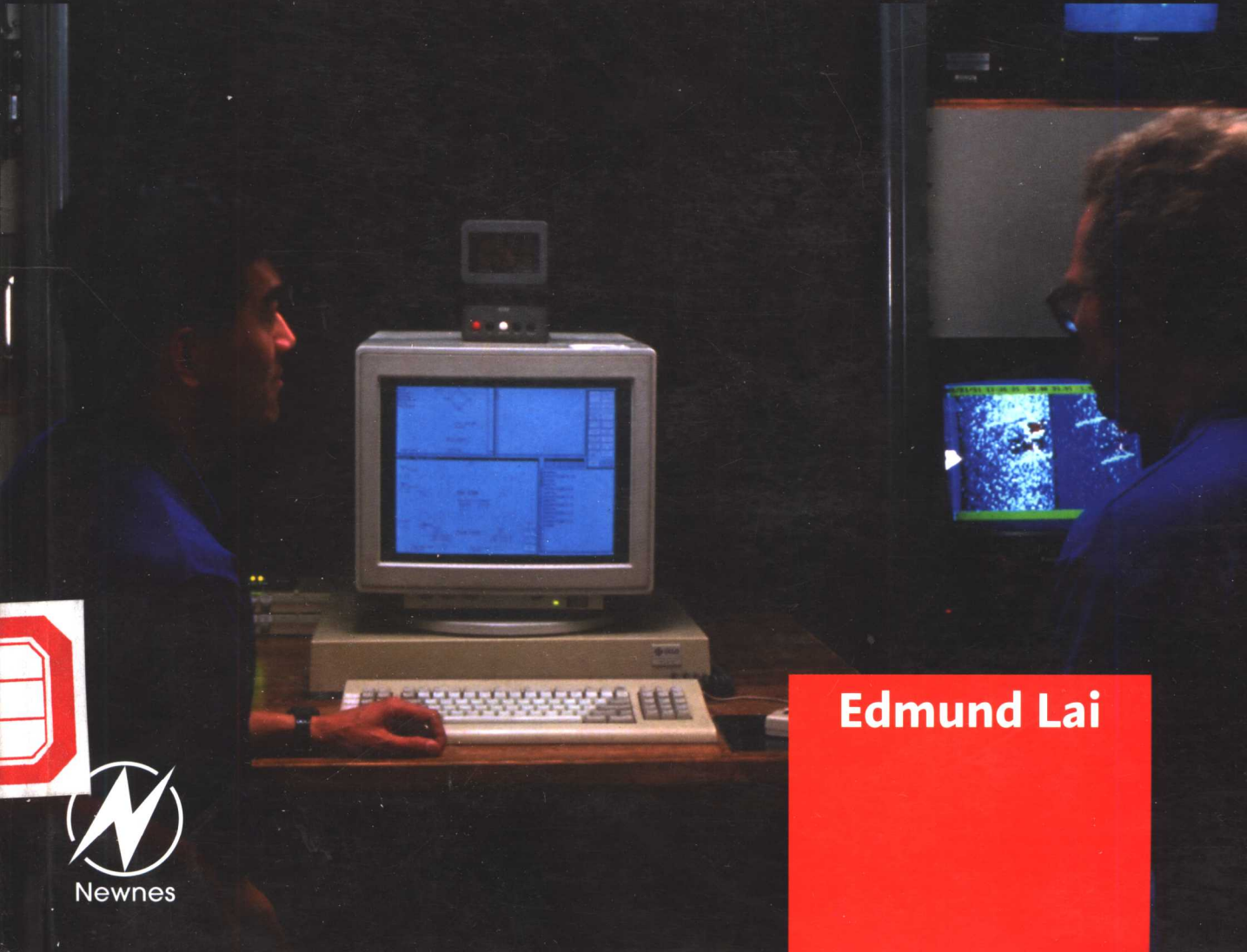


Practical

Digital Signal Processing

for Engineers and Technicians

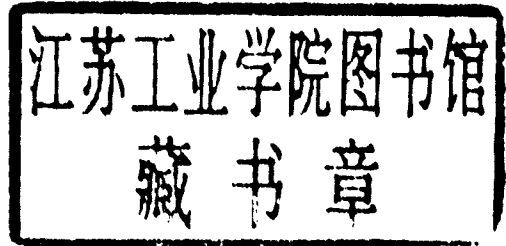


Edmund Lai



Practical Digital Signal Processing for Engineers and Technicians

Edmund Lai PhD, BEng; Lai and Associates, Singapore



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Newnes is an imprint of Elsevier



Newnes
An imprint of Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP
200 Wheeler Road, Burlington, MA 01803

First published 2004

Copyright © 2004, IDC Technologies. All rights reserved

No part of this publication may be reproduced in any material form (including photocopying or storing in any medium by electronic means and whether or not transiently or incidentally to some other use of this publication) without the written permission of the copyright holder except in accordance with the provisions of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London, England W1T 4LP. Applications for the copyright holder's written permission to reproduce any part of this publication should be addressed to the publisher

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 07506 57987

For information on all Newnes publications,
visit our website at www.newnespress.com

Typeset and Edited by Vivek Mehra, Mumbai, India

Printed and bound in Great Britain

Contents

1	Introduction	1
1.1	Benefits of processing signals digitally	1
1.2	Definition of some terms	2
1.3	DSP systems	3
1.4	Some application areas	4
1.5	Objectives and overview of the book	12
2	Converting analog to digital signals and vice versa	14
2.1	A typical DSP system	14
2.2	Sampling	15
2.3	Quantization	24
2.4	Analog-to-digital converts	34
2.5	Analog reconstruction	42
2.6	Digital-to-analog converters	46
2.7	To probe further	48
3	Time-domain representation of discrete-time signals and systems	50
3.1	Notation	50
3.2	Typical discrete-time signals	50
3.3	Operations on discrete-time signals	52
3.4	Classification of systems	54
3.5	The concept of convolution	55
3.6	Autocorrelation and cross-correlation of sequences	57
4	Frequency-domain representation of discrete-time signals	61
4.1	Discrete Fourier series for discrete-time periodic signals	62
4.2	Discrete Fourier transform for discrete-time aperiodic signals	63
4.3	The inverse discrete Fourier transform and its computation	64
4.4	Properties of the DFT	64
4.5	The fast Fourier transform	67
4.6	Practical implementation issues	71
4.7	Computation of convolution using DFT	74
4.8	Frequency ranges of some natural and man-made signals	78
5	DSP application examples	79
5.1	Periodic signal generation using wave tables	80
5.2	Wireless transmitter implementation	83
5.3	Speech synthesis	88
5.4	Image enhancement	91

5.5	Active noise control	94
5.6	To probe further	97
6	Finite impulse response filter design	98
6.1	Classification of digital filters	98
6.2	Filter design process	99
6.3	Characteristics of FIR filters	102
6.4	Window method	106
6.5	Frequency sampling method	128
6.6	Parks-McClelland method	134
6.7	Linear programming method	141
6.8	Design examples	142
6.9	To probe further	144
7	Infinite impulse response (IIR) filter design	145
7.1	Characteristics of IIR filters	146
7.2	Review of classical analog filter	147
7.3	IIR filters from analog filters	157
7.4	Direct design methods	165
7.5	FIR vs IIR	169
7.6	To probe further	170
8	Digital filter realizations	171
8.1	Direct form	171
8.2	Cascade form	179
8.3	Parallel form	181
8.4	Other structures	183
8.5	Software implementation	186
8.6	Representation of numbers	187
8.7	Finite word-length effects	191
9	Digital signal processors	204
9.1	Common features	204
9.2	Hardware architecture	206
9.3	Special instructions and addressing modes	215
9.4	General purpose microprocessors for DSP	224
9.5	Choosing a processor	224
9.6	To probe further	225

10	Hardware and software development tools	226
10.1	DSP system design flow	226
10.2	Development tools	231
Appendix A		238
Appendix B		242
Index		285

1

Introduction

Digital signal processing (DSP) is a field which is primarily technology driven. It started from around mid 1960s when digital computers and digital circuitry became fast enough to process large amounts of data efficiently.

When the term 'digital' is used, often it loosely refers to a finite set of distinct values. This is in contrast to 'analog', which refers to a continuous range of values. In digital signal processing we are concerned with the processing of signals which are discrete in time (sampled) and in most cases, discrete in amplitude (quantized) as well. In other words, we are primarily dealing with data sequences – sequences of numbers.

Such discrete (or digital) signals may arise in one of the following two distinct circumstances:

- The signal may be inherently discrete in time (and/or amplitude)
- The signal may be a sampled version of a continuous-time signal

Examples of the first type of data sequences include monthly sales figures, daily highest/lowest temperatures, stock market indices and students examination marks. Business people, meteorologists, economists, and teachers process these types of data sequences to determine cyclic patterns, trends, and averages. The processing usually involves filtering to remove as much 'noise' as possible so that the pattern of interest will be enhanced or highlighted.

Examples of the second type of discrete-time signals can readily be found in many engineering applications. For instance, speech and audio signals are sampled and then encoded for storage or transmission. A compact disc player reads the encoded digital audio signals and reconstructs the continuous-time signals for playback.

1.1 Benefits of processing signals digitally

A typical question one may ask is why process signals digitally? For the first type of signals discussed previously, the reason is obvious. If the signals are inherently discrete in time, the most natural way to process them is using digital methods. But for continuous-time signals, we have a choice.

Analog signals have to be processed by analog electronics while computers or microprocessors can process digital signals. Analog methods are potentially faster since the analog circuits process signals as they arrive in real-time, provided the settling time is fast enough. On the other hand, digital techniques are algorithmic in nature. If the computer is fast and the algorithms are efficient, then digital processing can be performed in 'real-time' provided the data rate is 'slow enough'. However, with the speed of digital logic increasing exponentially, the upper limit in data rate that can still be considered as real-time processing is becoming higher and higher.

The major advantage of digital signal processing is consistency. For the same signal, the output of a digital process will always be the same. It is not sensitive to offsets and drifts in electronic components.

The second main advantage of DSP is that very complex digital logic circuits can be packed onto a single chip, thus reducing the component count and the size and reliability of the system.

1.2 Definition of some terms

DSP has its origin in electrical/electronic engineering (EE). Therefore the terminology used in DSP are typically that of EE. If you are not an electrical or electronic engineer, there is no problem. In fact many of the terms that are used have counterparts in other engineering areas. It just takes a bit of getting used to.

For those without an engineering background, we shall now attempt to explain a few terms that we shall be using throughout the manual.

- **Signals**

We have already started using this term in the previous section. A signal is simply a quantity that we can measure over a period of time. This quantity usually changes with time and that is what makes it interesting. Such quantities could be voltage or current. They could also be the pressure, fluid level and temperature. Other quantities of interest include financial indices such as the stock market index. You will be surprised how much of the concepts in DSP has been used to analyze the financial market.

- **Frequency**

Some signals change slowly over time and others change rapidly. For instance, the (AC) voltage available at our household electrical mains goes up and down like a sine function and they complete one cycle in 50 times or 60 times a second. This signal is said to have a frequency of 50 or 60 hertz (Hz).

- **Spectrum**

While some signals consist of only a single frequency, others have a combination of a range of frequencies. If you play a string on the violin, there is a fundamental tone (frequency) corresponding to the musical note that is played. But there are other harmonics (integer multiples of the fundamental frequency) present. This musical sound signal is said to have a spectrum of frequencies. The spectrum is a frequency (domain) representation of the time (domain) signal. The two representations are equivalent.

- **Low-pass filter**

Filters let a certain range of frequency components of a signal through while rejecting the other frequency components. A low-pass filter lets the 'low-frequency' components through. Low-pass filters have a cutoff frequency below which the frequency components can pass through the filter. For

instance, if a signal has two frequency components, say 10 Hz and 20 Hz, applying a low-pass filter to this signal with a cutoff frequency of 15 Hz will result in an output signal, which has only one frequency component at 10 Hz; the 20 Hz component has been rejected by the filter.

- **Band-pass filter**

Band-pass filters are similar to low-pass filters in that only a range of frequency components can pass through it intact. This range (the passband) is usually above the DC (zero frequency) and somewhere in the mid-range. For instance, we can have a band-pass filter with a passband between 15 and 25 Hz. Applying this filter to the signal discussed above will result in a signal having only a 20 Hz component.

- **High-pass filter**

These filters allow frequency components above a certain frequency (cutoff) to pass through intact, rejecting the ones lower than the cutoff frequency.

This should be enough to get us going. New terms will arise from time to time and they will be explained as we come across them.

1.3 DSP systems

DSP systems are discrete-time systems, which means that they accept digital signals as input and output digital signals (or information extracted). Since digital signals are simply sequences of numbers, the input and output relationship of a discrete-time system can be illustrated as in Figure 1.1. The output sequence of sample $y(n)$ is computed from the input sequence of sample $x(n)$ according to some rules, which the system (H) defines.

There are two main methods by which the output sequence is computed from the input sequence. They are called sample-by-sample processing and block processing respectively. We shall encounter both types of processing in later chapters. Most systems can be implemented with either processing method. The output obtained in both cases should be equivalent if the input and the system H are the same.

1.3.1 Sample-by-sample processing

With the sample-by-sample processing method, normally one output sample is obtained when one input sample is presented to the system.

For instance, if the sequence $\{y_0, y_1, y_2, \dots, y_n, \dots\}$ is obtained when the input sequence $\{x_0, x_1, x_2, \dots, x_n, \dots\}$ is presented to the system. The sample y_0 appears at the output when the input x_0 is available at the input. The sample y_1 appears at the output when the input x_1 is available at the input, etc.

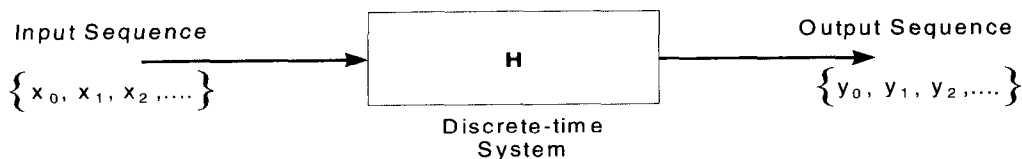


Figure 1.1
A discrete-time system

The delay between the input and output for sample-by-sample processing is at most one sample. The processing has to be completed before the next sample appears at the input.

1.3.2 Block processing

With block processing methods, a block of signal samples is being processed at a time. A block of samples is usually treated as a vector, which is transformed, to an output vector of samples by the system transformation H .

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \end{bmatrix} \xrightarrow{H} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \end{bmatrix} = \mathbf{y}$$

The delay between input and output in this case is dependent on the number of samples in each block. For example, if we use 8 samples per block, then the first 8 input samples have to be buffered (or collected) before processing can proceed. So the block of 8 output samples will appear at least 8 samples after the first sample x_0 appears. The block computation (according to H) has to be completed before the next block of 8 samples are collected.

1.3.3 Remarks

Both processing methods are extensively used in real applications. We shall encounter DSP algorithms and implementation that uses one or the other. The reader might find it useful in understanding the algorithms or techniques being discussed by realizing which processing method is being used.

1.4 Some application areas

Digital signal processing is being applied to a large range of applications. No attempt is made to include all areas of application here. In fact, new applications are constantly appearing. In this section, we shall try to describe a sufficiently broad range of applications so that the reader can get a feel of what DSP is about.

1.4.1 Speech and audio processing

An area where DSP has found a lot of application is in speech processing. It is also one of the earliest applications of DSP. Digital speech processing includes three main sub-areas: encoding, synthesis, and recognition.

1.4.1.1 Speech coding

There is a considerable amount of redundancy in the speech signal. The encoding process removes as much redundancy as possible while retaining an acceptable quality of the remaining signal. Speech coding can be further divided into two areas:

- Compression – a compact representation of the speech waveform without regard to its meaning.
- Parameterization – a model that characterizes the speech in some linguistically or acoustically meaningful form.

The minimum channel bandwidth required for the transmission of an acceptable quality of speech is around 3 kHz, with a dynamic range of 72 dB. This is normally referred to as telephone quality. Converting into digital form, a sampling rate of 8 k samples per second

with a 12-bit quantization (2^{12} amplitude levels) is commonly used, resulting in 96 k bits per second of data. This data rate can be significantly reduced without affecting the quality of the reconstructed speech as far as the listener is concerned. We shall briefly describe three of them:

- **Companding or non-uniform quantization**

The dynamic range of speech signals is very large. This is due to the fact that voiced sounds such as vowels contain a lot of energy and exhibit wide fluctuations in amplitude while unvoiced sounds like fricatives generally have much lower amplitudes. A compander (compressor–expander) compresses the amplitude of the signal at the transmitter end and expands it at the receiver end. The process is illustrated schematically in Figure 1.2. The compressor compresses the large amplitude samples and expands the small amplitude ones while the expander does the opposite.

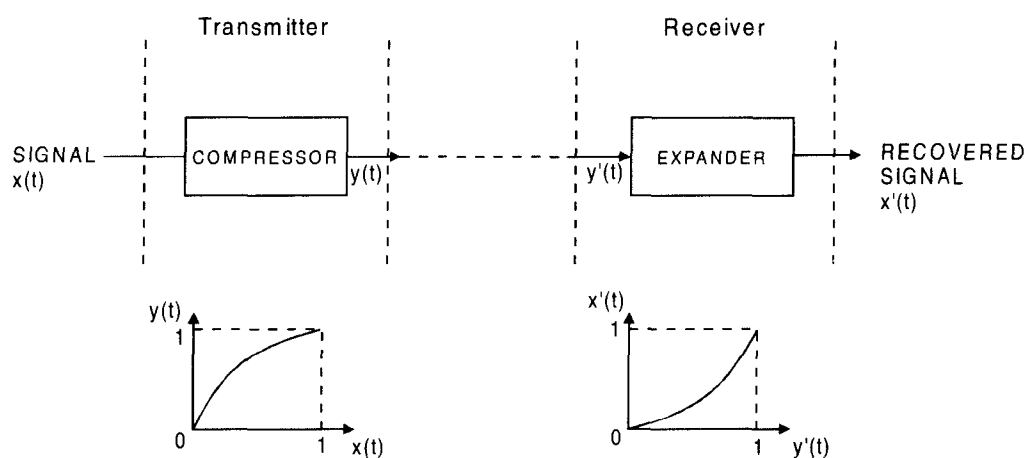


Figure 1.2
Schematic diagram showing the companding process

The μ -law compander (with $\mu = 255$) is a North American standard. A-law companding with $A = 87.56$ is a European (CCITT) standard. The difference in performance is minimal. A-law companding gives slightly better performance at high signal levels while μ -law is better at low levels.

- **Adaptive differential quantization**

At any adequate sampling rate, consecutive samples of the speech signal are generally highly correlated, except for those sounds that contain a significant amount of wideband noise. The data rate can be greatly reduced by quantizing the difference between two samples instead. Since the dynamic range will be much reduced by differencing, the number of levels required for the quantifier will also be reduced.

The concept of differential quantization can be extended further. Suppose we have an estimate of the value of the current sample based on information from the previous samples, then we can quantize the difference between the current sample and its estimate. If the prediction is accurate enough, this difference will be quite small.

Figure 1.3 shows the block diagram of an adaptive differential pulse code modulator (ADPCM). It takes a 64 kbits per second pulse code modulated (PCM) signal and encodes it into 32 kbit per second adaptive differential pulse code modulated (ADPCM) signal.

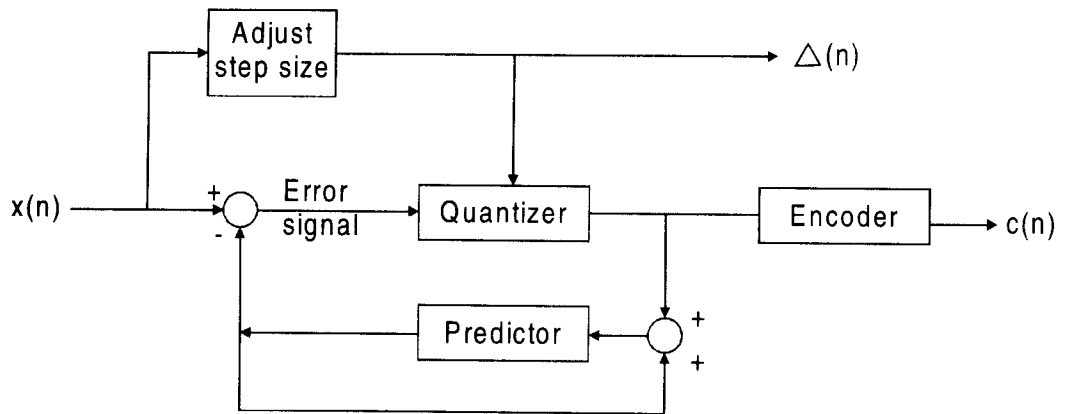


Figure 1.3
Block diagram of an adaptive differential pulse code modulator

• **Linear prediction**

The linear predictive coding method of speech coding is based on a (simplified) model of speech production shown in Figure 1.4.

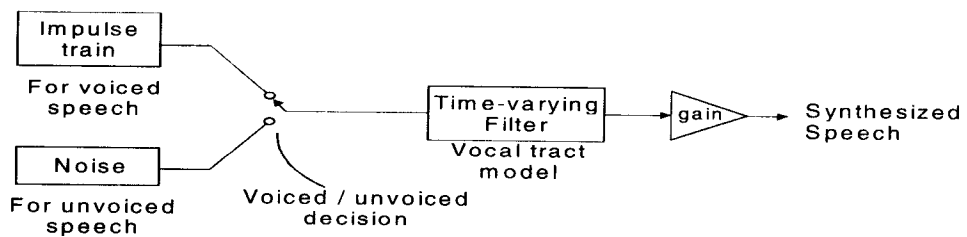


Figure 1.4
A model of speech production

The time-varying digital filter models the vocal tract and is driven by an excitation signal. For voiced speech, this excitation signal is typically a train of scaled unit impulses at pitch frequency. For unvoiced sounds it is random noise.

The analysis system (or encoder) estimates the filter coefficients, detects whether the speech is voiced or unvoiced and estimates the pitch frequency if necessary. This is performed for each overlapping section of speech usually around 10 milliseconds in duration. This information is then encoded and transmitted. The receiver reconstructs the speech signal using these parameters based on the speech production model. It is interesting to note that the reconstructed speech is similar to the original perceptually but the physical

appearance of the signal is very different. This is an illustration of the redundancies inherent in speech signals.

1.4.1.2 Speech synthesis

The synthesis or generation of speech can be done through the speech production model mentioned above. Although the duplication of the acoustics of the vocal tract can be carried out quite accurately, the excitation model turns out to be more problematic.

For synthetic speech to sound natural, it is essential that the correct allophone be produced. Despite the fact that different allophones are perceived as the same sound, if the wrong allophone is selected, the synthesized speech will not sound natural. Translation from phonemes to allophones is usually controlled by a set of rules. The control of timing of a word is also very important. But these rules are beyond the realm of DSP.

1.4.1.3 Speech recognition

One of the major goals of speech recognition is to provide an alternative interface between human user and machine. Speech recognition systems can either be speaker dependent or independent, and they can either accept isolated utterances or continuous speech. Each system is capable of handling a certain vocabulary.

The basic approach to speech recognition is to extract features of the speech signals in the training phase. In the recognition phase, the features extracted from the incoming signal are compared to those that have been stored. Owing to the fact that our voice changes with time and the rate at which we speak also varies, speech recognition is a very tough problem. However, there are now commercially available some relatively simple small vocabulary, isolated utterance recognition systems. This comes about after 30 years of research and the advances made in DSP hardware and software.

1.4.2 Image and video processing

Image processing involves the processing of signals, which are two-dimensional. A digital image consists of a two dimensional array of pixel values instead of a one dimensional one for, say, speech signals. We shall briefly describe three areas of image processing.

1.4.2.1 Image enhancement

Image enhancement is used when we need to focus or pick out some important features of an image. For example, we may want to sharpen the image to bring out details such as a car license plate number or some areas of an X-ray film. In aerial photographs, the edges or lines may need to be enhanced in order to pick out buildings or other objects. Certain spectral components of an image may need to be enhanced in images obtained from telescopes or space probes. In some cases, the contrast may need to be enhanced.

While linear filtering may be all that is required for certain types of enhancement, most useful enhancement operations are non-linear in nature.

1.4.2.2 Image restoration

Image restoration deals with techniques for reconstructing an image that may have been blurred by sensor or camera motion and in which additive noise may be present. The blurring process is usually modeled as a linear filtering operation and the problem of

image restoration then becomes one of identifying the type of blur and estimating the parameters of the model. The image is then filtered by the inverse of the filter.

1.4.2.3 Image compression and coding

The amount of data in a visual image is very large. A simple black-and-white still picture digitized to a 512×512 array of pixels using 8 bits per pixel involves more than 2 million bits of information. In the case of sequences of images such as in video or television images, the amount of data involved will be even greater. Image compression, like speech compression, seeks to reduce the number of bits required to store or transmit the image with either no loss or an acceptable level of loss or distortion. A number of different techniques have been proposed, including prediction or coding in the (spatial) frequency domain. The most successful techniques typically combine several basic methods. Very sophisticated methods have been developed for digital cameras and digital video discs (DVD).

Standards have been developed for the coding of both image and video signals for different kinds of applications. For still images, the most common one is JPEG. For high quality motion video, there is MPEG and MPEG-2. MPEG-2 was developed with high definition television in mind. It is now used in satellite transmission of broadcast quality video signals.

1.4.3 Adaptive filtering

A major advantage of digital processing is its ability of adapting to changing environments. Even though adaptive signal processing is a more advanced topic, which we will not cover in this course, we shall describe the basic ideas involved in adaptive signal processing and some of its applications.

A basic component in an adaptive digital signal processing system is a digital filter with adjustable filter coefficients – a time-varying digital filter. Changing the characteristics of a filter by a change in the coefficient values is a very simple operation in DSP. The adaptation occurs through an algorithm which takes a reference (or desired) signal and an error signal produced by the difference between the current output of the filter and the current input signal. The algorithm adjusts the filter coefficients so that the averaged error is minimized.

1.4.3.1 Noise cancellation

One example of noise cancellation is the suppression of the maternal ECG component in fetal ECG. The fetal heart rate signal can be obtained from a sensor placed in the abdominal region of the mother. However, this signal is very noisy due to the mother's heartbeat and fetal motion.

The idea behind noise cancellation in this case is to take a direct recording of the mother's heartbeat and after filtering of this signal, subtract it off the fetal heart rate signal to get a relatively noise-free heart rate signal. A schematic diagram of the system is shown in Figure 1.5.

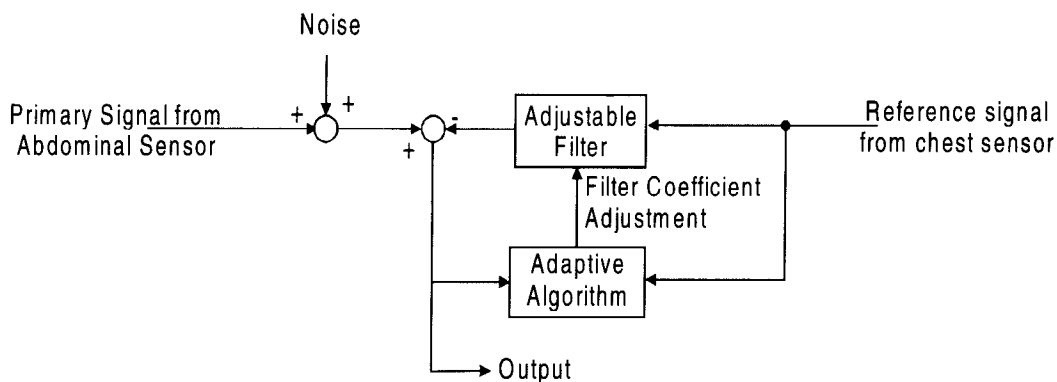


Figure 1.5
An adaptive noise cancellation system

There are two inputs: a primary and a reference. The primary signal is of interest but has a noisy interference component, which is correlated with the reference signal. The adaptive filter is used to produce an estimate of this interference or noise component, which is then subtracted off the primary signal. The filter should be chosen to ensure that the error signal and the reference signal are uncorrelated.

1.4.3.2 Echo cancellation

Echoes are signals that are identical to the original signals but are attenuated and delayed in time. They are typically generated in long distance telephone communication due to impedance mismatch. Such a mismatch usually occurs at the junction or hybrid between the local subscriber loop and the long distance loop. As a result of the mismatch, incident electromagnetic waves are reflected which sound like echoes to the telephone user.

The idea behind echo cancellation is to predict the echo signal values and thus subtract it out. The basic mechanism is illustrated in Figure 1.6. Since the speech signal is constantly changing, the system has to be adaptive.

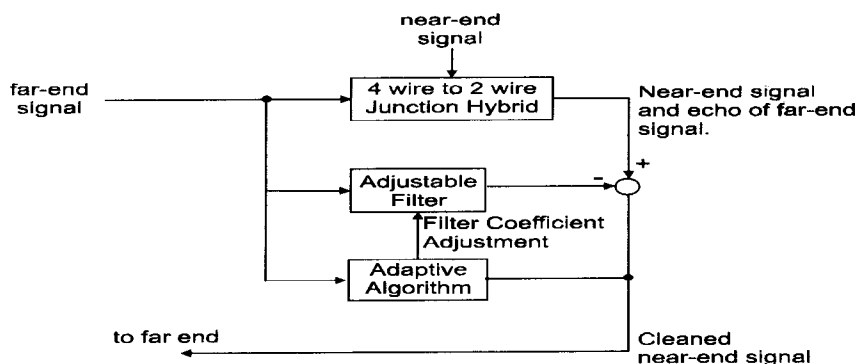


Figure 1.6
An adaptive echo cancellation system

1.4.3.3 Channel equalization

Consider the transmission of a signal over a communication channel (e.g. coaxial cable, optical fiber, wireless). The signal will be subject to channel noise and dispersion caused,

for example, by reflection from objects such as buildings in the transmission path. This distorted signal will have to be reconstructed by the receiver.

One way to restore the original signal is to pass the received signal through an equalizing filter to undo the dispersion effects. The equalizer should ideally be the inverse of the channel characteristics. However, channel characteristics typically drift in time and so the equalizer (a digital filter) coefficients will need to be adjusted continuously. If the transmission medium is a cable, the drift will occur very slowly. But for wireless channels in mobile communications the channel characteristics change rapidly and the equalizer filter will have to adapt very quickly.

In order to 'learn' the channel characteristics, the adaptive equalizer operates in a training mode where a pre-determined training signal is transmitted to the receiver. Normal signal transmission has to be regularly interrupted by a brief training session so that the equalizer filter coefficients can be adjusted. Figure 1.7 shows an adaptive equalizer in training mode.

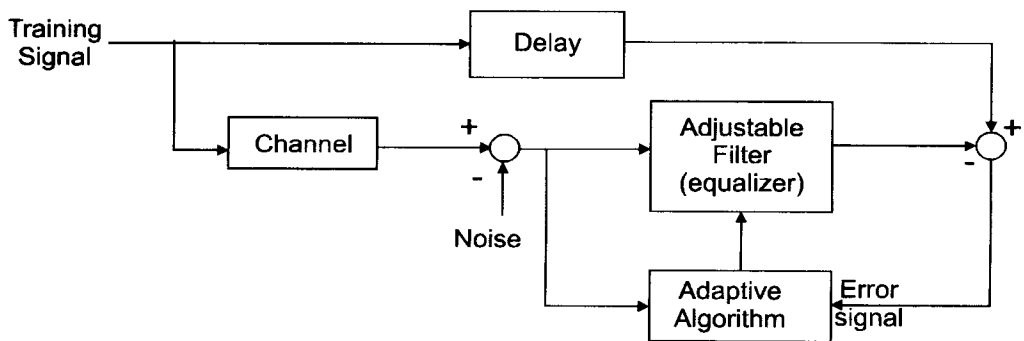


Figure 1.7
An adaptive equalizer in training mode

1.4.4 Control applications

A digital controller is a system used for controlling closed-loop feedback systems as shown in Figure 1.8. The controller implements algebraic algorithms such as filters and compensatory to regulate, correct, or change the behavior of the controlled system.

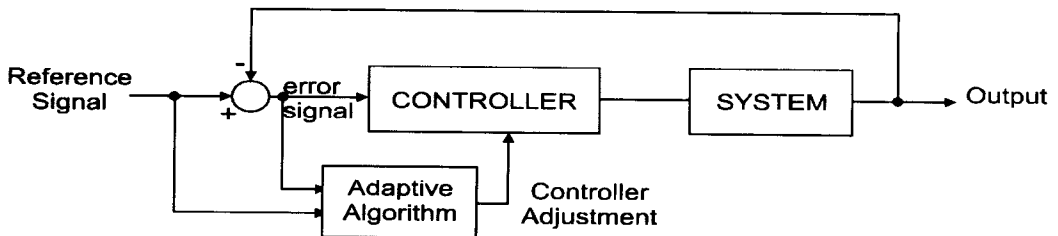


Figure 1.8
A digital closed-loop control system

Digital control has the advantage that complex control algorithms are implemented in software rather than specialized hardware. Thus the controller design and its parameters can easily be altered. Furthermore, increased noise immunity is guaranteed and parameter

drift is eliminated. Consequently, they tend to be more reliable and at the same time, feature reduced size, power, weight and cost.

Digital signal processors are very useful for implementing digital controllers since they are typically optimized for digital filtering operations with single instruction arithmetic operations. Furthermore, if the system being controlled changes with time, adaptive control algorithms, similar to adaptive filtering discussed above, can be implemented.

1.4.5 Sensor or antenna array processing

In some applications, a number of spatially distributed sensors are used for receiving signals from some sources. The problem of coherently summing the outputs from these sensors is known as beamforming. Beyond the directivity provided by an individual sensor, a beamformer permits one to 'listen' preferentially to wave fronts propagating from one direction over another. Thus a beamformer implements a spatial filter. Applications of beamforming can be found in seismology, underwater acoustics, biomedical engineering, radio communication systems and astronomy.

In cellular mobile communication systems, smart antennas (an antenna array with digitally steerable beams) are being used to increase user capacity and expand geographic coverage. In order to increase capacity, an array, which can increase the carrier to interference ratio (C/I) at both the base station and the mobile terminal, is required. There are three approaches to maximizing C/I with an antenna array.

- The first one is to create higher gain on the antenna in the intended direction using antenna aperture. This is done by combining the outputs of each individual antenna to create aperture.
- The second approach is the mitigation of multipath fading. In mobile communication, fast fading induced by multipath propagation requires an additional link margin of 8 dB. This margin can be recovered by removing the destructive multipath effects.
- The third approach is the identification and nulling of interferers. It is not difficult for a digital beamformer to create sharp nulls, removing the effects of interference.

Direction of arrival estimation can also be performed using sensor arrays. In the simplest configuration, signals are received at two spatially separated sensors with one signal being an attenuated, delayed and noisy version of the other. If the distance between the sensors is known, and the signal velocity is known, then the direction of arrival can be estimated. If the direction does not change, or changes very slowly with time, then it can be determined by cross-correlating the two signals and finding the global maximum of the cross-correlation function. If the direction changes rapidly, then an adaptive algorithm is needed.

1.4.6 Digital communication receivers and transmitters

One of the most exciting applications of DSP is in the design and implementation of digital communication equipment. Throughout the 1970s and 80s radio systems migrated from analog to digital in almost every aspect, from system control to source and channel coding to hardware technology. A new architecture known generally as 'software radio' is emerging. This technology liberates radio-based services from dependence on hardwired characteristics such as frequency band, channel bandwidth, and channel coding.