MECHANICS AND MODELLING FOR THE MIDDLE EAR

W. Robert J. Funnell[1,2], Nima Maftoon[1] & Willem F. Decraemer[3]

1. Department of BioMedical Engineering, McGill University, Montréal, Canada

2. Department of Otolaryngology – Head & Neck Surgery, McGill University, Montréal, Canada

3. Department of Biomedical Physics, University of Antwerp, Antwerpen, Belgium

1. INTRODUCTION

This document covers two general areas: mechanics for middle-ear modelling, and modelling for middle-ear mechanics. It is intended mainly for people who do not have advanced training in mathematics, mechanics and modelling but who wish to understand what is being done in the modelling of middle-ear mechanics as reviewed by Funnell et al. (2013).

Some readers may find that at some point the material becomes too difficult to follow in detail – they are encouraged to continue reading, if only to capture some of the flavour. On the other hand, some readers may find that the material seems too elementary – they also are encouraged to continue reading, because the topics do gradually become more substantial.

Section 2 is a review of some of the principles of mechanics that are relevant to the middle ear, from simple levers to viscoelasticity, continuum dynamics and non-linearity. Section 3 is a review of some general concepts related to modelling. It briefly mentions a number of different types of models, and Sections 3.5, 3.6 and 3.7 deal specifically with two-port modelling, lumped-circuit modelling and finite-element modelling in somewhat more detail.

2. MECHANICS

This section gradually introduces some of the basic concepts of mechanics that are required for understanding the mechanical behaviour of the ear. ‘Mechanics’ is taken here to include acoustics. The sequence starts with statics, moves through dynamics, first without and then with energy dissipation, then on to continuum mechanics, and finishes with a brief look at non-linearity. (Although particular textbooks are cited in various places below for further information, the reader will undoubtedly be able to find comparable information in many other undergraduate textbooks of physics and engineering.)

2.1 Statics: simple machines

Statics is the study of mechanical systems in which all velocities are zero, or at least negligible. (Strictly speaking, one should say that all relative velocities within the system are zero. An overall constant velocity of the system relative to the centre of the galaxy, for example, is irrelevant.) In real life, no system is truly static. However, if a system changes slowly enough that relative velocities and accelerations can be ignored, or if one waits long enough after a change that transient effects die out, then a system can be treated as though it were static.

Time does not enter into the analysis of static systems, and the mechanical variables consist of *force* and *displacement*. Each of these is a *vector* quantity, involving both a magnitude and a direction, as opposed to *scalar* quantities, which do not involve a direction. A vector can be described in, for example, *polar coördinates*, that is, as a magnitude and a direction (e.g., two miles northwest), or *Cartesian coördinates*, that is, as orthogonal (mutually perpendicular) magnitudes (e.g., four blocks

north and three blocks west). In our three-dimensional (3-D) world, forces and displacements are 3-D vectors. A common convention is to write vectors in bold face, and to use orthogonal x, y and z axes: $\boldsymbol{f} = f_x \boldsymbol{u}_x + f_y \boldsymbol{u}_y + f_z \boldsymbol{u}_z$ where $\boldsymbol{u}_x$, $\boldsymbol{u}_y$ and $\boldsymbol{u}_z$ are unit vectors in the x, y and z directions and f_x, f_y and f_z are the three components of the vector force $\boldsymbol{f}$.

Static analysis involves the balancing of forces and of force components. An example of such an analysis is that of a lever (Figure 1), one of the 'simple machines' described by Archimedes around 250 BC (e.g., Chondros, 2007). In this case the forces acting on the lever produce a *moment*, or torque, around the axis at the fulcrum. The moment is given by $M = f\,d$ where f is the magnitude of the force and d is the perpendicular distance between the axis and the line of action of the force. (In this simple example the moment has only one component, perpendicular to the image plane, but in general it is a 3-D vector.) The principle of the lever is that a moment applied to the lever must be balanced by an equal and opposite moment in order for the lever to be in equilibrium, that is, for it to have zero velocity. This moment balance can be expressed as $f_1 d_1 = f_2 d_2$. In the case of the wedge, another of the simple machines of Archimedes, one needs to balance the different components of the force vectors.

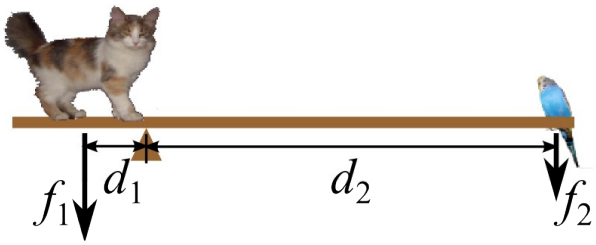

Figure 1: Lever and fulcrum, with a heavy weight and a light weight.

As another example of a static analysis, consider a thumbtack (Figure 2) being pushed into a wall. The function of a thumbtack is most easily understood in terms of *pressure*, with $p = f / A$ where A is the area of a flat surface and f is the magnitude of a force applied perpendicular to that surface. The pressure p is a scalar but the force and the area are both actually vectors; the 'direction' of a surface is the direction perpendicular to the surface. In the case of the thumbtack, a pressure p_1 applied to the large surface (with area A_1) results in a pressure p_2 at the small tip (with area A_2). Since the forces at the two ends of the thumbtack must be equal and opposite as long as it does not move, $p_1 A_1 = p_2 A_2$. Therefore $p_2 = p_1 (A_1 / A_2)$ and the thumbtack exerts a much larger pressure on the wall than could be exerted by the thumb on the tack. This analysis should in general be done with vectors, of course, and once the tack breaks into the wall then things become more complicated. (This use of the term 'pressure' apparently originated with Pascal (1623-1662) in the study of fluids (Pascal, 1937). A hydraulic press or jack also involves two different surface areas but, unlike the thumbtack, relies on the principle that the pressure is the same at all points in a static fluid.)


Figure 2: Thumbtack.

These particular examples, the lever and the thumbtack, are presented here because they are often invoked to explain the basic mechanism of the middle ear as involving an ossicular lever ratio and an eardrum/footplate area ratio (although such a description is very much oversimplified: Funnell, 1996).

2.2 Statics: springs

The systems discussed above consisted of rigid parts. The simplest example of a non-rigid body is the ideal spring, the behaviour of which was described by Hooke (1678). In modern terms Hooke's law can be written as

$$f = k\,u \tag{1}$$

where f is the force exerted on the free end of the spring, u is the displacement there, and the constant k is the *stiffness* of the spring. Sometimes it is convenient to use the *compliance* $C = 1/k$. Equation 1 can

be viewed as representing a balance of two forces, with *ku* being an imaginary force exerted by the spring. The work put into stretching or compressing a spring is said to be stored in the spring as *potential energy*. Work is calculated by multiplying force by the distance through which it acts; using calculus, the potential energy can be calculated by integrating the product of the applied force and the resulting displacement of the end of the spring:

$$E_p = W = \int_0^u F \,\mathrm{d}x = \int_0^u k\,x \,\mathrm{d}x = \tfrac{1}{2} k u^2. \tag{2}$$

Fluids, whether gases like air or liquids like water, do not have fixed shapes to which they return when external forces are removed. Thus, the archetypical spring is made from a solid material. However, if a fluid like air is compressed by a piston in a rigid-walled cavity, it effectively does have a fixed shape and it also can act like a spring. For a fixed amount of an ideal gas at a constant temperature, Boyle's law states that the product $P\,V$ is constant, where P is the internal pressure of the gas and V is the volume (Boyle, 1662). (Boyle's experiments were quite challenging, and his presentation of his results was obscure by modern standards, as discussed by West (1999, 2005).) Consider a piston of surface area A compressing a gas at pressure P_0 in a cylinder of length l_0. Suppose that the piston is displaced inward by distance u so that the new length becomes $l_1 = l_0 - u$, and the gas pressure and volume become P_1 and V_1. Boyle's law requires that $P_0 V_0 = P_1 V_1$, which means that $P_0(l_0 A) = P_1(l_1 A)$, so $P_1 = P_0(l_0/l_1)$. Therefore, with a little algebra it can be seen that the incremental pressure $P_1 - P_0$ is given by

$$p = P_1 - P_0 = P_0(l_0/l_1) - P_0 = P_0(l_0/l_1 - 1) = P_0\left(\frac{l_0 - l_1}{l_1}\right) = (P_0/l_1)u. \tag{3}$$

Since the additional force required to produce this displacement of the piston must be equal to the force exerted by p on the piston area A, the required force is $f = A\,p = (A\,P_0/l_1)u$. This is actually not as nice as it may appear since the right-hand side contains a combination of the original pressure and the new length, but if the piston displacements are assumed to be very small ($u \ll l_0$) then $l_1 \approx l_0$ and

$$f = k\,u \tag{4}$$

where $k \approx A\,P_0/l_0$. Equation 4 is the same as equation 1 for an ideal spring. The restriction to small displacements is not unreasonable, since a real-world spring will also no longer be so simply described when displacements become large compared with its length.

Equation 4 can also be expressed in terms of pressure rather than force:

$$p = \frac{f}{A} = \frac{1}{A} k\,u = \frac{1}{A}\frac{A\,P_0}{l_0} u = \frac{P_0}{A\,l_0} A\,u = \frac{P_0}{V_0} U \tag{5}$$

where V_0 is the initial volume of the enclosed air and $U = A\,u$ is the *volume displacement*, that is, the volume displaced by the surface area A in moving through distance u. This is the form used for acoustical circuit models (Section 3.6).

As will be discussed below, the middle-ear air cavities are often modelled as springs in this sense, and the ossicular ligaments are also often modelled as springs.

2.3 Dynamics: masses

Newton's three laws of motion from 1687 (Newton, 1729, Book I, pp. 19–21) underlie classical dynamics. In particular, the second law can be expressed as $f = m\,a$ where f is a force exerted on an object with mass m, and a is the resulting acceleration of the object. Again this can be viewed as a balance of forces, with ma being conceived of as an inertial force. If the mass changes then the law can be generalized in terms of *momentum*, the product of mass and velocity.

The acceleration is the rate of change of the velocity v, in the same way that velocity is the rate of change of displacement. The acceleration, velocity and displacement are all functions of time, indicated sometimes by writing '(t)' as in $a(t)$, v(t) and $u(t)$. In the notation of differential equations, $v = \dfrac{\mathrm{d}u(t)}{\mathrm{d}t}$ and $a = \dfrac{\mathrm{d}v(t)}{\mathrm{d}t} = \dfrac{\mathrm{d}^2u(t)}{\mathrm{d}t^2}$. The *derivative* $\dfrac{\mathrm{d}u(t)}{\mathrm{d}t}$ can be thought of as the ratio of an infinitesimal change in u (which varies with time) to an infinitesimal change in time, that is, the rate of change of u(t):

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{\Delta u}{\Delta t}. \tag{6}$$

The acceleration is the *second derivative* of $u(t)$, that is, the derivative of the derivative.

The work put into getting a mass to a certain velocity can be thought of as being stored in the form of *kinetic energy*. Analogous to Equation 2, the kinetic energy is found by integration to be

$$E_k = \int F\,\mathrm{d}x = \int F\,v\,\mathrm{d}t = \int m\frac{\mathrm{d}v}{\mathrm{d}t}v\,\mathrm{d}t = \int m\,v\,\mathrm{d}v = \tfrac{1}{2}\,m\,v^2. \tag{7}$$

In the middle ear the mass of the stapes is an obvious candidate for the straightforward application of $f = m\,a$, as long as it moves back and forth in a straight line and doesn't rock 'n' roll.

2.4 Dynamics: masses and springs

Consider the case of a mass attached to one end of a spring, with the other end of the spring fixed. Neglecting the effect of gravity or any other external force, the force acting on the mass is now the ku of equation 1, and Newton's second law gives this 'homogeneous' (with no forcing function) second-order linear differential equation:

$$-k\,u = m\frac{\mathrm{d}^2u}{\mathrm{d}t^2}. \tag{8}$$

One way to solve such an equation, that is, to find some function $u(t)$ that causes the two sides to always be equal, is to guess. In this case, the magic of sinusoidal functions is used.

Given a right-angled (90°) triangle as shown in Figure 3a, the sine function $\sin(\theta)$ is defined as y/r where y is the length of the side opposite the angle θ and r (a radius) is the length of the *hypotenuse*, that is, the side opposite to the right angle. It can be seen that $\sin(0°) = 0$, $\sin(45°) = \sqrt{2}$ (from the theorem of Pythagoras, ca. 550 BC), and $\sin(90°) = 1$. As θ sweeps from 0° through 90°, 180°, 270° and back to 0° (= 360°) the value of $\sin(\theta)$ varies smoothly from 0 to 1, back to 0, down to −1 and back to 1 (Figure 3b). The cosine function $\cos(\theta)$ is defined as x/r where x is the length of the side adjacent to the angle θ. As θ sweeps from 0° around and back to 0°, the curve of $\cos(\theta)$ looks like a shifted

version of sin(θ), starting at 1: cos(θ) = sin(θ+90°). This is described as a *phase shift* of 90°. Note that the value of cos(θ) is greatest at θ=0° where the slope of sin(θ) is greatest; that the value of cos(θ) is zero at θ=90° where the slope of sin(θ) is zero; and so on. This suggests that the derivative of sin(θ) is the same as cos(θ) and it is in fact not very difficult to prove both that

$$\frac{\mathrm{d}\sin(\theta)}{\mathrm{d}\theta}=\cos(\theta) \tag{9}$$

and that

$$\frac{\mathrm{d}\cos(\theta)}{\mathrm{d}\theta}=-\sin(\theta). \tag{10}$$

It can also be seen that

$$\frac{\mathrm{d}^2\sin(\theta)}{\mathrm{d}\theta^2}=-\sin(\theta). \tag{11}$$

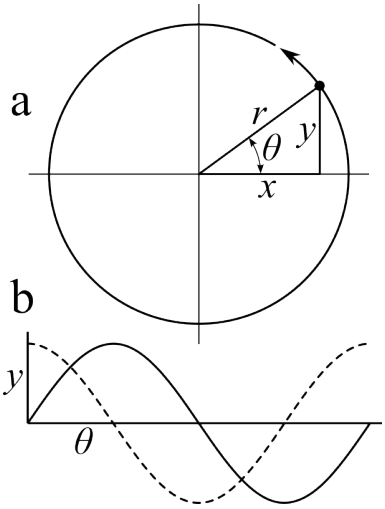

Figure 3: Definition of sine function. (a) Illustration of circular nature of sine function. (b) Shapes of functions sin(θ) (solid line) and cos(θ) (dashed line).

The angle argument of sin( ) and cos( ) is often expressed in *radians* rather than in degrees, with 2π rad = 360°; 1 rad is the angle which, when projected onto a circle from the centre, covers an arc equal in length to the radius of the circle. In other words, the value of an angle in radians is the length of the arc subtended by the angle divided by the radius. As a ratio of two lengths, the unit of radian is dimensionless.

It is now convenient (another 'guess') to write $\theta=\omega t$. This ω is interpreted as the *angular speed* or *angular frequency* (in radians/second, or just s^{-1}) with which θ cycles around the circle from 0 to 2π radians; $\omega=2\pi f$ where f is the frequency in cycles/second (Hz). After θ has been replaced by ωt, by the 'chain rule' of differentiation Equation 9 becomes

$$\frac{\mathrm{d}\sin(\omega t)}{\mathrm{d}t}=\frac{\mathrm{d}(\omega t)}{\mathrm{d}t}\frac{\mathrm{d}\sin(\omega t)}{\mathrm{d}(\omega t)}=\omega\cos(\omega t) \tag{12}$$

and Equation 11 becomes

$$\frac{\mathrm{d}^2\sin(\omega t)}{\mathrm{d}t^2}=-\omega^2\sin(\omega t). \tag{13}$$

Returning now to Equation 8, one guesses that $u(t)=\sin(\omega t)$, and the use of Equation 13 leads to

$$-k\sin(\omega t)=-\omega^2 m\sin(\omega t). \tag{14}$$

This equation can be reduced to

$$k=\omega^2 m \tag{15}$$

or

$$\omega=\sqrt{\frac{k}{m}}. \tag{16}$$

Thus, the 'lucky guess' for $u(t)$ has led to the finding that a sinusoidal motion of the mass at one particular frequency, given by Equation 16, satisfies the differential equation governing the ideal mass-spring system. This frequency is called the *natural frequency*, or *resonance frequency*, of the system. It is the frequency at which the system vibrates most easily, and the frequency at which it would continue to vibrate (in the absence of damping) even with no applied force. A resonance frequency is the basis for the usefulness of a tuning fork. The concept of resonance is often used to describe the behaviour of the middle ear, but that system is obviously much more complicated than a single mass and spring.

The fact that a sinusoidal signal can be completely described by an amplitude and a frequency is the simplest case of the complementary use of descriptions in the *time domain* and in the *frequency domain*. This concept will be used more fully below.

2.5 Energy dissipation

2.5.1 Introduction

In the discussion so far, there has been no mechanism for energy dissipation, but in real life every system has some energy dissipation. This dissipation can arise in many different ways, most or all of which involve heat generation due to complex interactions among molecules and atoms.

Dissipative effects are difficult or impossible to compute 'from first principles', and experimental observations are typically approximated with simple empirical formulas. Classical dry friction *between* two bodies is characterized as a *kinetic friction* (or Coulomb friction), with an opposing force proportional to the normal (perpendicular) force holding the bodies in contact and independent of the relative velocity of the two bodies, and as a larger *static friction* when the relative velocity is zero (e.g., Palmer, 1949). In fluids, the classical *viscous friction* involves an opposing force (e.g., by the walls of a pipe on the fluid flowing through it) proportional to velocity. There will also be energy dissipation *within* a deforming solid, due to some sort of *internal friction*. Various combinations and refinements of these types of friction models have been developed.

The effects of energy dissipation are discussed below in five parts. The first four parts deal with quasi-static conditions where mass and inertia are negligible: the first part deals with fluids, and then the next three parts deal with dissipation within solids. The fifth part deals with vibrations.

2.5.2 Viscosity in fluids

Fluids resist flow as a result of *viscosity*, which results mainly from molecular diffusion and, in liquids, from intermolecular forces. An applied force causes a fluid to move (e.g., in a pipe) with a velocity that, for a simple *Newtonian fluid*, is proportional to the force (Newton, 1729, Book II, p. 184). The constant of proportionality (usually denoted by μ or η) is the viscosity (specifically, the *dynamic* viscosity) of the fluid. When the force is removed, the fluid does not return to its initial configuration, so the work that was done by the force is lost and energy has been dissipated, unlike the case of a spring which returns its potential energy when the force is removed. Liquids generally have much higher viscosities than gases do, and a liquid like maple syrup has a much higher viscosity than water does. Corresponding to the notion of an ideal spring is the idealized mechanical element called a *dashpot*, thought of as a cylindrical reservoir filled with a viscous fluid in which a somewhat loosely fitting piston moves up and down (cf. Figure 9 below). The viscosity of the fluid squeezed between the piston and the cylinder walls dominates the behaviour of the device.

2.5.3 Viscoelasticity: creep and relaxation in solids

When a force is applied to an ideal spring, the spring will instantaneously take on a new length. In real life, this motion would be affected by the need to accelerate some mass, although the effect might be very small. The motion would also be affected by energy dissipation, but for a metal spring this would be very small. For some materials, however, the effect of energy dissipation can be quite substantial. Such materials are said to be *viscoelastic*, as opposed to just elastic. One phenomenon characteristic of viscoelastic solids is known as *creep*: if one suddenly applies and then maintains a force, there is a sudden deformation followed by a gradual further approach to a final, larger deformation. A related behaviour is called *relaxation*: if one applies whatever force is required to suddenly obtain a specific deformation and then holds that deformation, it will be found that the force required to hold it decreases gradually, approaching some lower but non-zero value.

One very simple way of attempting to represent such behaviour is the *Maxwell* model (Maxwell, 1867), effectively a series combination of an ideal Hookean spring and an ideal Newtonian dashpot (Figure 4a). In the case of creep, the spring contributes the desired instantaneous initial deformation and the dashpot contributes a subsequent further deformation. However, because the dashpot is in series with the spring, the combination will continue to deform without limit if the applied force is maintained, causing the combination to act more like a fluid than like a solid. In the case of relaxation, the behaviour is as required except that the force approaches zero. (Maxwell's analysis was for gases.)

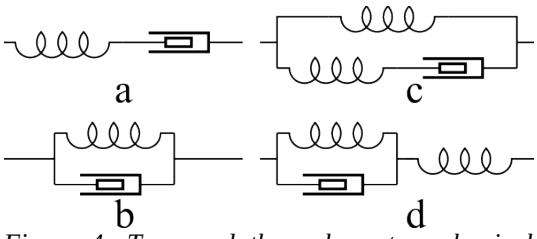

Figure 4: Two- and three-element mechanical models. (a) Maxwell. (b) Kelvin-Voigt. (c) Standard linear solid. (d) Alternative standard linear solid.

An alternative simple model (Figure 4b) is a parallel combination of a spring and dashpot (Voigt, 1892). In the case of creep, application and holding of a constant force will now cause a gradual approach to a final configuration, but no instantaneous deformation will occur at the beginning. In the case of relaxation, applying an instantaneous deformation to the dashpot (and thus to the parallel combination) will require an infinite force, but the force to hold the deformation will then instantly drop to a non-zero value. (Voigt's analysis was for friction in metals.)

The presentation of the differential equations of Maxwell and Voigt in terms of mechanical components actually came later (Poynting & Thomson, 1902; Zener, 1948) and was not universally appreciated (Tanner & Walters, 1998, p. 30). Reiner (1954) denoted the series spring-dashpot model by H — N and the parallel combination by $H \mid N$. The parallel combination has been variously referred to as the *Voigt* model, as the *Kelvin* model with reference to Thomson (1878), and as the *Kelvin-Voigt* model, and Tanner & Walters (1998) have suggested that perhaps it should be credited to Meyer (1874).

It can be seen that neither of these two simple models is adequate to describe creep and relaxation of a solid. The next step in complexity (Figure 4c,d) is to either add a spring in parallel with the Maxwell model ($H \mid M$) (Poynting & Thomson, 1902; Zener, 1948), or add a spring in series with the Kelvin-Voigt model (H — K). The $H \mid M$ model has been variously referred to as the *Poynting-Thomson* model; as the *Zener* model; as the *Kelvin* model by Fung (1993, §2.11); and often as the *standard linear solid* (SLS) following Zener (1948). The name Poynting-Thomson has sometimes been used to distinguish the H — K model from the $H \mid M$ model (e.g., Sobotka, 1984; Schiessel et al., 1995), but this seems to have been based on a misinterpretation of the way Poynting & Thomson (1902) drew their model. (The issue of names is not helped by the potential confusion between W. Thomson, Lord Kelvin; J. Thomson, W. Thomson's father; and J.J. Thomson of Poynting & Thomson.) In any case, the $H \mid M$ and H — K models are equivalent (e.g., Reiner, 1971, p. 179; Schiessel et al., 1995) and they both produce reasonable behaviour for both creep and relaxation.

With the SLS model, the gradual change of either deformation or force, produced for a constant

applied force (creep) or applied deformation (relaxation), respectively, is described by the exponential $e^{-t/\tau}$, where τ is a characteristic time constant. For biological materials like muscle and connective tissue, this time constant might be on the order of ten to twenty seconds. It turns out, however, that with real materials the gradual change often cannot be well described with a single time constant. One reason is that the phenomenon may be caused by multiple physical mechanisms, each with its own time constant. Each such time constant τ can be associated with a corresponding frequency $f = 1/\tau$, and the relaxation can be described in the frequency domain as a *relaxation spectrum* with narrow peaks corresponding to the different time constants. Moreover, each physical mechanism may be best described, not by a single time constant, but by a broad peak in the relaxation spectrum. For example, Zener (1948, p. ii) showed a 'typical' relaxation spectrum for a metal exhibiting six peaks over a range of frequencies spanning twenty orders of magnitude (i.e., a factor of 10^{20}). Five of the peaks were relatively narrow but one (due to grain boundaries) was very broad, itself covering three orders of magnitude. Relaxation mechanisms spanning large frequency ranges are also typical of the long-chain polymers found in biological materials. The mechanisms range from very fast ones, involving individual atoms, to much slower ones involving large-scale interactions among the long chains (e.g., Ferry, 1980; McLeish, 2008).

One way of dealing with relaxation spectra with multiple time constants is to use a *generalized SLS* model, that is, a parallel or series combination of multiple SLS models. Combining one or two such models per decade over the range of time constants that is required gives a fairly smooth relaxation spectrum. An alternative is to define a continuous spectrum with an infinite number of time constants (e.g., Neubert, 1963; Fung, 1993, §§7.6.5–7.6.7). Other mathematical approaches may also be used, such as correlation functions, stretched exponentials like $e^{-(t/\tau)^{\alpha}}$, power laws like $(t/\tau)^{-\beta}$, or *fractional calculus* in which derivatives like $\mathrm{d}/\mathrm{d}t$ are replaced by 'fractional' derivatives like $\mathrm{d}^{\beta}/\mathrm{d}t^{\beta}$ (e.g., Schiessel et al., 1995).

It is challenging to experimentally characterize a material exhibiting time constants that span many orders of magnitude. An interesting case is the pitch-drop experiment started by Parnell in 1927 (Edgeworth et al., 1984). Although pitch at room temperature acts like a hard solid, shattering when hit sharply, over periods of years it actually flows. In Parnell's set-up the pitch forms drops that separate and fall once every ten years or so.

Materials like pitch blur the dividing line between solids and fluids. Reiner (1964) proposed the *Deborah number*

$$D = \frac{\text{time of relaxation}}{\text{time of observation}} \tag{17}$$

to quantify the distinction and to make explicit the importance of how short or long the time of measurement is. If D is small, the material will appear to be a fluid, but if D is very large then the material will appear to be a solid.

Rheology is replete with observations of odd material behaviour such as *shear thinning*, *shear thickening*, *thixotropy* and *anti-thixotropy* (e.g., Tanner & Walters, 1998). One example is the behaviour of the synovial fluid that occurs in synovial joints. It is a highly viscous liquid, but when squeezed between a flat surface and a convex surface it does not completely flow away so as to allow the two surfaces to touch. A narrow layer remains, effectively acting like a solid (Ogston & Stanier, 1953).

2.5.4 Viscoelasticity: hysteresis

In the previous section the loads were applied as step functions. It is also common experimentally to apply ramp functions, with either force or displacement linearly increasing with time. This can be done cyclically using triangle-wave functions. For an ideal spring, the resulting plot of displacement against force, or force against displacement, is a straight line with the unloading curve superimposed on the loading curve. For a viscoelastic material, however, the loading and unloading curves form a loop. This phenomenon is referred to as *hysteresis* (Figure 5). The area within the hysteresis loop is a measure of how much energy is dissipated.

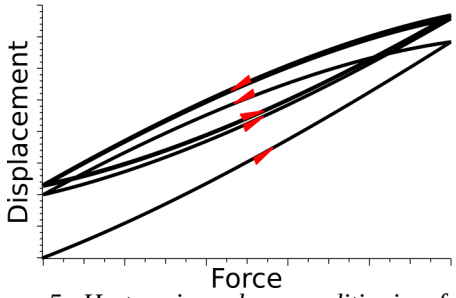

Figure 5: Hysteresis and preconditioning for a standard linear solid, loaded and unloaded with several cycles of a triangular wave. The first loading starts from (0,0) and the first unloading does not return to that point. After one cycle the curves converge to a final shape.

In general, when a sequence of loadings and unloadings is started, the hysteresis loop will gradually change shape over the first few (or several) cycles, gradually converging to a final shape. This is referred to as *preconditioning*. If one does this loading and unloading until the loop converges, then if one starts the loading and unloading again after a long pause the loops will again gradually change until they converge to a stable form. If the timing or amplitude of the applied cyclical load changes, the loops will again change shape. An example for simple single-time-constant viscoelasticity is shown in Figure 5. The actual shapes of the curves depend on how quickly the loads are applied and removed.

It is extremely important to precondition the material samples when attempting to experimentally characterize a viscoelastic material, in order to obtain reproducible results. Ideally the experimental conditions (including rate and amplitude but also things like temperature and humidity) should be the same for the preconditioning cycles as for the measurement cycles, and there should be no pause between them.

The effects of hysteresis and preconditioning are important in interpreting attempts to measure the material properties of different parts of the middle ear. They may also be important in understanding tympanometry. Normally material properties are measured on preconditioned samples, but there is no opportunity for preconditioning in the clinical application of large quasi-static tympanometric pressures, and tympanometric measurements have been found to depend on the direction and rate of the applied pressure changes and the number of measurement runs (e.g., Creten & Van Camp, 1974; Vanpeperstraete et al., 1979; Osguthorpe & Lam, 1981).

2.5.5 Damage

There is a somewhat different form of hysteresis known as the *Mullins effect*. In the idealized form of this effect, all of the hysteresis and preconditioning takes place the first time the material is loaded past the previous maximum loading. The effect is particularly associated with *filled* rubbers, that is, rubbers with a filling of small particles, such as natural rubber with carbon black filler, or silicone rubber filled with silica particles. In this case the phenomenon is associated with damage that is irreversible (e.g., Ogden, 2004; Peña et al., 2009). The distinction between the reversible response changes associated with SLS models and the irreversible changes associated with damage may not be clear-cut. Breakage of intermolecular bonds and changes of molecular configuration may be involved in both cases, with the bonds and configurations subsequently recovering easily, more slowly, or not at all. Some materials, including but not limited to living tissues, can heal even macroscopic damage (e.g., Wietor & Sijbesma, 2008; Canadell et al., 2011).

Another phenomenon caused by irreversible damage is *plasticity*. In this case, when a load is applied to a material, it responds linearly up to a *yield point* and then softens irreversibly. When the load is removed, the material returns linearly to a configuration different from the original one. Usually

associated with stiff materials and small deformations, this behaviour is found in metals (e.g., when bending a spoon) and also in bone (e.g., Charlebois et al., 2010).

2.5.6 Damped vibrations

Once the mass-spring system of Section 2.4 was set into motion, it would continue to vibrate at its natural frequency forever even with no further driving force applied. Energy would be cyclically swapped between the potential energy of the compressed spring and the kinetic energy of the moving mass. In the real world, however, energy dissipation would cause the vibrations to decay with time. A common way of adding energy dissipation to a mass-spring system is with the viscous friction of a dashpot. Its effect on vibrations is called *damping*. An everyday application is found in the shock absorbers of a car. Another example is seen in the fact that the vibrations of a tuning fork die out after a few seconds.

Including a viscous damper involves adding to Equation 8 a term proportional to velocity. Since energy is dissipated when the system moves, it is also necessary to add a forcing function to prevent the motion from dying out, giving

$$k\,u(t)+d\frac{\mathrm{d}u(t)}{\mathrm{d}t}+m\frac{\mathrm{d}^2u(t)}{\mathrm{d}t^2}=f(t) \tag{18}$$

where d is the damping parameter. (The use of d for the damping parameter invites confusion with the differential operator d; c is more commonly used for damping but may be confused with the speed of sound.) It is again assumed that $u(t)$ is sinusoidal: $u(t)=A\sin(\omega t)$. (Unlike the guess for solving Equation 8, this guess involves an amplitude A, which is required because of the presence here of the forcing function f.) With this guess, Equation 18 becomes

$$A\Big(k\sin(\omega t)+d\cos(\omega t)-\omega^2 m\sin(\omega t)\Big)=f(t). \tag{19}$$

The mixture of sine and cosine terms makes this more difficult to handle than Equation 14 was. A convenient method of approaching it involves the use of *complex numbers*. A complex number is an ordered pair of numbers, with a *real part* and an *imaginary part*. If the real part of a complex number is a and the imaginary part is b, then the complex number can be thought of as $a+ib$, where i is defined as the square root of -1. (Since the square of any ordinary ('real') number must be positive, i must clearly be 'imaginary' since $i^2 = -1$.) Equivalently, the complex number can be thought of as a point on the *complex plane*, with a horizontal coördinate of a and a vertical coördinate of b. One can treat $u(t)$ as a complex number and guess that $u(t)=A\big(\cos(\omega t)+i\sin(\omega t)\big)$. Even simpler, one can use the amazing fact that $e^{i\theta}=\cos(\theta)+i\sin(\theta)$ (Euler, 1748, p. 148; Sandifer, 2007), where e is defined as the remarkable number for which

$$\frac{\mathrm{d}e^x}{\mathrm{d}x}=e^x. \tag{20}$$

Substituting $u(t)=U\,e^{i\omega t}$ and $f(t)=F\,e^{i\omega t}$ (where U and F are both complex numbers) into Equation 18 yields

$$k\,U\,e^{i\omega t}+i\,\omega d\,U\,e^{i\omega t}-\omega^2 m\,U\,e^{i\omega t}=F\,e^{i\omega t}. \tag{21}$$

The quantity $F/U=k+i\,\omega d-\omega^2 m$ is the *dynamic stiffness*, a complex, frequency-dependent analogue

of the simple spring stiffness *k*. The real part is $k - \omega^2 m$ and the imaginary part is ωd. It is sometimes convenient to work with the complex velocity *V* rather than the displacement *U*. In that case *F*/*V* is the complex *impedance Z*. The inverse is the complex *admittance* (or *mobility*) $Y = 1/Z$.

At very low frequencies, $\omega^2 m$ is much less than *k* and the behaviour of the mass-spring system is *stiffness-dominated*, acting almost like a simple spring: the displacement is *in phase* with the applied force (the displacement is inward when the force is inward, and outward when the force is outward) and the magnitude of the displacement is almost constant. At very high frequencies, $\omega^2 m$ is much greater than *k* and the system is *mass-dominated*, acting almost like a simple mass: as indicated by the minus sign, the phase of the displacement is opposite to the phase of the applied force (it is 180°, or π radians, out of phase: the displacement is outward when the force is inward, and inward when the force is outward) and the magnitude of the displacement decreases (it is inversely proportional to the frequency squared).

At the undamped natural frequency $\omega = \sqrt{k/m}$, the stiffness and mass terms cancel each other out and the magnitude of the dynamic stiffness becomes small—very small if the damping (energy dissipation) is small. Figure 6 shows a plot of the *frequency response* (the magnitude and phase of the displacement for a constant magnitude and phase of the driving force) for various values of the damping parameter *d*. The resonance peak is very sharp for the lowest damping and the transition from being in phase to being in opposite phase is very abrupt. The resonance peak disappears completely for the highest damping. The damping has an effect only in the vicinity of the resonance, where increased damping reduces the magnitude of the displacements; above and below that region the damping has no effect on the overall level of the displacement curve. It is fairly common to use the words ‘damp’ and ‘damping’ when referring to an overall attenuation due, for example, to contractions of the tensor tympani (e.g., Hallpike, 1935). This usage is consistent with the general meaning of the verb ‘damp’ (OED, 2011, s.v. damp, v. 1.a) but is not consistent with the technical usage (OED, 2011, s.v. damp, v. 1.c).

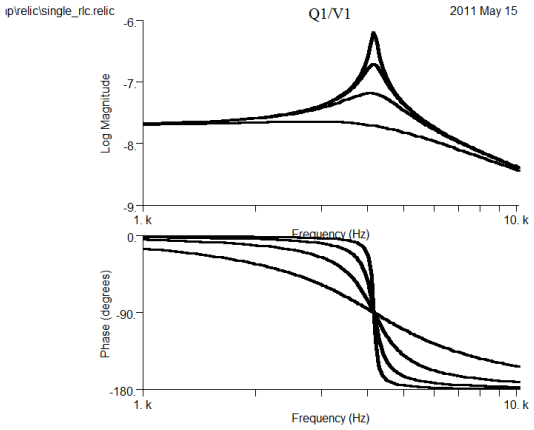

Figure 6: Frequency response for a simple mass-spring-dashpot system for four different values of the damping.

Notice that Figure 6 is plotted with a logarithmic frequency scale and a logarithmic magnitude scale. This is common in auditory studies, in part because it is well matched to human perception, for which both loudness and pitch are detected approximately logarithmically. A log-log scale for magnitude has the convenient effect that a mass-dominated roll-off proportional to frequency squared appears as a straight line, with a slope of −12 dB/octave. Similarly, a stiffness-dominated increase of velocity with frequency and a mass-dominated decrease of velocity with frequency have slopes of +6 and −6 dB/octave, respectively. (The logarithmic unit *dB* is a *decibel*, or a tenth of a *bel*. A ratio of 10 corresponds to 10 dB for a quantity of power (e.g., Friedland et al., 1961, p. 207). For non-power quantities like displacements and voltages, a ratio of 10 corresponds to 20 dB, and 6 dB corresponds to a ratio of approximately 2. An *octave* corresponds to a frequency ratio of 2.)

Real-world systems will usually include more than a single mass, spring and dashpot. A system with many such elements connected together will generally have a much more complicated frequency response than the one shown in Figure 6. The resonance peaks will be smeared together more or less, depending on the amount of damping. At the lowest frequencies the displacement will always become stiffness-dominated, and thus almost independent of frequency.

One approach to analyzing the frequency response of a system with a combination of multiple

masses, springs and dashpots is to express the frequency responses of individual mass-spring-dashpot *branches* as polynomial functions of $s = i\omega$. For example, the left-hand side of Equation 21 would become $k + s\,d + s^2 m$. One can then combine the resulting polynomials of different branches algebraically. It can be shown that the result is always a *rational* polynomial, that is, a ratio of two polynomials. One then attempts to find the roots of each of the two polynomials, that is, those complex values of s that cause the polynomial to equal zero. The roots of the denominator polynomial are called the *poles* of the system and correspond to the resonances discussed above. The roots of the numerator polynomial are called the *zeroes* of the system and correspond to *anti*-resonances, where the system response is small. In general the poles and zeroes are complex numbers and they are sometimes visualized as a *pole-zero plot* on the complex plane with the real part of s along the horizontal axis and the imaginary part along the vertical axis (e.g., Friedland et al., 1961, p. 202). If a pole lies on the vertical axis, then it corresponds to an undamped natural frequency.

So far in this section the discussion has been limited to the modelling of viscous damping, in which the force is proportional to velocity. This is the type of damping that has been used for the middle ear, but other forms exist. Coulomb friction (Section 2.5.1) is an important source of damping in some dynamic systems (e.g., Beards, 1996, pp. 39–42). In another variant (Beards, 1996, pp. 43-45), the damping parameter d in Equation 18 is divided by ω, so the dynamic stiffness becomes $k + i\,d - \omega^2 m$. This is referred to as *structural damping* or *hysteretic damping*; unfortunately, neither name is particularly distinctive. In some areas this formulation matches the frequency characteristics of the experimental data better than viscous damping does.

2.6 Rigid-body analysis

In the dynamic analysis so far, the spatial extents of the masses, springs and dashpots have not been taken into account. Each element has been represented by a single displacement, a single parameter (e.g., stiffness) and a single applied force, all of them scalar. No consideration has been given to positions, sizes and directions. In *rigid-body analysis* the extents of objects are taken into account, but the objects are still not considered to be deformable. Taking extents into account re-introduces the notions of lever arms and force vectors that were involved in the analysis of simple machines (Section 2.1). In addition, the concepts of *centre of mass*, *moment of inertia* and *angular momentum* arise. All depend on the size and shape of an object. The moment of inertia is analogous to the mass of a point mass as considered in Section 2.3, but for use with force moments as described in Section 2.1. In a system of connected rigid bodies, each part of the system can apply both forces and moments in specified directions and at specified locations on other parts of the system. The analysis uses the two laws of motion of Euler (1752; Euler Archive, n.d.), which together are a generalization of Newton's second law.

The three ossicles of the middle ear are obvious candidates for rigid-body analysis, although there is evidence that the manubrium does not always act as a rigid body (Decraemer et al., 1994). Even if the ossicles themselves are rigid, their motions are controlled by many surrounding structures that cannot be treated in that way.

2.7 Statics of a continuum

In the mechanical analyses considered so far, no account has been taken of deformation of an object: each object or part of a system was considered either to have no spatial extent at all or to be rigid. Clearly, real-life objects have spatial extent which affects their behaviour, and they can deform. The study of such systems is *continuum mechanics*. In the case of the middle ear, the most obvious structure for which spatial extent is likely to be important is the eardrum.

As an example of taking spatial extent into account, consider an elastic rod of material such as rubber, with length l and cross-sectional area A. If a perpendicular force F is applied to the ends, the rod will stretch by, say, amount s. Our experience tells us that, for a given material and a given force F, a rod with a small cross-sectional area will stretch more than one with a large cross-sectional area, and that a long rod will stretch more than a short one. For most materials it turns out experimentally that, at least for small deformations, s is proportional to l and inversely proportional to A. Thus,

$$s = \frac{F\,l}{A\,E} \tag{22}$$

where E is *Young's modulus* (Young, 1807, p. 137), a measure of the stiffness of the material. It is convenient to define *stress* as $\sigma = F/A$ and *strain* as $\epsilon = s/l$, leading to $\epsilon = \sigma/E$. (In different contexts one may meet somewhat different definitions of stress and strain, under various names.)

To analyze the effect of a longitudinal force F applied to a rod of length l for which the cross-sectional area A changes along its length, consider a short length Δl which stretches by amount Δs in response to the force. Because the forces applied to every part of the rod must be balanced, the force applied to this small piece of the rod is also F. Therefore

$$\Delta s = \frac{F\,\Delta l}{A\,E}. \tag{23}$$

As Δl is made infinitesimally small this leads to the differential equation

$$\frac{\mathrm{d}s}{\mathrm{d}l} = \frac{F}{E}\,A(l), \tag{24}$$

leading by integration to the overall stretch

$$s = \frac{F}{E}\int_0^l A(l)\,\mathrm{d}l. \tag{25}$$

It can be seen that consideration of the extent of an object has led to a differential equation with a spatial variable, as opposed to the time variable seen previously for dynamic analyses. In both cases these are *ordinary* differential equations because there is only one variable for which derivatives are involved. Clearly, however, a dynamic analysis of a spatially extended object will lead to a differential equation with derivatives of both spatial and temporal variables. This is called a *partial* differential equation, and the symbol ∂ is used instead of d to indicate differentiation by one variable while the other variables are held constant.

If a material is *incompressible*, that is, if the volume of a block of the material cannot change, then clearly the block must become narrower if an applied longitudinal force makes it become longer. On the other hand, if the material consists of uncoupled longitudinal fibres, it will not become narrower in response to a longitudinal force. The extent to which this narrowing occurs can be characterized by the *Poisson's ratio* ν (nu). For most materials this ratio lies between zero (as for uncoupled fibres) and 0.5 (for an incompressible material), although in some cases it can be negative (e.g., Greaves et al., 2011).

For three dimensions the simple $\epsilon = \sigma/E$ relationship given above must be generalized. For a 3-D block there are *axial* (or *normal*) stresses, each acting perpendicular to two opposite faces of the block, and the corresponding strains. There are also *shear* stresses, with equal and opposite forces acting parallel to each of two opposite faces, causing shear strains, with a shear modulus G analogous to E. Since the block has three pairs of opposite faces, there are thus three axial stresses (σ_x, σ_y and σ_z), three shear stresses (τ_x, τ_y and τ_z), three axial strains (ϵ_x, ϵ_y and ϵ_z) and three shear strains (γ_x, γ_y and γ_z). If the

material is *isotropic*, that is, if it responds to forces in the same way regardless of direction, then the simple $\epsilon = \sigma / E$ can be generalized as

$$\begin{pmatrix} \epsilon_x \\ \epsilon_y \\ \epsilon_z \\ \gamma_x \\ \gamma_y \\ \gamma_z \end{pmatrix} = \begin{pmatrix} 1/E & -\nu/E & -\nu/E & 0 & 0 & 0 \\ -\nu/E & 1/E & -\nu/E & 0 & 0 & 0 \\ -\nu/E & -\nu/E & 1/E & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/G & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/G & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/G \end{pmatrix} \begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_x \\ \tau_y \\ \tau_z \end{pmatrix}. \tag{26}$$

The six strains and six stresses can be further generalized to the 3×3 arrays

$$\begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{pmatrix} \text{ and } \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \tag{27}$$

and the array relating them is 9×9. These arrays are referred to as *tensors*. A tensor can be thought of as a generalization of the notion of a vector. Like a vector, a tensor is related to geometry. Like the special *scalar-product* (or *dot-product*) and *vector-product* (or *cross-product*) operations associated with vectors (**a**·**b**, **a**×**b**), there are special scalar-product and *tensor-product* operations associated with tensors (**A**:**B**, **A**⊗**B**). Just as taking the vector product of two vectors creates another vector, so too does taking the tensor product of two tensors create another tensor. Various relations among the components of a tensor are enforced by the geometrical and physical interpretation, including symmetries such as those seen in Equation 26 for the special case of isotropy.

Crystalline materials will be anisotropic, with symmetry depending on their crystal structure (e.g., cubic), while amorphous materials will be isotropic (e.g., Callen, 1960, sec. 13.6). Fibrous materials may be more or less isotropic, depending on how the fibres are organized. Fibrous materials are an example of *composite* materials, comparable to concrete with reinforcing bars.

2.8 Dynamics of a continuum

As an example of how to analyze the dynamics of a continuous system, consider a plane (that is, flat) *membrane*, defined here to be a thin structure that has negligible inherent stiffness but that resists a transverse force by virtue of being kept under tension. The typical example is the head of a drum. Analysis of a vibrating membrane is a classical example of the analysis of physical systems that are defined mathematically as *boundary-value problems*, that is, continuous systems with certain constraints imposed by *boundary conditions*.

The differential equation governing a uniform, plane membrane under tension, with no energy dissipation, is

$$T \nabla^2 w - \sigma \frac{\mathrm{d}^2 w}{\mathrm{d}t^2} = 0 \tag{28}$$

where $w(x,y,t)$ is the membrane displacement as a function of spatial position and of time; σ is the area density (mass per unit area); ω is angular frequency; T is tension; and p is applied pressure (e.g., Morse, 1981, p. 174). The Laplacian operator ∇^2 here is an abbreviation for

$$\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2} \tag{29}$$

(in three dimensions there would also be a *z* term) so Equation 28 is a partial differential equation with derivatives taken with respect to the two spatial variables *x* and *y*, and also with respect to time *t*. Equation 28 is the *wave equation*, a continuum version of Equation 8 with a mass-acceleration term and a linear restoring-force term. The second derivatives of the restoring force arise because a given point in the continuum is acted on by almost-balanced forces exerted by the points around it.

If the membrane is assumed to vibrate sinusoidally then Equation 28 becomes

$$T\nabla^2 w+\sigma\omega^2 w=0. \tag{30}$$

This is known as the *Helmholtz equation* and is a continuum version of Equation 15. The boundary condition is that the membrane displacement is zero at the boundary. For particularly simple boundary shapes, it is possible to solve this equation analytically using various tricks such as a judicious choice of coördinate system (e.g., Cartesian coördinates for a rectangular membrane, polar coördinates for a circular membrane) and clever separation of variables (e.g., breaking the *x*-*y* problem into an *x* problem and a *y* problem and then superimposing the solutions). When it can be solved, it results in multiple solutions, each with its own frequency (recall that sinusoidal vibration is assumed) and pattern of vibration. These are the natural frequencies and modes of the membrane, corresponding to the natural frequency found in Section 2.4.

Note that a ‘membrane’ in the sense used here has no intrinsic bending stiffness, so it can neither hold a shape nor resist a force unless a tension is applied to it. In the case of the eardrum, there is no strong evidence that it is normally under tension (e.g., Funnell & Laszlo, 1982) and much modelling has ascribed much or all of its behaviour to bending stiffness. There is thus room for confusion in referring to it as the tympanic ‘membrane’. A thin structure controlled by inherent bending stiffness is referred to as a *plate*. The differential equation for a flat plate is more difficult to solve than that for a membrane. The resulting natural modes of vibration are similar in form to those for a membrane with the same boundary shape, but the corresponding natural frequencies are different. In real life the material of a drum head must have some non-zero bending stiffness, and a plate may have some tension applied to it because of the nature of the fixation at the boundary. Such mixed problems are obviously more difficult to analyze.

The eardrum is of course not flat. A plate with curvature is referred to as a *shell*. The curvature causes a shell to act more stiffly than if it were flat. It also makes analysis much more difficult.

2.9 Travelling waves and standing waves

If a plate (or a stretched membrane) is tapped at one point, a wave will spread out from that point, subject to the same differential equation as discussed above but with the addition of energy dissipation. This *travelling wave* will carry energy across the surface of the plate. This is easily seen in real life in the waves that result from dropping a pebble into a pond. A one-dimensional version of this is a stretched string. If an extremely (*semi-infinitely*) long ideal string, with no energy dissipation, is plucked at one end ($x = 0$) with the waveform $F(-ct)$, a wave with that shape will travel along the string with speed *c*: $w=F(x-ct)$ (d’Alembert, 1749a, 1749b).

If the string has a finite length, when the wave hits the end it will be reflected and travel back: $w'=F(x+ct)$. The two waves will be superimposed for a while after the reflection. The reflected wave will itself be reflected when it reaches the first end, and travel back toward the other end. If, instead of being plucked, the end of the string is vibrated up and down sinusoidally, then the resulting wave that

travels along the string will be sinusoidal in both position and time: $w = A\sin(k\,x - \omega t)$. The reflected wave will be $w' = A\sin(k\,x + \omega t)$. In the steady state these forward-travelling and backward-travelling waves will be superimposed, giving $w = A\sin(k\,x - \omega t) + A\sin(k\,x + \omega t)$. By a trigonometric identity, it turns out that this gives $w = 2\,A\cos(\omega t)\sin(k\,x)$. This equation describes a *standing wave*, that is, a wave for which the spatial variation is independent of time (depending only on $k\,x$). For the spatial term sin(k x) to satisfy the boundary conditions that the string is fixed at both ends (i.e., sin(k 0) = 0 and sin(k *L) = 0*, where L is the length of the string), $k\,L$ must be an integer multiple of π. In the real world the waves generally die out more or less quickly because of energy dissipation as they are travelling and at the terminations. The result can be viewed as a combination of travelling and standing waves and may be characterized by a *standing-wave ratio*. It is instructive and entertaining to play with a simulation of these effects (e.g., "Wave on a String," 2011).

If there is a gradient of the geometry or material properties of a medium (e.g., an increasing width or thickness of a plate from one end to the other), it can exhibit *pseudo-travelling waves* that look like travelling waves but do not actually carry energy. The basilar membrane of the inner ear is a good example of this, and the eardrum may also be.

2.10 Longitudinal and transverse waves

The vibrations discussed in the previous section were *transverse*, that is, perpendicular to the direction of travel of the wave. Vibrations of strings, bars, membranes and plates are (primarily) transverse. Transverse waves, or *shear* waves, can also occur within 3-D solids. Electromagnetic waves (e.g., light waves) are also in this category. The waves in a Slinky toy, on the other hand, are *longitudinal* (or *compression*) waves. Sound waves in air are also longitudinal. The wave equation for longitudinal sound waves in air is the same as the wave equation for transverse waves on a string (e.g., Morse, 1981, pp. 221–222).

The speed of waves in both solids and fluids is $c = \sqrt{stiffness\,/\,mass}$ where *stiffness* is some relevant measure of stiffness, and *mass* is some measure of mass. The speed of sound in air, for example, is given by $c = \sqrt{P_0\,\gamma_c/\rho}$ (ibid.) where P_0 is the equilibrium pressure of the air, γ_c is the ratio of the specific heat of air at constant pressure to that at constant volume, and ρ is the equilibrium mass density of the air. The product $P_0\,\gamma_c$ provides the stiffness term.

Both transverse and longitudinal waves can exhibit standing waves. Acoustical standing waves in the external ear canal (cf. Richmond et al., 2011) are an important example for middle-ear studies.

2.11 Non-linearity

2.11.1 Linearity and non-linearity

Everything that has been said so far has involved *linear* systems, that is, systems for which the output is linearly related to the input. By definition this means that if *input*$_1$ causes *output*$_1$, and *input*$_2$ causes *output*$_2$, then the response to (*input*$_1$ + *input*$_2$) will be (*output*$_1$ + *output*$_2$). More concisely, if the system is represented by $y = F(x)$, then $F(x_1 + x_2) = F(x_1) + F(x_2)$. Setting $x_2 = x_1$ immediately leads to $F(2\,x_1) = 2\,F(x_1)$. One can then easily show that $F(3\,x) = 3\,F(x)$, $F(4\,x) = 4\,F(x)$, and so on. Generalizing from the natural numbers (1, 2, 3 …*)* to all real numbers (with a certain amount of blind faith) gives $F(\alpha x) = \alpha F(x)$ for any constant α. (An alternative to blind faith is to include this multiplicative requirement as part of the definition of linearity.)

For a system with dynamics, linearity does not imply that a particular input waveform will lead to an output with the same waveform, but it does imply that if the input waveform is scaled then the output waveform will be scaled accordingly. Furthermore, neglecting *initial conditions* and *transients*, linearity does imply that a *sinusoidal* input will lead to the same waveform at the output, that is, to a sinusoidal output at the same frequency, although possibly with a time delay. This assumes that the system is *time invariant*, that is, that none of its parameters or structure changes with time.

Linearity greatly simplifies the mathematical analysis of a system, and it is remarkable how many real-world systems behave linearly or almost so. This can be attributed to the possibility of approximating many general functions as *Taylor series*, and the fact that many systems of interest are operating close to equilibrium (e.g., Callen, 1960, sec. 16.4) so the behaviour can be approximated using only the lowest terms of the relevant Taylor-series expansion, neglecting the higher-order, non-linear terms. The outer and middle ear, for example, is practically linear for comfortable sound pressure levels. However, many important systems are not linear, and the middle ear itself is decidedly non-linear when responding to (1) the large quasi-static pressures involved in Eustachian-tube blockage, in tympanometry and in the rapid altitude changes experienced in airplanes and express elevators; (2) explosions and impact; (3) middle-ear-muscle contractions and chewing; and (4) surgical manipulation.

Non-linearity in the response to intense sound leads to distortion of the input. For example, Figure 7 shows the result of distorting a sine wave by a particular *clipping* non-linearity. It happens that, roughly speaking, any time-domain function can be represented as a *Fourier series*, that is, as the sum of a number (possibly infinite) of sine and cosine functions of different frequencies. (Strictly speaking, the time-domain function must be *periodic*, but as long as it is not infinite in duration one can always pretend that it repeats itself. This use of sines and cosines is quite amazing, but they are not the only family of functions that can be used as *basis functions* in this way.) Thus, a distorted output signal can be viewed as a combination of frequencies that were not present in the input. This is known as *harmonic distortion*. The new frequencies can be either greater than the original frequency (overtones) or, less well known, less than the original frequency (undertones) (e.g., Dallos, 1966; Huang et al., 2010).

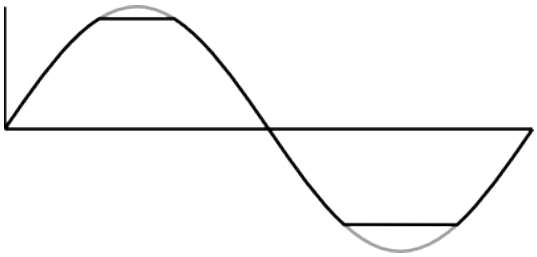

Figure 7: Clipping of a sine wave. The original sine wave (grey) has been clipped at different levels on the positive and negative sides.

2.11.2 Types of non-linearity

In the modelling of mechanical systems, non-linear effects can be divided into *geometrical effects* and *material effects*. Geometrical non-linearities occur when the geometry changes because of large displacements. For example, the force due to a pressure is always perpendicular to the surface to which the pressure is applied, so if the surface is displaced and changes its orientation significantly, the direction of the force will change and the original equations describing the system will no longer be valid. This causes a non-linear response. Another example of a geometrical effect is *contact*: when displacements cause two separate parts of a system to come into contact, the equations change dramatically and the response becomes very non-linear.

Large displacements can occur even when local strains remain small everywhere. However, if the strains themselves become large, then material non-linearities will usually arise because the stress-strain relationships of all materials become non-linear for large enough strains: a material may become

stiffer as it is stretched more and more, like a rubber band or a strip of eardrum, or it may become softer as it fails.

2.11.3 Hyperelasticity

Young's modulus and Poisson's ratio are obviously not sufficient to describe such non-linear behaviour, and a popular generalization is the *hyperelastic* (or Green) formulation. This assumes that the mechanical properties of a material can be defined in terms of a *strain-energy function*, *W*, which represents the potential energy stored in the material as a result of deformation (e.g., Ogden, 1984, p. 205). This function is a generalization of the potential energy of a spring (Equation 2) and its existence may be necessary for any physically reasonable material model (e.g., Carroll, 2009). The strain-energy function is often expressed in terms of λ_1, λ_2 and λ_3, the *principal stretches* (stretch being the ratio of stretched length to unstretched length), or in terms of the three *strain invariants* (so called because they do not change when the coördinate system changes):

$$\begin{aligned} I_1 &= \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \\ I_2 &= \lambda_1^2\lambda_2^2 + \lambda_2^2\lambda_3^2 + \lambda_3^2\lambda_1^2. \\ I_3 &= \lambda_1^2\lambda_2^2\lambda_3^2 \end{aligned} \tag{31}$$

For incompressible materials it can be shown that $\lambda_1\lambda_2\lambda_3 = 1$. Much of the work on hyperelastic modelling has been done for rubber, with a mix of phenomenological and structural approaches, the former focused on conveniently fitting experimental data and the latter on building models based on the long-chain polymeric nature of rubber. What follows is a very brief chronological summary of a few of the many hyperelastic formulations that have been proposed, to give some idea of the possibilities. Note that in the equations for *W* all of the symbols on the right-hand sides (except for the *I*'s and λ's, and *i* and *j*) represent material parameters.

The simplest hyperelastic strain-energy function is the Neo-Hookean proposed by Rivlin (1948a):

$$W = \tfrac{1}{6}E(I_1 - 3). \tag{32}$$

The *Mooney* (or *Mooney-Rivlin*) strain-energy function, with two material parameters, was proposed by Mooney (1940) and expressed in terms of the strain invariants by Rivlin (1948b):

$$W = C_1(I_1 - 3) + C_2(I_2 - 3). \tag{33}$$

This form has been generalized in various ways in order to better fit experimental data. For example, Rivlin & Saunders (1951) first generalized it by creating the *polynomial* strain-energy function, a doubly infinite series:

$$W = \sum_{i=0,\, j=0}^{\infty} C_{ij}(I_1 - 3)^i (I_2 - 3)^j, \; C_{00} = 0. \tag{34}$$

They then concluded by generalizing it in the following form:

$$W = C(I_1 - 3) + f(I_2 - 3) \tag{35}$$

where *f*( ) is some function. Some quite elaborate forms have been proposed, such as that of Alexander (1968):

$$W = \frac{\mu}{2}\left(C_1 \int \mathrm{e}^{k(I_1-3)^2}\,\mathrm{d}I_1 + C_2 \ln\left(\frac{(I_2-3)+\gamma}{\gamma}\right) + C_3(I_2-3)\right). \tag{36}$$

Veronda & Westmann (1970), fitting experimental data from skin, proposed what may have been the first strain-energy function not designed for rubber:

$$W = c_1\left(\mathrm{e}^{\beta(I_1-3)}-1\right) + c_2(I_2-3). \tag{37}$$

Ogden (1982, 1984) proposed the following frequently used generalization:

$$W = \sum_{i=1}^{N} \frac{\mu_i}{\alpha_i}\left(\lambda_1^{\alpha_i} + \lambda_2^{\alpha_i} + \lambda_3^{\alpha_i} - 3\right) \tag{38}$$

where the α_i coefficients are not restricted to being integers. Arruda & Boyce (1993), based on an 'eight-chain' model of the polymer, proposed the following form which looks complicated but actually has very few parameters:

$$W = n k \Theta\left(\frac{1}{2}(I_1-3) + \frac{1}{20N}(I_1^2-3^2) + \frac{11}{1050N^2}(I_1^3-3^3) + \frac{19}{7000N^3}(I_1^4-3^4) + \cdots\right). \tag{39}$$

Marlow (2003) took a quite different approach, using a non-parametric representation of W calculated directly from uniaxial experimental measurements. The goal of all of these approaches is to find a model that can be fitted to uniaxial experimental data and then used for other deformation modes such as biaxial and torsional.

2.11.4 Incompressibility and compressibility

All the strain-energy functions shown above are for incompressible materials. Incompressibility is quite a good approximation for rubber and also for soft biological tissues. When necessary, compressibility is usually handled by assuming that one can just add another independent term to the strain-energy function.

2.11.5 Isotropy and anisotropy

All the strain-energy functions above are for isotropic materials. Isotropy is a very good approximation for rubber and for some soft tissues, but not for the anisotropic behaviour of tissues with highly organized fibres. If the fibres are parallel and are embedded in an isotropic ground substance, the tissue can be considered to be *transversely isotropic*. Among a number of early works describing models of fibrous tissues, Lanir (1979) took into account the distribution of fibre orientations and Decraemer et al. (1980) took into account the distribution of fibre lengths. The mechanisms assumed to contribute to tissue stiffness were quite different. Lanir (1983) appears to have been the first to formulate such a material law using a strain-energy function. Weiss et al. (1996) treated the fibre stiffness empirically using an exponential function, and used the Mooney-Rivlin strain-energy function (Eqn. 33) for the ground substance in which the fibres are embedded. Holzapfel et al. (2000) similarly split the strain-energy function into an isotropic part for the ground substance and an anisotropic part for the fibres, for which there were two different orientations:

$$\begin{aligned} W_{\text{iso}} &= \frac{c}{2}(I_1-3) \\ W_{\text{aniso}} &= \frac{k_1}{2k_2}\sum_{i=4,6}\left(\mathrm{e}^{k_2(I_i-1)^2}-1\right) \end{aligned} \quad . \tag{40}$$

Here I_4 and I_6 are additional strain invariants that account for the orientations of the two families of fibres. Gasser et al. (2006) extended this to allow for distributions of fibre orientations.

For a recent review in more depth, see Holzapfel & Ogden (2010).

2.11.6 Non-linear viscoelasticity

Non-linear viscoelasticity, or visco-hyperelasticity, the combination of viscoelasticity (as discussed in Sections 2.5.3 and 2.5.4) with non-linearity, makes for considerable complexity. See, for example, the review by Wineman (2009), and the comparison of two computational approaches by Charlebois et al. (2013). In practice it is often handled by means of *pseudo-elasticity*, using a hyperelastic material with artificially different properties for the loading and unloading situations, to obtain the effect of hysteresis. Motallebzadeh et al. (2013) showed, however, how a truly non-linear viscoelastic approach can match both the loading and unloading curves for strips of eardrum with a single set of parameters.

3. MODELLING

3.1 Why model?

In general, modelling can be used for purposes of understanding, prediction and control (e.g., Haefner, 2005). Construction and exploration of a model can concisely summarize what is known about a system, and can lead to a better understanding of how it works. Once a model has been validated to some extent, it can be used for prediction. Finally, models can be used for purposes of control; in the context of the middle ear this could mean the design of prostheses and implants.

In the 1980's, quantitative modelling had already led to the feeling that computational science had joined theoretical and experimental science and was 'rapidly approaching its two older sisters in importance and intellectual respectability' (Lax, 1986). It is true that 'the fact that some computer code produces a pattern that looks like a snowflake … doesn't necessarily tell us that the ingredients of that code have anything to do with the natural phenomenon it seems to depict' (Langer, 1999) and some models may contain intentional *fictions* (as the word is used by Winsberg, 2010). Even so, however, such a model, once confidence in its validity has been gained, can be used to probe how a system works in order to build understanding. Models can be used to examine and manipulate variables that are not accessible (*observable*) or manipulable (*controllable*) in the real system. A quantitative model can often be used to perform computational experiments that are not feasible with the real system: too expensive, too small, too big, too fast, too slow, too many confounding factors, and so on.

3.2 Caveat lector

Readers of modelling research should beware. Zwislocki (2002) stated that 'the history of auditory research is full of examples of unrealistic mathematical and conceptual models that ignore existing experimental evidence and contradict fundamental physical laws'. To repeat the statement here is not to belittle the role of modelling but rather to encourage a certain skepticism. Moreover, even if a model

seems fundamentally sound, ‘most mathematical models published in peer-reviewed journals are not reproducible: they contain the authors’ errors of commission and omission, augmented by the errors introduced by editors and typesetters. Therefore, an exactly reproducible model is a rarity.’ (Chizeck et al., 2009). Errors of commission and omission can involve:

- inappropriate approximations, e.g., linearity for an obviously non-linear system
- inadequate approximations, e.g., finite-element mesh too coarse or time step too large
- wrong geometry, boundary conditions, applied loads or material properties
- model defects, e.g., connectivity errors in finite-element meshes
- wrong units (cf. Euler et al., 2001)
- software errors
- typographical and arithmetic errors

Even in the absence of such errors, one must continually keep in mind the assumptions, simplifications and constraints of a model when drawing conclusions from it. One must keep in mind also that, just as any scientific theory can be disproven but not proven, so a model may be shown to be wrong but cannot be shown to be definitively correct (e.g., Anderson & Papachristodoulou, 2009).

3.3 Approaches to modelling

In this section some of the general approaches to modelling are briefly introduced and contrasted.

Models, like measurements, can be either *qualitative* or *quantitative*. Measurements can be classified (Stevens, 1946) as using *nominal scales* (e.g., hockey-player sweater numbers), *ordinal scales* (e.g., hockey-player ranks in scoring statistics), *interval scales* (hockey-player birth dates) or *ratio scales* (hockey-player ages or numbers of goals). There is some question as to whether nominal scales involve measurement at all. In any case, quantitative modelling requires the use of interval or ratio scales for measuring the parameters and variables of the models, and is the only kind of modelling considered here.

Analytical models are those for which the behaviour can be determined using mathematical tools like algebra and calculus, with very little arithmetic. All of the mathematical descriptions in Section 2 above are actually analytical models. One method for developing approximate analytical models is the *asymptotic* technique used by Rabbitt & Holmes (1986, 1988) for the outer and middle ear. It involves cleverly formulating an analysis using a sum of a series of terms, and then making an approximation by judiciously keeping only the lowest-order, simplest terms. This is conceptually related to the use of Taylor series and Fourier series (Section 2.11.1). Such models can be quite realistic but their formulations are dependent on every assumption and approximation made during the analysis, sometimes making it difficult either to modify assumptions or to refine approximations, and making it essentially impossible to use the details of actual measured geometries.

Numerical models involve less analysis and more arithmetic than analytical models. There is really a continuum from mostly analytical to mostly numerical. The analytical part of the treatment can handle continuous time and space variables, whereas the numerical part must deal with discrete time and space. The discretization itself involves approximations in addition to the approximations made in representing a real-world system analytically. Even the purest analytical model requires discretization and arithmetic when it comes to displaying the model’s behaviour graphically or in tabular form. The

more complicated is the system being modelled, the more approximations are required in the analytical phase, and the sooner the numerical phase becomes necessary.

Non-parametric models represent a system simply by an array of data, like a look-up table. For example, the middle ear might be represented by its input/output transfer function (e.g., stapes velocity over canal sound pressure) sampled over some range of frequencies. Such a model would be the result of certain kinds of system-identification procedures, and may be useful for some purposes, but is very inflexible. The model might be converted into a *parametric* model by fitting some mathematical curve to the data, the parameters of the fitted curve being the parameters of the model.

Black-box models are characterized only by their inputs and outputs, and have no internal structure related to the components and configuration of the system being modelled. For example, if the fitted curve in the previous paragraph were just a polynomial chosen because it has enough degrees of freedom to be able to match the measured frequency response adequately, it would constitute a black-box model. Again, such models may be useful but are inflexible. *Structural* models (sometimes called white-box models) have some internal structure established by consideration of the system being modelled. The mathematical descriptions in Sections 2.1 to 2.3 might be considered as parametric black-box models but Section 2.4 presents a parametric structural model.

Lumped-parameter models neglect the spatial extents of their components, as in Sections 2.1 to 2.6. (The rigid-body analysis in Section 2.6 takes spatial extent into account only in a very limited way.) The continuum analyses in Sections 2.7 to 2.10 represent *distributed-parameter* models.

Numerical, white-box, distributed-parameter models tend to be computationally expensive for systems of any degree of complexity. Since the 1960's and 1970's, the availability of increasingly powerful digital computers, and the development of sophisticated numerical techniques for structural engineering, have made such models more feasible.

Linear dynamic systems can be modelled in either the *time domain* (representing signals as temporal sequences of values) or the *frequency domain* (representing signals as combinations of different frequencies as described in Section 2.11.1). Although either the time-domain or the frequency-domain approach may be preferred in some cases, the two are practically equivalent and one can move from one to the other using *Fourier transforms* and *inverse Fourier transforms*. For example, simulations can be done in the frequency domain by calculating natural frequencies and modes of vibration. Frequency-domain responses to specified inputs can then be computed by combining those natural modes, and time-domain responses can be computed as inverse Fourier transforms. If the number of natural modes in the frequency range of interest is very large, as it often is for complex systems, it may be easier to compute a transient response (say, a step response) in the time domain and then use a Fourier transform to obtain the frequency response if desired.

Frequency-domain modelling is only valid for linear systems. In the case of non-linear systems, it is normally necessary to do the analysis in the time domain. The frequency domain may then be used to characterize harmonic distortion, that is, the amount of energy introduced at additional frequencies in response to a single-frequency input.

Multiscale models include phenomena across a range of different scales that must be treated differently. Winsberg (2010) described some multiscale techniques at an introductory level and discussed the philosophical implications, and White et al. (2009a, 2009b) presented a collection of articles on biomedical multiscale modelling. For example, Tang et al. (2009) linked results from full atomic-scale and molecular-scale modelling to a continuum theory for a fibre-reinforced composite, 'handshaking the fibril scale to the fiber and continuum scale' to simulate collagenous soft tissue.

3.4 Types of models

Two-port models and *circuit models* have often been used for the middle ear and are discussed in some detail in Sections 3.5 and 3.6, respectively. *Finite-element models* have also often been used for the middle ear. The finite-element method, one of a family of approaches to boundary-value problems (as

introduced in Section 2.8), is discussed in Section 3.7.

Multibody models, based on rigid-body analysis (Section 2.6) supplemented by various approaches to modelling the elastic parts of the system, have also been used for the middle ear (Eiber & Kauf, 1994; Hudde & Weistenhöfer, 1997; Eiber & Freitag, 2002; Volandri et al., 2012; Ihrle et al., 2013).

State-variable (or *state-space*) *models* are expressed as sets of first-order difference equations involving input, output and *state* variables (e.g., Chen, 2010, sec. 7.6). They can be non-linear, and they can be used as representations of circuit models.

Mass-spring models consist of networks of ideal springs with point masses where they are connected. Such models might be thought of as low-end finite-element models, or as circuit models gone wild. They are sometimes used when computational speed is very important, as in virtual-reality applications. Ho et al. (2012) presented such a model for the eardrum.

Discrete-element (or *discontinuum* or *particle*) *models* are formulated in terms of multiple interacting objects such as spheres or other shapes (e.g., Williams et al., 1985; Šmilauer et al., 2015). They have been used to simulate fluid flow, among many other things, and might be appropriate for modelling synovial fluid in the middle-ear joints. (The term ‘discrete-element’ has sometimes been used for multibody models (e.g., Schmitz & Piovesan, 2016).) Particle models can also be thought of as a special form of the more general *agent-based models* (e.g., Jennings, 2000, 2001).

Other types of models that are not considered further here are *fuzzy models*, *stochastic models*, *rule-based models* and *neural networks*.

3.5 Two-port models

Two-port models are also referred to as ‘quadripole’ (e.g., Belevitch, 1962), ‘four-pole’ (e.g., Egolf, 1977) or ‘two-terminal-pair’ (e.g., Friedland et al., 1961) models. A two-port model consists of a black box containing any linear system, and characterized by an input voltage and current and an output voltage and current. At each of the two ports, the current out of one terminal is assumed to be the same as the current into the other terminal. This equality makes sense when one thinks about the connection of the port to the rest of the universe.

Consider the inputs to the black box to be the two voltages (v), and the outputs to be the currents (i). Then, since the system within the black box is assumed to be linear, it follows that the outputs must be related to the inputs by equations of the form

$$\begin{aligned} i_1 &= y_{11} v_1 + y_{12} v_2 \\ i_2 &= y_{21} v_1 + y_{22} v_2 \end{aligned} \tag{41}$$

or

$$\begin{pmatrix} i_1 \\ i_2 \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{42}$$

where the y coefficients have the units of admittance. (It is implicitly assumed here that there are no energy sources within the black box.)

By considering the case when the output port is short-circuited (v_2=0), one obtains $y_{11} = i_1 / v_1$ and $y_{21} = i_2 / v_1$. Similarly, if the input port is short-circuited (v_1=0), one obtains $y_{22} = i_2 / v_2$ and $y_{12} = i_1 / v_2$.

Thus the coefficients can be thought of as *short-circuit admittances*.

One can equally well consider the inputs to be the currents and the outputs to be the voltages, leading to

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \end{pmatrix} \tag{43}$$

where the coefficients can be thought of as open-circuit impedances. One can express the relationships among the four quantities in other ways as well, including the use of the standard *transfer coefficients* to express the inputs in terms of the outputs:

$$\begin{pmatrix} v_1 \\ i_1 \end{pmatrix} = \begin{pmatrix} A & -B \\ C & -D \end{pmatrix} \begin{pmatrix} v_2 \\ i_2 \end{pmatrix}. \tag{44}$$

Such models are normally analyzed in the frequency domain, with the voltages, currents and coefficients all being functions of the complex frequency s. There are well-defined rules for handling series, parallel and cascade combinations of two-port networks.

Two-port models are sometimes convenient because they completely encapsulate the external behaviour of a system. The frequency-dependent coefficients can be based on an analytical model, either lumped or distributed; or on a numerical white-box model of any kind; or they can be non-parametric, based entirely on measured data. In any case, once the behaviour has been represented by the four coefficients, the model is then treated as a lumped-parameter black box when its interactions with other parts of a larger system are analyzed. Since two-port models have only one scalar flow variable and one scalar potential variable at each port, they are well suited for modelling experimental data like acoustical impedance and admittance, and single-point displacement or velocity measurements, but they are not well suited for dealing with spatial patterns such as eardrum vibration patterns or high-frequency sound-pressure variations across the eardrum.

3.6 Circuit models

A powerful theory for networks (circuits) of idealized elements was originally developed for electrical circuits made up of resistors, inductors and capacitors (e.g., Friedland et al., 1961). A typical circuit is shown in Figure 8. Each branch in the circuit is assumed to be completely defined by a single *flow* variable and a single non-flow *potential* variable, and corresponds to a *port* in the sense of the previous section. For an electrical circuit, the flow variable is current (i) and the potential variable is voltage (v). Current consists of the flow of electrical charge (q).

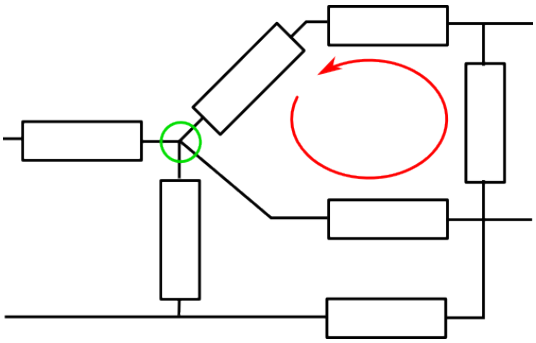

Figure 8: Example lumped circuit. The green circle denotes a node at which currents are summed, and the red arc denotes a loop around which voltages are summed.

The branches are connected at *nodes* and the connections form a set of *loops*. The interactions among the branches are completely defined by the circuit laws of Kirchhoff (1845):

- Current law: the sum of the currents at each node is zero. There will be some currents flowing into a node and some flowing out, and the net current will be zero. This is simply a statement of conservation of charge.
- Voltage law: the sum of the voltages around each loop is zero. This is a statement of

conservation of energy.

Each individual branch is taken to consist of an ideal electrical *resistor*, *capacitor* or *inductor* (or, in practice, of a series combination of one of each). The ideal resistor is defined by Ohm's law: $v = i\,R$, where R is the *resistance*. The ideal inductor and capacitor are defined by $v = L\dfrac{\mathrm{d}i}{\mathrm{d}t}$ and $v = \dfrac{1}{C}q = \dfrac{1}{C}\int i\,\mathrm{d}t$, respectively, where C is the *capacitance* and L is the *inductance*. Note that R, L and C here are all constant, not varying with time or anything else, and that the equations are all linear. Note also that there are no spatial variables at all, so these are lumped-parameter circuits. The voltages and currents can vary with time. If they do not, the circuit is referred to as *direct-current* (or DC) and the inductor voltages and capacitor currents must be zero. If the voltages and currents vary sinusoidally at some fixed frequency, the circuit is referred to as *alternating-current* (or AC). In general, the voltages and currents can vary in more complex ways, as they do for example when there is a transient fluctuation (e.g., voltage spike or outright power failure) in an AC power system.

In addition to the resistors, inductors and capacitors, one can also have voltage sources, current sources and *transformers*. A transformer can be thought of as two branches that are coupled in such a way that $\dfrac{v_1}{v_2} = \dfrac{i_2}{i_1} = r$ where r is the transformer ratio. This is an idealized form of the step-down transformers used to convert from high distribution voltages to low household voltages.

If there are b branches in the circuit then there are $2b$ unknowns (voltages and currents) to be found. Since a physical circuit must have a unique set of voltages and currents, there must therefore be $2b$ independent equations to be solved. The voltage-current relationships for the individual branches give b independent equations. For a circuit with n nodes, the Kirchhoff current law gives $n-1$ independent equations. There must therefore be $2\mathrm{b} - b - (n-1) = b - n + 1$ independent voltage-law equations (e.g., Friedland et al., 1961, sec. 4.3). For a simple circuit it is easy to identify the required number of independent equations. They are then solved simultaneously to obtain all of the currents and voltages in the circuit. This can be done either in the time domain or in the frequency domain. If solving in the time domain, it is relatively easy to generalize to circuit components that are non-linear and time-varying. If solving in the frequency domain, it is relatively easy to generalize to circuit components with parameters that vary with frequency.

Where all this becomes relevant to middle-ear mechanics is in the observation that the differential equations describing the idealized electrical circuit elements have the same form as the equations describing idealized springs, masses and dashpots as seen in Sections 2.2, 2.3 and 2.5.6. The analogy extends to idealized acoustical elements (e.g., Beranek, 1954). If voltage is taken to be analogous to mechanical force (or acoustical pressure) and current is taken to be analogous to mechanical velocity (or acoustical volume velocity), then Table 1 demonstrates the electrical, mechanical and acoustical equivalences.

Electrical		Mechanical		Acoustical	
v	voltage	f	force	p	pressure
i	current	u	velocity	U	volume velocity
R	resistance	R	resistance	R	resistance
L	inductance	m	mass	M	mass
C	capacitance	$1/k$	compliance (spring)	C	compliance (volume)
$v = i\,R$		$f = R\,u$		$p = R\,U$	
$v = L\dfrac{\mathrm{d}i}{\mathrm{d}t}$		$f = m\dfrac{\mathrm{d}u}{\mathrm{d}t}$		$p = M\dfrac{\mathrm{d}U}{\mathrm{d}t}$	
$v = \dfrac{1}{C}\int_{-\infty}^{t} i\,\mathrm{d}\tau$		$f = k\int_{-\infty}^{t} u\,\mathrm{d}\tau$		$p = \dfrac{1}{C}\int_{-\infty}^{t} U\,\mathrm{d}\tau$	

Table 1: Electrical, mechanical and acoustical analogies.

Figure 9 shows the mechanical and acoustical equivalents of a series resistor-inductor-capacitor branch. The mechanical components have been discussed already. For the acoustical components, the resistance (energy dissipation) is represented by a fine mesh: air flowing through it will lose energy through friction in the small holes. The acoustical mass is represented by a tube: air is supposed to move in it as a whole without either compression or friction. The acoustical compliance is represented as a rigid-walled cavity with the compliance determined by the volume of the cavity (cf. Equation 5).

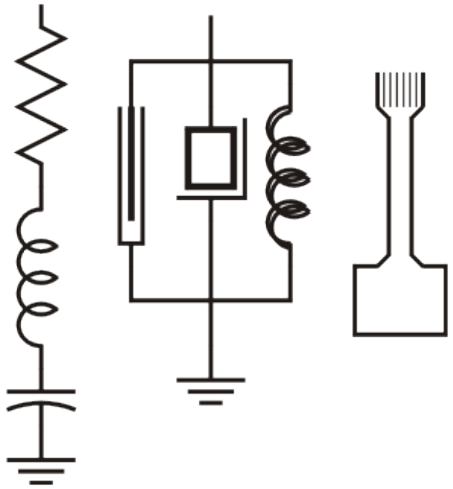

Figure 9: Symbols for electrical circuit elements (leftmost column) and the corresponding mechanical (centre) and acoustical (right) elements.

Note that the mechanical dashpot, mass and spring are shown as being in parallel and not in series. Since the series arrangement of electrical components means that each component has the same current, then, by the assumed association of force with voltage and velocity with current, the three mechanical components must be arranged so that they all have the same velocity. (The acoustical elements, on the other hand, will all have the same volume velocity when arranged physically in series, different from the arrangement of the mechanical elements even though the analogy seems closer.) Moreover, the electrical ground point, defined as having a potential of zero volts, corresponds to a mechanical ground corresponding to zero velocity, not force. Note also that in real electrical (and acoustical) systems it is easier to measure the non-flow variables, voltage and pressure, without disturbing the system, while in real mechanical systems it is easier to measure the flow variable, velocity. Because of this kind of discrepancy, the electrical/mechanical analogy is sometimes made the other way around, by associating voltage with velocity rather than with force, and current with force rather than with velocity. The electrical/acoustical analogy may also be inverted. There are advantages and disadvantages to the different approaches, and in fact the whole issue is more complicated than it first appears (Terhardt, 2000).

Note that the schematic symbol for a mechanical mass has a second part that does not correspond to anything physical. This is because each element needs two ends for connection to other elements. In the case of the acoustical compliance, however, the second (grounded) end is implicit as the outside of the cavity enclosure, corresponding to zero (usually atmospheric) pressure. If a real two-ended compliance element is required, it is possible to use a diaphragm placed across the cross-section of the mesh or tube

(Beranek, 1954, sec. 5.5). In the case of the outer and middle ear, a rigid-walled cavity is appropriate.

When one combines different types of system in the same model, for example, mechanical and acoustical parts for the middle ear, it is necessary to handle the differences in the physical quantities represented. This can be done using transformer components to implement the changes of measurement units. Often the transformers are not explicitly included in the model, the conversions being handled implicitly by scaling the component values appropriately.

Circuit models are based on ideal components, but in the real world even specially designed electrical components are not ideal: resistors have non-zero inductances, inductors have non-zero resistance, and so on. Mechanical components are likely to be even less ideal: springs and dashpots have non-zero masses, for example.

One must be especially careful about the frequency range over which a model is a good enough approximation. For a lumped approximation of a rigid-walled cavity, for example, the dimensions should be much less than the wavelength of sound. For a 5% error in the acoustical impedance of a cylindrical cavity, Beranek (1954) calculated a maximum length of $\lambda/16$, and a maximum radius of $10/f$ that corresponds to a maximum diameter of about $\lambda/17$ (for a speed of sound in air of about 345 m/s). For a typical middle-ear dimension of 10 mm, this corresponds to a maximum frequency of only about 2 kHz. This is a very conservative estimate since an error of 5% is only about 0.4 dB, much smaller than the errors typically found in middle-ear measurements, so a higher maximum frequency is usually accepted.

Onchi (1949) was the first to present a circuit model for the middle ear. He presented a mechanical circuit, and an equivalent electrical circuit ‘courtesy of Charles T. Molloy’, but did not provide component parameters or actually do computations with the model. Zwislocki (1957) presented an electrical circuit model and estimated parameter values from measurements of acoustical input impedance.

As with two-port models, circuit models may be appropriate for data like acoustical impedances and single-point displacements but are not well suited for dealing with spatial patterns of pressures and displacements. Hudde & Weistenhöfer (1997, 2000; 2006) have introduced the notion of 3-D generalized circuit models in which each velocity is replaced by six velocities (three translational and three rotational) and each force is similarly replaced by six forces. This can be viewed (Hudde & Weistenhöfer, 1997, p. 536) as a small-displacement specialization of rigid-body analysis (§2.6). The approach is more powerful than conventional circuit models but less general than the continuum methods discussed in the next section.

3.7 Finite-element models

3.7.1 Methods for solving boundary-value problems

This section is a brief introduction to some of the concepts involved in using the finite-element method to solve boundary-value problems. Many texts are available on this topic, the first having been that of Zienkiewicz & Cheung (1967).

The simplest boundary-value problems can be solved by writing down the differential equations from consideration of the physical system, and then solving the equations (with their associated boundary conditions) analytically, perhaps using separation of variables to reduce a multidimensional problem to several one-dimensional ones. Separation of variables is possible only for a few simple boundary shapes, however, and for arbitrary boundaries approximate analytical methods are required, such as variational techniques, perturbation techniques and conformal mapping. For more complicated shapes, or in the presence of inhomogeneities, numerical methods become necessary. The main numerical methods for solving partial differential equations are numerical integration, the finite-difference method, the finite-element method, and the boundary-element method. The first is generally

not used for multidimensional problems.

The *finite-difference method* consists of replacing the differential operators by difference operators in the differential equation for an entire system. Graphically this is equivalent to placing a regular grid over the system and considering only points at intersections of grid lines. The *finite-element method* involves conceptually dividing the system into a mesh of elements of finite size with geometries that are simple enough to permit analytical solution of the differential equations within individual elements. The mesh can be as regular or irregular as desired. Rather than needing to solve one complex problem all at once, one solves a series of simple problems. One takes into account inter-element effects by considering the various elements to be connected only at specific nodes on their boundaries. The *boundary-element method* is similar to the finite-element method but considers only elements on the boundary of the structure being modelled (e.g., Brebbia, 1984; Beer et al., 2008). The effects of the interior are taken into account implicitly by integrating over a surface rather than throughout a volume. This can dramatically reduce the computational expense, especially if the interior is homogeneous. The boundary-element method is often used for acoustical fields.

There are a number of features that make the finite-element method particularly attractive. First, in even its simplest form it handles irregular boundary shapes very conveniently. Second, it is relatively easy to handle inhomogeneities, nonlinearities, and complex geometries and boundary conditions systematically. Third, it is not necessary to analyze the behaviour of the whole system. All of the analysis is done on simple elements, and even there it is not necessary to write down the governing differential equations. The elements can be formulated using any method that is convenient, such as the variational method often used for plates and shells. Fourth, the finite-element method can be understood on a physical or intuitive basis. Its use is essentially an extension of the very common technique of analyzing mechanical or electrical systems as networks of interconnected discrete components.

3.7.2 Overview of the finite-element method

In the finite-element method the physical system to be analyzed is divided into a number of discrete elements which may or may not correspond to natural subdivisions of the actual structure. For example, for an assembly of plates and beams, the elements may consist of the individual plates and beams themselves, if these are easy enough to analyze individually. On the other hand, a single homogeneous plate of irregular shape may be considered to be composed of a number of triangular or rectangular plate elements, which together make up the overall irregular shape, but which individually are easy to analyze. Once one has divided the structure into elements, the mechanical behaviour of each element is analyzed, and its response to applied loads is expressed in terms of the displacements of its edges. This analysis is often based on the Rayleigh-Ritz procedure, which is discussed in Section 3.7.3. An example of a particularly simple element formulation is presented in Section 3.7.4; more complex elements are discussed briefly in Section 3.7.5.

The element analysis results in a matrix equation relating the behaviour of the element to the applied forces. The components of the matrix are functions of the shape and properties of the element. After the initial element analysis has been done, one need only substitute the shape and property definitions for each element into the formulæ obtained for the matrix components. In fact the user seldom needs to actually do the preliminary analysis at all, since such analyses have been published in the literature and included in computer programmes for a wide variety of element types.

Once the element matrix equations are ready, they are all combined together into one overall system matrix equation, as discussed in Section 3.7.6. (The boundary conditions are also included in the system matrix equation.) Since the behaviour of each element has been described in terms of its behaviour at its edges, and in fact at certain discrete nodes along its edges, this assembly of element matrices is simply a statement of the fact that a node shared by two elements must have the same

displacement when considered as part of either element, and of the assumption that the elements can interact only at these discrete nodes.

The actual solution of the system equations is discussed in Section 3.7.7.

3.7.3 Rayleigh-Ritz procedure

The Rayleigh-Ritz procedure was introduced by Ritz in 1908 and is a generalization (e.g., Finlayson, 1972, pp. 223–224; Leissa, 2005) of a technique used by Rayleigh (1877). It is a common procedure for formulating finite-element approximations. The following discussion is based on Stakgold (1968, pp. 332–335) and Silvester (1969).

The procedure is based on the theorem of minimum potential energy in mechanics. This theorem states that if one obtains a *functional* giving the potential energy of a system, then the *admissible* function that minimizes the functional is the solution of the system. (A functional here can be thought of as a function of a function. An admissible function is one that satisfies the boundary conditions of the boundary-value problem, as well as certain continuity conditions.)

The differential equation for a plane membrane under tension, vibrating sinusoidally with an applied load is

$$\nabla^2 w + \lambda^2 w = g. \tag{45}$$

This is Equation 30 with the addition of an applied pressure term g and with $\lambda^2 = \sigma \omega^2 / T$. It can be shown that the energy functional corresponding to Equation 45 is

$$F(w) = \tfrac{1}{2} \iint |\nabla w|^2 \, \mathrm{d}S + \tfrac{1}{2} \iint \lambda^2 w^2 \, \mathrm{d}S = \iint w g \, \mathrm{d}S \tag{46}$$

where the integrals are over the entire region of interest (e.g., over the triangular element considered in the next section). The three integral terms on the right-hand side represent energy due to membrane tension, inertia, and the applied pressure, respectively.

In practice it will be difficult or impossible to find the one function that actually minimizes the functional. Thus, one must limit the set of functions over which one will attempt to minimize the functional. The Rayleigh-Ritz procedure consists of restricting the search to a particularly simple subset of admissible functions, namely, the space of linear combinations of n independent admissible basis functions $w_1(\boldsymbol{x})$, … , $w_n(\boldsymbol{x})$. The value of n is chosen to be as small as is consistent with the accuracy required of the answer. The particular set of basis functions used is more or less arbitrary, as long as they are independent and admissible.

Since each of the basis functions is admissible, every function in the space of linear combinations will also be admissible. These functions can be expressed as

$$w(\boldsymbol{x}) = \sum_{i=1}^{n} c_i w_i(\boldsymbol{x}) \tag{47}$$

where the c_i are n constants defining $w(\boldsymbol{x})$. Now, minimizing the functional $F(w)$ over this set of functions evidently requires choosing the c_i such that F (which is now a function of the c_i) is minimal. Thus, taking the partial derivatives of F with respect to each c_i in turn and setting it to zero, one ends up with a set of n algebraic equations in c_i:

$$\frac{\partial}{\partial c_i} F\left(\sum_{j=1}^{n} c_j w_j \right) = 0, \, i = 1, 2, \ldots, n. \tag{48}$$

Thus, the boundary-value problem is reduced to the solution of n linear equations in n unknowns.

3.7.4 A simple element analysis

As an example of the type of analysis involved in the formulation of finite elements, this section will consider the analysis of a very simple case consisting of a triangular plane-membrane element using only three basis functions.

The functional of Equation 46 is the one to be minimized using the Rayleigh-Ritz procedure. As for what basis functions to use, the simplest choice is a set that permits the displacement field over the element to be a general linear function of x and y. For this, three basis functions are needed. There are two ways of formulating them, one based on Cartesian coördinates and the other on so-ca1led 'natural' coördinates. The ultimate result is the same, but in some circumstances one or the other formulation may be superior. Both methods are presented below.

The Cartesian-coördinate method involves choosing as basis functions the set $\{1, x, y\}$. The displacement at any point (x,y) in the element is then given by a linear combination of these three functions, namely,

$$w(x,y) = c_1 + c_2 x + c_3 y. \tag{49}$$

Following the Rayleigh-Ritz procedure, one then replaces w by this expression in the equation for $F(w)$, carries out the double integrations, and differentiates with respect to each of the c_i in turn. Setting the derivatives equal to zero then results in a set of algebraic equations for the c_i. The equations can be written in matrix notation as

$$\boldsymbol{A}\,\boldsymbol{c} = \boldsymbol{B}\,\boldsymbol{g}. \tag{50}$$

Lower-case bold is used here for vectors, and upper-case bold for matrices. The components of $\boldsymbol{g}$ represent the nodal values of the pressure field $g(x,y)$. The components of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are functions of the vertex coördinates of the triangular element, $(x_1,y_1,x_2,y_2,x_3,y_3)$, and of λ.

To be able to combine this equation with similar equations representing other elements, one must write it in terms of the displacements. Applying Equation 49 at each node, one obtains the transformation

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \tag{51}$$

or, in matrix notation,

$$\boldsymbol{w} = \boldsymbol{X}\,\boldsymbol{c}. \tag{52}$$

The w_i are the nodal displacements, and this equation simply represents the expression of the nodal displacements as linear combinations of the basis functions. Solving Equation 52 for $\boldsymbol{c}$ and substituting it into Equation 50 gives

$$\boldsymbol{A}\,\boldsymbol{X}^{-1}\boldsymbol{w} = \boldsymbol{B}\,\boldsymbol{g}. \tag{53}$$

Changing notation, one may write

$$\boldsymbol{K}_{\mathrm{e}}\boldsymbol{w} = \boldsymbol{B}\,\boldsymbol{g}. \tag{54}$$

Note that the right-hand side is actually a vector of nodal forces, so that one may also write the equation as

$$\boldsymbol{K}_e \boldsymbol{w} = \boldsymbol{f}. \tag{55}$$

The components of the matrix $\boldsymbol{K}_e$ are evidently ratios of forces over displacements, and $\boldsymbol{K}_e$ is known as the element stiffness matrix.

The natural-coördinate method (e.g., Zienkiewicz et al., 1969, p. 390 ff.; Desai & Abel, 1972, pp. 88–91) expresses the location of any point in the triangular element by the area coördinates (ζ_1, ζ_2, ζ_3), where $\zeta_i = A_i / A$; A is the total area of the triangle, and the A_i are as shown in Figure 10. (Note that the three ζ_i are not independent, since $\zeta_1 + \zeta_2 + \zeta_3 = 1$.) It can be shown that the relationship between Cartesian coördinates and these 'natural' coördinates is given by

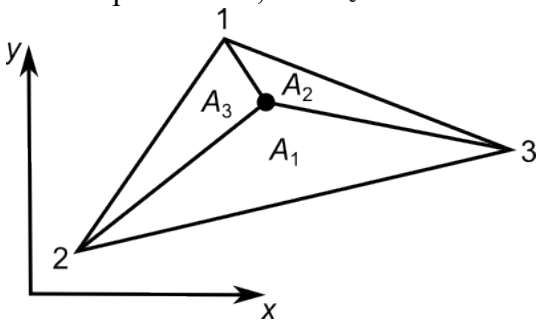

Figure 10: Areas used to define area coördinates of a triangle.

$$\begin{pmatrix} 1 \\ x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix} \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix} = \boldsymbol{X}' \boldsymbol{\zeta}. \tag{56}$$

The area coördinates $\{\zeta_1, \zeta_2, \zeta_3\}$ may now be used as the set of basis functions. It can easily be seen that in this case the coefficients c_i become the nodal displacements w_i. Following the Rayleigh-Ritz procedure again leads to an equation in the form of Equation 55.

3.7.5 More complex elements

In the simple example presented above, the basis functions for the displacements were linear, that is, linear interpolation was used to determine the displacements along the boundary and within the area of each element. One can, however, develop elements based on higher-order functions, for example, quadratic polynomials. Two points may be worth highlighting about such elements. First, the coefficients of the higher-order basis functions (the c_i) include, in addition to the displacements at the vertices of the triangle, the displacements at a number of extra nodes along the edges and in the interior of the triangle. For example, a third-order element has two extra nodes on each edge and one in its interior. These extra nodes need not add to the difficulty of defining the physical problem to the computer since they can easily be generated automatically.

The second point is to justify the use of more complex high-order elements instead of a larger number of simple low-order elements. The order of the system of equations to be solved is the same if one uses nine first-order elements, for example, instead of one third-order element. It can been shown, however, that the accuracy in the latter case will be considerably superior. This is analogous to the fact that if one wishes to use three degrees of freedom to approximate an arbitrary curve, then it is generally better to use one second-order polynomial than to use two straight-line segments.

The element in the previous section was a 2-D membrane element; 2-D elements for plates and shells are also used, and can be triangular or quadrilateral. Beams and pipes are examples of 1-D elements. In 3-D, elements can be tetrahedra (triangular pyramids) or hexahedra (brick-like). Special formulations can be used to couple different types of problems: mechanical, acoustical, thermal, electromagnetic, etc. Fluid-structure interactions (acoustical plus mechanical) are particularly relevant to middle-ear modelling.

3.7.6 Assembly of system equation

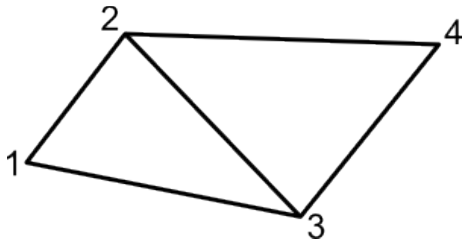

Figure 11: Two triangles to be assembled, as described in text.

The previous section outlined the analysis for a simple triangular element. Suppose that a system to be studied consists of two such elements interconnected as shown in Figure 3. For the first element, suppose that substituting the vertex coördinates and material properties into the formulæ obtained from the analysie3ws gives the stiffness (***S***) matrix components a_{ij} for the equations relating the nodal displacements w_i to the nodal loads f_i:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}. \tag{57}$$

(The f_i are the individual nodal loads due to a distributed load applied to the triangle.) Similarly, suppose that for the second element the following matrix equation is obtained:

$$\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} w_2 \\ w_3 \\ w_4 \end{pmatrix} = \begin{pmatrix} g_2 \\ g_3 \\ g_4 \end{pmatrix}. \tag{58}$$

Now, since the elements are connected at nodes 2 and 3, the w_2 and w_3 in the first equation are the same as the ones in the second equation. Therefore one can combine the two equations as follows:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22}+b_{11} & a_{23}+b_{12} & b_{13} \\ a_{31} & a_{32}+b_{21} & a_{33}+b_{22} & b_{23} \\ 0 & b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2+g_2 \\ f_3+g_3 \\ g_4 \end{pmatrix}. \tag{59}$$

In practice this can be done for thousands or millions of elements. The result is

$$\boldsymbol{K}\,\boldsymbol{w} = \boldsymbol{f}, \tag{60}$$

analogous to Equation 1, and similar to Equation 55 but with the much bigger system stiffness matrix ***K***.

The system boundary conditions now consist of prescribed values for some of the displacements. The constraints can be implemented in various ways.

Recall that the force terms on the right-hand side of this equation represent nodal forces due to surface forces (or body forces like gravity) acting on the various elements. One can also add further nodal force terms representing specified point loads on the system. The equation then represents the entire boundary-value problem, and is solved as discussed in the next section.

3.7.7 Solution of system equations

The actual nodal displacements for a static problem are obtained by solving Equation 60, that is, by *inverting* the matrix, giving

$$\boldsymbol{w} = \boldsymbol{K}^{-1} \boldsymbol{f}. \tag{61}$$

For undamped vibrations the force on the right-hand side of Equation 60 is replaced by a term involving frequency, mass and displacement (cf. Equation 14):

$$\boldsymbol{K}\,\boldsymbol{w} = \omega^2 \boldsymbol{M}\,\boldsymbol{w}. \tag{62}$$

This is treated as an *eigenvalue* problem: simplifying somewhat, an *eigenvector* $\boldsymbol{v}$ of a matrix $\boldsymbol{A}$ is defined as a vector such that multiplying the matrix and vector together results in a vector that is the same as the original vector except for a scale factor, λ: $\boldsymbol{A}\ \boldsymbol{v} = \lambda\ \boldsymbol{v}$. Comparing this with Equation 62, one can see that the eigenvectors of the matrix $\boldsymbol{M}^{-1}\boldsymbol{K}$ correspond to displacement vectors $\boldsymbol{w}$ and that the eigenvalues correspond to squares of the vibration frequencies ω. These frequencies are the natural frequencies of the system and the corresponding displacement vectors describe the natural mode shapes of the system. A system like the middle ear, which has many degrees of freedom and is very asymmetrical, will have large numbers of very closely spaced natural frequencies.

There are many ways of computing the inverse of a matrix and of computing its eigenvalues and eigenvectors, including both *direct* and *iterative* methods (e.g., Desai & Abel, 1972, pp. 19–24). The choice of method is a fine art but the user is generally rather insulated from the details by the finite-element software.

For a damped vibration problem the system of second-order differential equations is

$$\boldsymbol{K}\,\boldsymbol{w} + \boldsymbol{C}\,\dot{\boldsymbol{w}} + \boldsymbol{M}\,\ddot{\boldsymbol{w}} = \boldsymbol{f} \tag{63}$$

(cf. Equation 21). Since knowledge of the damping is often poor anyways, the calculations are often simplified by assuming *Rayleigh damping*, that is, assuming that $\boldsymbol{C} = \alpha\boldsymbol{K} + \beta\boldsymbol{M}$ where α and β are parameters relating the damping matrix to the stiffness and mass matrices.

The frequency response for this system can be computed by combining natural modes, but this tends to be inefficient for a system like the middle ear with large numbers of closely spaced modes. The usual method is to calculate the system response in the time domain, step by step. As a simple illustration, consider the first-order differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x) \tag{64}$$

where x is a function of time. This is solved numerically at discrete values of t, so the value of x at time t_i is denoted by x_i. The derivative dx/dt is approximated using the difference between the values of x at consecutive times:

$$\frac{x_{i+1} - x_i}{\Delta t} \tag{65}$$

where Δt is the time step selected for the calculation. The differential equation is thus converted into a *difference equation*. The choice of Δt is very important: making it very small requires a lot of computation, but making it too large can lead to inaccuracy and even instability.

	Implicit	**Explicit**
Solve completely at each time step?	Yes	No
Speed/time step	Slow	Fast
Step size	Large	Small
Stability	Unconditional	Conditional

Table 2: Comparison of implicit and explicit integration schemes.

The algorithms for solving a difference

equation may be classified as *explicit* or *implicit* (Table 2). The explicit method uses a *forward difference* at time t_i:

$$\frac{x_{i+1}-x_i}{\Delta t}=f\left(x_i\right). \tag{66}$$

This leads to an explicit equation for x_{i+1} in terms of the already known value for x_i:

$$x_{i+1}=x_i+f\left(x_i\right)\Delta t. \tag{67}$$

The implicit method uses a *backward difference* at time t_{i+1}:

$$\frac{x_{i+1}-x_i}{\Delta t}=f\left(x_{i+1}\right). \tag{68}$$

This leads to an equation for x_{i+1},

$$x_{i+1}=x_i+f\left(x_{i+1}\right)\Delta t, \tag{69}$$

which is said to be implicit because the unknown x_{i+1} appears as the argument of $f(\ )$ and not just by itself. The explicit method requires fewer calculations at each time step, but it requires very small steps in order to produce stable results. The implicit method requires more calculations at each time step but larger (and fewer) time steps can be used; if the time steps are too large, the method will still be stable but will have the effects of suppressing the higher-frequency components of the system response and of introducing a frequency shift. (This is sometimes referred to as low-pass filtering or as numerical damping.) There are also variations on these methods, including the use of *central differences* and the semi-implicit *symplectic* method.

If the system is non-linear, then the system stiffness matrix depends on the displacements and must be recalculated repeatedly as the system responds to the load. This is conceptually straightforward but numerically delicate and computationally expensive. The load is usually applied in small increments to avoid numerical problems, and the matrices and displacements are calculated iteratively at each time step until the results converge.

Once a time-domain response has been calculated, a Fourier transform can be used to calculate the frequency response if the system is linear.

3.7.8 Conclusion

Many choices are available for finite-element software, with costs ranging from zero to thousands of dollars. Examples of well-supported free software are FEBio (mrl.sci.utah.edu), Code Aster (www.code-aster.org), CalculiX (www.calculix.de) and Cast3M (www-cast3m.cea.fr).

A finite-element model comprises geometrical information, material properties, boundary conditions and applied loads. In our review of middle-ear models (Funnell et al., 2013) we discuss the handling of geometry and material properties, as well as the issues of model verification and validation.

4. ACKNOWLEDGEMENTS

This work has been funded by the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council (Canada) and the Research Fund of Flanders (Belgium). We also thank J. Lauzière for valuable contributions to the writing.

5. REFERENCES

d'Alembert, J.-B. le R. (1749a). Recherches sur la courbe que forme une corde tenduë mise en vibration. *Mémoires de l'Académie Royale des Sciences et des Belles-Lettres de Berlin*, *3*, 214–219.

d'Alembert, J.-B. le R. (1749b). Suite des recherches sur la courbe que forme une corde tenduë, mise en vibration. *Mémoires de l'Académie Royale des Sciences et des Belles-Lettres de Berlin*, *3*, 220–249.

Alexander, H. (1968). A constitutive relation for rubber-like materials. *International Journal of Engineering Science*, *6*(9), 549–563.

Anderson, J., & Papachristodoulou, A. (2009). On validation and invalidation of biological models. *BMC Bioinformatics*, *10*(1), 132 (13 pp.).

Arruda, E. M., & Boyce, M. C. (1993). Three-dimensional constitutive model for the large stretch behavior of rubber elastic materials. *Journal of the Mechanics and Physics of Solids*, *41*(2), 389–412.

Beards, C. F. (1996). *Structural vibration: Analysis and damping*. London: Arnold.

Beer, G., Smith, I. M., & Duenser, C. (2008). *The boundary element method with programming: for engineers and scientists*. Wien: Springer.

Belevitch, V. (1962). Summary of the history of circuit theory. *Proceedings of the IRE*, *50*(5), 848–855.

Beranek, L. L. (1954). *Acoustics*. New York: McGraw-Hill.

Boyle, R. (1662). *New experiments physico-mechanical touching the spring of the air, and its effects* (Second edition, whereunto is added A defence of the authors explication of the experiments, against the objections of Franciscus Linus and Thomas Hobbes.). Oxford: H. Hall for Tho. Robinson.

Brebbia, C. A. (1984). The boundary element method in engineering practice. *Engineering Analysis*, Eng. Anal. (UK), *1*(1), 3–12.

Callen, H. B. (1960). *Thermodynamics: An Introduction to the Physical Theories of Equilibrium Thermostatics and Irreversible Thermodynamics*. New York: Wiley.

Canadell, J., Goossens, H., & Klumperman, B. (2011). Self-healing materials based on disulfide links. *Macromolecules*, *44*(8), 2536–2541.

Carroll, M. M. (2009). Must elastic materials be hyperelastic? *Mathematics and Mechanics of Solids*, *14*(4), 369–376.

Charlebois, M., Jirásek, M., & Zysset, P. K. (2010). A nonlocal constitutive model for trabecular bone softening in compression. *Biomech Model Mechanobiol*, *9*(5), 597–611.

Charlebois, M., Motallebzadeh, H., & Funnell, W. R. J. (2013). Visco-hyperelastic law for finite deformations: a frequency analysis. *Biomechanics and Modeling in Mechanobiology*, *12*(4), 705–715.

Chen, C.-T. C. (2010). *Signals and Systems: A Fresh Look*. Retrieved from http://www.ee.sunysb.edu/~ctchen/

Chizeck, H., Butterworth, E., & Bassingthwaighte, J. (2009). Error detection and unit conversion. *IEEE Eng. Med. Biol. Mag.*, *28*, 50–58.

Chondros, T. G. (2007). Archimedes (287–212 BC). In M. Ceccarelli (Ed.), *Distinguished Figures in Mechanism and Machine Science: Their Contributions and Legacies Part I. History of Mechanism and Machine Science* (Vol. 1). Springer. Retrieved from http://lib.myilibrary.com?ID=106633

Creten, W. L., & Van Camp, K. J. (1974). Transient and quasi-static tympanometry. *Scandinavian Audiology*, *3*(1), 39–42.

Dallos, P. J. (1966). On the generation of odd-fractional subharmonics. *Journal of the Acoustical Society of America*, *40*(6), 1381–1391.

Decraemer, W. F., Khanna, S. M., & Funnell, W. R. J. (1994). Bending of the manubrium in cat under normal sound stimulation. *Optical and Imaging Techniques in Biomedicine, Proceedings of SPIE* (Vol. 2329, pp. 74–84).

Decraemer, W. F., Maes, M. A., & Vanhuyse, V. J. (1980). An elastic stress-strain relation for soft biological tissues based on a structural model. *J. Biomech.*, *13*(6), 463–468.

Desai, C. S., & Abel, J. F. (1972). *Introduction to the finite element method: A numerical method for engineering analysis*. New York: Van Nostrand Reinhold Co.

Edgeworth, R., Dalton, B. J., & Parnell, T. (1984). The pitch drop experiment. *European Journal of Physics*, Eur. J. Phys. (UK), *5*(4), 198–200.

Egolf, D. P. (1977). Mathematical modeling of a probe-tube microphone. *J. Acoust. Soc. Am.*, *61*(1), 200–205.

Eiber, A., & Freitag, H.-G. (2002). On simulation models in otology. *Multibody System Dynamics*, *8*, 197–217.

Eiber, A., & Kauf, A. (1994). Berechnete Verschiebungen der Mittelohrknochen unter statischer Belastung. *HNO*, *42*(12), 754–759.

Euler Archive. (n.d.). E177 -- Decouverte d'un nouveau principe de Mecanique. *The Euler Archive*. Retrieved May 15, 2011, from http://www.math.dartmouth.edu/~euler/pages/E177.html

Euler, E. E., Jolly, S. D., & Curtis, H. H. (2001). The failures of the Mars Climate Orbiter and Mars Polar Lander: A perspective from the people involved. *Proceedings of Guidance and Control 2001* (pp. 01–074, 24 pp.).

Euler, L. (1748). *Introductio in analysin infinitorum* (Vols. 1-2, Vol. 1). Lausanne: Bousquet.

Euler, L. (1752). Découverte d'un nouveau principe de Mecanique. *Mémoires de l'Académie Royale des Sciences et des Belles-Lettres de Berlin*, *6*, 185–217.

Ferry, J. D. (1980). *Viscoelastic properties of polymers*. John Wiley and Sons.

Finlayson, B. A. (1972). *The method of weighted residuals and variational principles: With application in fluid mechanics, heat and mass transfer*. New York: Academic Press.

Friedland, B., Wing, O., & Ash, R. B. (1961). *Principles of Linear Networks*. New York: McGraw-Hill.

Fung, Y. C. (1993). *Biomechanics: Mechanical Properties of Living Tissues* (2nd ed.). New York: Springer.

Funnell, W. R. J. (1996). Low-frequency coupling between eardrum and manubrium in a finite-element model. *J. Acoust. Soc. Am.*, *99*(5), 3036.

Funnell, W. R. J., & Laszlo, C. A. (1982). A critical review of experimental observations on ear-drum structure and function. *ORL J. Otorhinolaryngol. Relat. Spec*, *44*(4), 181–205.

Funnell, W. R. J., Maftoon, N., & Decraemer, W. F. (2013). Modeling of middle-ear mechanics. In S. Puria, A. N. Popper, & R. R. Fay (Eds.), *The Middle Ear: Science, Otosurgery, and Technology*, Springer Handbook of Auditory Research (pp. 171–210). New York, NY: Springer.

Gasser, T. C., Ogden, R. W., & Holzapfel, G. A. (2006). Hyperelastic modelling of arterial layers with distributed collagen fibre orientations. *Journal of The Royal Society Interface*, *3*(6), 15–35.

Greaves, G. N., Greer, A. L., Lakes, R. S., & Rouxel, T. (2011). Poisson's ratio and modern materials. *Nature Mater*, *10*(11), 823–837.

Haefner, J. W. (2005). *Modeling Biological Systems: Principles and Applications* (2nd ed.). New York: Springer.

Hallpike, C. S. (1935). On the function of the tympanic muscles. *Proc Roy Soc Med*, *28*(3), 226–232.

Ho, A. K., Alsaffar, H., Doyle, P. C., Ladak, H. M., & Agrawal, S. K. (2012). Virtual reality myringotomy simulation with real-time deformation: Development and validity testing. *The Laryngoscope*, *122*(8), 1844–1851.

Holzapfel, G. A., Gasser, T. C., & Ogden, R. W. (2000). A new constitutive framework for arterial wall mechanics and a comparative study of material models. *Journal of Elasticity*, *61*(1–3), 1–48.

Holzapfel, G. A., & Ogden, R. W. (2010). Constitutive modelling of arteries. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, *466*(2118), 1551–1597.

Hooke, R. (1678). *Lectures de potentia restitutiva, or Of spring: Explaining the power of springing bodies. To which are added some collections viz. A description of Dr. Pappins wind-fountain and force-pump. Mr. Young's observation concerning natural fountains. Some other considerations concerning that subject, Captain Sturmy's remarks of a subterraneous cave and cistern. Mr. G.T. observations made on the Pike of Teneriff, 1674. Some reflections and conjectures occasioned thereupon. A relation of a late eruption in the Isle of Palma*. London: The Bell in St. Paul's Church-Yard.

Huang, S., Dong, W., & Olson, E. S. (2010). Subharmonics and auditory nerve tuning curves in gerbil. Presented at the 33rd Midwinter Res. Mtg., Assoc. Res. Otolaryngol., Anaheim.

Hudde, H., & Weistenhöfer, C. (1997). A three-dimensional circuit model of the middle ear. *Acta Acustica united with Acustica*, *83*(3), 535–549.

Hudde, H., & Weistenhöfer, C. (2000). Circuit models of middle ear function. *The Function and Mechanics of Normal, Diseased and Reconstructed Middle Ears* (pp. 39–58). Presented at the 2nd Int. Symp. Middle-Ear Mechanics in Research and Otolaryngology, Boston: Kugler.

Hudde, H., & Weistenhofer, C. (2006). Key features of the human middle ear. *ORL J. Otorhinolaryngol. Relat. Spec.*, *68*(6), 324–328.

Ihrle, S., Lauxmann, M., Eiber, A., & Eberhard, P. (2013). Nonlinear modelling of the middle ear as an elastic multibody system — Applying model order reduction to acousto-structural coupled systems. *Journal of Computational and Applied Mathematics*, *246*, 18–26.

Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence*, *117*(2), 277–296.

Jennings, N. R. (2001). An agent-based approach for building complex software systems. *Communications of the ACM*, *44*(4), 35–41.

Kirchhoff, G. R. (1845). Ueber den Durchgang eines elektrischen Stromes durch eine Ebene, insbesondere durch eine kreisförmige. *Annalen der Physik und Chemie*, *140*(4), 497–514.

Langer, J. (1999). Computing in physics: Are we taking it too seriously? Or not seriously enough? *Physics Today*, *52*(7), 11, 13.

Lanir, Y. (1979). A structural theory for the homogeneous biaxial stress-strain relationships in flat collagenous tissues. *Journal of Biomechanics*, *12*(6), 423–436.

Lanir, Y. (1983). Constitutive equations for fibrous connective tissues. *Journal of Biomechanics*, *16*(1), 1–12.

Lax, P. D. (1986). Mathematics and computing. *Journal of Statistical Physics*, *43*, 749–756.

Leissa, A. W. (2005). The historical bases of the Rayleigh and Ritz methods. *Journal of Sound and Vibration*, *287*(4–5), 961–978.

Marlow, R. S. (2003). A general first-invariant hyperelastic constitutive model. *Constitutive models for rubber III: Proceedings of the Third European Conference on Constitutive Models for Rubber* (pp. 157–160). London, UK.

Maxwell, J. C. (1867). On the dynamical theory of gases. *Phil Trans Roy Soc London*, *157*, 49–88.

McLeish, T. (2008). A tangled tale of topological fluids. *Physics Today*, *61*(8), 40–45.

Meyer, O. E. (1874). Theorie der elastischen Nachwirkung [Theory of elastic aftereffect]. *Ann Phys Chem*, *227*(1), 108–119.

Mooney, M. (1940). Theory of large elastic deformation. *Journal of Applied Physics*, *11*(9), 582–592.

Morse, P. M. (1981). *Vibration and Sound*. New York: American Institute of Physics for the Acoustical Society of America.

Motallebzadeh, H., Charlebois, M., & Funnell, W. R. J. (2013). A non-linear viscoelastic model for the tympanic membrane. *J Acoust Soc Am*, *134*(6), 4427.

Neubert, H. K. P. (1963). Simple model representing internal damping in solid materials. *Aeronautical Quarterly*, *14*(2), 187–210.

Newton, I. (1729). *The Mathematical Principles of Natural Philosophy*. (A. Motte, Tran.). London: printed for Benjamin Motte.

OED. (2011). *Oxford English Dictionary*. Oxford University Press. Retrieved from www.oed.com

Ogden, R. W. (1982). Elastic deformations of rubberlike solids. *Mechanics of solids: The Rodney Hill 60th anniversary volume* (pp. 499–537). Oxford: Pergamon Press.

Ogden, R. W. (1984). *Non-linear elastic deformations*. Chichester, West Sussex, UK: Ellis Horwood.

Ogden, R. W. (2004). Mechanics of rubberlike solids. *Mechanics of the 21st Century* (pp. 263–274). Presented at the XXI International Congress of Theoretical and Applied Mechanics, Warsaw: Springer.

Ogston, A. G., & Stanier, J. E. (1953). The physiological function of hyaluronic acid in synovial fluid; viscous, elastic and lubricant properties. *J. Physiol. (Lond.)*, *119*(2–3), 244–252.

Onchi, Y. (1949). A study of the mechanism of the middle ear. *J. Acoust. Soc. Am.*, *21*(4), 404–410.

Osguthorpe, J. D., & Lam, C. (1981). Methodologic aspects of tympanometry in cats. *Otolaryngol. Head Neck Surg.*, *89*(6), 1037–1040.

Palmer, F. (1949). What about friction? *American Journal of Physics*, *17*(4), 181–187.

Pascal, B. (1937). *The physical treatises of Pascal: the equilibrium of liquids and the weight of the mass of the air, Pascal, Blaise, 1623-1662.* (I. H. B. Spiers & A. G. H. Spiers, Trans.). New York: Columbia Univ. Press.

Peña, E., Peña, J. A., & Doblaré, M. (2009). On the Mullins effect and hysteresis of fibered biological materials: A comparison between continuous and discontinuous damage models. *International Journal of Solids and Structures*, *46*(7–8), 1727–1735.

Poynting, J. H., & Thomson, J. J. (1902). *A text-book of physics: Properties of matter*. London: C. Griffin.

Rabbitt, R. D., & Holmes, M. H. (1986). A fibrous dynamic continuum model of the tympanic membrane. *J. Acoust. Soc. Am.*, *80*(6), 1716–1728.

Rabbitt, R. D., & Holmes, M. H. (1988). Three-dimensional acoustic waves in the ear canal and their interaction with the tympanic membrane. *J. Acoust. Soc. Am.*, *83*(3), 1064–1080.

Rayleigh, L. (Strutt, J. W.). (1877). *Theory of Sound* (Vols. 1-2, Vol. 1). London: Macmillan.

Reiner, M. (1954). Theoretical rheology. *Building materials: Their elasticity and inelasticity* (p. xv + 560 pp.). Amsterdam: North-Holland.

Reiner, M. (1964). The Deborah Number. *Phys. Today*, *17*(1), 62.

Reiner, M. (1971). *Advanced rheology*. London: Lewis.

Richmond, S. A., Kopun, J. g., Neely, S. T., Tan, H., & Gorga, M. P. (2011). Distribution of standing-wave errors in real-ear sound-level measurements. *J. Acoust. Soc. Am.*, *129*(6), 3134–3140.

Rivlin, R. S. (1948a). Large elastic deformations of isotropic materials. I. Fundamental concepts. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *240*(822), 459–490.

Rivlin, R. S. (1948b). Large elastic deformations of isotropic materials. IV. Further developments of the general theory. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *241*(835), 379–397.

Rivlin, R. S., & Saunders, D. W. (1951). Large elastic deformations of isotropic materials. VII. Experiments on the deformation of rubber. *Philosophical Transactions of the Royal Society of London. Series A (Mathematical and Physical Sciences)*, *243*(865), 251–288.

Sandifer, E. (2007, August). e, π and i: Why is “Euler” in the Euler identity? Mathematical Association of America. Retrieved from http://www.maa.org/news/howeulerdidit.html

Schiessel, H., Metzler, R., Blumen, A., & Nonnenmacher, T. F. (1995). Generalized viscoelastic models: their fractional equation with solutions. *J. Phys. A: Math. Gen.*, *28*, 6567–6584.

Schmitz, A., & Piovesan, D. (2016). Development of an Open-Source, Discrete Element Knee Model. *IEEE Transactions on Biomedical Engineering*, 1–1.

Silvester, P. (1969). High-order polynomial triangular finite elements for potential problems. *Int. J. Engng. Sci.*, *7*, 849–861.

Šmilauer, V., Catalano, E., Chareyre, B., Dorofeenko, S., Duriez, J., Dyck, N., Elias, J., Er, B., … Yuan, C. (2015). Yade Documentation 2nd ed. Retrieved from https://doi.org/10.5281/zenodo.34073

Sobotka, Z. (1984). *Rheology of materials and engineering structures*. Amsterdam: Elsevier.

Stakgold, I. (1968). *Boundary Value Problems of Mathematical Physics* (Vol. II). New York: Macmillan.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, New Series, *103*(2684), 677–680.

Tang, H., Buehler, M. J., & Moran, B. (2009). A constitutive model of soft tissue: From nanoscale collagen to tissue continuum. *Ann Biomed Eng*, *37*(6), 1117–1130.

Tanner, R. I., & Walters, K. (1998). *Rheology: An Historical Perspective*. Amsterdam: Elsevier.

Terhardt, E. (2000, March 10). Electromechanic analogy. *Electro-mechanic analogy*. Retrieved June 1, 2011, from http://www.mmk.ei.tum.de/persons/ter/top/emanalogy.html

Thomson, W. (1878). Elasticity. *The encyclopaedia britannica; A dictionary of arts, sciences, and general literature* (Ninth., Vols. 1-25, Vol. 7, pp. 796–825). New York: Scribner.

Vanpeperstraete, P. M., Creten, W. L., & Van Camp, K. J. (1979). On the asymmetry of susceptance tympanograms. *Scandinavian Audiology*, *8*, 173–179.

Veronda, D. R., & Westmann, R. A. (1970). Mechanical characterization of skin—Finite deformations. *Journal of Biomechanics*, *3*(1), 111–122, IN9, 123–124.

Voigt, W. (1892). Ueber innere Reibung fester Körper, insbesondere der Metalle. *Ann. Phys. Chem.*, *283*(12), 671–693.

Volandri, G., Di Puccio, F., Forte, P., & Manetti, S. (2012). Model-oriented review and multi-body simulation of the ossicular chain of the human middle ear. *Med Eng Phys*, *34*(9), 1339–1355.

Wave on a String. (2011, April 26). *PhET Interactive Simulations*. Retrieved from http://phet.colorado.edu/sims/wave-on-a-string/wave-on-a-string_en.html

Weiss, J. A., Maker, B. N., & Govindjee, S. (1996). Finite element implementation of incompressible, transversely isotropic hyperelasticity. *Computer Methods in Applied Mechanics and Engineering*, *135*(1–2), 107–128.

West, J. B. (1999). The original presentation of Boyle's law. *J. Appl. Physiol*, *87*(4), 1543–1545.

West, J. B. (2005). Robert Boyle's landmark book of 1660 with the first experiments on rarified air. *J. Appl. Physiol*, *98*(1), 31–39.

White, R. J., Peng, G. C. Y., & Demir, S. S. (2009a). Multiscale modeling of biomedical, biological, and behavioral systems (Part 1) [Introduction to the special issue]. *IEEE Engineering in Medicine and Biology Magazine*, *28*, 12–13.

White, R. J., Peng, G. C. Y., & Demir, S. S. (2009b). Multiscale modeling of biomedical, biological, and behavioral systems (part 2) [Introduction to special issue]. *IEEE Engineering in Medicine and Biology Magazine*, *28*, 8–9.

Wietor, J., & Sijbesma, R. P. (2008). A self-healing elastomer. *Angew Chem Int Ed*, *47*(43), 8161–8163.

Williams, J. R., Hocking, G., & Mustoe, G. G. W. (1985). Theoretical basis of the discrete element method. *Proc. NUMETA '85 Conference*. Presented at the MUMETA 85, Numerical Methods in Engineering: Theory and Applications, Swansea.

Wineman, A. S. (2009). Nonlinear viscoelastic solids :A review. *Mathematics and Mechanics of Solids*, Math. Mech. Solids (USA), *14*(3), 300–366.

Winsberg, E. (2010). *Science in the age of computer simulation*. Chicago: University of Chicago Press.

Young, T. (1807). *A course of lectures on natural philosophy and the mechanical arts.* (Vols. 1-2, Vol. 1). London: Johnson.

Zener, C. (1948). *Elasticity and Anelasticity of Metals*. Chicago: University Press.

Zienkiewicz, O. C., & Cheung, Y. K. (1967). *The Finite-Element Method in Structural and Continuum Mechanics*. London: McGraw-Hill.

Zienkiewicz, O. C., Irons, B. M., Ergatoudis, J., Ahmad, S., & Scott, F. C. (1969). Iso-parametric and associated element families for two- and three-dimensional analysis. In I. Holand & K. Bell

(Eds.), *Finite Element Methods in Stress Analysis* (pp. 383-432 of xiv+500 pp). Trondheim: Tapir.

Zwislocki, J. (1957). Some impedance measurements on normal and pathological ears. *J. Acoust. Soc. Am.*, *29*(12), 1312–1317.

Zwislocki, J. (2002). *Auditory sound transmission an autobiographical perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.

* Originally published 2012 Jul 6. Revised 2013 Aug 17: Added missing *i* in equation on p. 12; updated reference to Funnell et al. (2013) [was 2012]; changed formatting of Eqn. 20; corrected some line spacing in References. 2013 Oct 15: Minor formatting changes. 2014 Nov 17: Added text about mass-spring models and replaced Ho reference by a later one. 2016 Oct 4: Expanded text about rigid-body (now using term ‘multibody’), discrete-element & agent-based models. 2016 Oct 16: Expanded text about non-linear viscoelasticity, and added text about pseudo-travelling waves. 2017 Jul 23: Changed ‘block’ to ‘rod’ in §2.7. 2017 Aug 22: Added Neo-Hookean in §2.11.3; cosmetic changes in some equations. 2017 Oct 17: Added mention of transverse isotropy in §2.11.5. 2018 Jan 21: Use of italics for additional terms in §3.3. 2019 Nov 4: updated reference to Funnell et al. (2013); added reference to Greaves et al. 2020 Sep 1: rewording re air as a spring; added explicit eqn. no. in reference to eqn. 1; rewording re homogeneous eqns and natural frequencies; fixed formatting of affiliations; fixed a weird font. 2021 May 23: added ‘Two- and’ to caption of Figure 4; repositioned Figure 8; changed list of references to single spacing. 2021 May 24: use of ‘nospace’ for $\mathrm{d}/\Delta/\partial/\nabla$ in equations. 2021 May 25: minor changes in §2.5.3. 2023 Feb 11: replaced Figure 5. 2023 Nov 25: Added notes in §2.5.3 that the analyses of Maxwell and Voigt were for gases and metals, respectively. 2024 Nov 2: removed poor analogy from §2.3; created separate section for damage in §2.5, added mention of healing; miscellaneous formatting changes; revised Figure 3 and related text