Physics of diagnostic methods

This chapter discusses physics-based diagnostic technology—especially non-invasive methods. We begin with the classic method of auscultation, that is, listening to the characteristic sounds made by internal organs.

1. Auscultation

Doubtless the doctor-priests of the Greek healing shrines of Æsklepios, as well as the Egyptian priests of Thoth knew thousands of years ago that certain body sounds are normal whereas others signal the presence of disease. However, the science of auscultation did not come into its own until the invention of the stethoscope by René Laënnec[1] at the beginning of the 19th Century.

The first stethoscope was a hollow wooden tube with a bell-shaped end that was pressed against the chest or abdomen of the patient. It may have been invented to spare the modesty of female patients as well as improving the hygeine of the procedure. However, Laënnec realized that it provided the more important benefits of reducing background noise and amplifying the sound from the patient, thereby greatly increasing the chance of a correct diagnosis.

In particular the stethoscope made possible the diagnosis of such heart malfunctions as arrythmias, inflammation of the pericardium and damaged valves. The latter manifest themselves through "murmurs", resulting from the backflow of blood from the aorta to the left ventricle (or from the ventricle back into the atrium) through an incompletely closed valve.

The sounds produced in the air tubes and alveolar sacs of the lungs during breathing are modified by the presence of fluid, phlegm or inflammation, and can be detected by auscultation.

The technique of *percussion* was devised in 1761 by a Viennese physician, Leopold Auenbrugger. The son of a publican, Auenbrugger applied the method he had learned as a boy, of gauging the contents of a cask by tapping it and listening to the resultant echos, to the detection of fluid and other abnormalities in the chest cavity. This method, widely ignored during his lifetime, was greatly improved by the invention of the stethoscope.

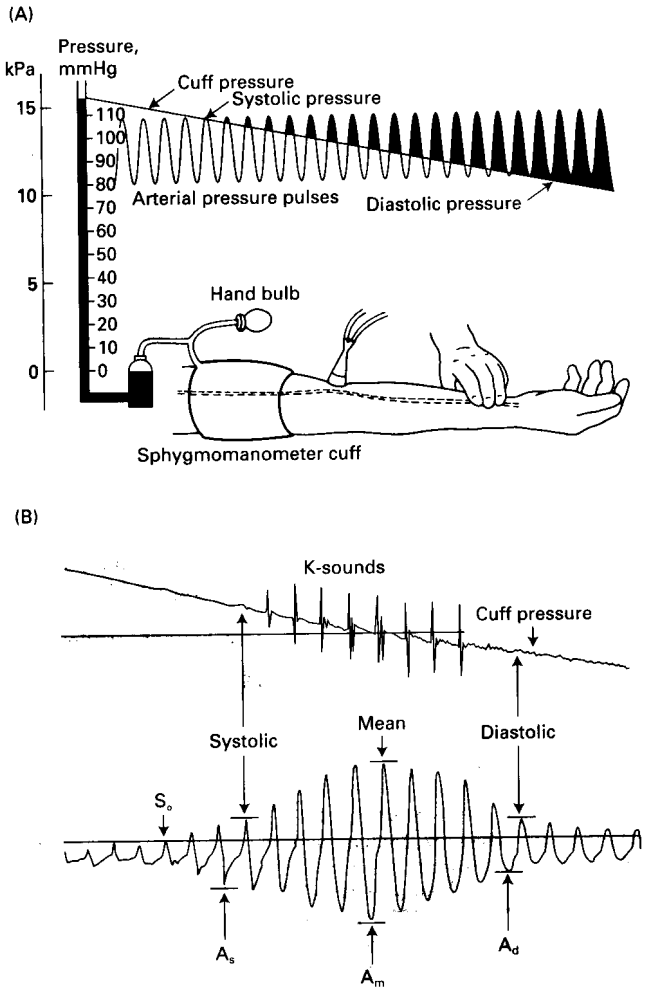
The standard binaural stethoscope was invented by G.P. Cammann, a New York physician, in the early 20th Century. It has survived with only minor improvements into the 21st Century. Today, however, electronically amplified models of the binaural stethoscope are widely available.

1. born Feb. 17, 1781, Quimper, France; died Aug. 13, 1826, Kerlouanec, France.

2. Blood pressure

Numerous pathological conditions are signalled by a change in blood pressure. Since there is a moderate range of blood pressures within a population of healthy individuals, and since a given healthy individual's pressures can vary widely based on time of day, activity level, current diet and psychological state[2], in order that blood pressure measurements have diagnostic value the physician must know the patient's normal levels under similar circumstances. This is one of the main reasons that healthy individuals should subject themselves to annual or biennial physical examinations.

Blood pressure is traditionally measured in millimeters of mercury[3]. The most precise blood pressure readings are obtained by inserting catheters into arteries and veins. However, the *sphygmomanometer*[4], shown to the right is accurate enough for many purposes and completely non-invasive. It was invented in 1896 by Italian physician Scipione Riva-Rocci, who combined an inflatable cuff with a mercury manometer. One records both the peak pressure produced by the heartbeat (*systolic* pressure) as well as the resting pressure (*diastolic* pressure). The measurement involves compressing a major artery (most often the brachial artery) until no pulse is heard by a stethoscope placed downstream (that is, in the portion of the artery distal to the point of compression). The external pressure required to do this is the systolic pressure. The cuff pressure is gradually reduced while listening to the pulse. Since the artery is only partially compressed, one hears a pulse that gets louder, then ceases. The pressure at the second cessation of sound is the diastolic pressure.



Elevated blood pressure—especially diastolic pressure—can signal the presence of various pathological conditions, generally related to narrowing of the arteries, or reduction of elasticity of the arterial walls. For example, some renal malfunctions cause elevated blood pressure. Excessive systolic pressure can cause hemorrhages in the brain, leading to one form of stroke.

2. For example, many patients exhibit "white coat syndrome" wherein having one's pressures measured by a nurse or assistant, in the venue of a doctor's office or hospital emergency room, can raise the systolic blood pressure by 10 mm or more.
3. The original pressure gauges employed columns of mercury, akin to barometers.
4. …from the Greek word *sphygmo-* ("of the pulse") and the French word *manometre* ("pressure meter").

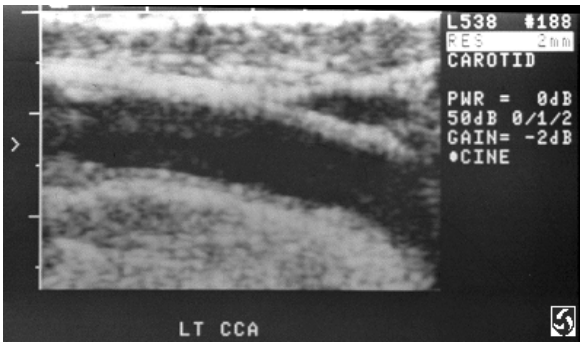
3. Ultrasound imaging

The use of short wavelength sound waves to image internal structures non-invasively has become one of the staple tools of modern medicine. This is an example of "spin-off"—the application of a technology developed for specialized purposes to a broader market. Ultrasound imaging is an outgrowth of sonar[5], developed originally to locate submarines in naval warfare; then applied to locating fish schools and depth sounding; and finally, to medical diagnosis.

The typical speed of sound in the body is about that in water, 1500 m/sec. In order to distinguish two features that are spatially separated by a distance D, any visualization method using waves must employ waves of wavelength $\lambda \leq D$. Thus, to "see" structures about 2 mm in size, the sound must have a frequency

$$f \approx \frac{u_s}{D} = 7.5\times10^5 \text{ Hz}.$$

Below we exhibit an example ultrasound image[6] of a human carotid artery at about 2 mm resolution,



exhibiting about 80% blockage.

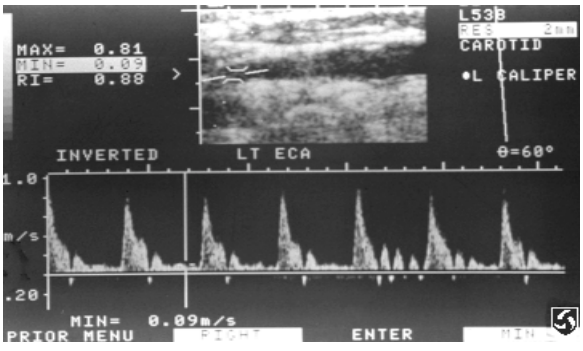
In order to examine a structure at a given depth with the body, the ultrasounic transducer emits brief pulses of sound—rather the way a bat echolocates its prey—and times the reflected pulses. By increasing the time interval before the receiver can "hear" the echo, the operator increases the round trip distance the sound has traveled, thereby increasing the depth. In this way, three-dimensional visualizations can be created.

A major virtue of ultrasound imaging is that sound reflected from a moving structure will have its frequency shifted up or down, according as the structure is moving toward or away from the receiver. This phenomenon is called the Doppler effect. The relative magnitude of the shift is

$$\frac{\Delta f}{f} \approx \frac{-\hat{r}\cdot\vec{v}}{u_s}$$

where $\hat{r}$ is the unit vector pointing from receiver to moving object, $\vec{v}$ is its velocity, and u_s is the speed of sound. (Thus an object moving toward the receiver has a *negative* velocity and so the relative change of pitch is upward.)

In particular, the Doppler effect permits the radiologist to measure the speed of flowing blood, or the speed of moving walls or valves in the heart. These are useful in diagnosing stenoses, as well as loss of flexibility of cardiovascular tissues. The figure below is an ultrasound scan of a carotid artery with superposed Doppler information exhibiting the markedly increased flow velocity near a stenosis.



5. Sonic navigation and ranging.
6. That is, the sound echos have been computer processed and displayed as an image.

4. Magnetic resonance imaging[7]

The phenomenon of nuclear magnetic resonance (NMR) has been known to physicists since the 1940's. Its application to medical diagnosis has been much more recent, beginning in the late 1980's. Much of the detailed technology that makes it possible to use NMR for noninvasive medical imaging lies in areas far beyond the scope of this course, specifically the development of suitable numerical algorithms and computers that can carry them out rapidly enough to convert large masses of NMR data into visual images of anatomical structures. However, the basic ideas are simple enough to explain here.

Atomic nuclei are bound states of neutrons and protons. Like electrons, these *nucleons* possess an intrinsic spin of ½ in units of the quantum of angular momentum, $h/2\pi$. Also like electrons, nucleons act like little dipole magnets. Isotopes with even numbers of each type of nucleon—that is, even-even nuclei—have intrinsic angular momentum zero, so their net magnetization also is zero. However, isotopes with an unpaired nucleon possess a net magnetization, that is, they can respond to an external magnetic field in the same manner that a compass needle responds to the Earth's magnetic field.

There are several pieces of physics necessary to understand how MRI works. First we recall Newton's Second Law of Motion,

$$\frac{d\vec{p}}{dt} = \vec{F}$$

describing the response of a particle to an external force acting on it. Here $\vec{p}$ is the momentum and $\vec{F}$ the force. To describe rotations, we consider an axis of rotation, passing through the origin of coordinates. Let $\vec{r}$ be the radius vector from the origin to the particle; then taking the vector product of $\vec{r}$ with both sides of Newton's Law gives

$$\vec{r}\times\frac{d\vec{p}}{dt} \equiv \frac{d}{dt}\left(\vec{r}\times\vec{p}\right) = \vec{r}\times\vec{F} \overset{df}{=} \vec{N}\,.$$

The vector product $\vec{r}\times\vec{p}$ is given the name *angular momentum*, and designated by the letter $\vec{J}$; the corresponding term

$$\vec{N} = \vec{r}\times\vec{F}$$

is called *torque*. (If we are discussing a rigid body we simply treat it as a collection of point masses, each acted on by a torque, and add up all the angular momenta and torques to get the total.)

So the rotational analogues of Newton's First and Second Laws are first, that, in the absence of external torques the (vector) angular momentum of a body or a collection of bodies remains constant in time. And conversely, the time rate of change of the angular momentum is equal to the (vector) torque.

Now we can apply this to an atomic nucleus with an intrinsic angular momentum and magnetic dipole moment $\vec{\mu}$, in an external magnetic field $\vec{B}$. First we note that the torque exerted on $\vec{\mu}$ by a magnetic field is[8]

$$\vec{N} = \vec{\mu}\times\vec{B}\,;$$

next we note that the magnetic moment of a quantum system is always proportional to its angular momentum:

$$\vec{\mu} = \kappa\vec{J}\,.$$

7. The dropping of the leading N from the acronym NMRI (nuclear magnetic resonance imaging) was done to assuage the fears of patients that they would be irradiated by ionizing radiation. That is, the present name, MRI, is pure PR.

8. This is most easily seen by pretending the magnetic dipole $\mu = Ga$ consists of equal and opposite magnetic charges $\pm G$, located at $\pm\mathbf{a}/2$ in an external magnetic field $\mathbf{B}$.

Then the equation of motion for the angular momentum becomes

$$\frac{d\vec{J}}{dt} = \vec{\mu}\times\vec{B} = \kappa\vec{J}\times\vec{B}\ .$$

The first thing we conclude is that the magnitude of $\vec{J}$ is a constant of the motion:

$$\vec{J}\cdot\frac{d\vec{J}}{dt} \equiv \frac{1}{2}\frac{dJ^2}{dt} = \kappa\vec{J}\cdot\vec{J}\times\vec{B} = 0\ ;$$

that is,

$$J^2 = \vec{J}\cdot\vec{J} = \text{constant}\ .$$

Next, if we take the magnetic field to define the $\hat{z}$ axis, we can decompose the equation of motion into components:

$$\frac{dJ_x}{dt} = \Omega J_y$$

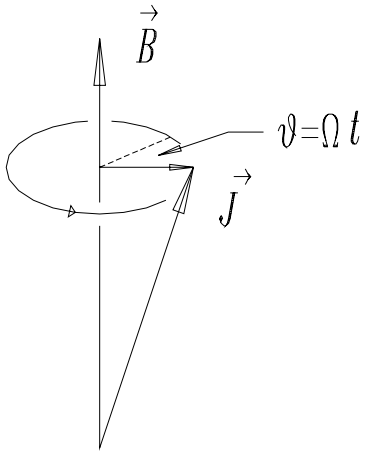
$$\frac{dJ_y}{dt} = -\Omega J_x$$

$$\frac{dJ_z}{dt} = 0\ ,$$

where $\Omega = \kappa B$ is called the *Larmor frequency*. We see that the z-component of the angular momentum is also a constant of the motion. Moreover, it is easy to see that

$$J_x = J_\perp \cos\Omega t$$
$$J_y = -J_\perp \sin\Omega t\ ,$$

or in other words, that the angular momentum vector consists of two parts: a constant projection along the direction of the applied magnetic field, and a perpendicular projection, whose length, $J_\perp$, remains constant, but whose direction rotates in the *x*–*y* plane with constant angular frequency Ω, as shown at the above right.



What is the Larmor frequency, quantitatively? For a free proton (and the human body contains a lot of those) we have (in Gaussian units)

$$\kappa = g_s \times 2.79 \times \frac{e}{2\,M_p\,c}$$

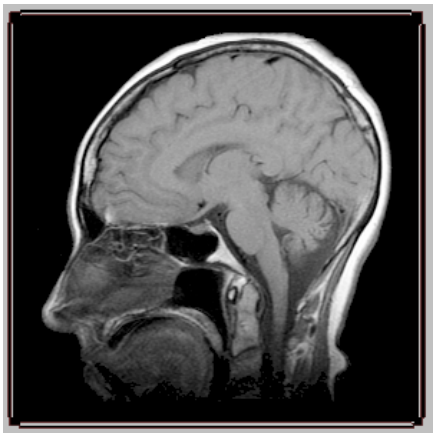
where e is the proton charge, 4.8×10^{-10} esu, M_p is the proton mass, and c is the speed of light. If the magnetic field is 10^3 Gauss, the ordinary frequency corresponding to the Larmor angular frequency is

$$f_L = \frac{\Omega}{2\pi} \approx 4\ \text{MHz}\ .$$

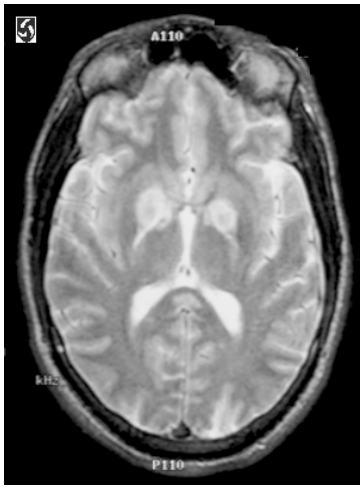
That is, it lies in the range of easily generated radio frequencies (RF). Thus if we superpose on the applied static magnetic field a small time-varying field at frequency f_L, a collection of nuclei in the same (static) magnetic field will tend to absorb energy resonantly. This phenomenon was, in fact, the technique originally used to measure the magnetic moments of many nuclei.

For us to apply NMR to MRI, that is, to the task of locating various isotopes within the body, and hence visualizing the internal structures of the body in three dimensions, the static magnetic field must vary spatially in three dimensions. (The variation must be known in advance, of course, for this method to work.) For technical reasons, the field must be rather large—4-6 Tesla (40,000–60,000 Gauss)—which means that it must be generated using superconducting magnets. The wavelength

of the applied RF field is still sufficiently long that it can be regarded as constant spatially (although sinusoidally varying in time) relative to the structures to be visualized. The MRI machine sweeps the frequency of the RF field over a range encompassing the Larmor frequency, and records the absorption as a function of frequency. Since a given frequency correlates with a small localized volume of the patient's body (by virtue of the known gradients of the static magnetic field), the degree of absorption reflects the spatial density



distribution of the particular isotopic species being scanned. Since different body structures contain different concentrations of—say—^{13}C, ^{17}O or protons, the NMR scan yields the information needed to visualize the interior of the body in three dimensions, with the choice of isotope providing "highlights" that emphasize regions of particular interest.



Yet another point to consider is the fact that healthy and diseased tissues exhibit subtle differences both in concentration and resonant frequency, of certain isotopes. Thus MRI permits the location of rather small tumors that cannot be visualized using other diagnostic procedures.

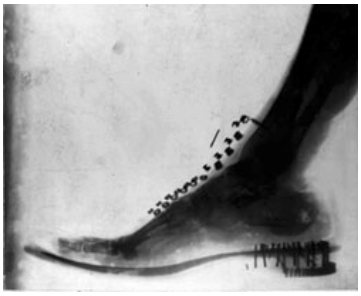
A collaboration between groups at Princeton University, SUNY Stony Brook, and The University of Virginia has demonstrated the feasibility of infusing patients with pre-aligned isotopes. That is, by externally aligning the spins (and hence the magnetic moments) of inert gases such as ^{3}He or ^{129}Xe, then adding them to the air in a patient's lungs, it has been possible to visualize the interior of the lungs with far better detail than any previous technique (other than dissection).

At present MRI remains a fairly expensive diagnostic procedure that HMOs are reluctant to pay for if there is an alternative method of equivalent quality. The reason for this is the complex electronics and superconducting magnet system, not to mention the elaborate software used to convert the raw data to a visual image in real time. An MRI system currently costs about 10–20 times more than a CT scanner or a lithotripter. Assuming a price of $\$10^7$ and a 10-year amortization period, it is not surprising that, assuming full-time utilization with no down-time (2,000 hours per year), the price of a 1-hour session of MRI with the services of a trained technician will run between $1,000 and $2,000.

However, one may anticipate that even without mass production, the major components of an MRI device will become more readily available and lower in price—especially if high temperature superconductors realize their present promise. That is, the cost of the machine will probably fall 10-fold within the coming decade. If this happens, not only will more medical centers acquire them, they will use MRI more routinely, since the greater part of their cost will become the time of the practitioner rather than amortization of a large initial investment.

5. X-rays and computerized tomography

Practically from the moment of their discovery X-rays have been used to visualize the interior of the human body.



Basically, X-rays are electromagnetic radiation whose wavelength lies in the range

$$1\ \text{Å} \leq \lambda \leq 10\ \text{Å}$$

They are generated by accelerating electrons in a vacuum, then letting them collide with a heavy metal target. This procedure generates both bremmstrahlung[9]—that is radiation emitted in consequence of the deceleration of the electrons—and line spectra caused by knocking out one of the innermost electrons of the target atoms.

Living tissues are nearly transparent to X-rays. Thus X-ray images of the interior of the body are basically shadow graphs, the shadows being cast by denser structures such as bone. The finite size of the emitting source limits the resolution of the X-ray image.

Several things can be done to improve the quality, and hence the diagnostic utility, of X-ray images. Since X-rays are more strongly scattered by heavy elements such as barium or iodine, radio-opaque dyes containing these materials can be placed within the organs being studied. The "barium meal", for example, is only too familiar to ulcer patients. Its ingestion renders opaque the upper gastrointestinal tract. Conversely, the colon can be studied by filling it (from below!) with a barium-containing paste.

The kidneys and urinary tract can be visualized in detail using an iodine-rich dye that is chemically tailored to accumulate rapidly in the kidneys and be secreted into the urine after injection into the bloodstream.

Finally, the angiogram is used to visualize the coronary arteries or sometimes the carotids. Here a dye is injected directly into the blood entering the artery under study using a catheter. The catheter is a long, flexible tube that is passed into the aorta *via* a portal inserted into either the femoral or brachial artery. The patient is given anticoagulents to reduce the risk of forming a clot that might lodge in the lungs or brain. The procedure is moderately risky in that between 0.1% and 1% of patients experience serious adverse effects, including stroke, heart attack or death.

Computerized tomography (CT scan) is an outgrowth of techniques originally developed to study the structure of matter in physics experiments. Rather than using a film or fluoroscope to convert the X-rays to a visible image, the X-rays are emitted from a highly collimated source and detected using a high resolution solid state detector. The source

9. German for "stopping radiation".

is moved around the patient's body in a known circular or spiral path, and the X-rays from the different source locations are detected continuously. Since what the source is emitting is known precisely, it is possible to reconstruct with reasonable precision the three-dimensional density distribution of the intervening matter. The mathematical algorithms for doing this are somewhat involved and require considerable computational capacity.

The reason the process is called "tomography" is that, having determined the 3D distribution of matter from the data, the computer can then be used to construct an image representing what could be seen if the body had been sliced along a given plane[10]. The orientation of the plane is of course controllable by the radiologist.

To recapitulate, the basic physics that makes it possible for X-rays to be used to diagnose conditions ranging from broken bones and dental caries, to tumors, arteriosclerosis and other ills flesh is heir to, is their ability to penetrate material composed of the lighter elements with little scattering. That is, since only a small fraction of the X-ray energy is absorbed in hydrogen and we are about 63% hydrogen[11] by elemental abundance, we can tolerate sufficiently large doses to permit accurate imaging with relatively small risk of developing cancer or radiation burns. The rest of X-ray diagnostic technique—opaque dyes, CT scans, *etc.* is the result of technological developments in solid state physics, applied mathematics and computer engineering.

6. Radioactive tracers

Both naturally occurring and artificially produced radioactive isotopes—but mainly the latter—have proven extremely valuable both in biological research and in diagnosis. Here we dwell on the latter application.

Some chemical elements have a special affinity for particular areas of the body; for example, technetium tends to concentrate selectively (and rapidly!) in the cardiac muscle. Thus, by administering small amounts of the isotope ^{99}Tc, whose half life is 6 hours, and whose primary decay mode is the emission of a 134 KeV γ ray, cardiologists can study the blood flow to various parts of the heart by detecting the emitted γ rays. The usual protocol involves first measuring Tc uptake by the resting heart; then the patient exercises on a treadmill until his pulse reaches a predetermined rate, at which point he is again given the Tc tracer and rescanned.

The attenuation coefficient μ for γ rays of 134 keV energy in water (*i.e.* in the human body!) is about 0.15 cm^2/gm. Multiplying by the density of water (1 gm/cm^3) and by the thickness x of tissue that must be traversed by the γ rays (about 15 cm in a person 1 foot thick), and finally exponentiating, we find

$$I(x = 15 \text{ cm}) = I_0\, e^{-\mu\rho x} \approx 0.1\, I_0\,;$$

that is, only 10% escape the body to reach the detectors. This is the typical situation for radioactive tracers in the body.

10. The Greek word tomos means cut or slice (and in fact the Engl;ish word atom comes from *atomos*: "cannot be cut"—atoms are the irreducible building blocks of matter, according to Demokritos).
11. This proportion must of course be modified for politicians, whose bodies are believed to contain much greater fractions of hydrogen and other gaseous substances. ☺