processing” (DSP) to the world of actual devices. Neglecting quantization will allow us to obtain very general results but care must be exercised: in the practice, actual implementations will have to deal with the effects of finite precision, sometimes with very disruptive consequences.

1.4 Communication Systems

Signals in digital form provide us with a very convenient abstract representation which is both simple and powerful; yet this does not shield us from the need to deal with an “outside” world which is probably best modeled by the analog paradigm. Consider a mundane act such as placing a call on a cell phone, as in Figure 1.9: humans are analog devices after all and they produce analog sound waves; the phone converts these into digital format,

⁽⁵⁾As long as we don't need to compute square roots; luckily, *linear* systems (which is what interests us) are made up only of sums and multiplications.

does some digital processing and then outputs an *analog* electromagnetic wave on its antenna. The radio wave travels to the base station in which it is demodulated, converted to digital format to recover the voice signal. The call, as a digital signal, continues through a switch and then is injected into an optical fiber *as an analog light wave*. The wave travels along the network and then the process is inverted until an analog sound wave is generated by the loudspeaker at the receiver's end.

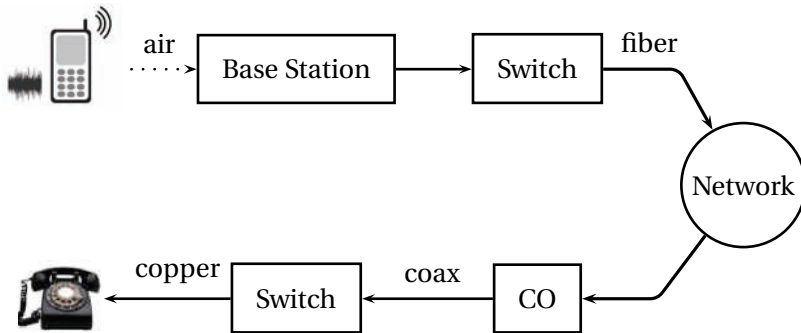


Figure 1.9 A prototypical telephone call and the associated transitions from the digital to the analog domain and back; processing in the blocks is done digitally while the links between blocks are over an analog medium.

Communication systems are in general a prime example of sophisticated interplay between the digital and the analog world: while all the processing is undoubtedly best done digitally, signal propagation in a medium (be it the the air, the electromagnetic spectrum or an optical fiber) is the domain of differential (rather than *difference*) equations. And yet, even when digital processing must necessarily hand over control to an analog interface, it does so in a way that leaves no doubt as to who's boss, so to speak: for, instead of transmitting an analog signal which is the reconstructed “real” function as per (1.4), we always transmit an analog signal which *encodes the digital representation of the data*. This concept is really at the heart of the “digital revolution” and, just like in the cassette tape example, it has to do with noise.

Imagine an analog voice signal $s(t)$ which is transmitted over a (long) telephone line; a simplified description of the received signal is

$$s_r(t) = \alpha s(t) + n(t)$$

where the parameter α , with $\alpha < 1$, is the *attenuation* that the signal incurs and where $n(t)$ is the noise introduced by the system. The noise function is of obviously unknown (it is often modeled as a Gaussian process, as we

will see) and so, once it's added to the signal, it's impossible to eliminate it. Because of attenuation, the receiver will include an amplifier with gain G to restore the voice signal to its original level; with $G = 1/\alpha$ we will have

$$s_a(t) = G s_r(t) = s(t) + G n(t)$$

Unfortunately, as it appears, in order to regenerate the analog signal we also have amplified the noise G times; clearly, if G is large (i.e. if there is a lot of attenuation to compensate for) the voice signal end up buried in noise. The problem is exacerbated if many intermediate amplifiers have to be used in cascade, as is the case in long submarine cables.

Consider now a digital voice signal, that is, a discrete-time signal whose samples have been quantized over, say, 256 levels: each sample can therefore be represented by an 8-bit word and the whole speech signal can be represented as a very long sequence of binary digits. We now build an analog signal as a *two-level* signal which switches for a few instants between, say, -1 V and $+1$ V for every "0" and "1" bit in the sequence respectively. The received signal will still be

$$s_r(t) = \alpha s(t) + n(t)$$

but, to regenerate it, instead of linear amplification we can use nonlinear thresholding:

$$s_a(t) = \begin{cases} +1 & \text{if } s_r(t) \geq 0 \\ -1 & \text{if } s_r(t) < 0 \end{cases}$$

Figure 1.10 clearly shows that as long as the magnitude of the noise is less than α the two-level signal can be regenerated perfectly; furthermore, the regeneration process can be repeated as many times as necessary with no overall degradation.

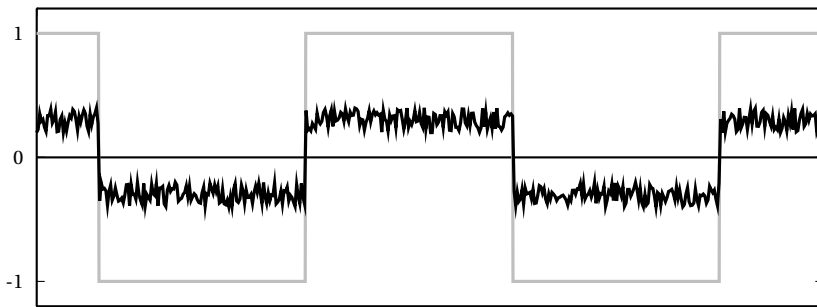


Figure 1.10 Two-level analog signal encoding a binary sequence: original signal $s(t)$ (light gray) and received signal $s_r(t)$ (black) in which both attenuation and noise are visible.

In reality of course things are a little more complicated and, because of the nature of noise, it is impossible to guarantee that some of the bits won't be corrupted. The answer is to use *error correcting codes* which, by introducing redundancy in the signal, make the sequence of ones and zeros robust to the presence of errors; a scratched CD can still play flawlessly because of the Reed-Solomon error correcting codes used for the data. Data coding is the core subject of Information Theory and it is behind the stellar performance of modern communication systems; interestingly enough, the most successful codes have emerged from group theory, a branch of mathematics dealing with the properties of closed sets of integer numbers. Integers again! Digital signal processing and information theory have been able to join forces so successfully because they share a common data model (the integer) and therefore they share the same architecture (the processor). Computer code written to implement a digital filter can dovetail seamlessly with code written to implement error correction; linear processing and nonlinear flow control coexist naturally.

A simple example of the power unleashed by digital signal processing is the performance of transatlantic cables. The first operational telegraph cable from Europe to North America was laid in 1858 (see Fig. 1.11); it worked for about a month before being irrecoverably damaged.⁽⁶⁾ From then on, new materials and rapid progress in electrotechnics boosted the performance of each subsequent cable; the key events in the timeline of transatlantic communications are shown in Table 1.1. The first transatlantic *telephone* cable was laid in 1956 and more followed in the next two decades with increasing capacity due to multicore cables and better repeaters; the invention of the echo canceler further improved the number of voice channels for already deployed cables. In 1968 the first experiments in PCM digital telephony were successfully completed and the quantum leap was around the corner: by the end of the 70's cables were carrying over one order of magnitude more voice channels than in the 60's. Finally, the deployment of the first fiber optic cable in 1988 opened the door to staggering capacities (and enabled the dramatic growth of the Internet).

Finally, it's impossible not to mention the advent of *data compression* in this brief review of communication landmarks. Again, digital processing allows the coexistence of standard processing with sophisticated decision

⁽⁶⁾Ohm's law was published in 1861, so the first transatlantic cable was a little bit the proverbial cart before the horse. Indeed, the cable circuit formed an enormous RC equivalent circuit, i.e. a big lowpass filter, so that the sharp rising edges of the Morse symbols were completely smeared in time. The resulting intersymbol interference was so severe that it took hours to reliably send even a simple sentence. Not knowing how to deal with the problem, the operator tried to increase the signaling voltage ("crank up the volume") until, at 4000 V, the cable gave up.

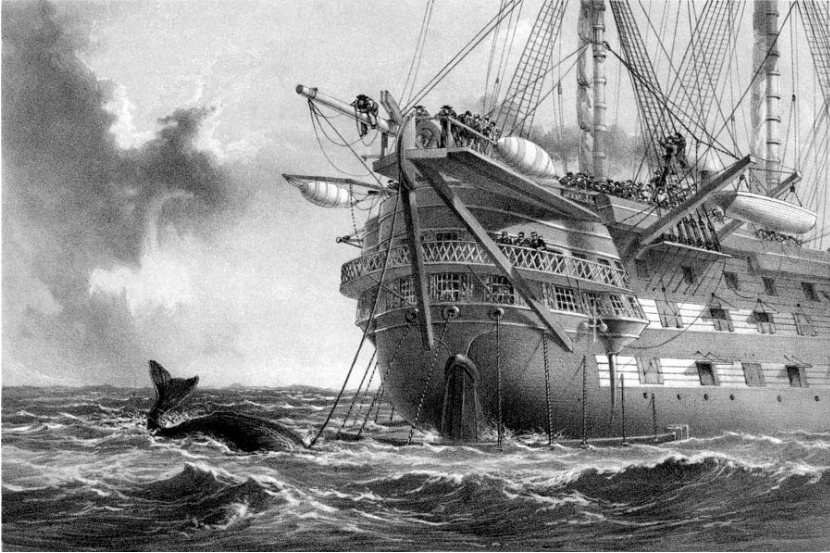


Figure 1.11 Laying the first transatlantic cable.

Table 1.1 The main transatlantic cables from 1858 to our day.

Cable	Year	Type	Signaling	Capacity
	1858	Coax	telegraph	a few words per hour
	1866	Coax	telegraph	6-8 words per minute
	1928	Coax	telegraph	2500 characters per minute
TAT-1	1956	Coax	telephone	36 [48 by 1978] voice channels
TAT-3	1963	Coax	telephone	138 [276 by 1986] voice channels
TAT-5	1970	Coax	telephone	845 [2112 by 1993] voice channels
TAT-6	1976	Coax	telephone	4000 [10,000 by 1994] voice channels
TAT-8	1988	Fiber	data	280 Mbit/s (~ 40,000 voice channels)
TAT-14	2000	Fiber	data	640 Gbit/s (~ 9,700,000 voice channels)

logic; this enables the implementation of complex *data-dependent* compression techniques and the inclusion of psychoperceptual models in order to match the compression strategy to the characteristics of the human visual or auditory system. A music format such as mp3 is perhaps the first example to come to mind but, as shown in Table 1.2, all communication domains have been greatly enhanced by the gains in throughput enabled by data compression.

Table 1.2 Uncompressed and compressed data rates.

Signal	Uncompressed Rate	Common Rate
Music	4.32 Mbit/s (CD audio)	128 Kbit/s (MP3)
Voice	64 Kbit/s (AM radio)	4.8 Kbit/s (cellphone CELP)
Photos	14 MB (raw)	300 KB (JPEG)
Video	170 Mbit/s (PAL)	700 Kbit/s (DivX)

1.5 How to Read this Book

This book tries to build a largely self-contained development of digital signal processing theory *from within discrete time*, while the relationship to the analog model of the world is tackled only after all the fundamental “pieces of the puzzle” are already in place. Historically, modern discrete-time processing started to consolidate in the 50’s when mainframe computers became powerful enough to handle the effective simulations of analog electronic networks. By the end of the 70’s the discipline had by all standards reached maturity; so much so, in fact, that the major textbooks on the subject still in use today had basically already appeared by 1975. Because of its ancillary origin with respect to the problems of that day, however, discrete-time signal processing has long been presented as a tributary to much more established disciplines such as Signals and Systems. While historically justifiable, that approach is no longer tenable today for three fundamental reasons: first, the pervasiveness of digital storage for data (from CDs to DVDs to flash drives) implies that most devices today are designed for discrete-time signals to start with; second, the trend in signal processing devices is to move the analog-to-digital and digital-to-analog converters at the very beginning and the very end of the processing chain so that even “classically analog” operations such as modulation and demodulation are now done in discrete-time; third, the availability of numerical packages like Matlab provides a testbed for signal processing experiments (both academically and just for fun) which is far more accessible and widespread than an electronics lab (not to mention affordable).

The idea therefore is to introduce discrete-time signals as a self-standing entity (Chap. 2), much in the natural way of a temperature sequence or a series of flood measurements, and then to remark that the mathematical structures used to describe discrete-time signals are one and the same with the structures used to describe vector spaces (Chap. 3). Equipped with the geometrical intuition afforded to us by the concept of vector space, we

can proceed to “dissect” discrete-time signals with the Fourier transform, which turns out to be just a change of basis (Chap. 4). The Fourier transform opens the passage between the time domain and the frequency domain and, thanks to this dual understanding, we are ready to tackle the concept of processing as performed by discrete-time linear systems, also known as filters (Chap. 5). Next comes the very practical task of designing a filter to order, with an eye to the subtleties involved in filter implementation; we will mostly consider FIR filters, which are unique to discrete time (Chaps 6 and 7). After a brief excursion in the realm of stochastic sequences (Chap. 8) we will finally build a bridge between our discrete-time world and the continuous-time models of physics and electronics with the concepts of sampling and interpolation (Chap. 9); and digital signals will be completely accounted for after a study of quantization (Chap. 10). We will finally go back to purely discrete time for the final topic, multirate signal processing (Chap. 11), before putting it all together in the final chapter: the analysis of a commercial voiceband modem (Chap. 12).

Further Reading

The Bible of digital signal processing was and remains *Discrete-Time Signal Processing*, by A. V. Oppenheim and R. W. Schaffer (Prentice-Hall, last edition in 1999); exceedingly exhaustive, it is a must-have reference. For background in signals and systems, the eponymous *Signals and Systems*, by Oppenheim, Willsky and Nawab (Prentice Hall, 1997) is a good start.

Most of the historical references mentioned in this introduction can be integrated by simple web searches. Other comprehensive books on digital signal processing include S. K. Mitra’s *Digital Signal Processing* (McGraw Hill, 2006) and *Digital Signal Processing*, by J. G. Proakis and D. K. Manolakis (Prentis Hall 2006). For a fascinating excursus on the origin of calculus, see D. Hairer and G. Wanner, *Analysis by its History* (Springer-Verlag, 1996). A more than compelling epistemological essay on the continuum is *Everything and More*, by David Foster Wallace (Norton, 2003), which manages to be both profound and hilarious in an unprecedented way.

Finally, the very prolific literature on current signal processing research is published mainly by the Institute of Electronics and Electrical Engineers (IEEE) in several of its transactions such as *IEEE Transactions on Signal Processing*, *IEEE Transactions on Image Processing* and *IEEE Transactions on Speech and Audio Processing*.