1
INTRODUCTION

Signal processing has a long and rich history. It is a technology that spans an immense set of disciplines including entertainment, communications, space exploration, medicine, and archaeology, just to name a few. Sophisticated signal processing algorithms and hardware are prevalent in a wide range of systems, from highly specialized military systems through industrial applications to low-cost, high-volume consumer electronics. Although we routinely take for granted the performance of home entertainment systems such as television and high-fidelity audio, these systems have always relied heavily on state-of-the-art signal processing. This is even more true today with the emergence of advanced television and multimedia entertainment and information systems. Furthermore, as communication systems become increasingly wireless, mobile, and multifunctional, the importance of sophisticated signal processing in these systems continues to grow. Overall, as we look to the future, it is clear that the role of signal processing in our society is accelerating, driven in part by the convergence of communications, computers and signal processing in both the consumer arena and in advanced industrial and government applications.

The field of signal processing has always benefited from a close coupling between theory, applications, and technologies for implementing signal processing systems. The growing number of applications and demand for increasingly sophisticated algorithms goes hand-in-hand with the rapid pace of device technology for implementing signal processing systems. By some estimates the processing capability of signal processing microprocessors is likely to increase by a factor of 200 or more over the next ten years. It seems clear that in many ways the importance and role of signal processing is accelerating and expanding.

Signal processing is concerned with the representation, transformation, and manipulation of signals and the information they contain. For example, we may wish to separate two or more signals that have somehow been combined, or we may want to enhance some signal component or some parameter of a signal model. In communications systems, it is generally necessary to do pre-processing such as modulation, signal conditioning, and compression prior to transmission over a channel and then to post process at the receiver. Prior to the 1960s, the technology for signal processing was almost exclusively continuous-time analog technology.[1] The rapid evolution of digital computers and microprocessors together with some important theoretical developments such as the fast Fourier transform algorithm (FFT) caused a major shift to digital technologies, giving rise to the field of digital signal processing. A fundamental aspect of digital signal processing is that it is based on processing sequences of samples. This discrete-time nature of digital signal processing technology is also characteristic of other signal processing technologies such as surface acoustic wave (SAW) devices, charge-coupled devices (CCDs), charge transport devices (CTDs), and switched-capacitor technologies. In digital signal processing, signals are represented by sequences of finite-precision numbers, and processing is implemented using digital computation. The more general term *discrete-time signal processing* includes digital signal processing as a special case, but also includes the possibility that sequences of samples (sampled data) are processed with other discrete-time technologies. Often the distinction between the terms discrete-time signal processing and digital signal processing is of minor importance, since both are concerned with discrete-time signals. While there are many examples in which signals to be processed are inherently sequences, most applications involve the use of discrete-time technology for processing continuous-time signals. In this case, a continuous-time signal is converted into a sequence of samples, i.e., a discrete-time signal. After discrete-time processing, the output sequence is converted back to a continuous-time signal. Real-time operation is often desirable for such systems, meaning that the discrete-time system is implemented so that samples of the output are computed at the same rate at which the continuous-time signal is sampled. Discrete-time processing of continuous-time signals in real time is commonplace in communication systems, radar and sonar, speech and video coding and enhancement, and biomedical engineering to name just a few. The compact disc player is a somewhat different example in which a processed form of the input is stored (on the compact disc) and final processing is carried out in real time when the output is desired. The compact disc recording and playback system relies on many of the signal processing concepts which we discuss in this book.

Much of traditional signal processing involves processing one signal to obtain another signal. Another important class of signal processing problems is *signal interpretation*. In such problems the objective of the processing is not to obtain an output signal but to obtain a characterization of the input signal. For example, in a speech recognition or understanding system, the objective is to interpret the input signal or extract information from it. Typically, such a system will apply digital preprocessing (filtering, parameter estimation, etc.) followed by a pattern recognition system to produce a symbolic representation such as a phonemic transcription of the speech. This symbolic output can in turn be the input to a symbolic processing system, such as a rule-based expert system, to

[1] In a general context, we typically refer to the independent variable as "time" even though in specific contexts the independent variable may take on any of a broad range of possible dimensions. Consequently. continuous time and discrete time should be thought of as generic terms referring to a continuous independent variable and a discrete independent variable, respectively.

provide the final signal interpretation. Still another and relatively new category of signal processing involves the symbolic manipulation of signal processing expressions. This type of processing is particularly useful in signal processing workstations and for the computer-aided design of signal processing systems. In this class of processing, signals and systems are represented and manipulated as abstract data objects. Object-oriented programming languages provide a convenient environment for manipulating signals, systems, and signal processing expressions without explicitly evaluating the data sequences and provide the basis for this class of processing. The sophistication of systems designed to do signal expression processing is directly influenced by the incorporation of fundamental signal processing concepts, theorems, and properties such as those that form the basis for this book. For example, a signal processing environment that incorporates the property that convolution in the time domain corresponds to multiplication in the frequency domain can explore a variety of rearrangements of filtering structures, including those involving the direct use of the discrete Fourier transform and the fast Fourier transform algorithm. Similarly, environments that incorporate the relationship between sampling rate and aliasing can make effective use of decimation and interpolation strategies for filter implementation. Similar ideas are currently being explored for implementing signal processing in network environments. In this type of environment, data can potentially be tagged with a high-level description of the processing to be done and the details of the implementation can be based dynamically on the resources available on the network.

The development of object-oriented environments for computer-aided system design and for signal processing on dynamically changing networks is still in its very early stages and any detailed discussion of it is beyond the scope of this text. However, it is important to recognize that the basic concepts that are the subject of this book should not be viewed as just theoretical in nature; they are likely to become an explicit integral part of computer-aided signal processing environments, workstations, and networks.

Many of the concepts and design techniques discussed in this text are now incorporated into the structure of sophisticated software systems such as Matlab. In many cases where discrete-time signals are acquired and stored in computers, these tools allow extremely sophisticated signal processing operations to be formed from basic functions. In such cases, it is not generally necessary to know the details of the underlying algorithm that implements the computation of an operation like the FFT, but it is essential to understand what is computed and how it should be interpreted. In other words, a good understanding of the concepts considered in this text is essential for intelligent use of the signal processing software tools that are now widely available.

Signal processing problems are not confined, of course, to one-dimensional signals. Although there are some fundamental differences in the theories for one-dimensional and multidimensional signal processing, much of the material that we discuss in this text has a direct counterpart in multidimensional systems. The theory of multidimensional digital signal processing is presented in detail in Dudgeon and Mersereau (1984), Lim (1989), and Bracewell (1986).[2] Many image processing applications require the use of two-dimensional signal processing techniques. This is the case in such areas as video coding, medical imaging, enhancement and analysis of aerial photographs, analysis

[2] Authors names and dates are used throughout the text to refer to books and papers listed in the Bibliography at the end of the book.

of satellite weather photos, and enhancement of video transmissions from lunar and deep-space probes. Applications of multidimensional digital signal processing to image processing are discussed in Andrews and Hunt (1977), Macovski (1983), Pratt (1991), Castleman (1996), Jain (1989), and Chellappa et al. (1998). Seismic data analysis as required in oil exploration, earthquake measurement, and nuclear test monitoring also utilizes multidimensional signal processing techniques. Seismic applications are discussed in Robinson and Treitel (1980) and Robinson and Durrani (1985).

Multidimensional signal processing is only one of many advanced and specialized topics that build on the fundamentals covered in this text. Spectral analysis based on the use of the discrete Fourier transform and the use of signal modeling is another particularly rich and important aspect of signal processing. We introduce many facets of this topic, focusing on the basic concepts and techniques relating to the use of the discrete Fourier transform. In addition to these techniques, a variety of spectral analysis methods rely in one way or another on specific signal models. For example, a class of high-resolution spectral analysis methods referred to as maximum entropy methods (MEM spectral analysis) is based on representing the signal to be analyzed as the response of a discrete-time linear time-invariant filter to either an impulse or to white noise. Spectral analysis is achieved by estimating the parameters (e.g., the difference equation coefficients) of the system and then evaluating the magnitude squared of the frequency response of the model filter. A thorough and detailed treatment of the issues and techniques of this approach to signal modeling and spectral analysis builds directly from the fundamentals in this text. Detailed discussions can be found in the texts by Kay (1988), Marple (1987), and Hayes (1996). Signal modeling also plays an important role in data compression and coding, and again the fundamentals of difference equations provide the basis for understanding many of these techniques. For example, one class of signal coding techniques, referred to as linear predictive coding (LPC), exploits the notion that if a signal is the response of a certain class of discrete-time filters, the signal value at any time index is a linear function of (and thus linearly predictable from) previous values. Consequently, efficient signal representations can be obtained by estimating these prediction parameters and using them along with the prediction error to represent the signal. The signal can then be regenerated when needed from the model parameters. This class of signal coding techniques has been particularly effective in speech coding and is described in considerable detail in Jayant and Noll (1984), Markel and Gray (1976), Rabiner and Schafer (1978), and Deller et al. (1993).

Another advanced topic of considerable importance is adaptive signal processing. Adaptive systems represent a particular class of time-varying and, in some sense, nonlinear systems with broad application and with established and effective techniques for their design and analysis. Again, many of these techniques build from the fundamentals of discrete-time signal processing covered in this text. Details of adaptive signal processing are given by Haykin (1996), and Widrow and Stearns (1985).

These represent only a few of the many advanced topics that extend from the topics covered in this text. Others include advanced and specialized filter design procedures, a variety of specialized algorithms for evaluation of the Fourier transform, specialized filter structures, and various advanced multirate signal processing techniques including filter banks and wavelet transforms.

It is often said that the purpose of a fundamental textbook should be to uncover rather than cover a subject, and in choosing the topics and depth of coverage in this book we have been guided by this philosophy. The preceding brief discussion of advanced topics and the Bibliography at the end of the book should be strongly suggestive of the rich variety of directions that these fundamentals begin to uncover.

HISTORICAL PERSPECTIVE

Discrete-time signal processing has advanced in uneven steps over a long period of time. Looking back at the development of the field provides a valuable perspective on fundamentals that will remain central to the field long into the future. Since the invention of calculus in the 17th century, scientists and engineers have developed models to represent physical phenomena in terms of functions of continuous variables and differential equations. Numerical techniques have been used to solve these equations when analytical solutions are not possible. Indeed, Newton used finite-difference methods that are special cases of some of the discrete-time systems that we present in this text. Mathematicians of the 18th century, such as Euler, Bernoulli, and Lagrange, developed methods for numerical integration and interpolation of functions of a continuous variable. Interesting historical research by Heideman, Johnson, and Burrus (1984) showed that Gauss discovered the fundamental principle of the fast Fourier transform (discussed in Chapter 9) as early as 1805–even before the publication of Fourier's treatise on harmonic series representation of functions.

Until the early 1950s, signal processing as we have defined it was typically done with analog systems that were implemented with electronic circuits or even with mechanical devices. Even though digital computers were becoming available in business environments and in scientific laboratories, they were expensive and had relatively limited capabilities. About that time, the need for more sophisticated signal processing in some application areas created considerable interest in discrete-time signal processing. One of the first uses of digital computers in digital signal processing was in oil prospecting, where seismic data could be recorded on magnetic tape for later processing. This type of signal processing could not generally be done in real time; minutes or even hours of computer time were often required to process only seconds of data. Even so, the flexibility of the digital computer and the potential payoffs made this alternative extremely inviting.

Also in the 1950s, the use of digital computers in signal processing arose in a different way. Because of the flexibility of digital computers, it was often useful to simulate a signal processing system on a digital computer before implementing it in analog hardware. In this way, a new signal processing algorithm or system could be studied in a flexible experimental environment before committing economic and engineering resources to constructing it. Typical examples of such simulations were the vocoder simulations carried out at Lincoln Laboratory and Bell Laboratories. In the implementation of an analog channel vocoder, for example, the filter characteristics affected the perceived quality of the coded speech signal in ways that were difficult to quantify objectively. Through computer simulations, these filter characteristics could be adjusted and the perceived quality of a speech coding system evaluated prior to construction of the analog equipment.

In all of these examples of signal processing using digital computers, the computer offered tremendous advantages in flexibility. However, the processing could not be done in real time. Consequently, a prevalent attitude was that the digital computer was being used to *approximate,* or *simulate,* an analog signal processing system. In keeping with that style, early work on digital filtering was very much concerned with ways in which a filter could be programmed on a digital computer so that with analog-to-digital conversion of the signal, followed by digital filtering, followed by digital-to-analog conversion, the overall system approximated a good analog filter. The notion that digital systems might, in fact, be practical for the actual real-time implementation of signal processing in speech communication, radar processing, or any of a variety of other applications seemed at the most optimistic times to be highly speculative. Speed, cost, and size were, of course, three of the important factors in favor of the use of analog components.

As signals were being processed on digital computers, researchers had a natural tendency to experiment with increasingly sophisticated signal processing algorithms. Some of these algorithms grew out of the flexibility of the digital computer and had no apparent practical implementation in analog equipment. Thus, many of these algorithms were treated as interesting, but somewhat impractical, ideas. The development of such signal processing algorithms made the notion of all-digital implementation of signal processing systems even more tempting. Active work began on the investigation of digital vocoders, digital spectrum analyzers, and other all-digital systems, with the hope that eventually such systems would become practical.

The evolution of a new point of view toward discrete-time signal processing was further accelerated by the disclosure by Cooley and Tukey (1965) of an efficient algorithm for computation of Fourier transforms. This class of algorithms has come to be known as the fast Fourier transform, or FFT. The FFT was significant for several reasons. Many signal processing algorithms that had been developed on digital computers required processing times several orders of magnitude greater than real time. Often this was because spectrum analysis was an important component of the signal processing and no efficient means were available for implementing it. The fast Fourier transform algorithm reduced the computation time of the Fourier transform by orders of magnitude, permitting the implementation of increasingly sophisticated signal processing algorithms with processing times that allowed interactive experimentation with the system. Furthermore, with the realization that the fast Fourier transform algorithms might, in fact, be implementable in special-purpose digital hardware, many signal processing algorithms that previously had appeared to be impractical began to appear to have practical implementations.

Another important implication of the fast Fourier transform algorithm was that it was an inherently discrete-time concept. It was directed toward the computation of the Fourier transform of a discrete-time signal or sequence and involved a set of properties and mathematics that was exact in the discrete-time domain–it was not simply an approximation to a continuous-time Fourier transform. This had the effect of stimulating a reformulation of many signal processing concepts and algorithms in terms of discrete-time mathematics, and these techniques then formed an exact set of relationships in the discrete-time domain. Following this shift away from the notion that signal processing on a digital computer was merely an approximation to analog signal processing techniques, there emerged a strong interest in discrete-time signal processing as an important field of investigation in its own right.

Another major development in the history of discrete-time signal processing occurred in the field of microelectronics. The invention and subsequent proliferation of the microprocessor paved the way for low-cost implementations of discrete-time signal processing systems. Although the first microprocessors were too slow to implement most discrete-time systems in real time, by the mid-1980s integrated circuit technology had advanced to a level that permitted the implementation of very fast fixed-point and floating-point microcomputers with architectures specially designed for implementing discrete-time signal processing algorithms. With this technology came, for the first time, the possibility of widespread application of discrete-time signal processing techniques.

FUTURE PROMISE

Microelectronics engineers continue to strive for increased circuit densities and production yields, and as a result, the complexity and sophistication of microelectronic systems are continually increasing. Indeed, complexity and capability of DSP chips have grown exponentially since the early 1980s and show no sign of slowing down. As wafer-scale integration techniques become highly developed, very complex discrete-time signal processing systems will be implemented with low cost, miniature size, and low power consumption. Consequently, the importance of discrete-time signal processing will almost certainly continue to increase and the future development of the field is likely to be even more dramatic than the course of development that we have just described. Discrete-time signal processing techniques are already promoting revolutionary advances in some fields of application. A notable example is in the area of telecommunications, where discrete-time signal processing techniques, microelectronic technology, and fiber optic transmission combine to change the nature of communication systems in truly revolutionary ways. A similar impact can be expected in many other areas of technology.

While discrete-time signal processing is a dynamic, rapidly growing field, its fundamentals are well formulated. Our goal in this book is to provide a coherent treatment of the theory of discrete-time linear systems, filtering, sampling, and discrete-time Fourier analysis. The topics presented should provide the reader with the knowledge necessary for an appreciation of the wide scope of applications for discrete-time signal processing and a foundation for contributing to future developments in this exciting field of technology.