Although synthesis has great potential payoff, the usefulness of such models to date has been very poor and has been limited to very simplistic movements. The major problem is that the models that have been proposed are not very valid; they lack the correct anthropometrics and degrees of freedom to make their predictions very useful. However, because of its potential payoff, it is important that students have an introduction to the process, in the hope that useful models will evolve as a result of what we learn from our minor successes and major mistakes.

1.4 REFERENCES

Bernstein, N. A. *The Coordination and Regulation of Movements.* (Pergaman Press. Oxford, UK, 1967).

DeLuca, C. J., R. A. LeFever, M. P. McCue, and A. P. Xenakis. "Control Scheme Governing Concurrently Active Motor Units During Voluntary Contractions," *J. Physiol.* **329**:129–142, 1982.

Henneman, E. and C. B. Olson. "Relations Between Structure and Function in the Design of Skeletal Muscle," *J. Neurophysiol.* **28**:581–598, 1965.

Winter, D. A. "Overall Principle of Lower Limb Support During Stance Phase of Gait," *J. Biomech.* **13**:923–927, 1980.

Winter, D. A. "Kinematic and Kinetic Patterns in Human Gait: Variability and Compensating Effects," *Human Movement Sci.* **3**:51–76, 1984.

Winter, D. A. "Biomechanics of Normal and Pathological Gait: Implications for Understanding Human Locomotor Control," *J. Motor Behav.* **21**:337–355, 1989.

2

KINEMATICS

2.0 HISTORICAL DEVELOPMENT AND COMPLEXITY OF PROBLEM

Interest in the actual patterns of movement of humans and animals goes back to prehistoric times and was depicted in cave drawings, statues, and paintings. Such replications were subjective impressions of the artist. It was not until a century ago that the first motion picture cameras recorded locomotion patterns of both humans and animals. Marey, the French physiologist, used a photographic "gun" in 1885 to record displacements in human gait and chronophotographic equipment to get a stick diagram of a runner. About the same time, Muybridge in the United States triggered 24 cameras sequentially to record the patterns of a running man. Progress has been rapid during this century, and we now can record and analyze everything from the gait of a child with cerebral palsy to the performance of an elite athlete.

The term used for these descriptions of human movement is *kinematics.* Kinematics is not concerned with the forces, either internal or external, that cause the movement, but rather with the details of the movement itself. A complete and accurate quantitative description of the simplest movement requires a huge volume of data and a large number of calculations, resulting in an enormous number of graphic plots. For example, to describe the movement of the lower limb in the sagittal plane during one stride can require up to 50 variables. These include linear and angular displacements, velocities, and accelerations. It should be understood that any given analysis may use only a small fraction of the available kinematic variables. An assessment of a running broad jump, for example, may require only the velocity and height of the body's center of mass. On the other hand, a mechanical power analysis of an

amputee's gait may require almost all the kinematic variables that are available.

2.1 KINEMATIC CONVENTIONS

In order to keep track of all the kinematic variables, it is important to establish a convention system. In the anatomical literature, a definite convention has been established, and we can completely describe a movement using terms such as *proximal, flexion,* and *anterior.* It should be noted that these terms are all relative, that is, they describe the position of one limb relative to another. They do not give us any idea as to where we are in space. Thus, if we wish to analyze movement relative to the ground or the direction of gravity, we must establish an absolute spatial reference system. Such conventions are mandatory when imaging devices are used to record the movement. However, when instruments are attached to the body, the data become relative, and we lose information about gravity and the direction of movement.

2.1.1 Absolute Spatial Reference System

Several spatial reference systems have been proposed. The one utilized throughout the text is the one often used for human gait. The vertical direction is Y, the direction of progression (anterior–posterior) is X, and the sideways direction (medial–lateral) is Z. Figure 2.1 depicts this convention. The positive direction is as shown. Angles must also have a zero reference and a positive direction. Angles in the XY plane are measured from 0° in the X direction, with positive angles being counterclockwise. Similarly, in the YZ plane, angles start at 0° in the Y direction and increase positively counterclockwise. The convention for velocities and accelerations follows correctly if we maintain the spatial coordinate convention:

$\dot{x}$ = velocity in the X direction, positive when X is increasing

$\dot{y}$ = velocity in the Y direction, positive when Y is increasing

$\dot{z}$ = velocity in the Z direction, positive when Z is increasing

$\ddot{x}$ = acceleration in the X direction, positive when $\dot{x}$ is increasing

$\ddot{y}$ = acceleration in the Y direction, positive when $\dot{y}$ is increasing

$\ddot{z}$ = acceleration in the Z direction, positive when $\dot{z}$ is increasing

The same applies for angular velocities and angular accelerations. A counterclockwise angular increase is a positive angular velocity ω. When ω is increasing, we calculate a positive angular acceleration, α.

An example taken from the data on a human subject during walking will illustrate the convention. The kinematics of the right leg segment (as viewed from the right side) and its center of mass were analyzed as follows

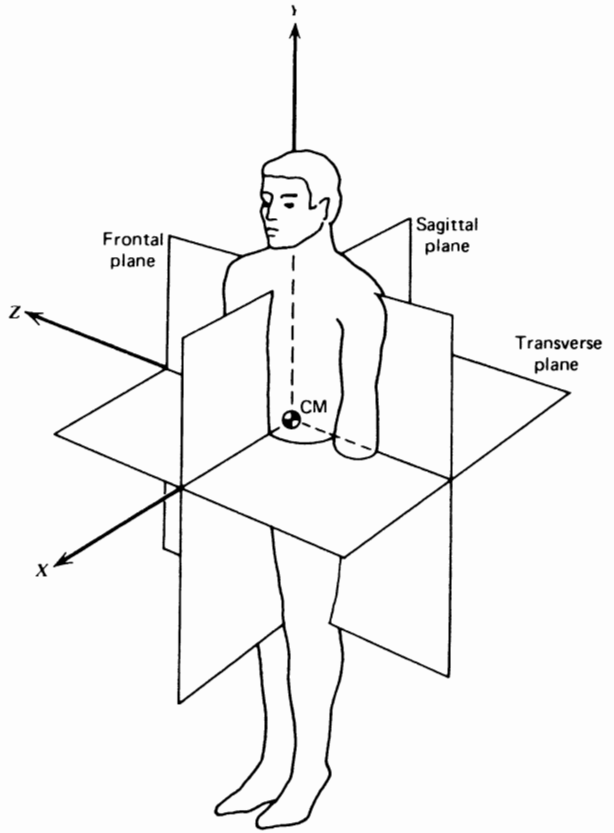

Figure 2.1 Spatial coordinate system used for all data and analysis.

$$\omega = -2.34 \text{ rad/s}, \alpha = 14.29 \text{ rad/s}^2, v_x = 0.783 \text{ m/s}$$

$$a_x = -9.27 \text{ m/s}^2, v_y = 0.021 \text{ m/s}, a_y = -0.31 \text{ m/s}^2$$

This means that the leg segment is rotating clockwise but is decelerating (accelerating in a counterclockwise direction). The velocity of the leg's center of mass is forward and very slightly upward, but is decelerating in the forward direction and accelerating downward.

2.1.2 Total Description of a Body Segment in Space

The complete kinematics of any body segment requires 15 data variables, all of which are changing with time:

1. Position (x, y, z) of segment center of mass
2. Linear velocity $(\dot{x}, \dot{y}, \dot{z})$ of segment center of mass

3. Linear acceleration ($\ddot{x}$, $\ddot{y}$, $\ddot{z}$) of segment center of mass
4. Angle of segment in two planes, θ_{xy}, θ_{yz}
5. Angular velocity of segment in two planes, ω_{xy}, ω_{yz}
6. Angular acceleration of segment in two planes, α_{xy}, α_{yz}

Note that the third angle data are redundant; any segment's direction can be completely described in two planes. For a complete description of the total body (feet + legs + thighs + trunk + head + upper arms + forearms and hands = 12 segments) movement in three-dimensional (3D) space required $15 \times 12 = 180$ data variables. It is no small wonder that we have yet to describe, let alone analyze, some of the more complex movements. Certain simplifications can certainly reduce the number of variables to a manageable number. In symmetrical level walking, for example, we can assume sagittal plane movement and can normally ignore the arm movement. The head, arms, and trunk (HAT) are often considered to be a single segment, and assuming symmetry, we need to collect data from one lower limb only. The data variables in this case (four segments, one plane) can be reduced to a more manageable 36.

2.2 DIRECT MEASUREMENT TECHNIQUES

2.2.1 Goniometers*

A goniometer is a special name given to the electrical potentiometer that can be attached to measure a joint angle. As such, one arm of the goniometer is attached to one limb segment, the other to the adjacent limb segment, and the axis of the goniometer is aligned to the joint axis. In Figure 2.2 we see the fitting of the goniometer to a knee joint along with the equivalent electrical circuit. A constant voltage E is applied across the outside terminals, and the wiper arm moves to pick off a fraction of the total voltage. The fraction of the voltage depends on the joint angle θ. Thus the voltage on the wiper arm is $v = kE\theta = k_1\theta$ volts. Note that a voltage proportional to θ requires a potentiometer whose resistance varies linearly with θ. A goniometer designed for clinical studies is shown fitted on a patient in Figure 2.3.

Advantages

1. A goniometer is generally inexpensive.
2. Output signal is available immediately for recording or conversion into a computer.
3. Planar rotation is recorded independent of the plane of movement of the joint.

*Representative paper: Finley and Karpovich, 1964

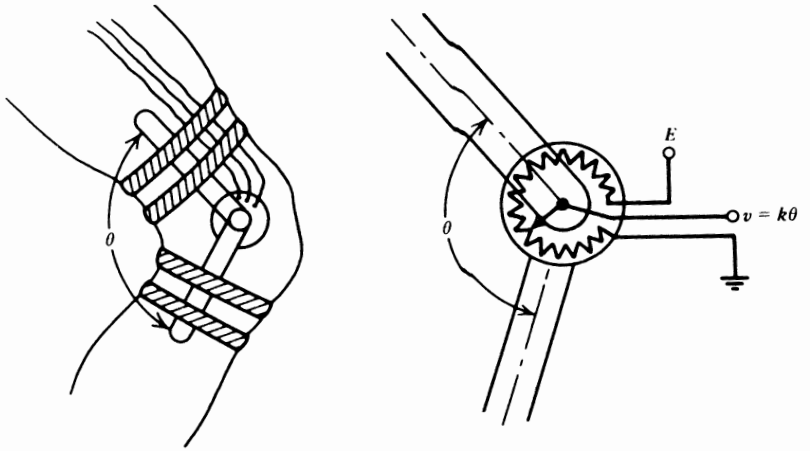

Figure 2.2 Mechanical and electrical arrangement of a goniometer located at the knee joint. Voltage output is proportional to the joint angle.

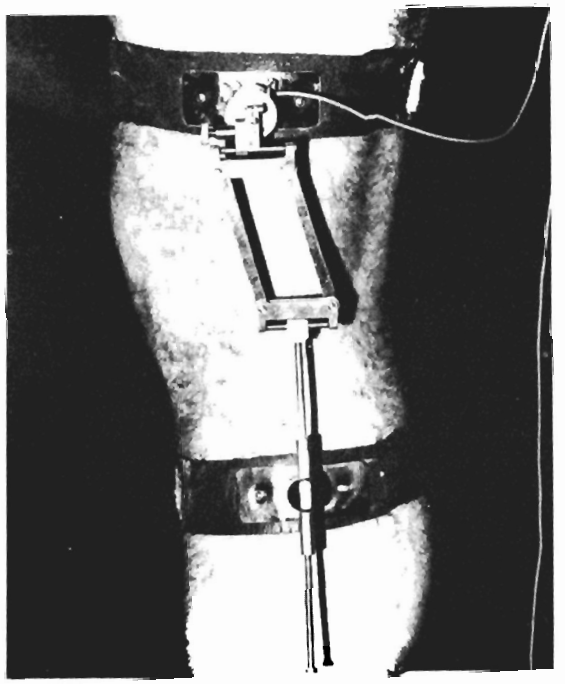

Figure 2.3 Electrogoniometer designed to accommodate changes in center of rotation of the knee joint shown here fitted on a patient. (Reproduced by permission of Orthodoc, McMaster Medical Center, Hamilton, Ont., Canada.)

Disadvantages

1. Relative angular data are given, not absolute angles, thus severely limiting the data's assessment value.
2. It may require an excessive length of time to fit and align, and the alignment over fat and muscle tissue can vary over the time of the movement.
3. If a large number are fitted, movement can be encumbered by the straps and cables.
4. More complex goniometers are required for joints that do not move as hinge joints.

2.2.1.1 Special Joint Angle Measuring Systems. More recently in the area of ergonomics, a special glove system has been developed to measure the kinematics of the fingers and the thumb. Figure 2.4 shows the construction of the glove transducer, which comprises a lightweight elastic glove with sensors on the proximal two joints of each finger and thumb plus a thumb abductor sensor. Each transducer is a loop of fiber optic cable with a constant infrared source and is etched in the region of the joint of interest. As the joint flexes, the fiber bends and light escapes; the greater the bend, the more light escapes. Thus, the flexion angular displacement is detected as a reduction in light intensity received by the detector and is precalibrated against the bending angle. A major use for such a system has been in the study of repetitive strain injuries (cf. Moore et al., 1991).

2.2.2 Accelerometers*

As indicated by its name, an accelerometer is a device that measures acceleration. Most accelerometers are nothing more than force transducers designed to measure the reaction forces associated with a given acceleration. If the acceleration of a limb segment is a and the mass inside is m, then the force exerted by the mass is $F = ma$. This force is measured by a force transducer, usually a strain gauge or piezoresistive type. The mass is accelerated against a force transducer that produces a signal voltage V, which is proportional to the force, and since m is known and constant, then V is also proportional to the acceleration. The acceleration can be toward or away from the face of the transducer; this is indicated by a reversal in sign of the signal. In most movements, there is no guarantee that the acceleration vector will act at right angles to the face of the force transducer. The more likely situation is depicted in Figure 2.5, with the acceleration vector having a component normal to the transducer and another component tangent to the transducer face. Thus the accelerometer measures the a_n component. Nothing is known

*Representative paper: Morris, 1973.

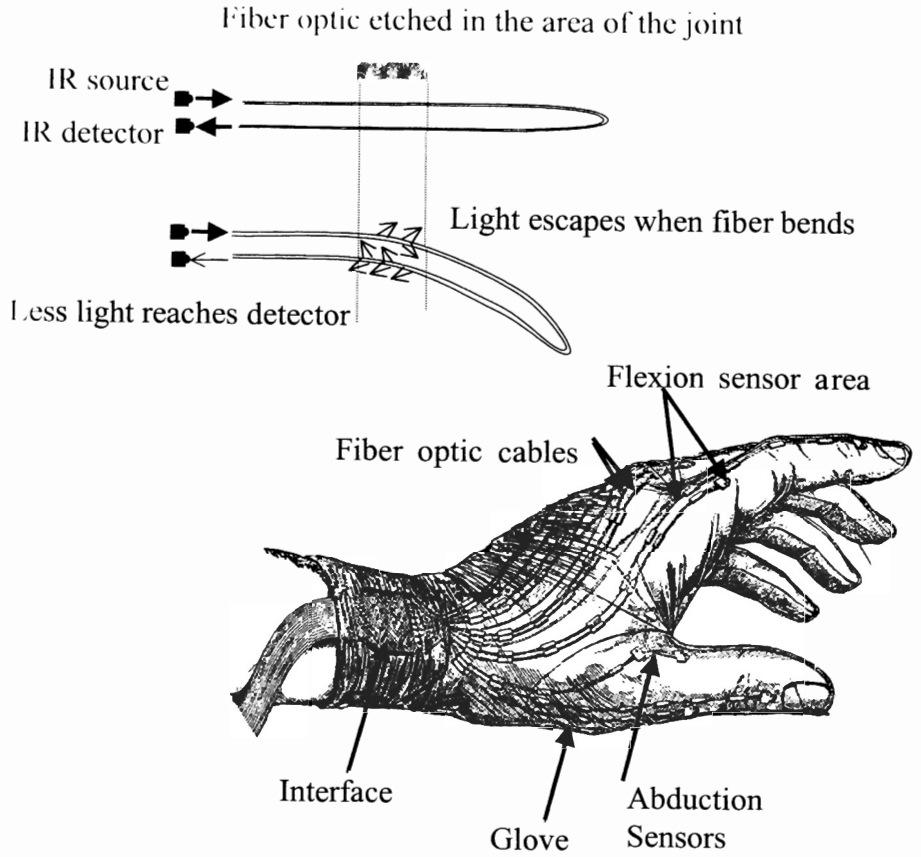

Figure 2.4 Construction and operation of a glove transducer to measure angular displacement of the fingers. Transducer is a loop of fiber optic cable; the amount of light returning to the detector decreases with increased finger flexion. Each cable is calibrated for angular displacement versus detected light intensity. (Courtesy of the Ergonomics Laboratory, Department of Kinesiology, University of Waterloo, Ont., Canada.)

about a_t or a unless a triaxial accelerometer is used. Such a 3D transducer is nothing more than three individual accelerometers mounted at right angles to each other, each one then reacting to the orthogonal component acting along its axis. Even with a triaxial accelerometer mounted on a limb, there can be problems because of limb rotation, as indicated in Figure 2.6. In both cases, the leg is accelerating in the same *absolute* direction as indicated by vector a. The measured acceleration component a_n is quite different in each case. Thus, the accelerometer is limited to those movements whose direction in space does not change drastically or to special contrived movements, such as horizontal flexion of the forearm about a fixed elbow joint.

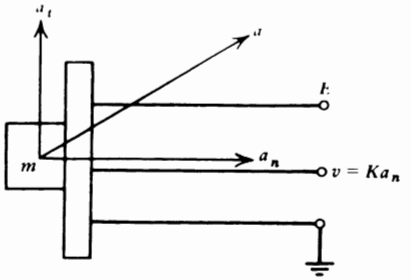

Figure 2.5 Schematic diagram of an accelerometer, showing acceleration with normal and tangential components. Voltage output is proportional to the normal component of acceleration a_n.

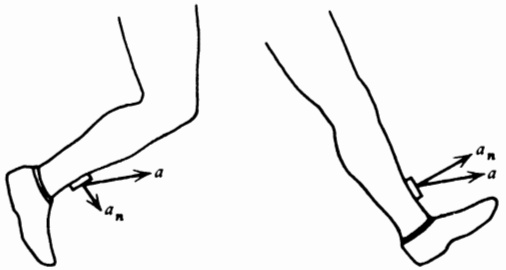

Figure 2.6 Two movement situations where the acceleration in space is identical but the normal component is quite different.

Less Important

A typical electric circuit of a piezoresistive accelerometer is shown in Figure 2.7. It comprises a half-bridge consisting of two equal resistors R_1. Within the transducer, resistors R_a and R_b change their resistances proportional to the acceleration acting against them. With no acceleration, $R_a = R_b = R_1$, and with the balance potentiometer properly adjusted, the voltage at terminal 1 is the same as that at terminal 2. Thus, the output voltage is $V = 0$. With the acceleration in the direction shown, R_b increases and R_a decreases; thus, the voltage at terminal 1 increases. The resultant unbalance in the bridge circuit results in voltage V proportional to the acceleration. Conversely, if the acceleration is upward, R_b decreases and R_a increases; the bridge unbalances in the reverse direction, giving a signal of the opposite polarity. Thus, over the dynamic range of the accelerometer, the signal is proportional to both the magnitude and the direction of acceleration acting along the axis of the accelerometer. However, if the balance potentiometer is not properly set, we have an unbalanced bridge and we could get a voltage–acceleration relationship like that indicated by the dashed lines.

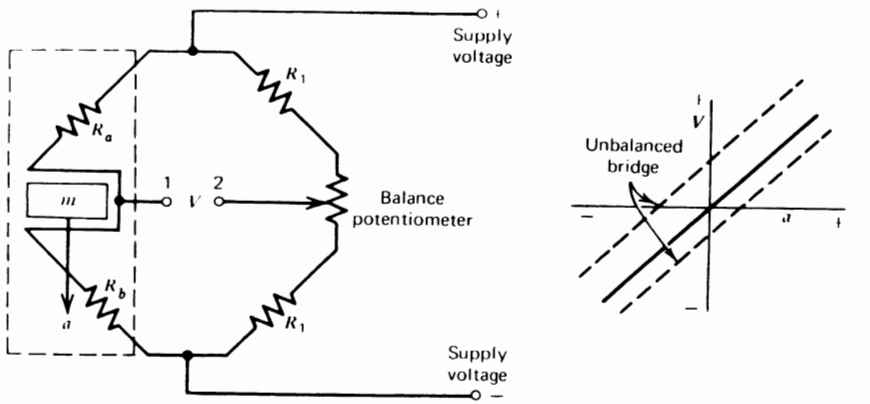

Figure 2.7 Electrical bridge circuit used in most force transducers and accelerometers. See text for detailed operation.

Advantages

1. Output signal is available immediately for recording or conversion into a computer.

Disadvantages

1. Acceleration is relative to its position on the limb segment.
2. Cost of accelerometers can be excessive if a large number are used; also the cost of the recorder or analog-to-digital converter may be high.
3. If a large number are used, they can encumber movement.
4. Many types of accelerometers are quite sensitive to shock and are easily broken.
5. The mass of the accelerometer may result in a movement artifact, especially in rapid movements or movements involving impacts.

2.3 IMAGING MEASUREMENT TECHNIQUES

The Chinese proverb "A picture is worth more than ten thousand words" holds an important message for any human observer, including the biomechanics researcher interested in human movement. Because of the complexity of most movements, the only system that can possibly capture all the data is an imaging system. Given the additional task of describing a dynamic activity, we are further challenged by having to capture data over an extended period of time. This necessitates taking many images at regular intervals during the event.

There are many types of imaging systems that could be used. The discussion will be limited to three different types: movie camera, television, and optoelectric types. Whichever system is chosen, a lens is involved; therefore a short review of basic optics is given here.

2.3.1 Review of Basic Lens Optics

A simple converging lens is one that creates an inverted image in focus at a distance v from the lens. As seen in Figure 2.8, if the lens–object distance is u, then the focal length f of the lens is

$$\frac{1}{f} = \frac{1}{v} + \frac{1}{u} \tag{2.1}$$

The imaging systems used for movement studies are such that the object–lens distance is quite large compared with the lens–image distance. Therefore,

$$\frac{1}{u} \approx 0, \frac{1}{f} = \frac{1}{v}, \text{ or } f = v \tag{2.2}$$

Thus, if we know the focal length of the lens system, we can see that the image size is related to the object size by a simple triangulation. A typical focal length is 25 mm, a wide-angle lens is 13 mm, and a telephoto lens is 150 mm. A zoom lens is just one in which the focal length is infinitely variable over a given range. Thus, as L increases, the focal length must increase proportionately to produce the same image size. Figure 2.9 illustrates this principle. For maximum accuracy, it is highly desirable that the image be as large as possible. Thus, it is advantageous to have a zoom lens rather than a series of fixed lenses; individual adjustments can be readily made for each movement to be studied, or even during the course of the event.

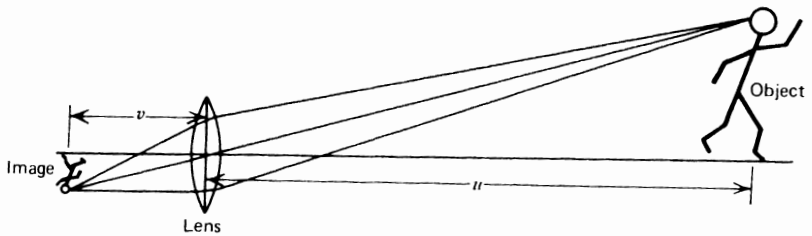

Figure 2.8 Simple focusing lens system showing relationship between object and image.

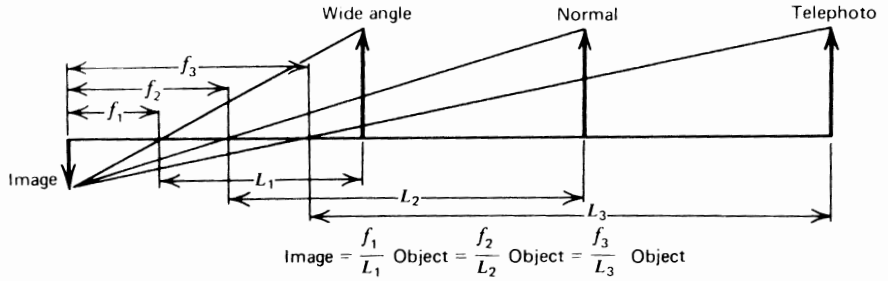

Figure 2.9 Differences in focal length of wide-angle, normal, and telephoto lenses result in an image that is the same size.

2.3.2 *f*-Stop Setting and Field of Focus

The amount of light entering the lens is controlled by the lens opening, which is measured by its f-stop (f means fraction of lens aperture opening). The larger the opening, the lower the f-stop setting. Each f-stop setting corresponds to a proportional change in the amount of light allowed in. A lens may have the following settings: 22, 16, 11, 8, 5.6, 4, 2.8, and 2. $f/22$ is 1/22 of the lens diameter and $f/11$ is 1/11 of the lens diameter. Thus $f/11$ lets in four times the light that $f/22$ does. The fractions are arranged so that each one lets in twice the light of the adjacent higher setting (e.g., $f/2.8$ is twice the light of $f/4$).

To keep the lighting requirements to a minimum, it is obvious that the lens should be opened as wide as possible with a low f setting. However, problems occur with the field of focus. This is defined as the maximum and minimum range of the object that will produce a focused image. The lower the f setting, the narrower the range over which an object will be in focus. For example, if we wish to photograph a movement that is to move over a range from 10 to 30 ft, we cannot reduce the f-stop below 5.6. The range set on the lens would be about 15 ft, and everything between 10 and 30 ft would remain in focus. The final decision regarding f-stop depends on the shutter speed of the movie camera and the film speed.

2.3.3 Cinematography*

Many different sizes of movie cameras are available; 8-mm cameras are the smallest. (They actually use 16-mm film, which is run through the camera twice, then split into two 8-mm strips after it is developed.) Then there are 16 mm, 35 mm, and 70 mm. The image size of 8 mm is somewhat small for

Representative paper: Eberhart and Inman, 1951

accurate measurements, while 35-mm and 70-mm movie cameras are too expensive to buy and operate. Thus, 16-mm cameras have evolved as a reasonable compromise, and most high-speed movie cameras are 16 mm. There are several types of 16-mm cameras available. Some are spring driven; others are motor driven by either batteries or power supplies from alternating current sources. Battery-driven types have the advantage of being portable to sites where power is not available.

The type of film required depends on the lighting available. The ASA rating is a measure of the *speed* of the film; the higher the rating, the less light is required to get the same exposure. 4-X reversal film with an ASA rating of 400 is a common type. Higher ASA ratings are also available and are good for a qualitative assessment of movement, especially faster moving sporting events. However, the coarse grain of these higher ASA films introduces inaccuracies in quantitative analyses.

The final factor that influences the lighting required is the shutter speed of the camera. The higher the frame rate, the less time is available to expose film. Most high-speed cameras have rotating shutters that open once per revolution for a period of time to expose a new frame of unexposed film. The arc of the opening, as depicted in Figure 2.10, and the speed of rotation of the shutter decide the exposure time. For example, at 60 frames per second, using a 3 factor shutter, the exposure time is 1/180 s. The amount of light entering will be the same as a normal (still) camera set to a speed of 1/180 s.

To make the final settings, we use an exposure meter to measure the light intensity on the human subject. For a given filming, the variables that are

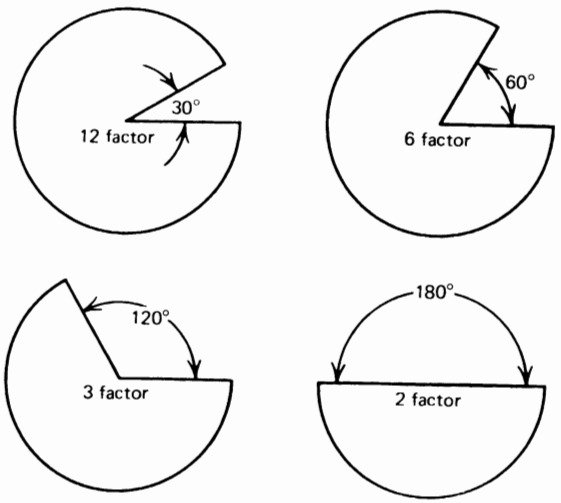

Figure 2.10 Various-factor shutters used in movie cameras. Film is exposed during the opening arc and is advanced while the shutter is closed.

preset are film ASA, shutter factor, and frame rate. The frame rate is set low enough to capture the desired event, but not too high to require extra lighting or result in film wastage. To understand the problem associated with the selection of an optimal rate, the student is referred to Section 2.5.3. The final variable to decide is the f-stop. The light meter gives an electrical meter reading proportional to the light intensity, such that when the film ASA and exposure time are set, the correct f-stop can be determined. Thus, with the movie camera set at the right frame rate, f-stop, and range, the filming is ready to commence.

2.3.4 Television*

The major difference between television and cinematography is the fact that television has a fixed frame rate. The name given to each television image is a field. In North America there are 60 fields per second, in Europe the standard is 50 fields per second. Thus, television has a high enough field rate for most movements, but probably too low for a quantitative analysis of rapid athletic events. The f-stop, focus, and lighting for television can be adjusted by watching the television monitor as the controls are varied. Many television cameras have electronic as well as optical controls that influence brightness and contrast, and some have built-in strobe lighting. Also, focus can be adjusted electronically as well as optically. The major advantage of television is the capability for instant replay, which serves both as a quality control check and as an initial qualitative assessment. Secondly, the television signal can be digitally converted by a "frame grabber" for immediate analysis.

*2.3.4.1 **Television Imaging Cameras.*** Some technical problems can result from the use of standard vidicon television cameras. The strong signal from a reflective marker produces a distinct circular image when the marker is not moving. However, when there is a rapid marker movement, the circular image blurs and produces a trailing edge. Thus, the triggering threshold for conversion into two levels (black and white) must be carefully set to ensure a circular digitized image. A more reliable way to get rid of the blur is to use a strobe system, which results in the exposure of the TV imaging tube for a millisecond or less. The strobe in effect acts as an electronic shutter. Strobe systems also eliminate a second problem associated with a continuously exposed imaging tube: skewing of the marker coordinates because of the time delay in the scanning from the top of the image to the bottom of the image. It takes about 15 ms to scan one TV field; thus, a head marker could be scanned 10 ms before a foot marker. The strobe system freezes all marker images at the same point in time in the same way that a movie camera does. Newer charge coupled diodes (CCD) cameras have mechanical or electronic shutter controls

*Representative paper: Winter et al., 1979.

that eliminate both blurring and skewing. A further development is the infrared camera, which does not use visible light and is not influenced by reflections from light sources other than those sources required to get the desired circular reflection from the markers. Figure 2.11 shows a typical infrared camera mounted permanently from the ceiling in a clinical gait laboratory. The active infrared lights form a "donut" shape about the camera lens and are pulsed at 120 Hz for a period of less than a millisecond. This camera is

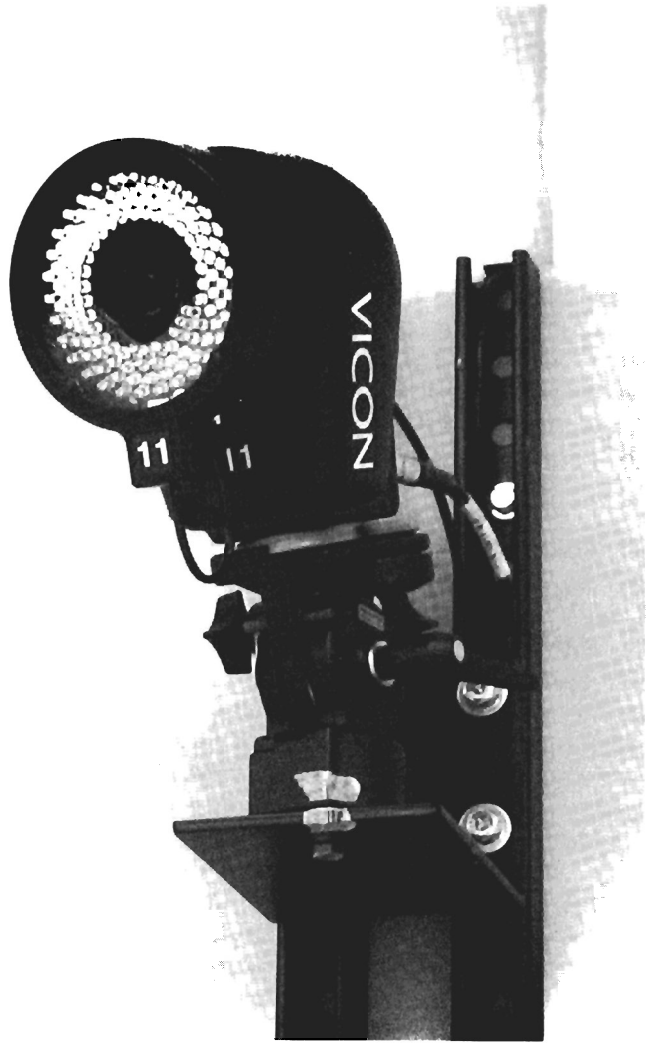

Figure 2.11 Typical infrared television camera mounted permanently in a clinical gait laboratory. The infrared lights form a "donut" shape around the lens and are pulsed for a short time each frame to freeze the image prior to scanning. (Courtesy of the Gait Analysis Laboratory, Connecticut Children's Medical Center, Hartford, CT.)

one of six to twelve cameras that could be mounted around the gait laboratory. Thus, the reflected infrared light from the markers is the only light that is picked up by the camera, and since it is a pulsed source, the marker images are "frozen" in time. Figure 2.12 depicts such an arrangement in a clinical gait laboratory along with the spherical reflective markers mounted on a young patient.

2.3.4.2 ***Historical Development of Television Digitizing Systems.*** Almost all of the movement analysis television systems were developed in university research laboratories. In the late 1960s, the first reports of television-based

Figure 2.12 Gait assessment of a cerebral palsy patient in a clinical gait laboratory. Infrared cameras are mounted on the ceiling and walls to capture the reflected light from the spherical reflective markers mounted on both sides of the body. (Courtesy of the Gait Analysis Laboratory, Connecticut Children's Medical Center, Hartford, CT.)

systems started to appear: at Delft University of Technology in The Netherlands (Furnée, 1967, according to Woltring, 1987) and at the Twenty-First Conference EMB in Houston, TX (Winter et al., 1968). The first published paper on an operational system was by Dinn et al. (1970) from the Technical University of Nova Scotia in Halifax, N.S., Canada. It was called CINTEL (Computer INterface for TELivision), and was developed for digitizing angiographic images at 4 bits (16 grey levels) to determine the time course of left ventricular volume (Trenholm et al., 1972). It was also used for gait studies at the University of Manitoba in Winnipeg, Man., Canada, where, with higher spatial resolution and a one-bit (black/white) conversion, the circular image of a reflective hemispheric ping pong ball attached on anatomical landmarks was digitized (Winter et al., 1972). With about 10 pixels within each marker image, it was possible by averaging their coordinates to improve the spatial precision of each marker from 1 cm (distance between scan lines of each field) to about 1 mm. The 3M Scotch® material that was used as reflective material has been used by most subsequent experimental and commercial systems.

Jarett et al. (1976) reported a system that detected the left edge of the image of a small reflective marker that occupied one or two scan lines. Unfortunately, the spatial precision was equal to the scan line distance, which is about 1 cm. This system was adopted and improved by VICON (VIdeo CONvertor) in their commercial system. Both left and right edges of the marker image were detected, and subsequently the detected points were curve fitted via software AMASS to a circle (Macleod et al., 1990). Based on the circle fit, the centroid was calculated. Other commercial systems, such as that developed by the Motion Analysis Corporation, use patented edge-detection techniques (Expert Vision). Shape recognition of the entire marker image, rather than edge detection, was used by the ELITE (Elaboratore di Immagini Televisive) system developed in Milan, Italy. A dedicated computer algorithm operating in real time used a cross-correlation pattern recognition technique based on size and shape (Ferrigno and Pedotti, 1985). This system uses all grey levels in the shape detection, thus improving the spatial resolution to 1/2800 of the viewing field. Considering the field height to be about 2.5 m, this represents a precision of about 0.9 mm.

2.3.5 Optoelectric Techniques*

In the past few years, there have been several developments in optoelectric imaging systems that have some advantages over cinematography and television. The first commercial system was developed by Northern Digital in Waterloo, Ont., Canada, and was called Watsmart. It was an active system that required the subject to wear tiny infrared lights on each desired anatomical landmark. The lights were flashed sequentially, and the light flash was detected on a special camera. The camera consisted of a standard lens focusing the light flash onto a special semiconductor diode surface. More recent development of this active system has evolved into a 3D camera system called OPTOTRAK. Such a system consists of three cameras mounted in line on a rigid frame, as shown in Figure 2.13. The left and right lenses are mounted to face slightly inward and their linear diode arrays are mounted horizontally. Thus, their scan of the pulsed light will define the location of a marker in a vertical plane. The middle lens is mounted with its diode array mounted vertically, and its scan will define a horizontal plane. Figure 2.14 depicts this arrangement. The left and right detectors each define the location of all markers in a vertical plane; the intersection of these two vertical planes is a vertical line. Thus, any markers on this vertical line will record the same signal on the left and right cameras. The middle camera has its lens facing directly ahead with its diode array mounted vertically. Thus, this camera will define all markers in a horizontal plane. The intersection of this horizontal plane with the vertical line defined by the other two cameras is a unique 3D point in space. Thus, as each infrared diode (IRED) pulses, its x, y, z coordinates in the global reference system (GRS) are recorded. The pulsed light from a

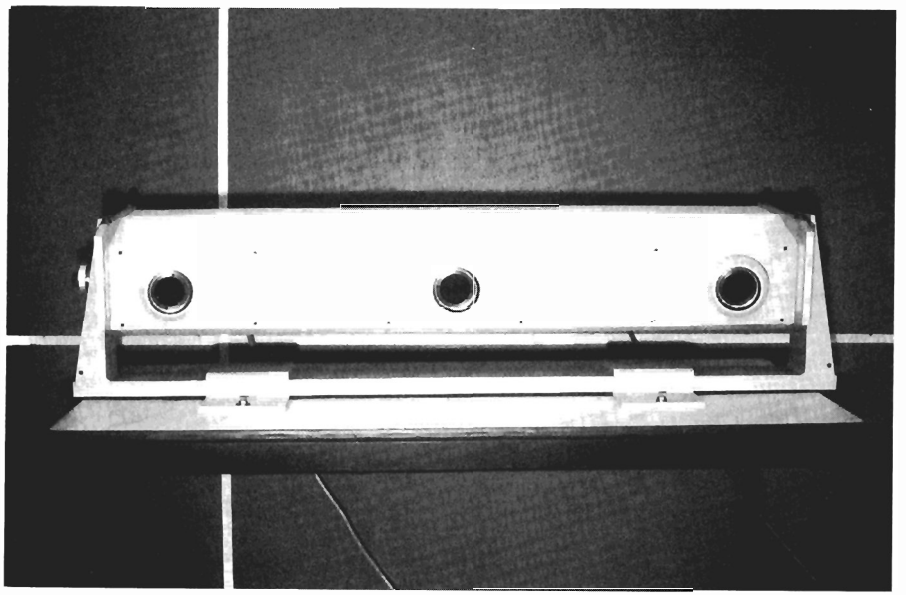

Figure 2.13 An OPTOTRAK system with three lenses, each with a linear diode array. The outside two lenses face slightly inward and each defines a vertical plane, while the middle lens defines a horizontal plane. See Figure 2.14 to see how these three diode arrays define a marker in 3D space. (Courtesy of Gait and Posture Laboratory, Department of Kinesiology, University of Waterloo, Waterloo, Ont., Canada.)

*Representative paper: Winter et al., 2003.

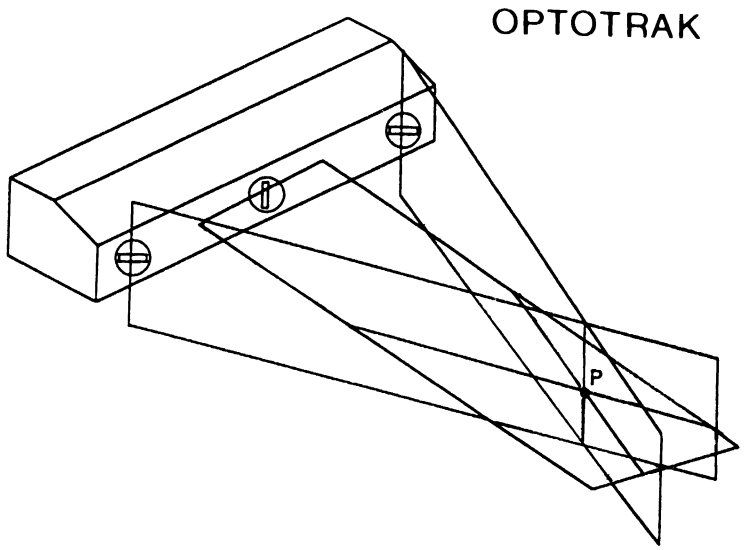

Figure 2.14 OPTOTRAK system with two outside lenses facing slightly inward with their diode arrays each defining markers in a vertical plane. Any marker on the intersection of these two planes will define all markers on this vertical line. The middle lens array defines all markers in a horizontal plane. Thus, the intersection of a marker on this horizontal plane with the vertical line will define the unique coordinates of a marker in 3D space.

second light source yields a different vertical line and horizontal plane, and thus a different set of x, y, z coordinates. There are some unique advantages of such an active system. There is no specialized software required (as in television) to identify which marker is which. Thus, in laboratories where the number and location of markers is changing day-to-day, there are no problems with marker labeling, which makes this system flexible for the changing research requirements. Also, because of the precision of the IRED array, the precision of the x, y, z coordinates is better than that of TV systems, which are constrained by the distance between the scan lines. The precision for an OPTOTRAK camera mounted as shown in Figure 2.13 at a distance of 4 m from the subject is 0.03 mm with noise = 0.015 mm (Gage et al., 2004). Possible disadvantages are the number of IREDs that can be mounted and the potential encumbrance of the cables connecting the power source to the active IREDs.

2.3.6 Advantages and Disadvantages of Optical Systems

Advantages

1. All data are presented in an absolute spatial reference system, in a plane normal to the optical axis of the camera.

2. Most systems (cine, TV) are not limited as to the number of markers used.
3. Encumbrance to movement is minimal for most systems that use lightweight reflective markers (cine, TV), and the time to apply the markers is minimal.
4. TV cameras and VCRs are reasonably inexpensive.
5. Cine and TV systems can be replayed for teaching purposes or for qualitative analysis of the total body movement.

Disadvantages

1. Most multiple-camera systems are expensive (cine, TV, optoelectric), as are the digitizing and conversion systems for all imaging sources.
2. For film, the turnaround time for development may be a problem, and the labor to digitize film coordinates may also be a constraint. The digitizing errors, however, are less than those from many commercial imaging systems.
3. Encumbrance and time to fit wired light sources (e.g., IREDs) can be prohibitive in certain movements, and the number of light sources is limited.
4. Some imaging systems (e.g., IREDs) cannot be used outside in daylight.

2.3.7 Summary of Various Kinematic Systems

Each laboratory must define its special requirements before choosing a particular system. A clinical gait lab may settle on TV because of the encumbrance of optoelectric systems and because of the need for a qualitative assessment, rapid turnaround, and for teaching. Ergonomic and athletic environments may require instant or near-instant feedback to the subject or athlete, thus dictating the need for an automated system. Basic researchers do not require a rapid turnaround and may need a large number of coordinates; thus, they may opt for movie cameras or an optoelectric system. And, finally, the cost of hardware and software may be the single limiting factor that may force a compromise as to the final decision.

2.4 DATA CONVERSION TECHNIQUES

2.4.1 Analog-to-Digital Converters

To students not familiar with electronics, the process that takes place during conversion of a physiological signal into a digital computer can be somewhat mystifying. A short schematic description of that process is now given. An electrical signal representing a force, an acceleration, an electromyographic (EMG) potential, or the like is fed into the input terminals of the analog-to-digital converter. The computer controls the rate at which the signal is sam

pled; the optimal rate is governed by the sampling theorem (see Section 2.5.3).

Figure 2.15 depicts the various stages in the conversion process. The first is a sample/hold circuit in which the analog input signal is changed into a series of short-duration pulses, each one equal in amplitude to the original analog signal at the time of sampling. (These times are specified by the computer operator.) The final stage of conversion is to translate the amplitude and polarity of the sampled pulse into digital format. This is usually a binary code in which the signal is represented by a number of bits. For example, a 12-bit code represents $2^{12} = 4096$ levels. This means that the original sampled analog signal can be broken into 4096 discrete amplitude levels with a unique code representing each of these levels. Each coded sample (consisting of 0s and 1s) forms a 12-bit "word," which is rapidly stored in computer memory for recall at a later time. If a 5-s signal were converted at a sampling rate of 100 times per second, there would be 500 data words stored in memory to represent the original 5-s signal.

2.4.2 Movie Conversion Techniques

As 16-mm movie cameras are the most common form of data collection, it is important to be aware of various coordinate extraction techniques. Each system that has evolved requires the projection of each movie frame on some form of screen. The most common type requires the operator to move a mechanical *xy* coordinate system until a point, light, or cross hair lies over the desired anatomical landmark. Then the *x* and *y* coordinates can be read

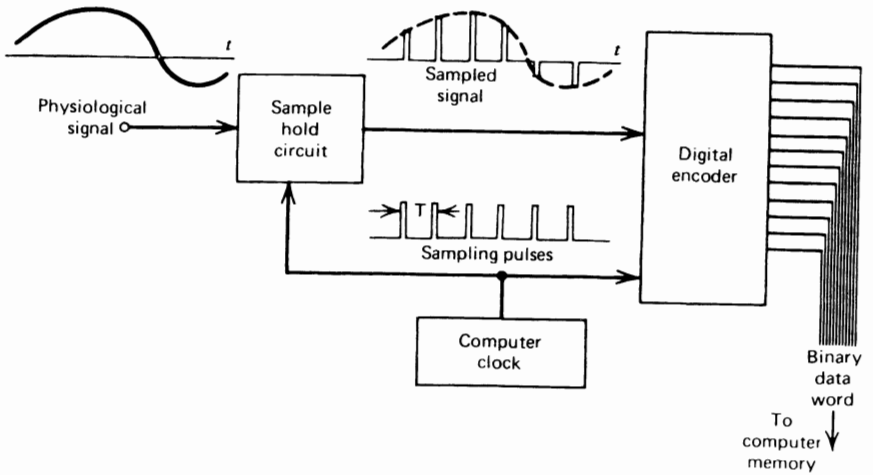

Figure 2.15 Schematic diagram showing the steps involved in an analog-to-digital conversion of a physiological signal.

off or transferred to a computer at the push of a button. Figure 2.16 shows the component parts of such a conversion system.

A second type of system involves the projection of the film image onto a special grid system. When the operator touches the grid with a special pen, the coordinates are automatically transferred into a computer. Both systems are limited to the speed and accuracy of the human operator. Our experience indicates that an experienced operator can convert an average of 15 coordinate pairs per minute. Thus a 3-s film record filmed at 50 frames per second could have five markers converted in 30 min.

The human error involved in this digitizing has been found to be random and quite small. For a camera 4 m from a subject, the root-mean-square (rms) noise" present in the converted data has been measured at 1–1.5 mm.

2.4.3 Television Conversion

Each of the commercial television systems has it own unique technique for identifying the presence of a marker and determining its centroid, and for labeling markers from a multi-camera system. Students are referred back to Section 2.3.4.2 for a summary of and references to the more common systems.

Figure 2.16 Typical arrangement for the microcomputer digitization of data coordinated from movie film. Foot pedal (not shown) allows operator to transfer coordinate data into a digital computer at the rate of about 10 coordinate pairs per minute. Digitizing error is about 1mm rms with the camera located 4 m from the subject.

2.5 PROCESSING OF RAW KINEMATIC DATA

2.5.1 Nature of Unprocessed Data

The converted coordinate data from film or television are called *raw* data. This means that they contain additive noise from many sources: electronic noise in optoelectric devices, spatial precision of the TV scan or film digitizing system, or human error in film digitizing. All of these will result in random errors in the converted data. It is therefore essential that the raw data be smoothed, and in order to understand the techniques used to smooth the data, an appreciation of harmonic (or frequency) analysis is necessary.

2.5.2 Harmonic (Fourier) Analysis

1. *Alternating Signals.* An alternating signal (often called ac, for alternating current) is one that is continuously changing with time. It may be periodic or completely random, or a combination of both. Also, any signal may have a dc (direct current) component, which may be defined as the bias value about which the ac component fluctuates. Figure 2.17 shows example signals.

2. *Frequency Content.* Any of these signals can also be discussed in terms of their frequency content. A sine (or cosine) waveform is a single frequency; any other waveform can be the sum of a number of sine and cosine wave.

Note that the Fourier transformation (Figure 2.18) of periodic signals has discrete frequencies, while nonperiodic signals have a continuous spectrum defined by its lowest frequency f_1 and its highest frequency f_2. To analyze a

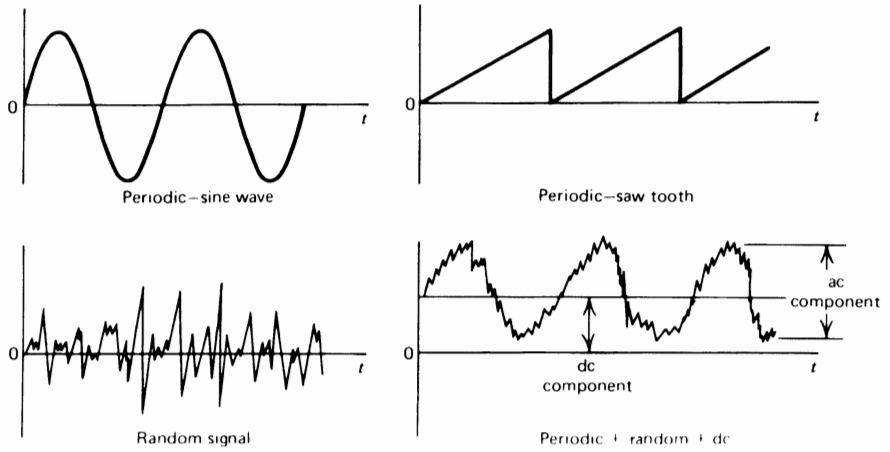

Figure 2.17 Time-related waveforms to demonstrate the different types of signals that may be processed

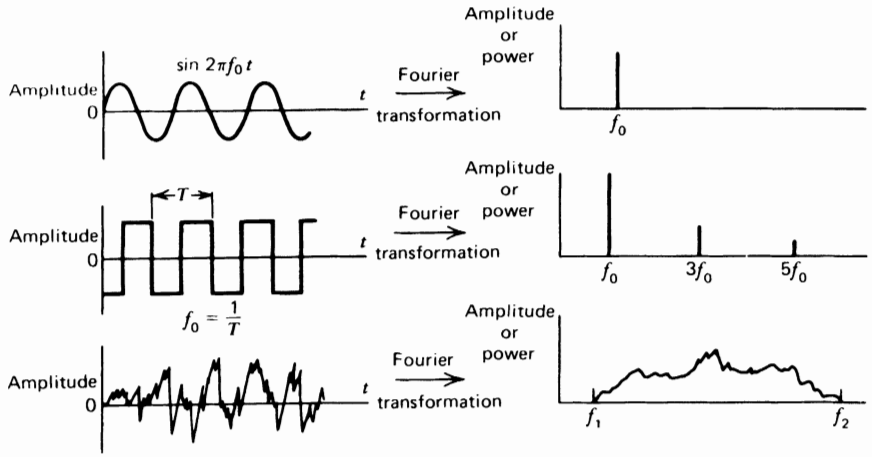

Figure 2.18 Relationship between a signal as seen in the time domain and its equivalent in the frequency domain.

periodic signal, we must express the frequency content in multiples of the fundamental frequency f_0. These higher frequencies are called *harmonics.* The third harmonic is $3f_0$ and the tenth harmonic is $10f_0$. Any perfectly periodic signal can be broken down into its harmonic components. The sum of the proper amplitudes of these harmonics is called a *Fourier series.*

Thus, a given signal $V(t)$ can be expressed as

$$V(t) = V_{dc} + V_1 \sin(\omega_0 t + \theta_1) + V_2 \sin(2\omega_0 t + \theta_2) + \cdots + V_n \sin(n\omega_0 t + \theta_n) \tag{2.3}$$

where $\omega_0 = 2\pi f_0$, and θ_n is the phase angle of the nth harmonic.

For example, a square wave of amplitude V can be described by the Fourier series

$$V(t) = \frac{4V}{\pi}\left(\sin \omega_0 t + \frac{1}{3}\sin 3\omega_0 t + \frac{1}{5}\sin 5\omega_0 t + \cdots\right) \tag{2.4}$$

A triangular wave of duration $2t$ and repeating itself every T seconds is

$$V(t) = \frac{2Vt}{T}\left[\frac{1}{2} + \left(\frac{2}{\pi}\right)^2 \cos \omega_0 t + \left(\frac{2}{3\pi}\right)^2 \cos 3\omega_0 t + \cdots\right] \tag{2.5}$$

Several names are given to the graph showing these frequency components: *spectral plots, harmonic plots,* or *spectral density functions.* Each shows the amplitude or power of each frequency component plotted against frequency:

the mathematical process to accomplish this is called a *Fourier transformation* or *harmonic analysis*. Figure 2.18 shows plots of time-domain signals and their equivalents in the frequency domain.

Care must be used when analyzing or interpreting the results of any harmonic analysis. Such analyses assume that each harmonic component is present with a constant amplitude and phase over the analysis period. Such consistency is evident in Equation (2.3), where amplitude V_n and phase θ_n are assumed constant. However, in real life each harmonic is not constant in either amplitude or phase. A look at the calculation of the Fourier coefficients is needed for any signal $x(t)$. Over the period of time T, we calculate:

$$a_n = \frac{2}{T}\int_0^T x(t)\cos n\omega_0 t\, dt \tag{2.6}$$

$$b_n = \frac{2}{T}\int_0^T x(t)\sin n\omega_0 t\, dt \tag{2.7}$$

$$c_n = \sqrt{a_n^2 + b_n^2}$$

$$\theta_n = \tan^{-1}\left(\frac{a_n}{b_n}\right) \tag{2.8}$$

It should be noted that a_n and b_n are calculated *average* values over the period of time T. Thus, the amplitude c_n and the phase θ_n of the nth harmonic are average values as well. A certain harmonic may be present only for part of the time T, but the computer analysis will return an average value assuming it is present over the entire time.

Figure 2.19 is presented to illustrate this assumption. It represents the Fourier reconstitution of the vertical trajectory of the heel of an adult walking his or her natural cadence. A total of nine harmonics are represented here because the addition of higher harmonics did not improve the curve of the original data. As can be seen, the harmonic reconstitution is visibly different from the original, sufficiently so as to cause reasonable errors in subsequent biomechanical analyses.

In spite of the negative comments about Fourier's reconstructions, there is considerable information in harmonic or Fourier analyses as to the bandwidth of the signal and the processing of that signal in both analog and digital systems. The first value of such analyses relates to the sampling theorem.

2.5.3 Sampling Theorem

Film and television are sampling processes. They capture the movement event for a short period of time, after which no further changes are recorded until the next field or frame. Playing a movie film back slowly demonstrates this

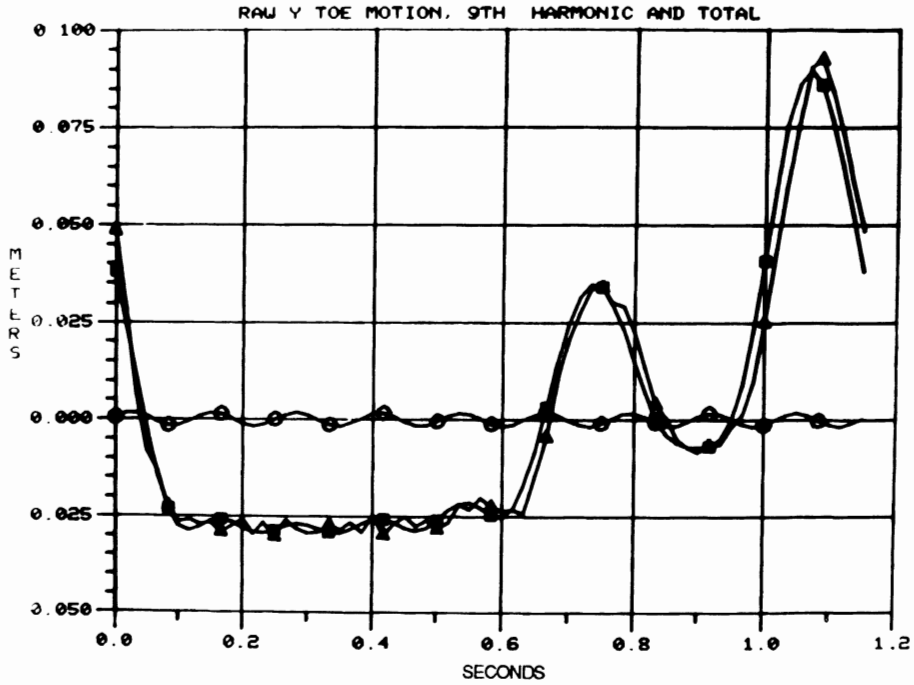

Figure 2.19 Fourier reconstruction of the vertical trajectory of a toe marker during one walking stride. The actual trajectory is shown by the open square, the reconstruction from the first 9 harmonics is plotted with open triangles, and the contribution of the 9th harmonic is plotted with open circles. The difference between the actual and the reconstructed waveforms is due to the lack of stationarity in the original signal.

phenomenon: the image jumps from one position to the next in a distinct step rather than a continuous process. The only reason film or television does not appear to jump at normal projection speeds (24 per second for film, 60 per second for television) is because the eye can retain an image for a period of about 1/15 s. The eye's short-term "memory" enables the human observer to average or smooth out the jumping movement.

In the processing of any time-varying data, no matter what their source, the sampling theorem must not be violated. Without going into the mathematics of the sampling process, the theorem states that "the process signal must be sampled at a frequency at least twice as high as the highest frequency present in the signal itself." If we sample a signal at too low a frequency, we get aliasing errors. This results in false frequencies, frequencies that were not present in the original signal, being generated in the sample data. Figure 2.20 illustrates this effect. Both signals are being sampled at the same interval T. Signal 1 is being sampled about 10 times per cycle, while signal 2 is being sampled less than twice per cycle. Note that the amplitudes of the samples

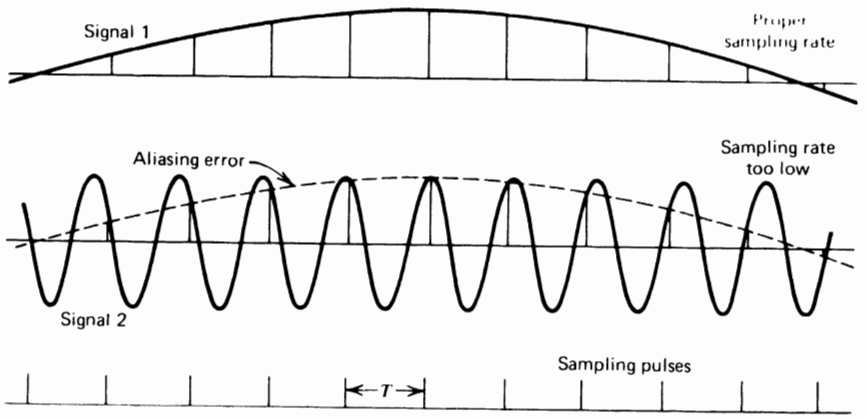

Figure 2.20 Sampling of two signals, one at a proper rate, the other at too low a rate. Signal 2 is sampled at a rate less than twice its frequency, such that its sampled amplitudes are the same as for signal 1. This represents a violation of the sampling theorem and results in an error called *aliasing*.

taken from signal 2 are identical to those sampled from signal 1. A false set of sampled data has been generated from signal 2 because the sample rate is too low—the sampling theorem has been violated.

The tendency of those using film is to play it safe and film at too high a rate. Usually there is a cost associated with such a decision. The initial cost is probably in the equipment required. A high-speed movie camera can cost four or five times as much as a standard model (24 frames per second). Or a special optoelectric system complete with the necessary computer can be a $70,000 decision. In addition to these capital costs, there are the higher operational costs of converting the data and running the necessary kinematic and kinetic computer programs. Except for higher speed running and athletic movements, it is quite adequate to use a standard movie or television camera. For normal and pathological gait studies, it has been shown that kinetic and energy analyses can be done with negligible error using a standard 24-frame per second movie camera (Winter, 1982). Figure 2.21 compares the results of kinematic analysis of the foot during normal walking, where a 50-Hz film rate was compared with 25 Hz. The data were collected at 50 Hz and the acceleration of the foot was calculated using every frame of data, then reanalyzed again, using every second frame of converted data. It can be seen that the difference between the curves is minimal; only at the peak negative acceleration was there a noticeable difference. The final decision as to whether this error is acceptable should not rest in this curve, but in your final goal. If, for example, the final anlaysis was a hip and knee torque analysis, the acceleration of the foot segment may not be too important, as is evident from another walking trial, shown in Figure 2.22. The minor differences in no way interfere with the general pattern of joint torques over the stride period and

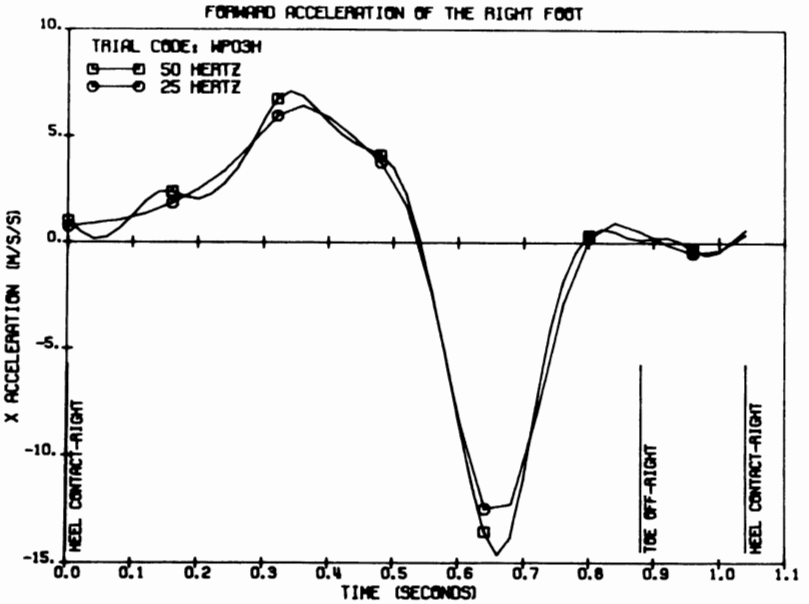

Figure 2.21 Comparison of the forward acceleration of the right foot during walking using the same data sampled at 50 Hz and at 25 Hz (using data from every second frame). The major pattern is maintained with minor errors at the peaks.

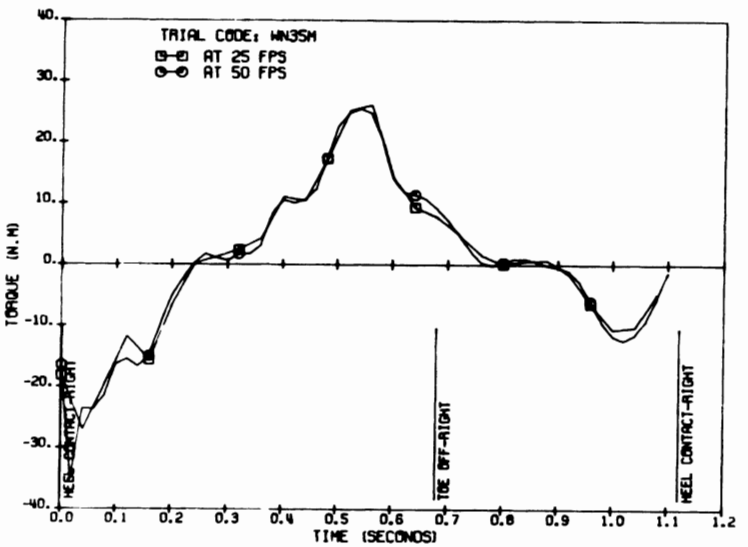

Figure 2.22 Comparison of the hip moment of force during level walking using the same data sampled at 50 Hz and at 25 Hz. The residual error is quite small because the joint reaction forces dominate the inertial contributions to the net moment of force.

the assessment of the motor patterns would be identical. Thus, for movements such as walking or for slower movements, an inexpensive camera at 24 frames per second appears to be quite adequate.

2.5.4 Signal Versus Noise

In the study of movement, the signal may be an anatomical coordinate that changes with time. For example, in running, the Y (vertical) coordinate of the heel will have certain frequencies that will be higher than those associated with the vertical coordinate of the knee or trunk. Similarly, the frequency content of all trajectories will decrease in walking compared with running. In repetitive movements, the frequencies present will be multiples (harmonics) of the fundamental frequency (stride frequency). When walking at 120 steps per minute (2 Hz), the stride frequency is 1 Hz. Therefore we can expect to find harmonics at 2 Hz, 3 Hz, 4 Hz, and so on. Normal walking has been analyzed by digital computer, and the harmonic content of the trajectories of seven leg and foot markers was determined (Winter et al., 1974). The highest harmonics were found to be in the toe and heel trajectories, and it was found that 99.7% of the signal power was contained in the lower seven harmonics (below 6 Hz). The harmonic analysis for the heel marker for 20 subjects is shown in Figure 2.23. Above the seventh harmonic, there was still some signal power, but it had the characteristics of "noise." Noise is the term used to describe components of the final signal that are not due to the process itself (in this case, walking). Sources of noise were noted in Section 2.5.1, and if the total effect of all these errors is random, then the true signal will have an added random component. Usually the random component is high frequency, as is borne out in Figure 2.23. Here we see evidence of higher frequency

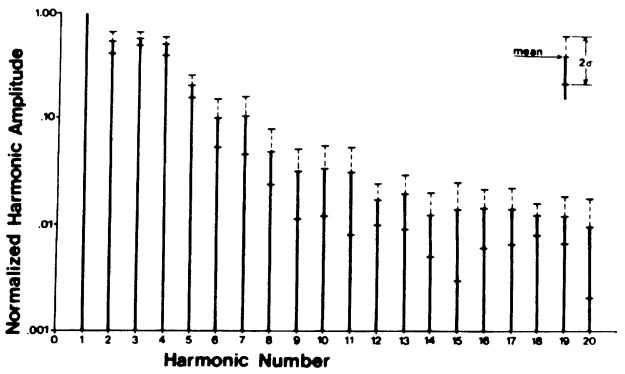

Figure 2.23 Harmonic content of the vertical displacement of a toe marker from 20 subjects during normal walking. Fundamental frequency (harmonic number 1) is normalized at 1.00. Over 99% of power is contained below the 7th harmonic. (Reproduced by permission from the *Journal of Biomechanics*.)

components extending up to the twentieth harmonic, which was the highest frequency analyzed. The presence of the higher frequency noise is of considerable importance when we consider the problem of trying to calculate velocities and accelerations. Consider the process of time differentiation of a signal containing additive higher frequency noise. Suppose the signal can be represented by a summation of N harmonics:

$$x = \sum_{n=1}^{N} X_n \sin (n\omega_0 t + \theta_n) \tag{2.9}$$

where ω_0 = fundamental frequency
n = harmonic number
X_n = amplitude of nth harmonic
θ_n = phase of nth harmonic

To get the velocity in the x direction V_x, we differentiate with respect to time:

$$V_x = \frac{dx}{dt} = \sum_{n=1}^{N} n\omega_0 X_n \cos (n\omega_0 t + \theta_n) \tag{2.10}$$

Similarly, the acceleration A_x is

$$A_x \frac{dV_x}{dt} = -\sum_{n=1}^{N} (n\omega_0)^2 X_n \sin (n\omega_0 t + \theta_n) \tag{2.11}$$

Thus, the amplitude of each of the harmonics increases with its harmonic number; for velocities they increase linearly, and for accelerations the increase is proportional to the square of the harmonic number. This phenomenon is demonstrated in Figure 2.24 where the fundamental, second, and third harmonics are shown along with their first and second time derivatives. Assuming that the amplitude x of all three components is the same, we can see that the first derivative (velocity) of harmonics increases linearly with increasing frequency. The first derivative of the third harmonic is now three times that of the fundamental. For the second time derivative, the increase repeats itself, and the third harmonic acceleration is now nine times that of the fundamental.

In the trajectory data for gait, x_1 might be 5 cm and x_{20} = 0.5 mm. The twentieth harmonic noise is hardly perceptible in the displacement plot. In the velocity calculation, the twentieth harmonic increases 20-fold so that it is now one-fifth that of the fundamental. In the acceleration calculation, the twentieth harmonic increases another factor of 20 and now is four times the magnitude of the fundamental. This effect is shown in Figure 2.25 which plots the acceleration of the toe during walking. The random-looking signal is the raw data differentiated twice. The smooth signal is the acceleration

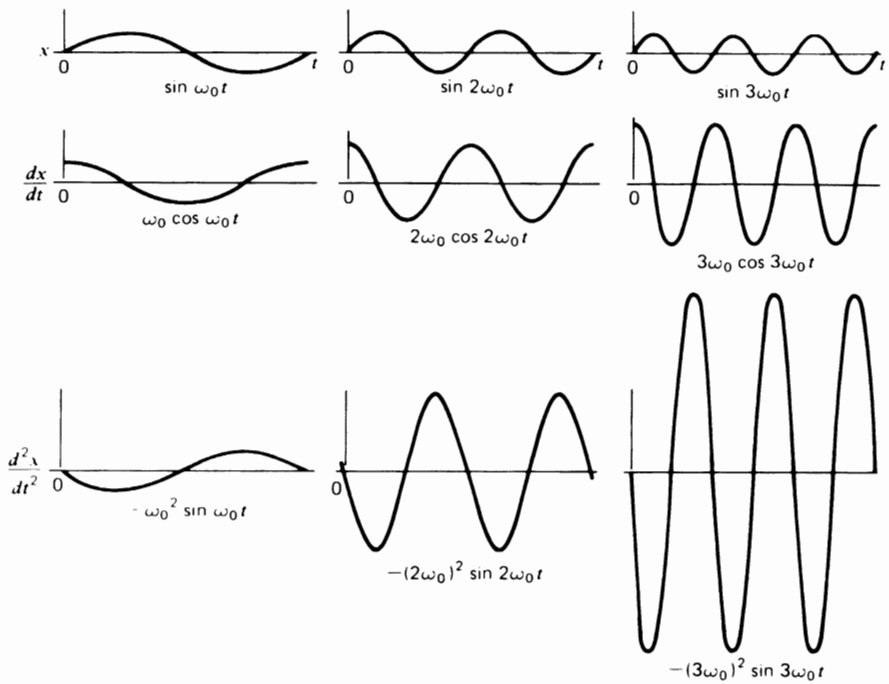

Figure 2.24 Relative amplitude changes as a result of time differentiation of signals of increasing frequency. First derivative increases amplitude proportional to frequency; second derivative increases amplitude proportional to frequency squared. Such a rapid increase has severe implications in calculating accelerations when the original displacement signal has high-frequency noise present.

calculated after most of the higher frequency noise has been removed. Techniques to accomplish this are now discussed.

2.5.5 Smoothing and Fitting of Data

The removal of noise can be accomplished in several ways. The aims of each technique are basically the same. However, the results differ somewhat.

2.5.5.1 Curve-Fitting Techniques. The basic assumption here is that the trajectory signal has a predetermined shape and that by fitting the assumed shape to a "best fit" with the raw data, a smooth signal will result. For example, it may be assumed that the data are a certain order polynomial:

$$x(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \cdots + a_n t^n \qquad (2.12)$$

By computer techniques, the coefficients $a_0, \cdots, a_n$ can be selected to give a best fit using such criteria as minimum mean square error. The complexity

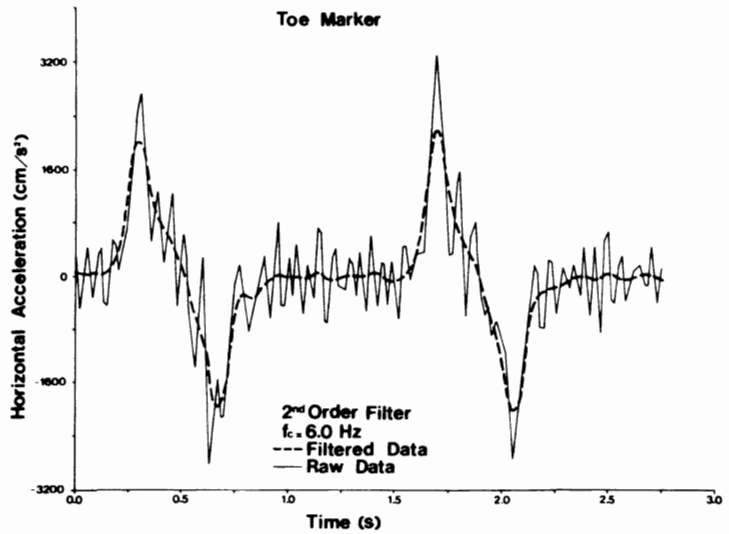

Figure 2.25 Horizontal acceleration of the toe marker during normal walking as calculated from displacement data from television. Solid line is acceleration based on unprocessed "raw" data; dotted line is that calculated after data have been filtered twice with a second-order low-pass filter. (Reproduced by permission from the *Journal of Biomechanics.*)

of the curve fitting can be quite restrictive in terms of computer time. A similar fit can be made assuming a certain number of harmonics. Reconstituting the final signal as a sum of N lowest harmonics,

$$x(t) = a_0 + \sum_{n=1}^{N} a_n \sin(n\omega_0 t + \theta_n) \qquad (2.13)$$

This model has a better basis, especially in repetitive movement, while the polynomial may be better in certain nonrepetitive movement such as broad jumping. However, there are severe assumptions regarding the consistency (stationarity) of a_n and θ_n, as was demonstrated in Figure 2.19. A third technique, spline curve fitting, is a modification of the polynomial technique. The curve to be fitted is broken into sections, each section starting and ending with an inflection point, with special fitting being done between adjacent sections. The major problem with this technique, other than computer time, is the error introduced by improper selection of the inflection points. These inflection points must be determined from the noisy data and, as such, are strongly influenced by the very noise that we are trying to eliminate.

2.5.5.2 Digital Filtering. Technological advances in digital filtering have opened up a much more promising and less restrictive solution to the noise reduction. The basic approach can be described by analyzing the frequency

spectrum of both signal and noise. Figure 2.26*a* shows a schematic plot of a signal and noise spectrum. As can be seen, the signal is assumed to occupy the lower end of the frequency spectrum and overlaps with the noise, which is usually higher frequency. Filtering of any signal is aimed at the selective rejection, or attenuation, of certain frequencies. In the above case, the obvious filter is one that passes, unattenuated, the lower frequency signals while at the same time attenuating the higher frequency noise. Such a filter, called a *low-pass filter,* has a frequency response as shown in Figure 2.26*b*. The frequency response of the filter is the ratio of the output $X_o(f)$ of the filter to its input $X_i(f)$ at each frequency present. As can be seen, the response at lower

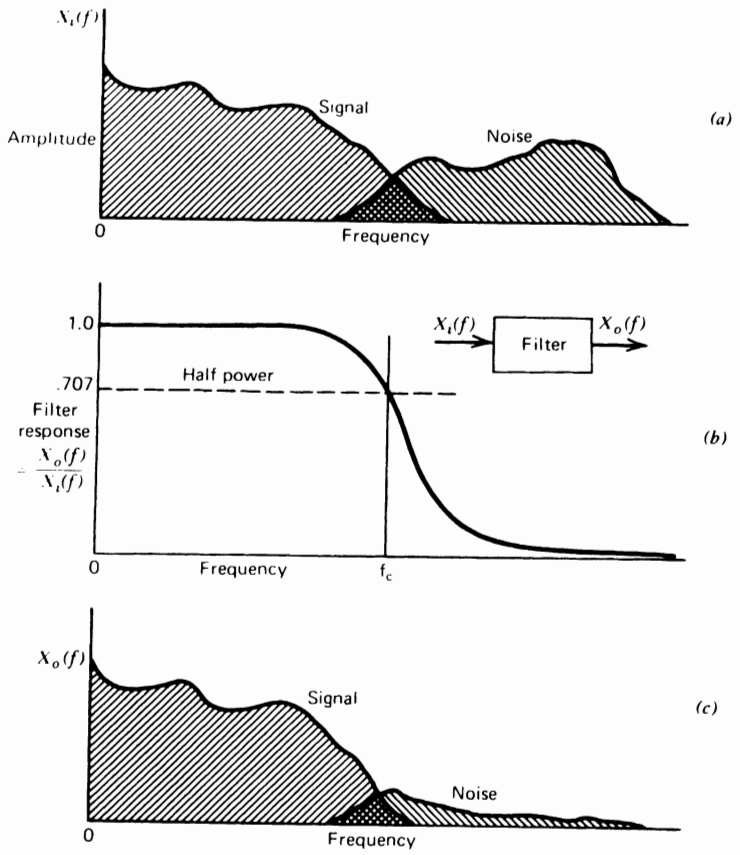

Figure 2.26 (*a*) Hypothetical frequency spectrum of a waveform consisting of a desired signal and unwanted higher frequency noise. (*b*) Response of low-pass filter $X_o(f)/X_i(f)$, introduced to attenuate the noise. (*c*) Spectrum of the output waveform, obtained by multiplying the amplitude of the input by the filter response at each frequency. Higher frequency noise is severely attenuated, while the signal is passed with only minor distortion in transition region around f

frequencies is 1.0. This means that the input signal passes through the filter unattenuated. However, there is a sharp transition at the cutoff frequency f_c so that the signals above f_c are severely attenuated. The net result of the filtering process can be seen by plotting the spectrum of the output signal $X_o(f)$ as seen in Figure 2.26*c*. Two things should be noted. First, the higher frequency noise has been severely reduced but not completely rejected. Second, the signal, especially in the region where the signal and noise overlap (usually around f_c) is also slightly attenuated. This results in a slight distortion of the signal. Thus, a compromise has to be made in the selection of the cutoff frequency. If f_c is set too high, less signal distortion occurs, but too much noise is allowed to pass. Conversely, if f_c is too low, the noise is reduced drastically, but at the expense of increased signal distortion. A sharper cutoff filter will improve matters, but at an additional expense. In digital filtering, this means a more complex digital filter and, thus, more computer time.

The theory behind digital filtering (Radar and Gold, 1967) will not be covered, but the application of low-pass digital filtering to kinematic data processing will be described in detail. First, it must be assessed what the signal spectrum is as opposed to the noise spectrum. This can readily be done, as is seen in the harmonic analysis presented previously in Figure 2.23. As a result of the previous discussion for these data on walking, the cutoff frequency of a digital filter should be set at about 6 Hz. The format of a recursive digital filter that processes the raw data in time domain is as follows:

$$X^1(nT) = a_0X(nT) + a_1X(nT - T) + a_2X(nT - 2T) + b_1X^1(nT - T) + b_2X^1(nT - 2T) \tag{2.14}$$

where X^1 = filtered output coordinates
X = unfiltered coordinate data
nT = nth sample
$(nT - T)$ = $(n - 1)$th sample
$(nT - 2T)$ = $(n - 2)$th sample
$a_0, \cdots, b_2, \cdots$ = filter coefficients

These filter coefficients are constants that depend on the type and order of the filter, the sampling frequency, and the cutoff frequency. As can be seen, the filter output $X^1(nT)$ is a weighted version of the immediate and past raw data plus a weighted contribution of past filtered output.

The order of the filter decides the sharpness of the cutoff. The higher the order, the sharper the cutoff, but the larger the number of coefficients. For example, a Butterworth-type low-pass filter of second order is to be designed to cut off at 6 Hz using film data taken at 60 Hz (60 frames per second). All that is required to determine these coefficients is the ratio of sampling fre-

quency to cutoff frequency. In this case it is 10. The design of such a filter would yield the following coefficients:

$$a_0 = 0.06746, \qquad a_1 = 0.13491, \qquad a_2 = 0.06746,$$
$$b_1 = 1.14298, \qquad b_2 = -0.41280$$

Note that the algebraic sum of all the coefficients equals 1.0000. This gives a response of unity over the passband. Tables 2.1 and 2.2 list the coefficients for use in a critically damped filter and a second-order Butterworth filter for various values of f_s/f_c. Note that the same filter coefficients could be used in many different applications, as long as the ratio f_s/f_c is the same. For example, an EMG signal sampled at 2000 Hz with cutoff desired at 400 Hz would have the same coefficients as one employed for movie film coordinates where the film rate was 30 Hz and cutoff was 6 Hz.

If more exact cutoff frequencies are required, the exact equations to calculate the coefficients for a Butterworth or a critically damped filter are as follows:

TABLE 2.1 Coefficients for Critically Damped Low-Pass Filter

f_s/f_c	a_0	a_1	a_2	b_1	b_2
4	0.25000	0.50000	0.25000	0.00000	0.00000
5	0.17708	0.35416	0.17708	0.31677	−0.02509
6	0.13397	0.26795	0.13397	0.53590	−0.07180
7	0.10565	0.21130	0.10565	0.69983	−0.12244
8	0.08579	0.17157	0.08579	0.82843	−0.17157
9	0.07121	0.14241	0.07121	0.93262	−0.21744
10	0.06014	0.12028	0.06014	1.01905	−0.25962
11	0.05152	0.10304	0.05152	1.09208	−0.29816
12	0.04466	0.08932	0.04466	1.15470	−0.33333
13	0.03910	0.07820	0.03910	1.20904	−0.36545
14	0.03453	0.06906	0.03453	1.25668	−0.39481
15	0.03073	0.06146	0.03073	1.29882	−0.42173
16	0.02753	0.05505	0.02753	1.33636	−0.44646
17	0.02480	0.04961	0.02480	1.37003	−0.46925
18	0.02247	0.04494	0.02247	1.40042	−0.49029
19	0.02045	0.04090	0.02045	1.42797	−0.50978
20	0.01869	0.03739	0.01869	1.45309	−0.52786
21	0.01716	0.03431	0.01716	1.47607	−0.54469
22	0.01580	0.03160	0.01580	1.49718	−0.56039
23	0.01460	0.02920	0.01460	1.51665	−0.57506
24	0.01353	0.02707	0.01353	1.53465	−0.58879
25	0.01258	0.02516	0.01258	1.55136	−0.60168

TABLE 2.2 Coefficients for Butterworth Low-Pass Filter

f_s/f_c	a_0	a_1	a_2	b_1	b_2
4	0.29289	0.58579	0.29289	0.00000	−0.17157
5	0.20657	0.41314	0.20657	0.36953	−0.19582
6	0.15505	0.31010	0.15505	0.62021	−0.24041
7	0.12123	0.24247	0.12123	0.80303	−0.28796
8	0.09763	0.19526	0.09763	0.94281	−0.33333
9	0.08042	0.16085	0.08042	1.05333	−0.37502
10	0.6746	0.13491	0.06746	1.14298	−0.41280
11	0.05742	0.11484	0.05742	1.21719	−0.44687
12	0.04949	0.09898	0.04949	1.27963	−0.47759
13	0.04311	0.08621	0.04311	1.33291	−0.50533
14	0.03789	0.07578	0.03789	1.37889	−0.53045
15	0.03357	0.06714	0.03357	1.41898	−0.55327
16	0.02995	0.05991	0.02995	1.45424	−0.57406
17	0.02689	0.05379	0.02689	1.48550	−0.59307
18	0.02428	0.04856	0.02428	1.51338	−0.61051
19	0.02203	0.04407	0.02203	1.53842	−0.62655
20	0.02008	0.04017	0.02008	1.56102	−0.64135
21	0.01838	0.03676	0.01838	1.58152	−0.65505
22	0.01689	0.03378	0.01689	1.60020	−0.66776
23	0.01557	0.03114	0.01557	1.61730	−0.67958
24	0.01440	0.02880	0.01440	1.63299	0.69060
25	0.01336	0.02672	0.01336	1.64746	0.70090

$$\omega_c = \frac{(\tan(\pi f_c/f_s))}{C} \tag{2.15}$$

where C is the correction factor for number of passes required, to be explained shortly. For a single-pass filter with the coefficients shown in Tables 2.1 and 2.2, $C = 1$.

$K = \sqrt{2}\omega_c$ for a Butterworth filter,
or, $2\omega_c$ for a critically damped filter

$$K_2 = \omega_c^2, \qquad a_0 = \frac{K_2}{(1 + K_1 + K_2)}, \qquad a_1 = 2a_0, \qquad a_2 = a_0$$

$$K_3 = \frac{2a_0}{K_2}, \qquad b_1 = -2a_0 + K_3$$

$b_2 = 1 - 2a_0 - K_2$ or $b_2 = 1 - a_0 - a_1 - a_2 - b_1$

As well as attenuating the signal, there is a phase shift of the output signal relative to the input. For this second-order filter there is a 90° phase lag at the cutoff frequency. This will cause a second form of distortion, called *phase distortion,* to the higher harmonics within the bandpass region. Even more phase distortion will occur to those harmonics above f_c, but these components are mainly noise, and they are being severely attenuated. This phase distortion may be more serious than the amplitude distortion that occurs to the signal in the transition region. To cancel out this phase lag, the once-filtered data can be filtered again, but this time in the reverse direction of time (Winter et al., 1974). This introduces an equal and opposite phase lead so that the net phase shift is zero. Also, the cutoff of the filter will be twice as sharp as that for single filtering. In effect, by this second filtering in the reverse direction we have created a fourth order zero-phase-shift filter, which yields a filtered signal that is back in phase with the raw data, but with most of the noise removed.

In Figure 2.27 we see the frequency response of a second-order Butterworth filter normalized with respect to the cutoff frequency. Superimposed on this curve is the response of the fourth-order zero-phase-shift filter. Thus, the new cutoff frequency is lower than that of the original single-pass filter; in this case, it is about 80% of the original. The correction factor for each additional pass of a Butterworth filter is $C = (2^{1/n} - 1)^{0.25}$, where n is the number of passes. Thus, for a dual pass, $C = 0.802$. For a critically damped filter, $C = (2^{1/2n} - 1)^{0.5}$; thus, for a dual pass, $C = 0.435$. This correction factor is applied to Equation (2.15) and results in the cutoff frequency for the original single-pass filter being set higher, so that after the second pass the

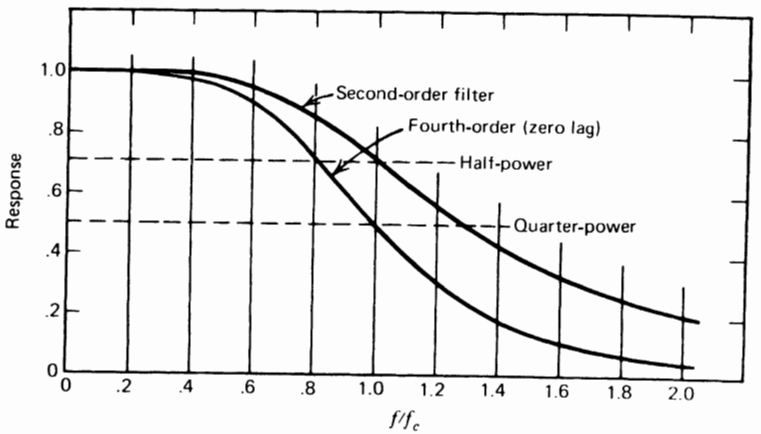

Figure 2.27 Response of a second-order low-pass digital filter. Curve is normalized at 1.0 at cutoff frequency f_c. Because of phase-lag characteristics of the filter, a second filtering is done in the reverse direction of time, which results in a fourth order zero lag filter.

desired cutoff frequency is achieved. The major difference between these two filters is a compromise in the response in the time domain. Butterworth filters have a slight overshoot in response to step or impulse type inputs, but they have a much shorter rise time. Critically damped filters have no overshoot, but suffer from a slower rise time. Because impulsive type inputs are rarely seen in human movement data, the Butterworth filter is preferred.

The application of one of these filters in smoothing raw coordinate data can now be seen by examining the data that yielded the harmonic plot of Figure 2.23. The horizontal acceleration of this toe marker, as calculated by finite differences from the filtered data, is plotted in Figure 2.25. Note how repetitive the filtered acceleration is and how it passes through the "middle" of the noisy curve, as calculated using the unfiltered data. Also, note that there is no phase lag in these filtered data because of the dual forward and reverse filtering processes.

2.5.5.3 ***Choice of Cutoff Frequency—Residual Analysis.*** There are several ways to choose the best cutoff frequency. The first is to carry out a harmonic analysis as depicted in Figure 2.23. By analyzing the power in each of the components, a decision can be made as to how much power to accept and how much to reject. However, such a decision assumes that the filter is ideal and has an infinitely sharp cutoff. A better method is to do a residual analysis of the difference between filtered and unfiltered signals over a wide range of cutoff frequencies. In this way, the characteristics of the filter in the transition region are reflected in the decision process. Figure 2.28 shows a theoretical

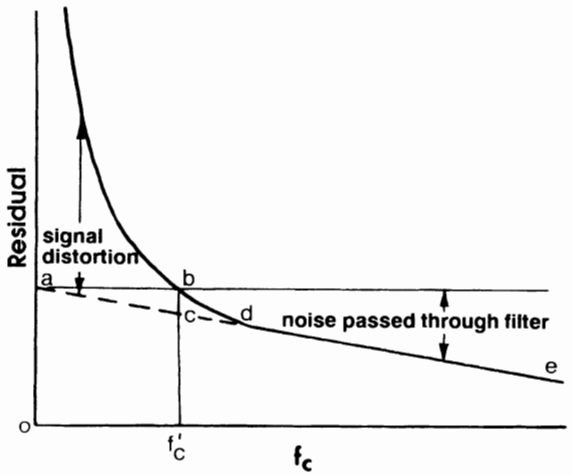

Figure 2.28 Plot of the residual between a filtered and an unfiltered signal as a function of the filter cutoff frequency. See text for the interpretation as to where to set the cutoff frequency of the filter.

plot of residual versus frequency. The residual at any cutoff frequency is calculated as follows for a signal of N sample points in time:

$$R(f_c) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X}_i)^2} \tag{2.16}$$

where X_i = raw data at ith sample
$\hat{X}_i$ = filtered data at the ith sample

If our data contained no signal, just random noise, the residual plot would be a straight line decreasing from an intercept at 0 Hz to an intercept on the abscissa at the Nyquist frequency (0.5 f_s). The line *de* represents our best estimate of that noise residual. The intercept a on the ordinate (at 0 Hz) is nothing more than the rms value of the noise, because $\hat{X}_i$ for a 0-Hz filter is nothing more than the mean of the noise over the N samples. When the data consist of true signal plus noise, the residual will be seen to rise above the straight (dashed) line as the cutoff frequency is reduced. This rise above the dashed line represents the signal distortion that is taking place as the cutoff is reduced more and more.

The final decision is where f_c should be chosen. The compromise is always a balance between the amount of signal distortion versus the amount of noise allowed through. If we decide that both should be equal, then we simply project a line horizontally from a to intersect the residual line at b. The frequency chosen is f_c^1, and at this frequency the signal distortion is represented by bc. This is also an estimate of the noise that is passed through the filter. Figure 2.29 is a plot of the residual of four markers from one stride of gait data, and both vertical and horizontal coordinates were analyzed. As can be seen, the straight regression line that represents the noise is essentially the same for both coordinates on all markers. This tells us that the noise content, mainly introduced by the human digitizing process, is the same for all markers. This regression line has an intercept of 1.8 mm, which indicates that the rms of the noise is 1.8 mm. In this case, the cine camera was 5 m from the subject and the image was 2 m high by 3 m wide. Thus the rms noise is less than one part in 1000.

Also, we see distinct differences in the frequency content of different markers. The residual shows the more rapidly moving markers on the heel and ball to have power up to about 6 Hz, while the vertical displacements of the rib and hip markers were limited to about 3 Hz. Thus, through this selection technique, we could have different cutoff frequencies specified for each marker displacement.

2.5.6 Comparison of Some Smoothing Techniques

It is valuable to see the effect of several different curve fitting techniques on the same set of noisy data. The following summary of a validation experiment

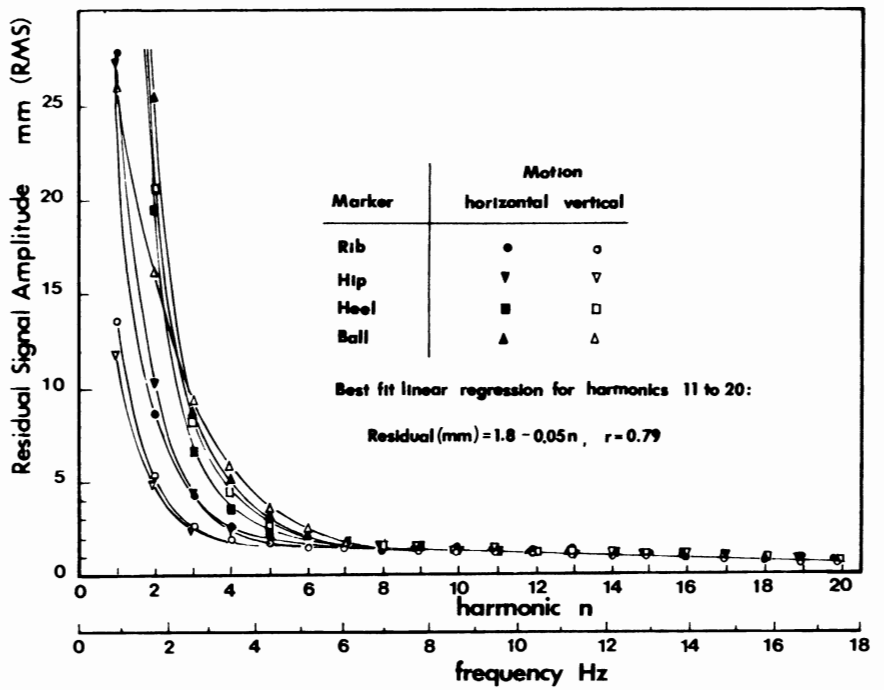

Figure 2.29 Plot of the residual of four markers from a walking trial; both vertical and horizontal displacement data. Data were digitized from movie film with the camera 5 m from the subject.

which was conducted to compare (Pezzack et al., 1977) three commonly used techniques, illustrates the wide differences in the calculated accelerations.

Data obtained from the horizontal movement of a lever arm about a vertical axis were recorded three different ways. A goniometer on the axis recorded angular position, an accelerometer mounted at the end of the arm gave tangential acceleration and thus angular acceleration, and cinefilm data gave image information that could be compared with the angular and acceleration records. The comparisons are given in Figure 2.30. Figure 2.30*a* compares the angular position of the lever arm as it was manually moved from rest through about 130° and back to the original position. The goniometer signal and the lever angle as analyzed from the film data are plotted and compare closely. The only difference is that the goniometer record is somewhat noisy compared with the film data.

Figure 2.30*b* compares the directly recorded angular acceleration, which can be calculated by dividing the tangential acceleration by the radius of the accelerometer from the center of rotation, with the angular acceleration as calculated via the second derivative of the digitally filtered coordinate data (Winter et al., 1974). The two curves match extremely well, and the finite-difference acceleration exhibits less noise than the directly recorded acceler-

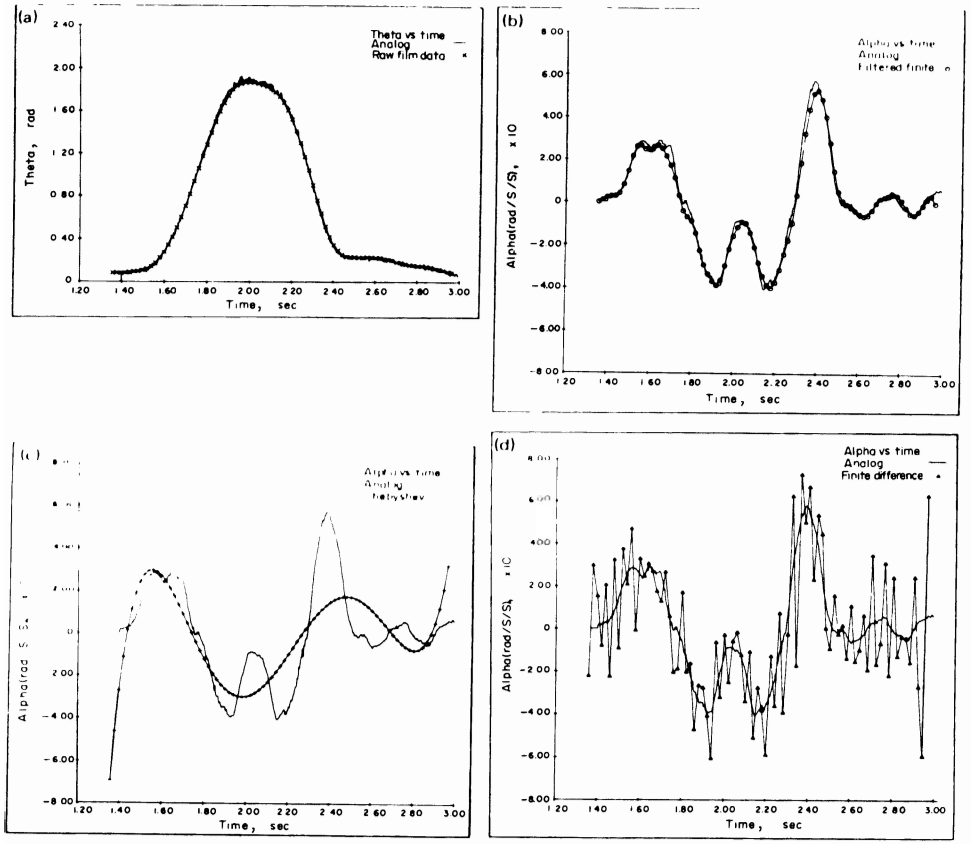

Figure 2.30 Comparison of several techniques used to determine the acceleration of a movement based on film displacement data. (*a*) Displacement angle of a simple extension/flexion as plotted from film and goniometer data. (*b*) Acceleration of movement in (*a*) as measured by accelerometer, calculated from film coordinates after digital filtering. (*c*) Acceleration as determined from a 9th-order polynomial fit of the displacement data compared with directly recorded acceleration. (*d*) Acceleration as determined by finite-difference techniques compared with measured curve. (Reproduced by permission from the *Journal of Biomechanics.*)

ation. Figure 2.30*c* compares the directly recorded acceleration with the calculated angular acceleration using a polynomial fit on the raw angular data. A ninth-order polynomial was fitted to the angular displacement curve to yield the following fit:

$$\theta(t) = 0.064 + 2.0t - 35t^2 + 210t^3 - 430t^4 + 400t^5 - 170t^6 + 25t^7 + 2.2t^8 - 0.41t^9 \quad \text{rad} \tag{2.17}$$

Note that θ is in radians and t in seconds. To get the curve for angular acceleration, all we need to do is take the second time derivative to yield

$$\alpha(t) \quad 70 + 1260t - 5160t^2 + 8000t^3 - 5100t^4 + 1050t^5 + 123t^6 - 29.5t^7 \quad \text{rad/s}^2 \tag{2.18}$$

This acceleration curve, compared with the accelerometer signal, shows considerable discrepancy, enough to cast doubt on the value of the polynomial fit technique. The polynomial is fitted to the displacement data in order to get an analytic curve, which can be differentiated to yield another smooth curve. Unfortunately, it appears that a considerably higher order polynomial would be required to achieve even a crude fit, and the computer time might become too prohibitive.

Finally, in Figure 2.30*d* we see the accelerometer signal plotted against angular acceleration as calculated by second-order finite-difference techniques. The plot speaks for itself—the accelerations are too noisy to mean anything.

2.6 CALCULATION OF ANGLES FROM SMOOTHED DATA

2.6.1 Limb-Segment Angles

Given the coordinate data from anatomical markers at either end of a limb segment, it is an easy step to calculate the absolute angle of that segment in space. It is not necessary that the two markers be at the extreme ends of the limb segment, so long as they are in line with the long-bone axis. Figure 2.31 shows the outline of a leg with seven anatomical markers in a four-segment three-joint system. Markers 1 and 2 define the thigh in the sagittal plane. Note that by convention all angles are measured in a counter clockwise direction starting with the horizontal equal to 0°. Thus θ_{43} is the angle of the leg in space and can be calculated from

$$\theta_{43} = \arctan \frac{y_3 - y_4}{x_3 - x_4} \tag{2.19}$$

or, in more general notation,

$$\theta_{ij} = \arctan \frac{y_j - y_i}{x_j - x_i} \tag{2.20}$$

As has already been noted, these segment angles are absolute in the defined spatial reference system. It is therefore quite easy to calculate the joint angles from the angles of the two adjacent segments.

2.6.2 Joint Angles

Each joint has a convention for describing its magnitude and polarity. For example, when the knee is fully extended, it is described as 0° flexion, and

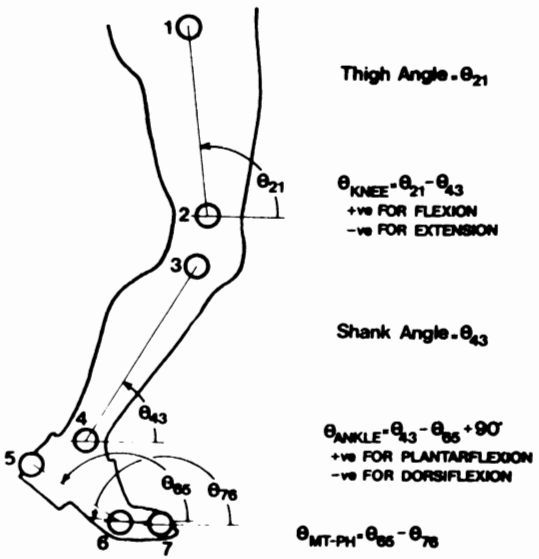

Figure 2.31 Marker location and limb and joint angles as defined using an established convention. Limb angles in the spatial reference system are defined using counterclockwise from the horizontal as positive. Thus, angular velocities and accelerations are also positive in a counterclockwise direction in the plane of movement, which is essential for consistent use in subsequent kinetic analysis. Convention for joint angles (which are relative) is subject to wide variations among researchers, but the convention must be clarified.

when the leg moves in a posterior direction relative to the thigh, the knee is said to be in flexion. In terms of the absolute angles described previously,

$$\text{knee angle} = \theta_k = \theta_{21} - \theta_{43}$$

If $\theta_{21} > \theta_{43}$, the knee is flexed; if $\theta_{21} < \theta_{43}$, the knee is extended.

The convention for the ankle is slightly different in that 90° between the leg and the foot is boundary between plantarflexion and dorsiflexion. Therefore,

$$\text{ankle angle} = \theta_a = \theta_{43} - \theta_{65} + 90°$$

If θ_a is positive, the foot is plantarflexed; if θ_a is negative, the foot is dorsiflexed.

2.7 CALCULATION OF VELOCITY AND ACCELERATION

2.7.1 Velocity Calculation

As was seen in Section 2.5.4, there can be severe problems associated with the determination of velocity and acceleration information. For the reasons

outlined, we will assume that the raw displacement data have been suitably smoothed by digital filtering and we have a set of smoothed coordinates and angles to operate upon. To calculate the velocity from displacement data, all that is needed is to take the finite difference. For example, to determine the velocity in the x direction, we calculate $\Delta x/\Delta t$, where $\Delta x = x_{i+1} - x_i$, and Δt is the time between adjacent samples x_{i+1} and x_i.

The velocity calculated this way does not represent the velocity at either of the sample times. Rather, it represents the velocity of a point in time halfway between the two samples. This can result in errors later on when we try to relate the velocity-derived information to displacement data, and both results do not occur at the same point in time. A way around this problem is to calculate the velocity and accelerations on the basis of $2\Delta t$ rather than Δt. Thus, the velocity at the ith sample is

$$Vx_i = \frac{x_{i+1} - x_{i-1}}{2\Delta t} \tag{2.21}$$

Note that the velocity is at a point halfway between the two samples, as depicted in Figure 2.32. The assumption is that the line joining x_{i-1} to x_{i+1} has the same slope as the line drawn tangent to the curve at x_i.

2.7.2 Acceleration Calculation

Similarly, the acceleration is

$$Ax_i = \frac{Vx_{i+1} - Vx_{i-1}}{2\Delta t} \tag{2.22}$$

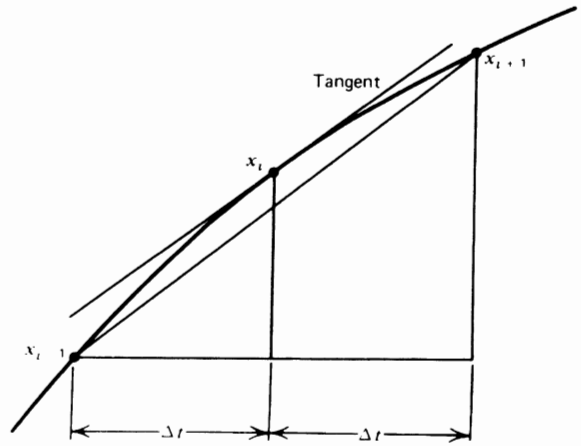

Figure 2.32 Finite difference techniques for calculated slope of curve in the ith sample point

Note that Equation (2.22) requires displacement data from samples $i + 2$ and $i - 2$; thus, a total of five successive data points go into the acceleration. An alternative and slightly better calculation of acceleration uses only three successive data coordinates and utilizes the calculated velocities halfway between sample times:

$$Vx_{i+1/2} = \frac{x_{i+1} - x_i}{\Delta t} \tag{2.23a}$$

$$Vx_{i-1/2} = \frac{x_i - x_{i-1}}{\Delta t} \tag{2.23b}$$

Therefore,

$$Ax_i = \frac{Vx_{i+1/2} - Vx_{i-1/2}}{\Delta t} = \frac{x_{i+1} - 2x_i + x_{i-1}}{\Delta t^2} \tag{2.23c}$$

2.8 PROBLEMS BASED ON KINEMATIC DATA

1. From Tables A.1 and A.2 in Appendix A, plot the vertical displacement of the raw and filtered data (in centimeters) for the greater trochanter (hip) marker for frames 1 to 30. Use a vertical scale as large as possible so as to identify the noise content of the raw data. In a few lines, describe the results of the smoothing by the digital filter.
2. Using filtered coordinate data (Table A.2), plot the vertical displacement of the heel marker from TOR (frame 1) to the next TOR (frame 70).
   (a) Estimate the instant of heel-off during midstance. (*Hint:* Consider the elastic compression and release of the shoe material when arriving at your answer.)
   (b) Determine the maximum height of the heel above ground level during swing. When does this occur during the swing phase? (*Hint:* Consider the lowest displacement of the heel marker during stance as an indication of ground level.)
   (c) Describe the vertical heel trajectory during the latter half of swing (frames 14–27), especially the four frames immediately prior to HRC.
   (d) Calculate the vertical heel velocity at HRC.
   (e) Calculate from the horizontal displacement data the horizontal heel velocity at HCR.
   (f) From the horizontal coordinate data of the heel during the first foot flat period (frames 35–40) and the second foot flat period (frames 102–106), estimate the stride length.
   (g) If one stride period is 69 frames, estimate the forward velocity of this subject.
3. Plot the trajectory of the trunk marker (rib cage) over one stride (frames 28–97).
   (a) Is the shape of this trajectory what you would expect in walking?
   (b) Is there any evidence of conservation of mechanical energy over the stride period? (That is, is potential energy being converted to kinetic energy and vice versa?)
4. Determine the vertical displacement of the toe marker when it reaches its lowest point in late stance and compare that with the lowest point during swing, and thereby determine how much toe clearance took place.
   Answer: y_{toe} (fr.13) = 0.0485 m, y_{toe} (fr.66) = 0.0333 m, clearance = 0.152 m = 1.52 cm.
5. From the filtered coordinate data (Table A.2), calculate the following and check your answer with that listed in the appropriate listings (Tables A.2, A.3, and A.4).
   (a) The velocity of the knee in the X direction for frame 10.
   (b) The acceleration of the knee in the X direction for frame 10.
   (c) The angle of the thigh and leg in the spatial reference system for frame 30.
   (d) From (c) calculate the knee angle for frame 30.
   (e) The absolute angular velocity of the leg for frame 30 (use angular data, Table A.3).
   (f) Using the tabulated vertical velocities of the toe, calculate its vertical acceleration for frames 25 and 33.
6. From the filtered coordinate data in Table A.2, calculate the following and check your answer from the results tabulated in Table A.3.
   (a) The center of mass of the foot segment for frame 80.
   (b) The velocity of the center of mass of the leg for frame 70. Give the answer in both coordinate and polar form.

2.9 REFERENCES

Dinn, D. E., D. A. Winter, and B. G. Trenholm. "CINTEL-Computer Interface for Television," *IEEE Trans. Computers* **C-19**:1091–1095, 1970.

Eberhart, H. D. and V. T. Inman. "An Evaluation of Experimental Procedures Used in a Fundamental Study of Human Locomotion," *Ann. NY Acad. Sci.* **5**:1213–1228, 1951.

Ferrigno, G. and A. Pedotti. "ELITE: A Digital Dedicated Hardware System for Movement Analysis Via Real Time TV Signal Processing," *IEEE Trans. Biomed. Eng.* **32**:943–950, 1985.