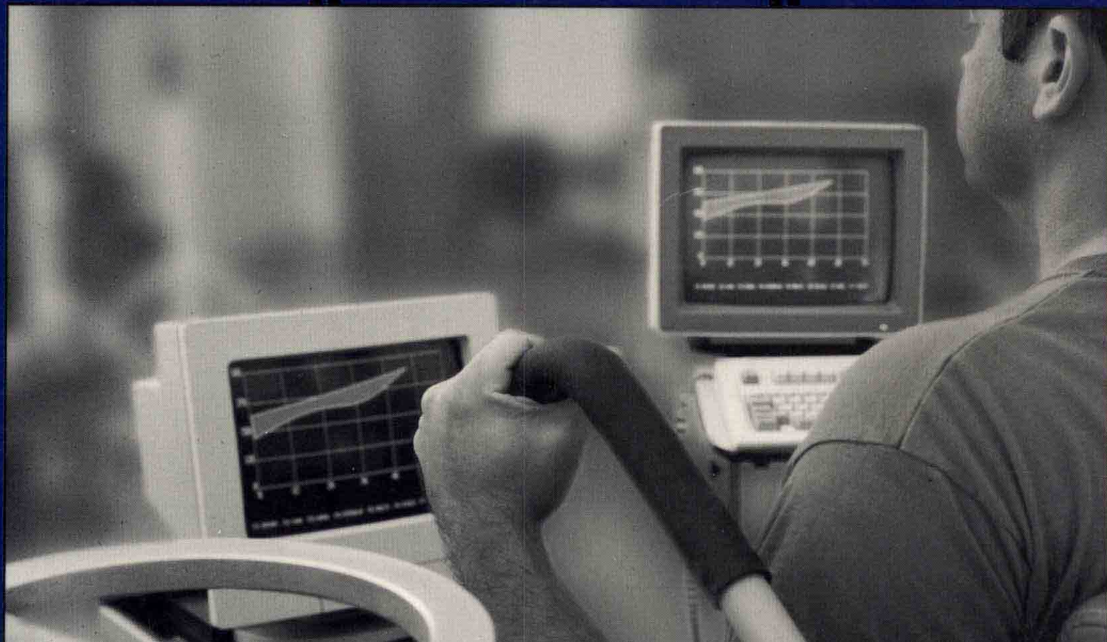
Computerized Data Acquisition and Analysis for the Life Sciences

A Hands-On Guide



Simon S. Young

Computerized Data Acquisition and Analysis for the Life Sciences

A Hands-on Guide

SIMON S. YOUNG
Schering-Plough Research Institute

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2001

Printed in the United States of America

Typeface Melior 10/13.5 pt *System* 3B2 [KW]

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

Library of Congress Cataloging-in-Publication Data
Young, Simon S., 1958–
Computerized data acquisition and analysis for the life sciences: a hands-on guide/
Simon S. Young
p. cm.
Includes bibliographical references
ISBN 0-521-56281-3 (hardback)
ISBN 0-521-56570-7 (paperback)
1. Life sciences – Data processing. 2. Automatic data collection systems. I. Title.
QH324.2.Y68 2001
570′.285 – dc21 00-045436

ISBN 0 521 56281 3 hardback
ISBN 0 521 56570 7 paperback

Preface

There has been a small revolution in data acquisition systems for scientific use in recent years. It was not so long ago that hardware and processing power were expensive and software nonexistent. Every program had to be written from scratch in a low-level language and, if you did not want your program to take all night to run, it usually included large chunks of assembler code as well. Nowadays there are many data acquisition systems to choose from, with good hardware and vastly improved software running on fast personal computers. Many such systems are purchased with the aim of using them as the primary method of recording experimental data in the laboratory. Unfortunately, it is not always as easy to set up the system as the manufacturers would have us believe and getting the machine to do what you want it to do with your apparatus can be a daunting and frustrating task. This book was written to help people collect and analyse experimental data using digital data acquisition systems and is particularly for those whose field of expertise is not engineering or computing. The book explains how data acquisition systems work, how to set them up to obtain the best performance and how to process the data. Terms which may be unfamiliar are explained so that manufacturer's specifications and literature become clearer and the differences between systems can be appreciated. The topics covered in the book are general but the examples are slanted towards the life sciences because this is often the discipline in which people have the greatest trouble setting up data acquisition systems. For people with an engineering background some of the material will be familiar but there are still useful tips and short cuts.

Program listings are not included in the book. There are some practical problems associated with publishing programs: they have to be cross-platform and written in several languages if they are to be useful to everyone. But the main reason is that I would strongly urge you to forget about low-level coding altogether. Buy a high-level language designed for signal processing (LabView, IDL, Matlab, etc.) and spend your programming time assembling algorithms from the building blocks that these programs provide, not reinventing the wheel. If you absolutely have to write code, use a library of prewritten functions (such as those sold by National Instruments) that have been tested and debugged for you.

This book would not have been possible without the help and encouragement of many people. The original idea arose out of the author's work with graduate students at the Universities of Cambridge, Guelph and Waterloo when it became clear that there was a need for a book on data acquisition and analysis written for non-engineers. Dr. Robin Gleed provided the initial encouragement that launched the project. Data for some of the examples used in Chapter 6 were provided by Dr. Jeff Thomason, Dr. Janet Douglas, and Dr. Zoe Dong. Dr. Richard Hughson ran the research project that led to the development of TSAS. Dr. Craig Johnson reviewed early versions of the manuscript and provided data for some of the examples. Dr. Robin Smith and Dr. Michael Penn of Cambridge University Press were invaluable at all stages of writing and editing the book. And most of all, I am indebted to my wife, Janet Douglas, for endless patience, encouragement, and manuscript reviews.

Contents

chapter one

The Bare Essentials

This chapter outlines the key tasks associated with setting up and using a data acquisition system. It is essentially a checklist with pointers to the sections of the book that cover the topics in more detail.

DEFINE THE VARIABLES TO BE MEASURED

The first task is to decide what specific physiological effects are to be recorded. For example, suppose the effects of a drug on cardiovascular function are being investigated. There is quite a range of variables that can be used as indicators of cardiovascular function: heart rate, blood pressure, cardiac output, systemic vascular resistance, rate of change of pressure (dp/dt), end diastolic volume, ejection fraction, myocardial fibre shortening, etc. Some of these are relatively easy to measure and some are quite involved. Often the starting point is to replicate a published method and this considerably shortens the development process.

CONVERT THE PHYSIOLOGICAL EFFECT TO A VOLTAGE

All data acquisition systems fundamentally only measure one thing: voltage. A transducer is needed (Chapter 4) to convert the variable(s) of interest (e.g. vascular pressure) into a representative voltage. Each transducer will also need an amplifier, and this is usually specific for the particular type of transducer. Even if biopotentials (which are already voltages) are being measured, one cannot just hook up an electrode to the input of a data acquisition system – it is very unlikely to work well. An amplifier specific for the type of biopotential (intracellular potential, ECG, EEG, EMG, etc.) is needed in order to obtain usable results. In addition, if recordings are being made from human subjects, an isolation amplifier is essential for safety.

SCALE THE SIGNAL

The main analogue-to-digital converter (ADC) has a range of voltages that it can accept. This range is typically ±10 V, ±5 V, or ±1 V and nearly all systems allow the range to be changed. The aim is to use a good part of this range (giving good resolution) without exceeding it (which severely distorts the signal). Matching the signal range to the ADC range is performed by setting the gain of the transducer amplifier so that the maximum expected signal is about half or two thirds of the ADC range. For example, if an ECG signal is being recorded, the maximum amplitude expected is about 3 mV (the height of a large R wave). An ECG preamplifier typically has a gain of ×1000, giving a maximum output of 3 V. Setting the ADC range to ±5 V would be a good choice, although even ±10 V would not be bad. Another example is a blood pressure transducer. These typically have a range of 0–300 mmHg, with a corresponding output from their amplifier of 0–10 V. Setting the ADC range to ±10 V will use the full range of the transducer. If the blood pressure exceeds 300 mmHg, which is unlikely, the transducer will not give a reliable reading, so using the full range of the ADC is fine.

SELECT A UNIPOLAR OR BIPOLAR INPUT RANGE

For most signals the input should be set to bipolar. If the input is set to bipolar for all signals, the system will function normally. The only disadvantage is that there is a potential loss of resolution if the signal is guaranteed never to be negative. In such cases setting the input to unipolar mode will improve the resolution by a factor of up to 2. On the other hand, if the input is set to unipolar and the signal goes negative, it will be severely distorted by the recording process. Chapter 2 has the details.

CHOOSE SINGLE-ENDED OR DIFFERENTIAL INPUTS

In most cases the inputs should be set to single-ended. Differential inputs are better at rejecting noise, particularly if the signal is small or if it is connected to the data acquisition system by long cables. On the other hand, differential inputs are more complex to wire up and in differential mode you usually only have half the number of inputs available compared to single-ended mode. Again, Chapter 2 has the details.

TABLE 1.1 SUGGESTED SAMPLING RATES FOR DIFFERENT TYPES OF BIOLOGICAL SIGNALS

Signal being recorded	Sampling rate (Hz)
Temperature, ionic composition, other variables that do not change much over a period of a few seconds	5–20
Blood pressure (except intravascular catheter-tip transducers), gas composition (using a fast analyser such as mass spectrometer), respiratory gas temperature, joint angle (goniometry), ECG, EEG	200–400
EMG, evoked potentials, action potentials	600–1,000
Intravascular catheter-tip transducers; all studies involving impacts such as force plate studies	2,000–3,000
Action potentials, transmembrane currents (patch and intracellular)	10,000–20,000

CHOOSE THE SAMPLING RATE AND ANTI-ALIASING FILTERS

Setting the sampling rate too low will distort the data ('aliasing'). Setting it too high will not do any harm, the data will just take up more storage space. Chapter 3 goes into the details. For now, Table 1.1 gives some very rough guidelines for sampling rates for different types of biological signals.

If possible, try to include a low-pass filter (an 'anti-aliasing' filter) in between the transducer amplifier and the ADC. The 'corner frequency' of the filter should be no higher than about 1/4 of the sampling rate: if the sampling rate is 400 samples/s, for example, set the filter corner frequency to below 100 Hz.

SIZE OF DATA FILES

The amount of data that will be collected can be calculated from the following equation:[1]

$$\text{amount of data (bytes)} = 2 \times \text{no. of channels} \times \text{sampling rate} \times \text{recording length (seconds)}$$

For example, recording from four channels at 200 Hz for 30 min will generate about 2.7 Mbyte of data. Some means of storing and archiving these data is needed.

CALIBRATION

Nearly all transducers require calibration before use. The calibration process relates the voltage recorded by the data acquisition system to the quantity measured by the transducer (pressure, temperature, etc.). The vast majority of transducers and transducer/amplifiers are *linear.* This means that a plot of the voltage recorded by the data acquisition system against the quantity being measured is a straight line (Figure 1.1). A linear system is particularly easy to calibrate because the straight line relating voltage to physical effect can be defined by just two points, and such a calibration process is known as a *two-point calibration.* Nonlinear transducers require a more complex calibration procedure, which is covered in Chapter 4.

To perform a two-point calibration, two known values of the variable (spanning most of the range of interest) are applied to the transducer with a reference system. The corresponding voltage is recorded both times. For example, a temperature transducer that is to be used to record mammalian body temperature might be calibrated at 35°C and 45°C. Two water baths at approximately these temperatures are set up. The exact temperature of each bath is measured with a precision thermometer and at the same time the voltage produced by the transducer (the latter is usually taped to the thermometer for this process) is recorded. Two corresponding pairs of values (temperature and voltage) are thus recorded: T_{hi}, T_{low} and V_{hi}, V_{low}. A calibration graph (Figure 1.2) now has to be constructed from these two points so that any voltage can be converted back into the corresponding temperature.

The straight line relating temperature and voltage (Figure 1.2) is usually defined by its *scale* and *offset.* The general equation of a line is

$$y = mx + c$$

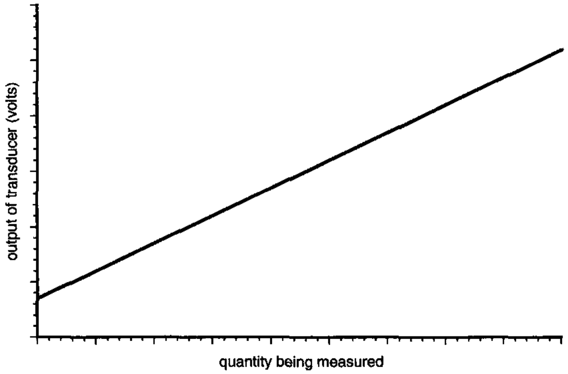
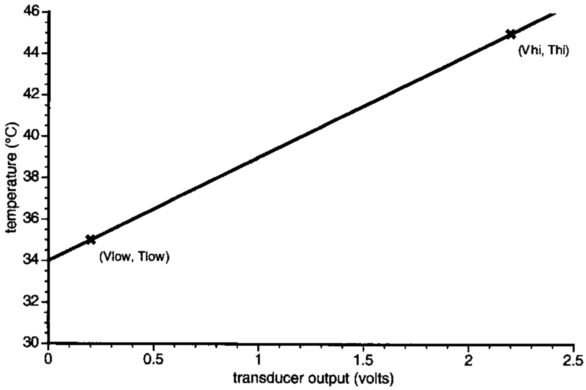


Figure 1.1 The response of a linear transducer and amplifier. A plot of the output of the transducer (in V) against the quantity being measured (temperature, pressure, etc.) is a straight line if the transducer is linear.

Figure 1.2 Two-point calibration of a linear temperature transducer. Two temperature points are selected that span the range of interest, and the output of the transducer at each temperature is measured. The scale and offset of the line joining the two points are calculated, and from this any measured voltage can be converted into a corresponding temperature. Note that the graph has temperature (the independent variable) on the *y*-axis rather than the more usual *x*-axis.



where m is the scale and c the offset. For the calibration graph (Figure 1.2) the equation of the line is given by the following equation:[2]

$$\text{temperature} = (\text{scale} \times \text{voltage}) + \text{offset}$$

The scale and offset are calculated from

$$\text{scale} = \frac{T_{\text{hi}} - T_{\text{low}}}{V_{\text{hi}} - V_{\text{low}}}$$

$$\text{offset} = T_{\text{hi}} - (\text{scale} \times V_{\text{hi}})$$

Now any voltage V can be converted to a corresponding temperature T with

$$T = (\text{scale} \times V) + \text{offset}$$

A data acquisition system is used to record the voltage produced by the temperature transducer. After the data have been collected, each voltage value has to be converted into a corresponding temperature using this equation. Many data acquisition systems calculate the scale and offset from a pair of calibration values and automatically calibrate data after they have been collected.

The raw data recorded by the data acquisition system are in their most compact form (2 bytes per sample). Once the data are calibrated they are stored in floating point format, which typically occupies 5 bytes per sample. It is therefore common to save the data in uncalibrated form, but if this is done it is essential to ensure that the calibration coefficients (scale and offset) are recorded along with the data.

DATA ANALYSIS

Recording reliable data is half of the process of data acquisition and analysis. The second (and more interesting!) part is converting the raw data into useful numbers, a topic that is covered in Chapter 5.

chapter two

How a Data Acquisition System Works

A data acquisition system for a personal computer is made up from several parts. Firstly, there will be a circuit board (often called a *card*) covered with electronic components. This is the heart of the system and it contains several sections of circuitry, each of which performs a specific function (analogue-to-digital conversion, sample and hold, variable gain amplification and multiplexing). There is also some circuitry to allow the computer to talk to the analogue-to-digital converter, and some optional extras such as memory.

The circuit board usually plugs into a socket inside the computer. This socket is wired into the bus, the main 'information highway' of the computer. Specific installation instructions come with the board but you will need to turn the computer off and open the case. The sockets are soldered directly to the main circuit board of the computer. Depending on the type of computer, there may be a single socket or a row of them, and some may already have other circuit boards plugged in to them. There may also be more than one type of socket. For example, the standard bus format used in personal computers is the PCI bus, which is the one used by peripherals such as A/D boards and network cards. The computer may also have an advanced graphic socket, which is used for high-performance video cards and has a different shape compared to the PCI sockets. Older PC format computers used the AT bus for peripherals and pre-power PC Macintoshes used the now-obsolete Nubus format. New PC format computers often retain one or two AT style sockets so that they are compatible with older equipment. You can still buy A/D boards that use the AT bus but use PCI format boards if you can, because the performance is better and they can be used in both PC and Macintosh computers. All plug-in cards have a metal bracket on one end with connectors mounted on it. When the board is fitted into the computer the bracket lines up with a slot in the back of the case, so that the connector pokes through and allows cables to be plugged in when the case is closed. Standard peripheral boards provide the necessary circuitry to allow the computer to exchange information with a printer, a network, and so on. The data acquisition board does much the same thing: it allows the computer to gather information on the world around it by measuring temperature, pressure, light levels, and so forth. To measure temperature, for instance, a sensor is used that produces a voltage proportional to its temperature. The computer cannot measure this voltage directly and the purpose of the data acquisition board is to convert the voltage into a number, something the computer can work with. Thus a data acquisition circuit board acts as a 'translator' between the computer (which works with numbers) and transducers and other devices, which work with voltages.

Sometimes the circuit board does not plug into a socket inside the computer but is enclosed in a box that sits near to the computer. These external circuit boards also need a power supply, which may be inside the box or external. The external power supplies are usually the 'brick' type that plug directly into the wall socket and connect to the A/D board with a long cable (a wonderful term for these adapters is 'wall warts'!). There are two main reasons why external A/D boards are used. Firstly, they are less affected by the electrical noise from the computer, which is important when making very precise measurements. The Instrunet system from GW Instruments uses this approach. The second reason is that in some cases an internal board cannot be used. For example, the Macintosh iMac computer does not have any internal PCI sockets.

All external A/D boards have to connect to the computer using one of the external sockets on the back of the latter. There are A/D systems available that will fit on to virtually any of the sockets on the back of a PC or Macintosh but the performance varies considerably depending on which socket is used. The newest A/D systems use either the Universal Serial Bus (USB) or the Firewire (IEEE 1394) connector. Both USB and Firewire are fast connections that allow high-speed data acquisition, and they are compatible with both PC and Macintosh computers. The older SCSI connector allows medium-speed data acquisition but it is becoming difficult to find SCSI-based A/D systems. The slowest systems are those that use the serial or parallel printer ports. These systems are not of much use for general-purpose data acquisition but have a niche market for applications where a small, cheap data logger is needed.

Standard desktop computers are fine for data acquisition systems that are going to be used in the laboratory and are the cheapest way to buy computing power. If the system has to be portable, however, laptop computers are essential. The problem with laptops is finding space to fit the A/D board. One solution is to use an external A/D board that plugs into one of the ports on the back of the computer. The other solution, which is a lot neater and easier to carry around, is to use an

A/D board that fits into the PCMCIA socket. National Instruments makes a good range of A/D cards in PCMCIA format, which are compatible with both PC and Macintosh laptops.

The data acquisition circuit board acts as a translator between the computer and the world of transducers and voltages. The board must therefore be connected to both worlds. The connection between the computer and the data acquisition board is made when the board is plugged in or, in the case of boards outside the computer, via a short cable. A connection is also needed to the transducers on the 'voltage' side of the data acquisition board. Soldering wires from the transducers directly to the circuit board is very awkward and inflexible. The usual way is to use some sort of *break-out box*. This is a plastic or metal box with a row of terminals or connectors on it. Each terminal is connected to the data acquisition circuit board by its own piece of wire and all the wires are either laid side-by-side in a wide, flat ribbon cable or bundled together into a single thick cable. This cable goes between the break-out box and the data acquisition board, and is usually a few feet long so that the break-out box is not too close to the computer. Computers give off quite a lot of radiofrequency interference that can be picked up by sensitive amplifiers, and it is a good idea to keep transducer amplifiers a few feet away from the computer and monitor. At the computer end the cable goes through a slot in the back of the computer's case and then plugs into the circuit board.

Using a break-out box is a lot more convenient than connecting the transducers directly to the data acquisition board. Each transducer is connected to one socket or terminal on the break-out box so that eight or sixteen transducers may be connected to one box. The sockets make it easy to rearrange the connections for a different experiment. Using a break-out box also protects your computer and data acquisition board from damage if someone trips over the transducer cable.

Figure 2.1 shows the main parts of a data acquisition/analysis system, and the dotted line encloses the sections that are on the data acquisition circuit board. Each component of the system is now examined in detail.

ANALOGUE-TO-DIGITAL CONVERTER

The analogue-to-digital converter (ADC) is the heart of the data acquisition system. It converts a voltage applied to its input terminals into a number which can be read by the computer. Figure 2.2 shows what an ADC does. The ADC is designed to operate over some specific *range* of voltage, which is often 0 to 10 V or −5 V to +5 V. This range is divided into a number of evenly spaced levels separated by small, equal steps.[1] A common number of levels used is 4,096, which therefore have 4,095 steps between them. If the voltage range is 0 to 10 V then each step is