Sound waves and the peripheral auditory system

David Meredith
Department of Computing,
City University, London.
dave@titanmusic.com
www.titanmusic.com

BMus/BSc in Music, Part Two Module: Music Perception and Cognition
Department of Music, City University, London
Tuesday, 6 May 2003.

1. Physical properties of sound waves (Moore, 1989, pp. 1–3; Whelan and Hodgson, 1978, pp. 91–94,98,296–297)

- Sound waves can be either
  - *periodic*, e.g., tuning fork, violin; or
  - *non-periodic*, e.g., sonic boom, hand-clap.
- Sound waves can be either
  - *progressive*, e.g., tuning fork, violin; or
  - *stationary*, e.g., inside sounding organ pipe.
- The *source* of a sound wave is a vibrating object.
- The *disturbance* in a sound wave is a variation in pressure and density of the medium.
- Sound waves are *mechanical waves*—they can only be transmitted through a material medium (not a vacuum).

1. Physical properties of sound waves (Moore, 1989, pp. 1–3; Whelan and Hodgson, 1978, pp. 91–94,98,296–297)

1. I'm David Meredith and I'm going to be giving you 10 lectures on music perception and cognition.
2. You should each have a copy of this 'extra information' sheet and I just want to make sure that you all understand everything on it.
3. [READ THROUGH SHEET AND ASK IF THERE ARE ANY QUESTIONS.]
4. [TAKE ROLL CALL AND TRY TO LEARN NAMES]
5. This week I'm going to review the physics of sound waves, and then I'm going to talk a bit about the anatomy and physiology of the auditory system.
6. As reading for today's lecture, I would recommend Chapters 2, 3 and 7 from Pierce's (1992) *The Science of Musical Sound*, Chapter 2 of Dowling and Harwood's (1986) *Music Cognition*, Chapter 1 of Moore's (1989) *An Introduction to the Psychology of Hearing* and Chapters 1 and 2 of Campbell and Greated's (1987) *The Musician's Guide to Acoustics*.
7. You don't have to read all of these. If you just look at the relevant bits from, say, two of the books, you should be fine.
8. Sound waves can be either *periodic* or *non-periodic*.
9. (a) A periodic sound wave is produced by a vibrating source such as a tuning fork which gives rise to a succession of disturbances called a *wavetrain*.
   (b) A non-periodic sound wave consists of just a single wavefront such as that which occurs in the shock wave emitted by an aircraft when it breaks the sound barrier or in the wave that results from a hand-clap.
10. Sound waves can be either *progressive* or *stationary*.

(a) A progressive wave (also known as a *travelling* wave) is one in which energy is transferred from one location (the source) to another. The wave generated by a tuning fork is an example of a progressive sound wave.

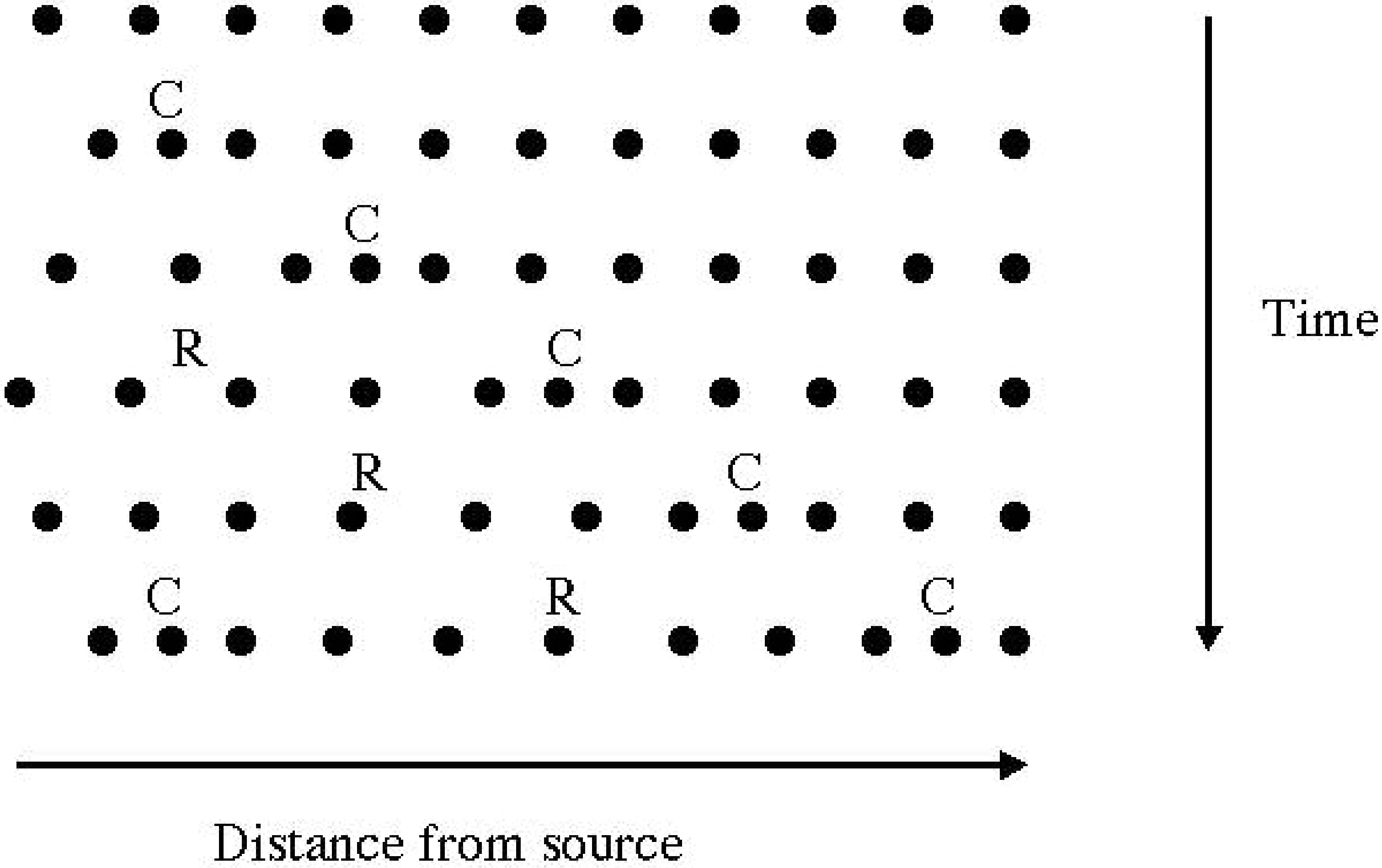
(b) In a stationary or *standing* wave, there is no *net* transfer of energy from one place to another. However, over time, the energy in a standing wave is repeatedly converted from kinetic energy into potential energy and back again. The sound wave *inside* a sounding organ pipe is an example of a standing or stationary sound wave.

11. A sound wave is usually produced by a vibrating object which sends out waves through the medium (usually air) in which it is immersed.

12. The *disturbance* in a sound wave is a variation in the pressure and density of the medium through which the wave passes.

13. A sound wave can therefore only be transmitted through a *material* medium which may be solid, liquid or gas. Sound waves are therefore mechanical waves and, unlike electromagnetic waves, they cannot be transmitted through a vacuum.

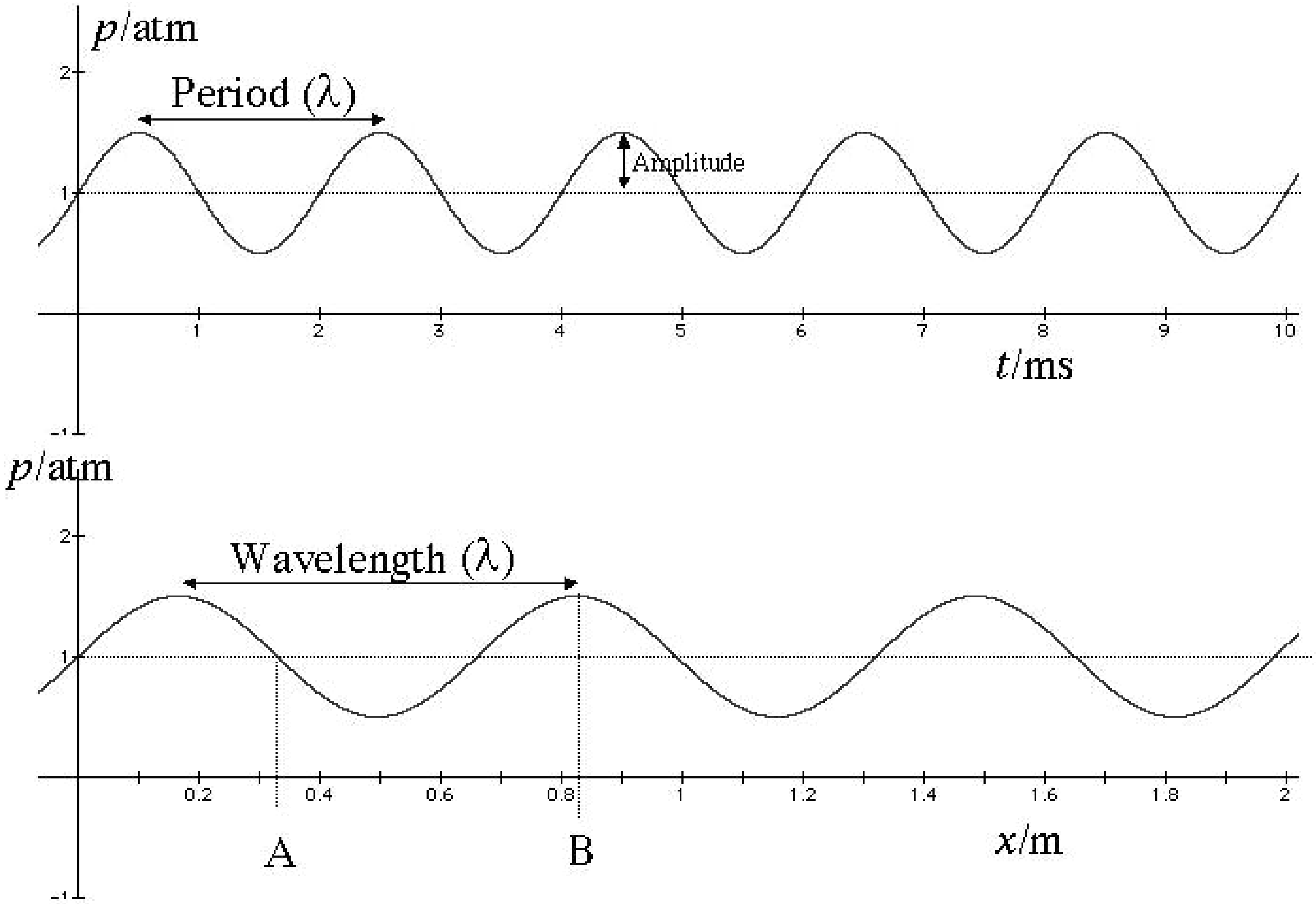
2. Oscillations of molecules in a sound wave



2. Oscillations of molecules in a sound wave

1. When a sound wave travels through a medium, what happens at a particular point in space is that the molecules in the medium are alternately squeezed together (compressed) and spread out (or rarefied).
2. This means that at a particular point in space in a sound wave, the pressure and density of the medium oscillate around their normal, undisturbed values because the molecules at that point are alternately more and less densely packed together than normal.
3. Even in the absence of a sound wave, gas molecules exhibit constant random motion. However, when a sound wave is present, an oscillatory motion is superimposed on this random motion.
4. In a sound wave, the direction of this oscillatory motion is parallel to the direction of propagation of the wave so sound waves are *longitudinal* mechanical waves.
5. You must remember, however, that the molecules in the medium *oscillate* around an average position and do not advance with the wave: it is the *disturbance* that is propagated, not the material of the medium itself.

3. Waveforms and simple tones



3. Waveforms and simple tones

1. The *waveform* of a wave is a graph showing either
   (a) the pressure variation at a particular point in space plotted against time (as shown in the top graph); or
   (b) the pressure variation at a particular instant in time plotted against position in space (as shown in the bottom graph).
2. One of the simplest types of sound wave is one whose waveform is sinusoidal, as shown here. Such a sound is called a simple tone or a pure tone.
3. Simple tones are not only simple mathematically, they also evoke particularly simple responses in the auditory system (Moore, 1989, p. 2).
4. This graph here shows the pressure variation at a point in space plotted against time for a simple tone.
5. The time that the pressure variation at a given point takes to complete one cycle is called the *period* of the wave and the number of periods completed per unit time is called the *frequency* of the wave.
6. The frequency of a periodic sound wave whose waveform is more complex than that of a simple tone is usually called its *periodicity*.
7. The frequency of a wave is usually expressed in hertz (Hz), 1 Hz being equal to 1 cycle per second.
8. The *amplitude* of a sound wave at a point is the difference between the mean pressure at that point and the maximum pressure at that point, as shown here in this graph. In other words, the amplitude of a wave is its *maximum* pressure variation.
9. Beware that in some texts (e.g., Moore, 1989, p. 2), the term 'amplitude' is incorrectly used to denote the pressure variation and the term 'maximum amplitude' is used to denote the maximum pressure variation (i.e., the amplitude).
10. This graph here shows the pressure variation at a particular instant in time plotted against position in space for a simple tone.

11. The *phase* at a given point in space and a given instant in time is the number of cycles through which the wave has advanced relative to some fixed point in time, expressed in angular measure. The phase is usually given in radians. A full cycle therefore corresponds to 2π radians, half a cycle corresponds to π radians and so on.

12. For example, in this waveform, the point B is lagging 3/4 of a cycle behind point A. Point B therefore exhibits a phase lag of $\frac{3}{4} \times 2\pi = \frac{3\pi}{2}$ relative to point A.

13. The distance between two consecutive points in the wave with the same phase (e.g., two consecutive wave crests) is called the *wavelength* of the wave.

14. The speed of a mechanical wave is the speed at which an observer would have to travel in order to see the wave as being stationary. The speed of a wave is equal to the number of wavelengths per unit time. It is therefore equal to the product of the wave's frequency and wavelength.

4. Power, intensity, pitch and tones

- *Power* of a wave is amount of energy transmitted per unit time.

$$\text{Power} \propto (\text{frequency})^2 \times (\text{amplitude})^2 \times \text{speed}$$

- *Intensity* is energy transmitted per unit time per unit area of the wavefront

$$\text{Intensity} = \frac{\text{Power}}{\text{Area of wavefront}}$$

- *Loudness* depends on intensity and sensitivity of listener to frequency of the sound.
- Periodic sounds with a periodicity between 20Hz and 5000Hz evoke sensation of pitch.
- Pitch is that perceptual attribute of a sound in terms of which it may be ordered on a musical scale.
- A *tone* is a sound that has a pitch.

4. Power, intensity, pitch and tones

1. A progressive wave such as the sound wave produced by a musical instrument transfers energy and momentum from the source to places around it.

2. The amount of energy transmitted by a wave in unit time is called the *power* of the wave.

3. The *power* of a wave is proportional to the square of the frequency, the square of the amplitude and the wave speed.

4. The *intensity* of a wave is the amount of energy transmitted by the wave perpendicularly to its wavefront in unit time per unit area. In other words, it is the power per unit area of the wavefront. The loudness of a sound depends upon the intensity of the sound and the sensitivity of the listener to the frequency of the sound.

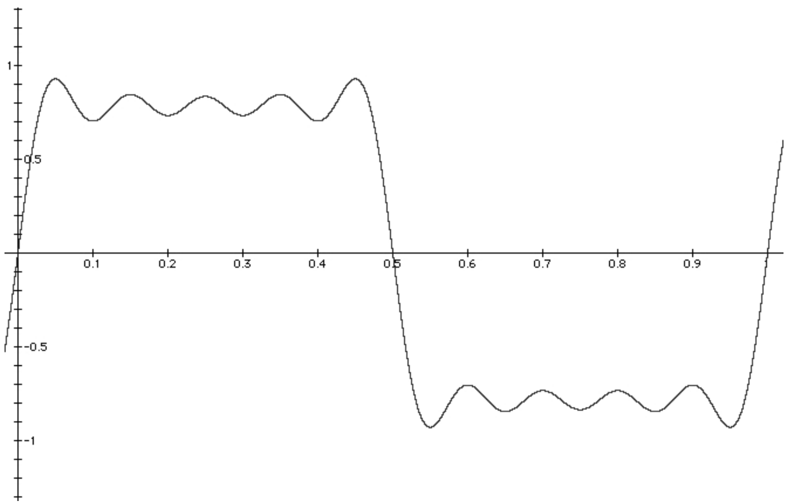
5. Any periodic sound wave with a frequency greater than 20Hz and less than about 5000Hz will evoke a sensation of pitch.

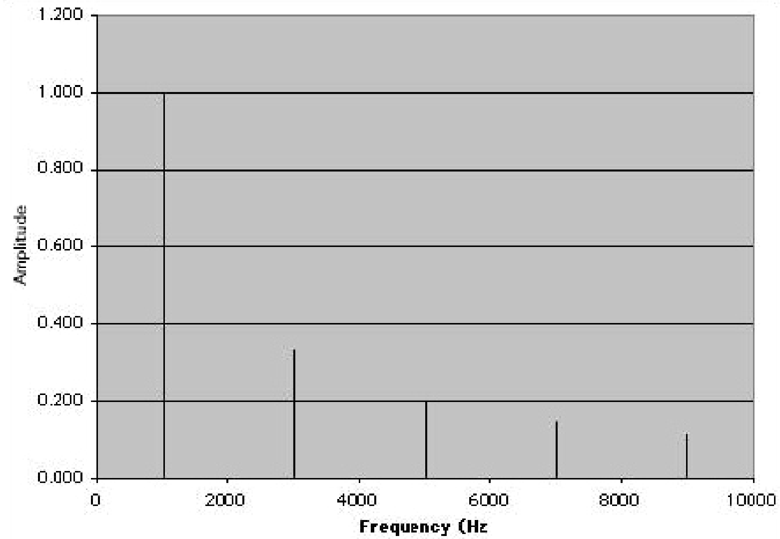
6. Moore (1989, p. 3) defines pitch to be 'that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale.'

7. Any sound that evokes a sensation of pitch is called a *tone*. Although most tones are periodic, not all of them are: some non-periodic tones are perceived to have pitch.

8. The pitch of a tone is related to its frequency or periodicity. The pitch of a simple tone is usually indicated by specifying its frequency in Hz. The pitch of any other tone S is usually indicated by giving the frequency of a simple tone whose pitch would be perceived to be the same as that of S.

5. Fourier analysis and spectral representations (Moore, 1989, pp. 3–7)

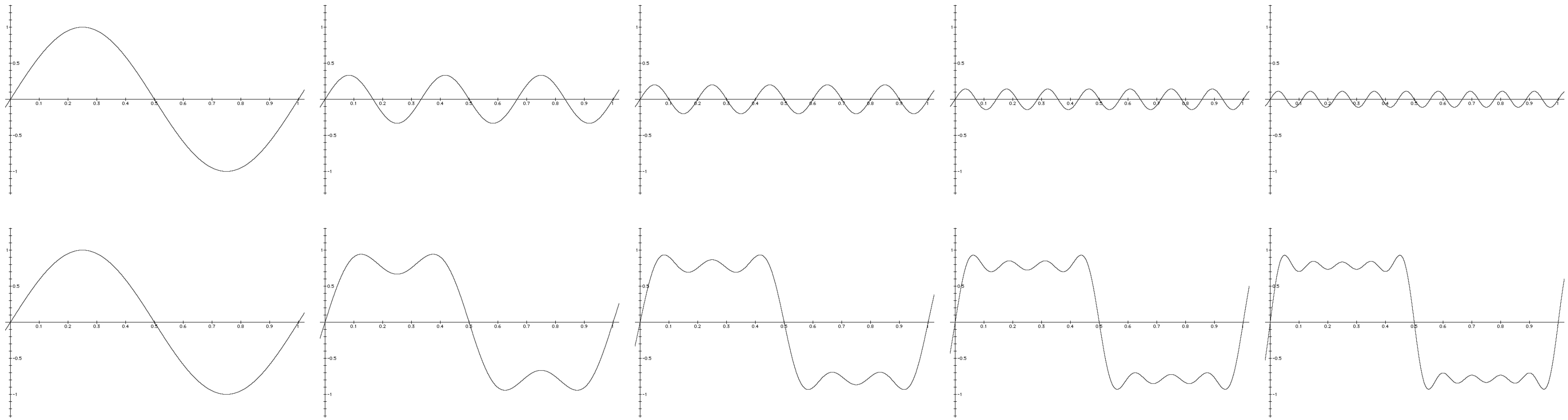




5. Fourier analysis and spectral representations (Moore, 1989, pp. 3–7)

1. Any sound can be represented by a waveform that plots pressure variation against time or spatial position.
2. However, when the sounds are more complicated than sinusoidal simple tones, it is usually more useful to represent the wave by its so-called *Fourier spectrum*.
3. This graph on the right shows the Fourier spectrum that corresponds to this waveform here on the left.
4. The Fourier spectrum is so-called because it is based on a theorem discovered by the French mathematician Jean Baptiste Joseph Fourier (1768–1830) who proved that any complex waveform (with certain restrictions) can be expressed as the sum of a series of sinusoidal simple tones with particular frequencies, amplitudes and phases.
5. Beware that Pierce (1992, p. 41) seems to think that Fourier analysis was invented by the French social reformer, Fran cois Marie Charles Fourier (1772–1837) who was an almost exact contemporary of the mathematician.
6. So this spectrum on the right tells us that we can construct this waveform on the left by adding together five simple tones whose frequencies are 1000Hz, 3000Hz, 5000Hz, 7000Hz and 9000Hz.
7. Moreover, it tells us that the relative amplitude of each of these sinusoidal simple tones must be as shown on the graph. That is, the tone at 3000Hz must have an amplitude that is 1/3 that of the tone at 1000Hz and so on.
8. When a complex waveform is expressed in this way it is called a *Fourier analysis* of the waveform and each sinusoidal simple tone in the summation is called a *Fourier component* or *partial* of the complex waveform.
9. When the frequencies of all the Fourier components of a complex tone are integer multiples of the frequency of the lowest component, the complex tone is called a *harmonic complex tone*, each of the Fourier components of the tone is called a *harmonic* and the lowest harmonic is called the *fundamental*.
10. The fundamental frequency of a harmonic complex tone is equal to the periodicity of the tone.
11. The Fourier component of a harmonic complex tone whose frequency is n times that of the fundamental frequency is called the nth harmonic of the tone. So, in this case, the component at 3000Hz is the third harmonic of the complex tone.

6. Constructing a harmonic complex tone

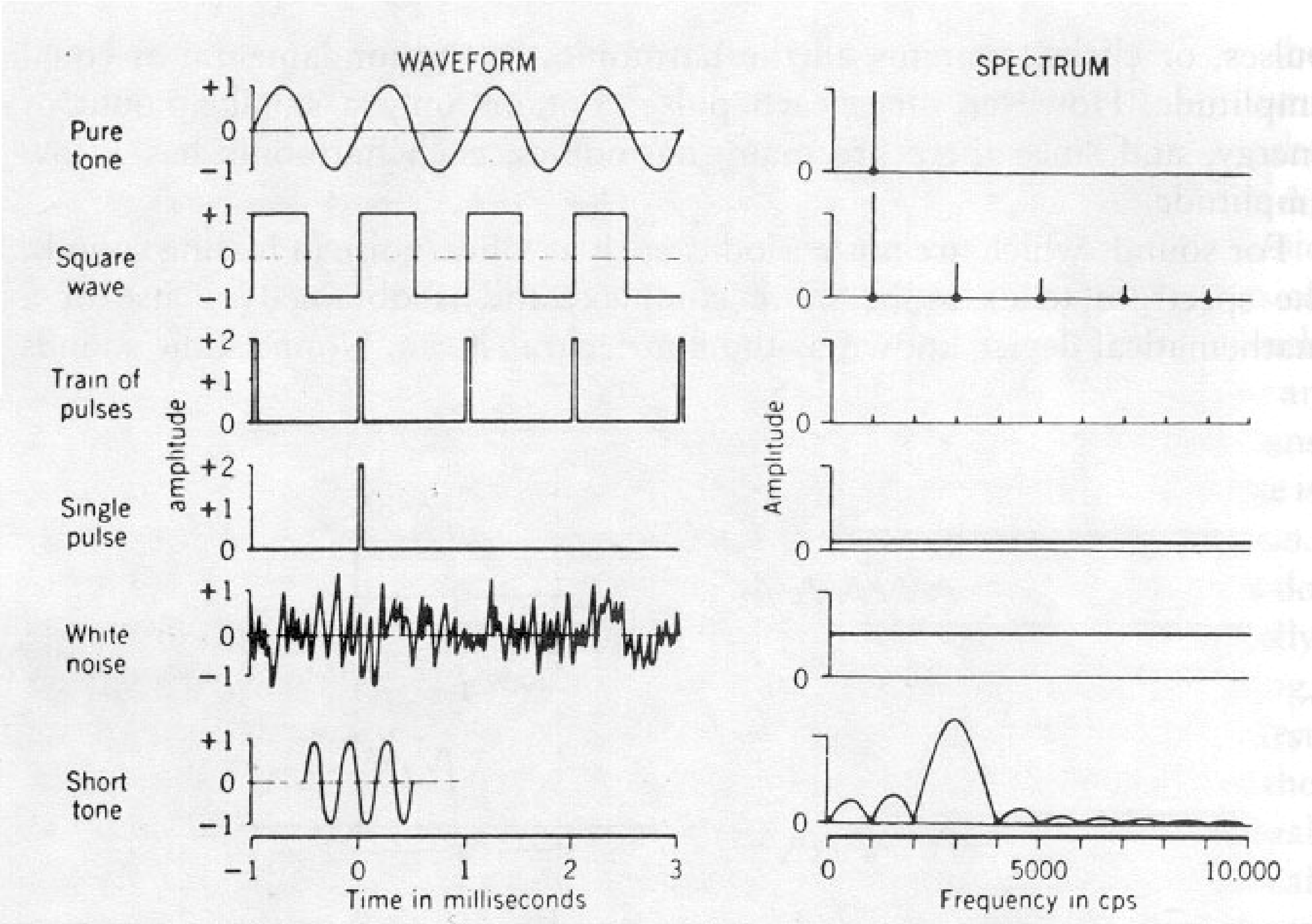


6. Constructing a harmonic complex tone

1. This figure shows how a harmonic complex tone can be built up by adding together sinusoidal components whose frequencies are all integer multiples of a fundamental frequency.
2. This particular example shows how a square wave can be built up by adding together odd-numbered harmonics of a fundamental, each harmonic having an amplitude that is inversely proportional to its harmonic number. So the third harmonic has an amplitude that is 1/3 of the amplitude of the fundamental, the fifth harmonic has an amplitude that is 1/5 of that of the fundamental and so on.
3. Note how all the harmonics in the diagram have the same phase at time zero.
4. Let's hear what each of these tones sound like.
5. [PLAY EXAMPLES OF BUILDING UP A SQUARE WAVE]
6. Fourier's theorem tells us that we can build up any complex tone by adding together the right combination of simple tones. This is very similar to the situation with light where we can produce light of any given colour by adding together different amounts of the primary colours.
7. Our visual system is incapable of performing the reverse process—that is, we cannot attend to just one of the component hues in a beam of light. If lights of two different frequencies (colours) are mixed, we see light of a single colour that corresponds to the mixture of frequencies it contains, we cannot choose to perceive each of the component hues separately.
8. However, when we listen to a complex harmonic tone we can choose either to hear the tone as a unitary percept with a particular timbre, loudness and pitch; *or* we can attend to each of the low harmonics of the tone individually.
9. For example, if I play you the third harmonic of this complex tone on its own and then I play you the complex tone, you should be able to focus your attention on the third harmonic in the complex tone.
10. Our ability to do this was observed by Mersenne in 1636 and formalized in Ohm's Acoustical Law.

11. Our ability to hear out the low harmonics of a complex tone implies that our auditory system performs a kind of Fourier analysis on the incoming sound signal.

7. Examples of Fourier spectra



7. Examples of Fourier spectra

1. We can represent a Fourier analysis of a complex sound by means of a *spectrum*, like the ones shown here.
2. The Fourier spectrum of a complex sound is a graph that shows the amplitude and frequency of each of the sinusoidal Fourier components of the sound. Note that a Fourier spectrum tells us nothing about the phases of the sinusoidal components. This information is usually represented separately in another type of spectrum called a *phase spectrum*.
3. Here we have six waveforms with their corresponding spectra taken from Moore (1989, p. 6).
4. Each of the top three spectra contains a set of isolated vertical lines indicating that the energy falls at highly specific, discrete frequencies. Such a spectrum is called a *line spectrum*. In fact, a sound only has a perfect line spectrum if it lasts forever!
5. So, an everlasting simple tone would have a perfect line spectrum containing just one vertical line at the frequency of the tone, as shown here.
6. A perfect and everlasting square wave would have a perfect line spectrum containing vertical lines at all the odd-numbered multiples of the fundamental frequency, with the amplitude of the nth harmonic being proportional to $1/n$, as shown here.
7. The third example shows a pulse train—that is, an infinite series of equally spaced clicks. This is a harmonic complex tone that contains harmonics at all multiples of the fundamental, all with the same amplitude. Note that because there is energy at every integer multiple of the fundamental, each harmonic only has a very small amplitude.
8. Both the brief click and the white noise (which sounds like a hiss) have continuous flat spectra indicating that they contain energy at all possible frequencies. However, the phase spectrum of the click is different from that of the white noise. For the white noise, the phase at which each component begins is random whereas for the click, all the components have phase $\pi/2$ at time zero. This means that at time zero, the pressure variation is at a maximum for all the frequency components.

9. The sixth example shows the spectrum for a simple tone that does not last forever. Such a tone is called a *tone pulse* or a *tone burst*. Note that whereas the everlasting simple tone has a line spectrum, the sinusoidal tone burst has a continuous spectrum, consisting of a peak at the frequency of the steady-state portion of the burst and surrounded by smaller and smaller peaks spreading out away from this frequency.

10. As the tone burst gets shorter and shorter, the spectrum gets more and more spread out until eventually you get the click here which has a completely flat spectrum.

8. The measurement of sound level (Moore, 1989, pp. 7–10)

- Intensity is energy transmitted per second perpendicularly through 1m^2 of the wavefront. That is,

$$\text{Intensity} = \frac{\text{Power}}{\text{Area of wavefront}}.$$

- Auditory system can deal with huge range of intensity (e.g., gunshot is 10 000 000 000 000 times intensity of quietest detectable sound).
- We generally use a *logarithmic* scale for intensity.
- If we have two sounds with intensities I_1 and I_2 then the *sound level* of I_1 is

$$\log_{10}(I_1/I_2) \text{ Bels}$$

  greater than I_2.
- For example, if $I_1 = 100I_2$ then

$$\log_{10}(I_1/I_2) = \log_{10}(100) = 2 \text{ Bels}.$$

- If we have two sounds with intensities I_1 and I_2 then the *sound level* of I_1 is

$$10\log_{10}(I_1/I_2) \text{ decibels (dB)}$$

  greater than I_2.
- For example, if $I_1 = 100I_2$ then

$$10\log_{10}(I_1/I_2) = 10\log_{10}(100) = 20 \text{ dB}.$$

- Increase in level of 10 dB corresponds to multiplying intensity by 10.
- Increase in level by 3 dB corresponds to doubling intensity.
- If $I_1 = I_2/10$ then

$$10\log_{10}(I_1/I_2) = 10\log_{10}(.1) = -10 \text{ dB}.$$

  That is, the sound level of I_1 is 10 dB less than that of I_2.

8. The measurement of sound level (Moore, 1989, pp. 7–10)

1. We've seen that the intensity of a sound—that is, the amount of energy transmitted by the sound per second through 1 square metre of the wavefront—is proportional to the square of the maximum pressure variation.

2. Our auditory systems can deal with a huge range of sound intensities. For example, the sound of a gunshot at close range has roughly 10 000 000 000 000 times the intensity of the quietest sound that we can hear. The sound of someone shouting at close range has roughly 10 000 000 times the intensity of a whisper.

3. If we have a sound with a low intensity and we increase its intensity linearly over time—say, for example, by $10^{-12}\mathrm{W/m}^2$ every second—then the loudness of the sound will be perceived to increase rapidly at first but the speed at which the loudness increases will become slower and slower as the sound becomes louder and louder.

4. Because of this, it is more useful to use a logarithmic scale for expressing intensities than a linear one.

5. So, if we have two sounds with intensities I_1 and I_2, then we express the difference between the intensities by the logarithm to the base 10 of the ratio of I_1 to I_2, thus $\log_{10}(I_1/I_2)$.

6. This expression gives the difference between the intensities in Bels: we say that the *level* of I_1 is $\log_{10}(I_1/I_2)$ Bels greater than that of I_2. So, for example, if I_1 is ten times I_2, then the *level* of I_1 is 1 Bel greater than that of I_2. If I_1 is 100 times I_2, then the level of I_1 is 2 Bels greater than that of I_2 and so on. Usually, we actually express level differences in *decibels* (dB), one decibel being one tenth of a Bel.

7. So, if we have two sounds with intensity I_1 and I_2 then we say that the level of I_1 is $10\log_{10}(I_1/I_2)$ decibels greater than that of I_2. An increase in level of 10dB corresponds to an increase in intensity by a factor of 10. A doubling of the intensity corresponds to an increase in level of 3 dB.

8. When I_1 is actually less intense than I_2, then the expression$10\log_{10}(I_1/I_2)$ is negative. For example, if I_2 is ten times as intense as I_1 then the level of I_1 is -10dB greater than I_2, or, equivalently, 10dB less than I_2.

9. Sound pressure level and sensation level

- To express *absolute* sound levels, we need to define a standard reference intensity.
- Most commonly used standard reference intensity is 10^{-12} watts per square metre (W/m^2) which corresponds to pressure variation of $2 \times 10^{-5} N/m^2$ or 20 μPa (micropascal).
- The sound level of a sound relative to 10^{-12} watts per square metre is called the *sound pressure level* (SPL) of the sound.
- If SPL of a sound with intensity I is 60 dB SPL, then this tells us that

$$10 \log_{10}(I/(10^{-12})) = 60 \Rightarrow \log_{10}(I/(10^{-12})) = 6 \Rightarrow \frac{I}{10^{-12}} = 10^6 \Rightarrow I = 10^{-6} W/m^2.$$

- O db SPL is close to human absolute threshold for 1000Hz tone (actually about 6.5 dB SPL on average).
- *Sensation level* of a sound is the intensity of the sound relative to the absolute threshold for that sound for a given individual, expressed in dB.

9. Sound pressure level and sensation level

1. So far we've only learnt how to express intensity ratios and level differences in decibels. If we want to express an absolute intensity I using decibels then we have to define some standard reference intensity I_0 and then find the number of decibels corresponding to the intensity ratio I/I_0.

2. The standard reference intensity most commonly used is 10^{-12} watts per square metre ($\mathrm{W/m^2}$) which corresponds to a pressure of $2 \times 10^{-5}\mathrm{N/m^2}$ or 20 μPa (micropascal). The sound level of a sound relative to this standard intensity is called the *sound pressure level* or SPL of the sound.

3. For example, the absolute intensity I of a sound whose sound pressure level is 60 dB SPL is given by

$$10 \log_{10}(I/(10^{-12})) = 60$$

   which implies that

$$\frac{I}{10^{-12}} = 10^6$$

   which implies that

$$I = 10^{-6}\mathrm{W/m^2}.$$

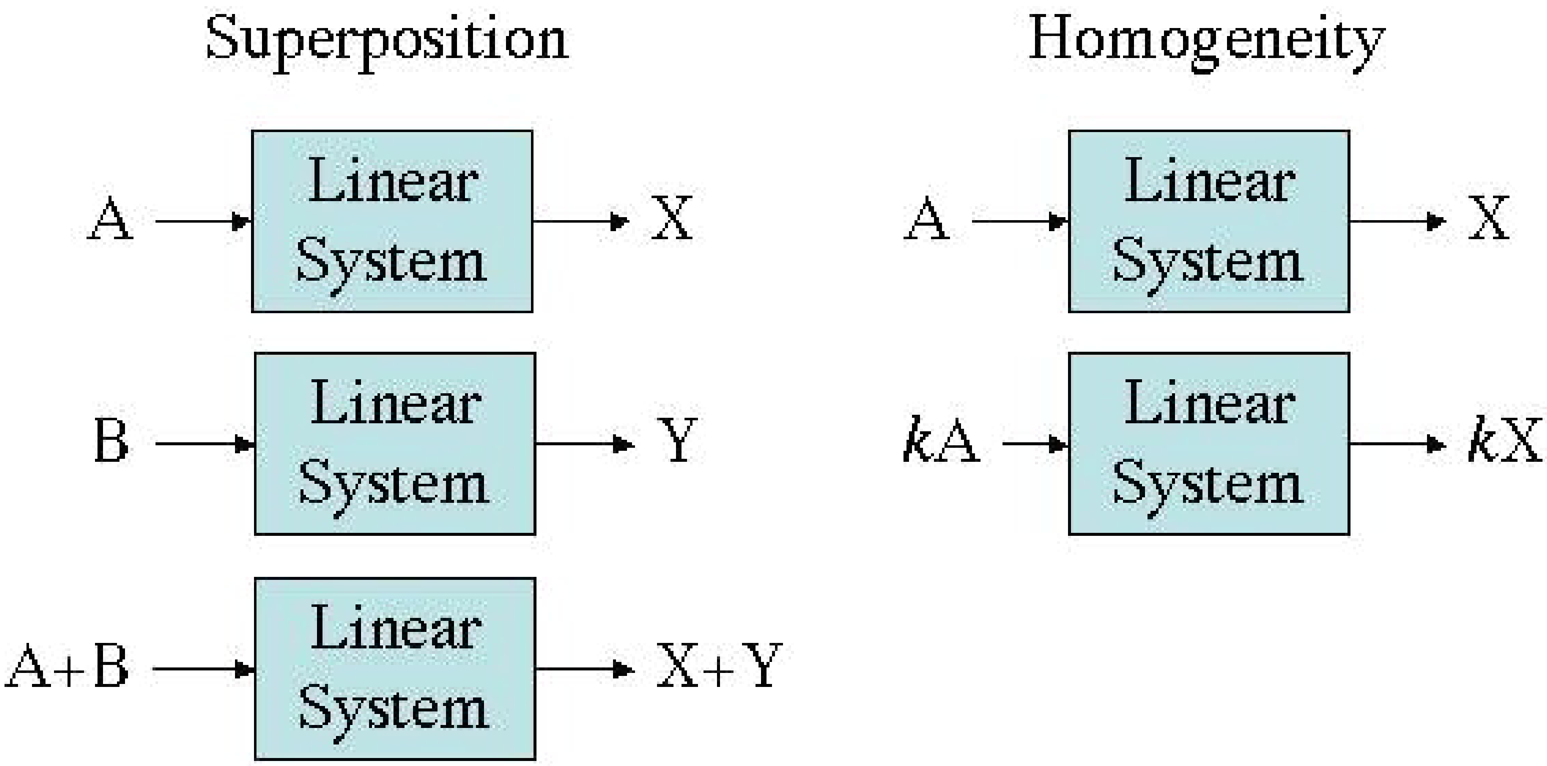
4. The standard reference intensity of $10^{-12}\mathrm{W/m^2}$ which corresponds to O dB SPL was chosen because it is very close to the absolute human threshold for a 1000Hz sinusoidal simple tone.

5. The absolute human threshold for a sound is the minimum detectable level of the sound in the absence of any other external sound (Moore, 1989, p. 8). In fact, the average human absolute threshold at 1000Hz is about 6.5 dB SPL.

6. Sometimes, instead of using the standard reference intensity of $10^{-12}\mathrm{W/m^2}$ we use the absolute threshold of the sound being measured for a particular individual. When the sound level is expressed in this way it is called a *sensation level* (SL). For example, if a sound is said to have a level of 60dB SL for a given subject, this means that the level of the sound is 60dB above the absolute threshold of that sound for that subject.

7. The physical intensity that corresponds to a given sensation level therefore varies from sound to sound and from subject to subject.

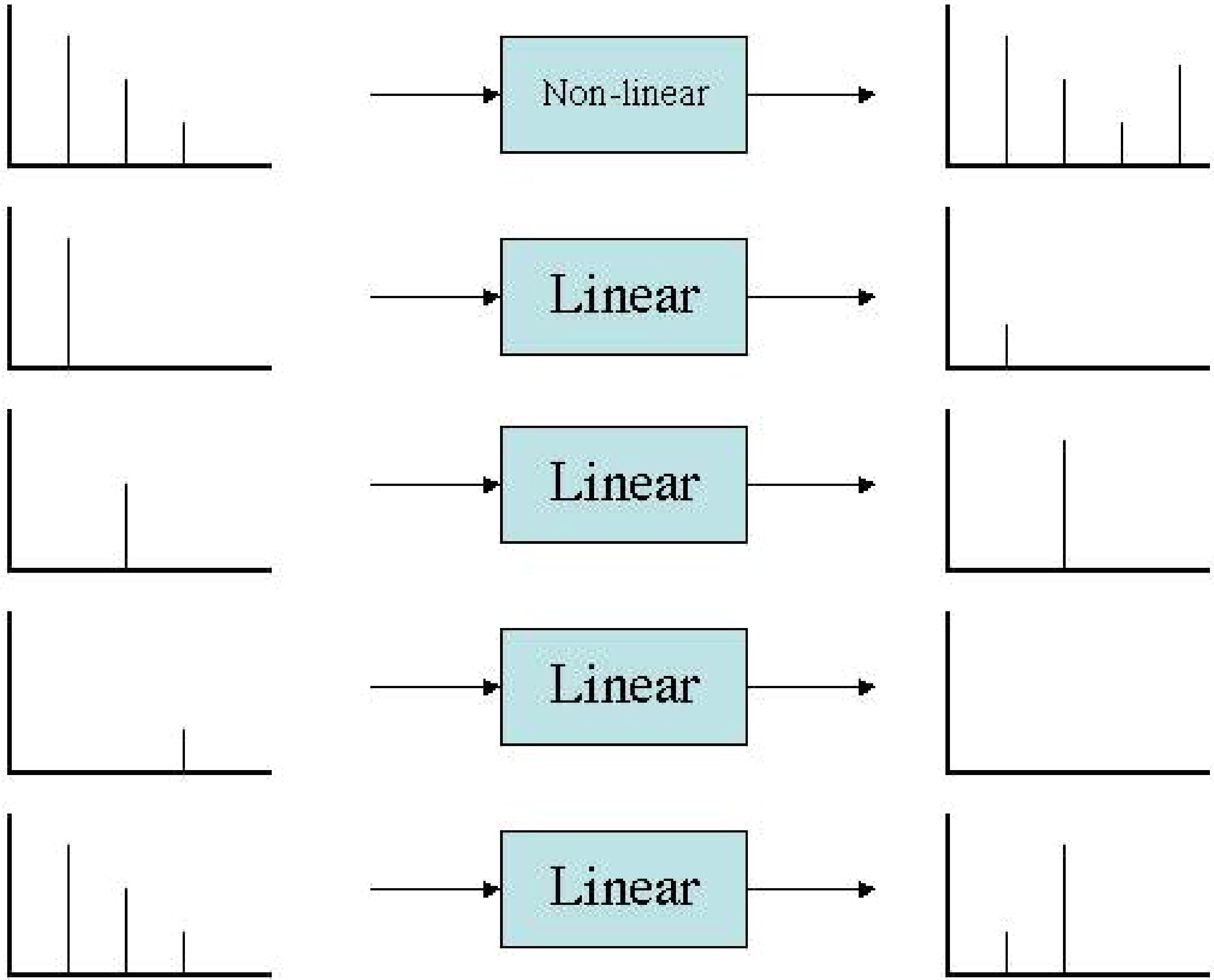
10. Linearity (Moore, 1989, pp. 10–11)



10. Linearity (Moore, 1989, pp. 10–11)

1. Before we learn about the peripheral auditory system, we need to know what is meant by the term *linearity*.
2. The auditory system is often modelled as a system consisting of a number of stages, each stage having an input and an output.
3. The input to a given stage is the output from the previous stage and the output of a particular stage provides the input to the next stage.
4. A given stage is said to be *linear* if it satisfies two conditions known as the conditions of *superposition* and *homogeneity*.
5. An input-output system is said to satisfy the condition of *superposition* if the output of the system for a number of independent inputs presented simultaneously is always equal to the sum of the outputs obtained when each input is presented in isolation.
6. For example, if X is the output of a system when the input is A and the system generates the output Y when the input is B, then the output when A and B are presented simultaneously will be $X + Y$ if the system is linear.
7. An input-output system is said to satisfy the condition of *homogeneity* if multiplying the magnitude of an input A by a factor k causes the magnitude of the output to be multiplied by k but otherwise causes no change in the output.

11. Linearity in acoustic systems



11. Linearity in acoustic systems

1. If the input to a linear system is a complex sound with Fourier components at a number of different frequencies, then the output of the system will not contain any components with frequencies that were not present in the input signal.

2. So if an acoustic system generates output containing frequencies that are not present in the input, then this is a sure sign that the system is non-linear.

3. This implies that if the input to a linear acoustic system is a sinusoidal simple tone, then the output of the system will be a sinusoidal simple tone with the same frequency as the input tone. The amplitude and phase of the output tone may, however, be different from those of the input tone.

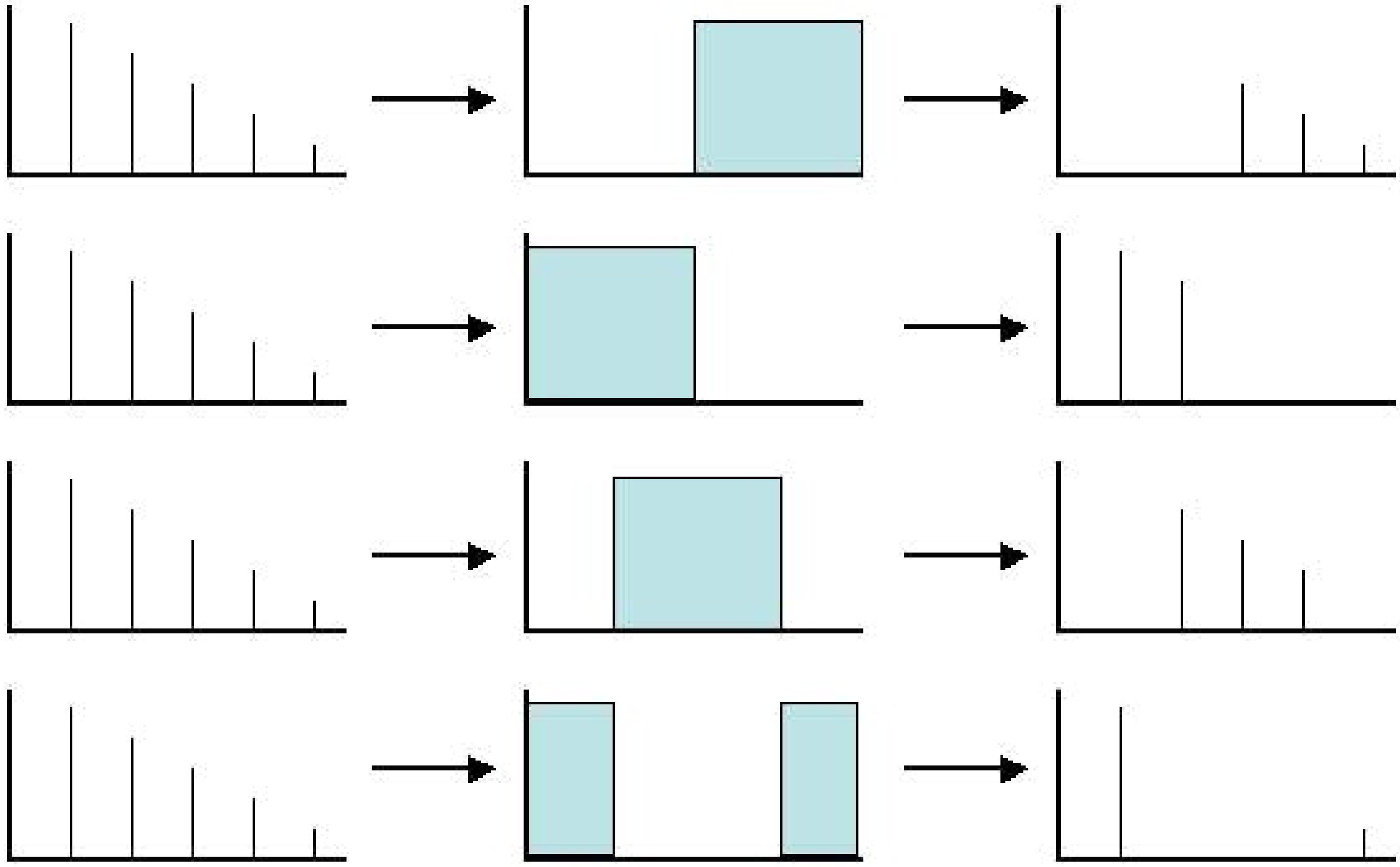
4. A linear system may respond differently to sinusoids of different frequencies, however. For example, it may reduce the amplitude to zero for any sinusoid with a frequency above a certain value.

5. This implies that if the input signal to a linear system is a complex tone containing Fourier components at various different frequencies, then the amplitudes of these Fourier components may not all be changed in the same way. If this happens, then the waveform of the output of the system will not be the same as the waveform of the input. For example, a square wave could be converted into a sinusoidal simple tone by a linear system that reduces to zero the amplitude of all frequencies above the fundamental of the square wave.

6. If an acoustic system is linear, then we can predict how it will respond to any complex sound provided we know how it responds to a sinusoidal input tone of any frequency. All we have to do is add together the responses of the system for all the sinusoidal Fourier components of the complex tone.

7. However, if an acoustic system is non-linear, then its response to a complex tone cannot in general be predicted from its responses to the individual Fourier components of the complex tone. This means that we cannot produce a precise model of a non-linear system simply by finding out how it responds to sinusoidal tones of different frequencies.

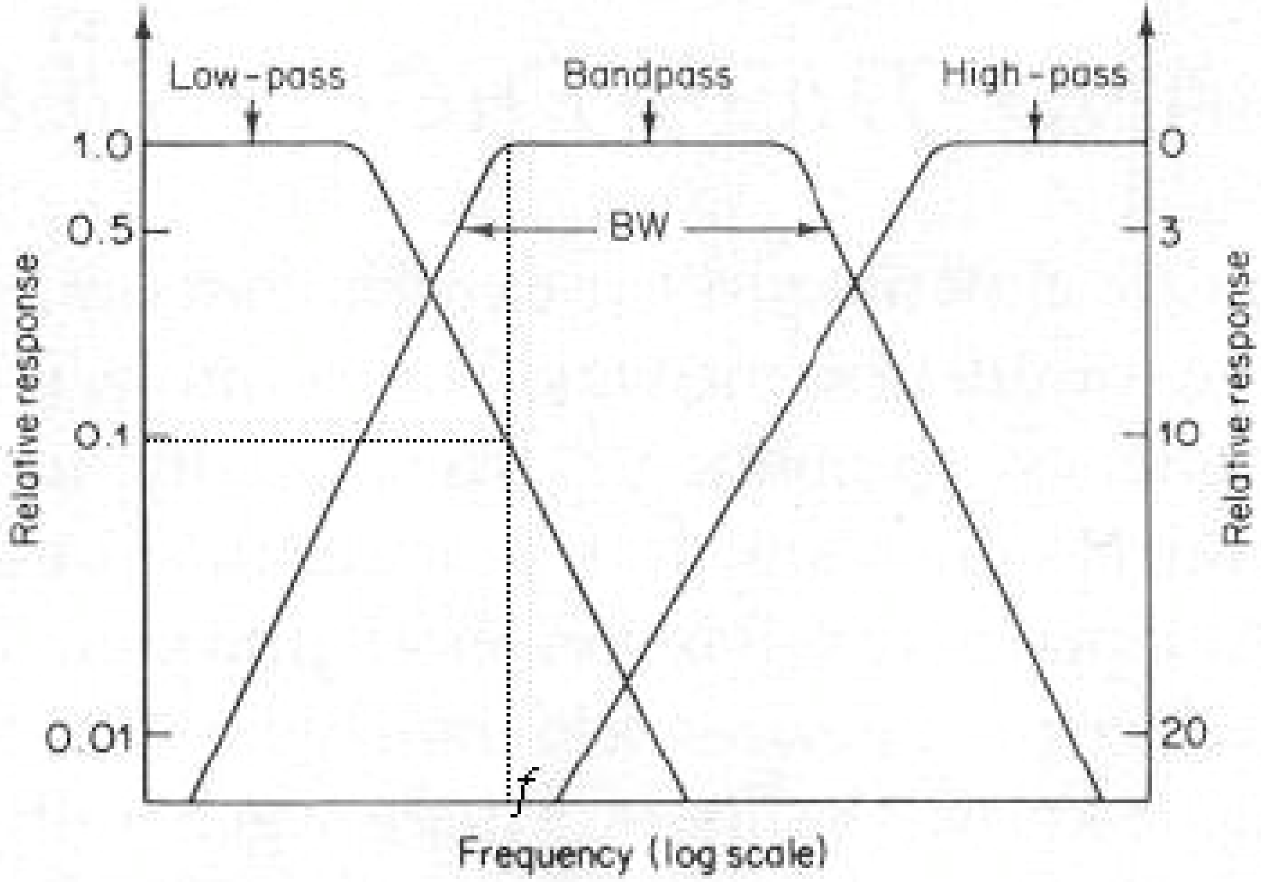
12. Filters and their properties (Moore, 1989, pp. 11–15; Roads, 1996, pp. 185–193)



12. Filters and their properties (Moore, 1989, pp. 11–15; Roads, 1996, pp. 185–193)

1. One particularly important type of linear acoustic device is a *filter*.
2. When the input to a filter is a sinusoid, the output is a sinusoid with the same frequency and an amplitude that is either equal to or less than that of the input sinusoid.
3. The amount by which the amplitude of the input sinusoid is reduced or *attenuated* by a filter depends upon the frequency of the sinusoid.
4. There are four basic types of filter: highpass filters, bandpass filters, lowpass filters and bandstop filters.
5. A highpass filter removes all the Fourier components in the input sound whose frequencies are less than a certain cutoff frequency, and leaves unchanged all the components with frequencies above the cutoff frequency.
6. A lowpass filter removes all the Fourier components in the input sound whose frequencies are greater than a certain cutoff frequency, and leaves unchanged all the components with frequencies below the cutoff frequency.
7. A bandpass filter has two cutoff frequencies. It removes all components in the input with frequencies outside the range defined by these cutoff frequencies and leaves unchanged all components within this range. The frequency half way between the cutoff frequencies of a bandpass filter is called the *centre frequency*. The difference between the two cutoff frequencies of a bandpass filter is called the *bandwidth* of the filter.
8. A bandstop filter also has two cutoff frequencies but it does the opposite of a bandpass filter: it removes all the components in the input signal with frequencies between the cutoff frequencies and leaves unchanged all components with frequencies outside of the range defined by the cutoff frequencies.
9. The range of frequencies that are left unchanged by a filter is called the *passband* of the filter and the range of frequencies that are attenuated is called the *stopband*.
10. The concept of a bandpass filter is particularly important for understanding the function of the inner ear because one component of the inner ear, called the *basilar membrane*, is often likened to a bank of bandpass filters, each filter having a different centre frequency.

13. Real filters



13. Real filters

1. Ideally, a filter would reduce to zero the amplitude of all components with frequencies outside the passband. However, in practice it is not possible to build filters with perfectly sharp cutoff frequencies. Instead, there will be a range of frequencies around each cutoff within which the amplitudes of components are reduced but not eliminated completely.

2. This range of frequencies within which the power of the input is reduced but not completely eliminated is called the *transition band*.

3. We can describe the behaviour of a filter by using a graph called a *filter response curve* or *filter characteristic*. This figure here shows some typical characteristics for bandpass, lowpass and highpass filters.

4. The filter characteristic gives, for each frequency f on the horizontal axis, the ratio of the output power to the input power when the input is a sinusoid with frequency f. For example, when a sinusoid with the frequency f in the diagram is passed through this bandpass filter, it is left unchanged. However, when it is passed through this lowpass filter, its power is reduced to about $\frac{1}{10}$th of its original value—that is, its power is reduced by 10dB.

5. In an ideal filter with sharp cutoffs, the cutoff frequencies are clearly defined. In a real, non-ideal filter, the cutoff frequencies are usually defined to be those frequencies at which the power of the output is half the power of the input, or, equivalently, 3dB less than the power of the input. So for these filters here, the cutoff frequencies would be as marked in the diagram.

6. The bandwidth of a bandpass filter is equal to the difference between the cutoff frequencies. When these cutoff frequencies are defined to be the frequencies at which the power is reduced by 3dB, the bandwidth is called the *−3 dB bandwidth* or the *3 dB down bandwidth* or the *half-power bandwidth*. It is also often called simply the *3 dB bandwidth*.

7. When the filter response curve axes are both logarithmic as they are here, it is often the case that the filter characteristic is approximately a straight sloping line within the transition band, as shown in this diagram. We

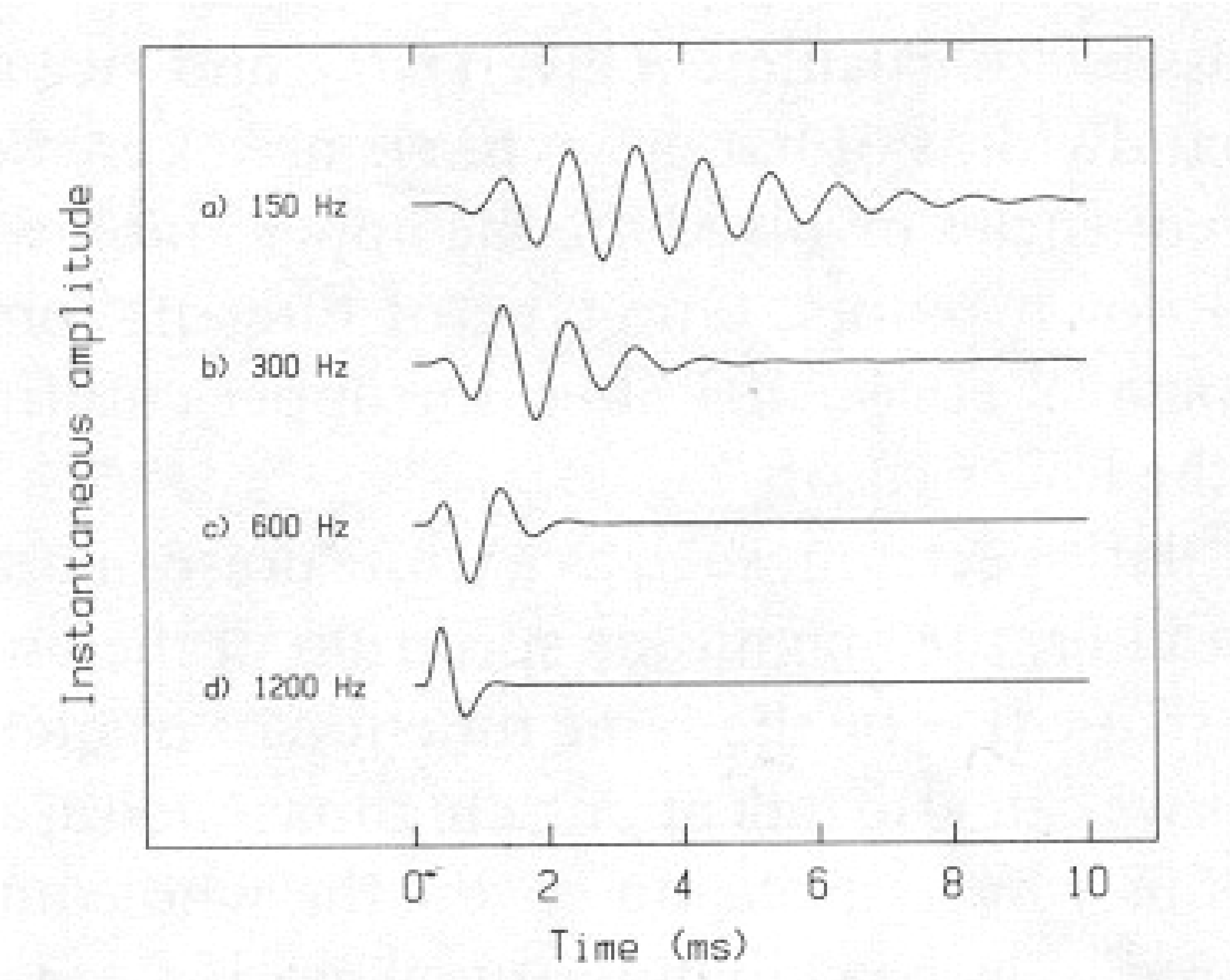
can therefore represent the sharpness of the cutoff frequencies of a filter by giving the slope of this line within the transition band.

8. On this graph, equal distances along the vertical axis correspond to equal level changes in dB and equal distances along the horizontal axis correspond to equal frequency *ratios*.

9. To express the slope of this line, we therefore have to choose a particular frequency ratio as the unit for the horizontal axis and the commonest ratio to choose is 2:1, that is, an octave.

10. The sharpness of the cutoff frequencies of a filter is therefore usually represented by the slope of the filter characteristic within the transition band, expressed in dB/octave. For example, if a lowpass filter has a slope of 24 dB/octave, this means that the ratio of the input to output power outside the passband decreases by 24 dB each time the frequency is doubled.

14. Impulse response



14. Impulse response

1. If we pass a signal with a flat spectrum such as a white noise or a click through a filter, then the spectrum of the output signal will be the same shape as the filter characteristic.

2. If we alter the spectrum of a sound by filtering it, this will cause a corresponding alteration in the waveform of the sound and a change in the way it is perceived.

3. For example, if we pass white noise through a narrow bandpass filter, then the waveform of the output will resemble a sinusoid with fluctuating amplitude. The output sound will have a pitch that is similar to that of a sinusoid with a frequency equal to the centre frequency of the filter.

4. If we pass a brief click or *impulse* through a filter then the output is called the *impulse response* of the filter and because the Fourier spectrum of the impulse is flat, the Fourier spectrum of the output will have the same shape as the filter characteristic.

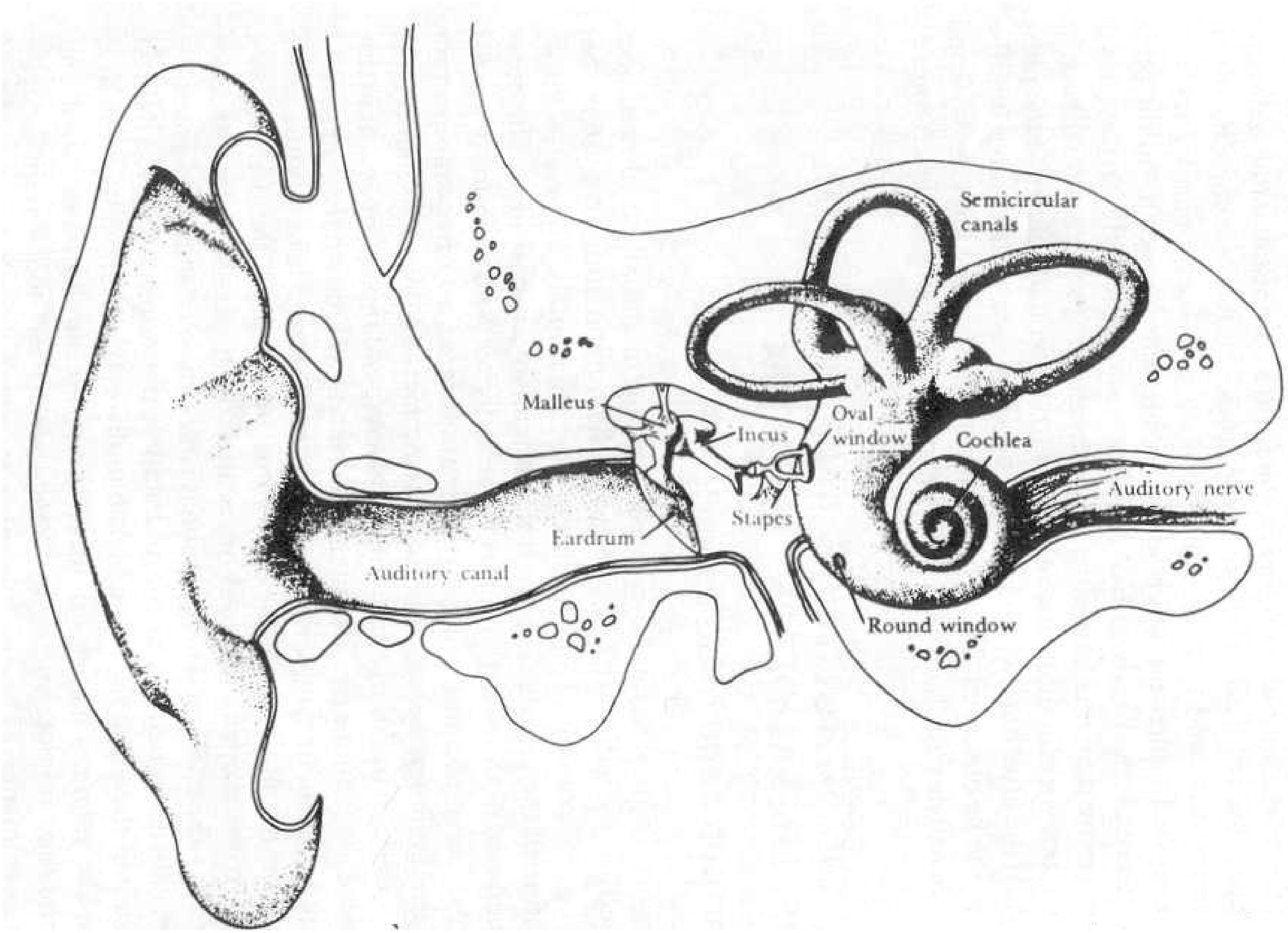
5. This figure shows the waveforms of the impulse responses obtained from four different bandpass filters. Each filter has a centre frequency at 1000Hz but they all have different bandwidths.

6. For the narrowest filter, the output resembles a sinusoidal tone burst at the centre frequency of the filter that rises and then falls in amplitude. This is called a 'ringing response'.

7. As the filter bandwidth increases, the oscillations in the impulse response become less regular, the duration of the response gets shorter and the output becomes more and more like the input click.

8. This demonstrates that if you use a bank of bandpass filters to perform a Fourier-like analysis of an input signal, then if you use very narrow filters in order to get high frequency resolution, you lose the ability to respond quickly to a changing input signal because the responses of the filters will last a long time. On the other hand, if you use wide filters, then you will have better time resolution but the frequency resolution will be reduced.

15. The peripheral auditory system



15. The peripheral auditory system

1. This figure shows the human peripheral auditory system.
2. Conventionally, this system is divided into the outer, middle and inner ear [SHOW ON DIAGRAM].
3. The outer ear consists of the *pinna* which is the bit that protrudes from the side of the head, and the ear canal or *auditory meatus*.
4. The human pinna actually has a quite significant effect on the incoming sound, particularly at high frequencies, and contributes to our ability to localize sounds—that is, tell where they are coming from.
5. In other mammals such as dogs, for example, the pinna is much more well developed and can be directed towards the sound source. This can be used to amplify quieter sounds by 'funnelling' more of the sound energy down into the ear. You can experience this amplification effect yourself by cupping your hands over your ears so that the opening is directed towards a sound source.
6. The incoming sound travels down the auditory meatus and causes the eardrum or *tympanic membrane* to vibrate.
7. The tympanic membrane is the interface between the outer ear and the middle ear. The vibrations of the tympanic membrane are transmitted through the middle ear by three small bones, called the *auditory ossicles*, to a membrane-covered opening in the bony wall of the inner ear. This opening is called the *oval window*.
8. The three ossicles are called the *malleus* (or hammer), the *incus* (or anvil) and the *stapes* (or stirrup). The lightest of these is the stapes and it is this bone which makes contact with the oval window. The auditory ossicles are the smallest bones in the human body.
9. The oval window is the interface between the middle ear and the inner ear. The inner ear consists of a fluid-filled, spiral-shaped structure called the cochlea.
10. The middle ear seems to have evolved to ensure that the vibrations of the tympanic membrane are efficiently transmitted to the fluids inside the cochlea.

11. If the sound were to impinge directly onto the oval window, then most of it would be reflected back because the fluids in the cochlea are much more dense than the air—we say that the *acoustical impedance* of the fluids in the cochlea is much greater than that of the air.

12. The force exerted on the tympanic membrane is equal to the pressure times the area. This force is transmitted by the auditory ossicles onto the oval window. However, the area of the oval window is roughly 1/25th of the area of the tympanic membrane, so all the force on the tympanic membrane is concentrated onto the much smaller area of the oval window. If this were the whole story, then the pressure on the oval window would be 25 times that on the tympanic membrane. But it is not the whole story because the auditory ossicles act as a system of levers that approximately doubles this pressure.

13. So the pressure on the oval window is about 50 times the pressure on the tympanic membrane, but the distance moved by the stapes is only about half that of the tympanic membrane.

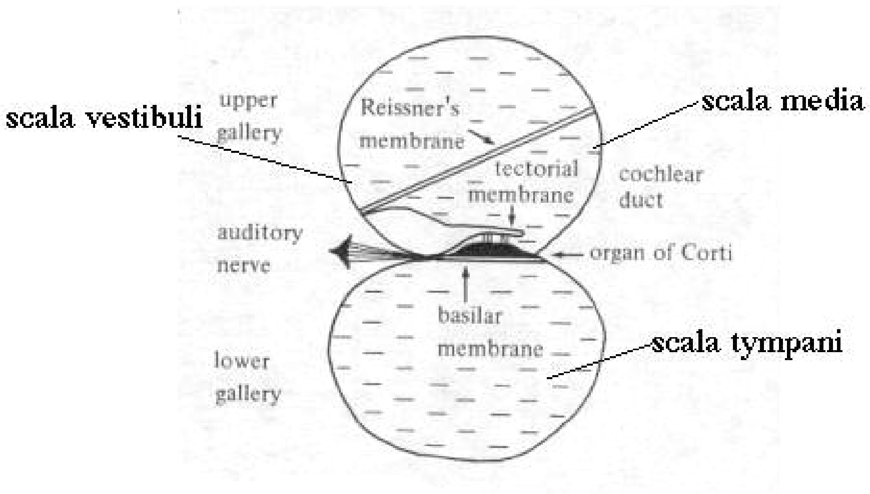
14. Recall that the loudness of a sound is related to its intensity which is proportional to the square of the pressure amplitude. If the pressure at the oval window is 50 times that at the tympanic membrane, then the intensity is multiplied by 2500 times by the middle ear. In other words, the middle ear increases the sound pressure level by over 30 dB. This extra intensity ensures that a much larger fraction of the sound energy is transmitted to the inner ear.

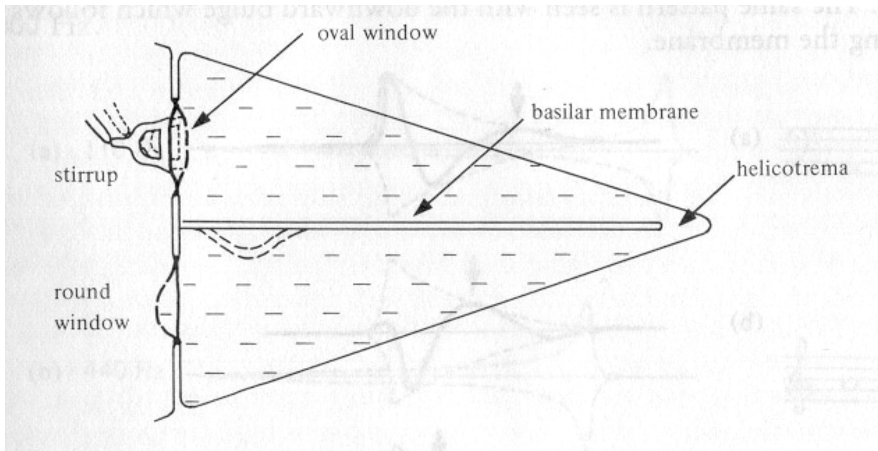
15. When the middle ear is exposed to very loud sounds at frequencies below about 1000Hz, a muscle attached to the stapes automatically draws the bone slightly away from the oval window. This is called the *middle ear reflex* or the *acoustic reflex* (Morgan and Dirks, 1975).

16. This reflex seems to have evolved in order to protect the inner ear from sudden loud noises. However, as the reflex takes about 1/10 second to act, it cannot protect the inner ear from very sudden noises such as a gunshot.

17. It has also been suggested that the middle ear reflex has evolved because it reduces the audibility of one's own speech—it has been shown that the reflex is activated just before vocalization.

16. The inner ear

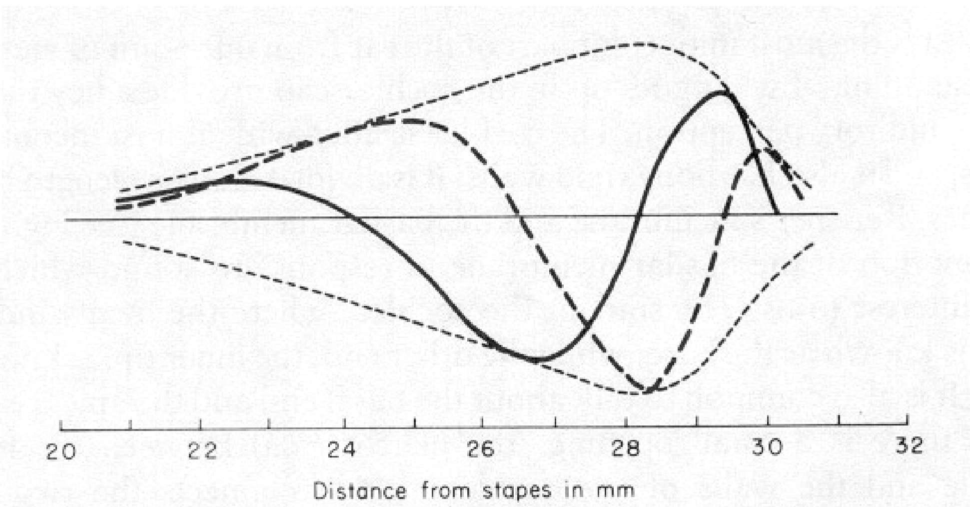


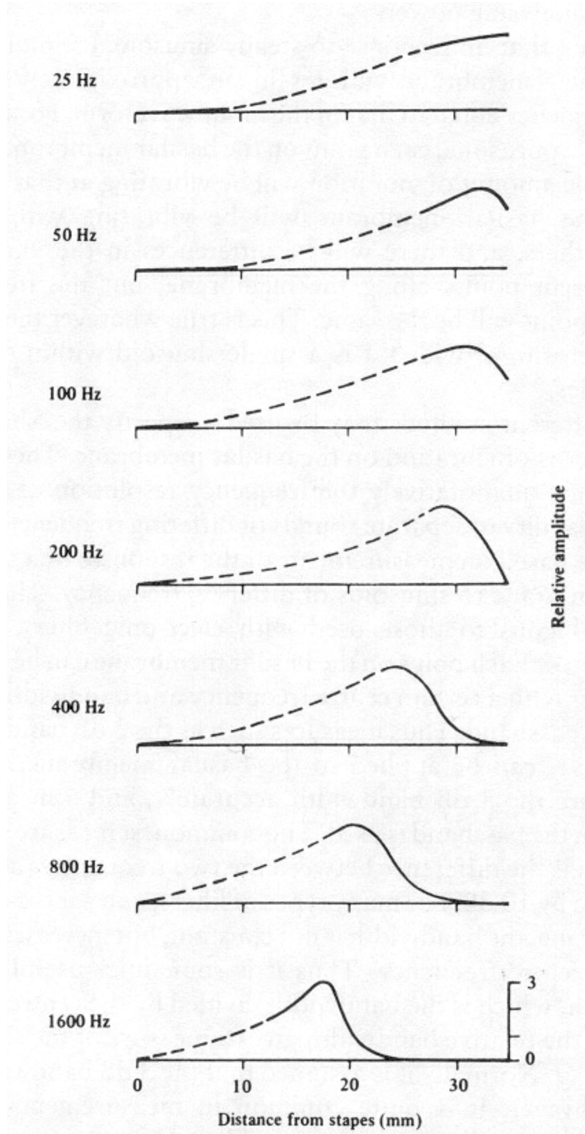


16. The inner ear

1. The sound is transmitted by the middle ear into the inner ear via the oval window.
2. The inner ear consists of the *cochlea* which is a spiral-shaped tube with bony walls filled with incompressible fluid.
3. When we take a cross-section across the cochlea, we see that it is divided into three chambers by two membranes that run almost all the way along the tube, as shown in this diagram.
4. The oval window opens into the top chamber which is called the *scala vestibuli* or *upper gallery*.
5. The scala vestibuli is separated from the middle chamber by a very thin membrane called Reissner's membrane.
6. The middle chamber is called the *scala media* or *cochlear duct* and this chamber is separated from the lowest of the three chambers by a rather more substantial membrane called the *basilar membrane*.
7. The lowest of the three chambers is called the *scala tympani* or the *lower gallery*.
8. The end of the cochlea where the oval window is situated is called the *base* or *basal end* and the opposite end of the cochlea is called the *apex* or *apical end*.
9. At the basal end of the scala vestibuli (the upper gallery), we find the oval window and at the basal end of the scala tympani (the lower gallery), we find another membrane covered opening called the *round window*.
10. At the apical end of the cochlea, the scala vestibuli and the scala tympani meet and the gap through which they communicate is called the *helicotrema*.
11. When the stapes is pushed inward against the oval window by an incoming sound, this displaces some of the fluid in the scala vestibuli and, because the basilar membrane is not rigid, it is pushed downward at the basal end. This in turn displaces fluid in the scala tympani which pushes the round window outward.

17. Patterns of vibration on the basilar membrane



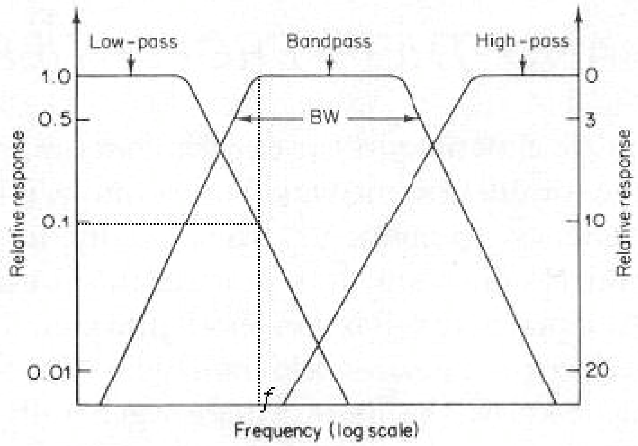


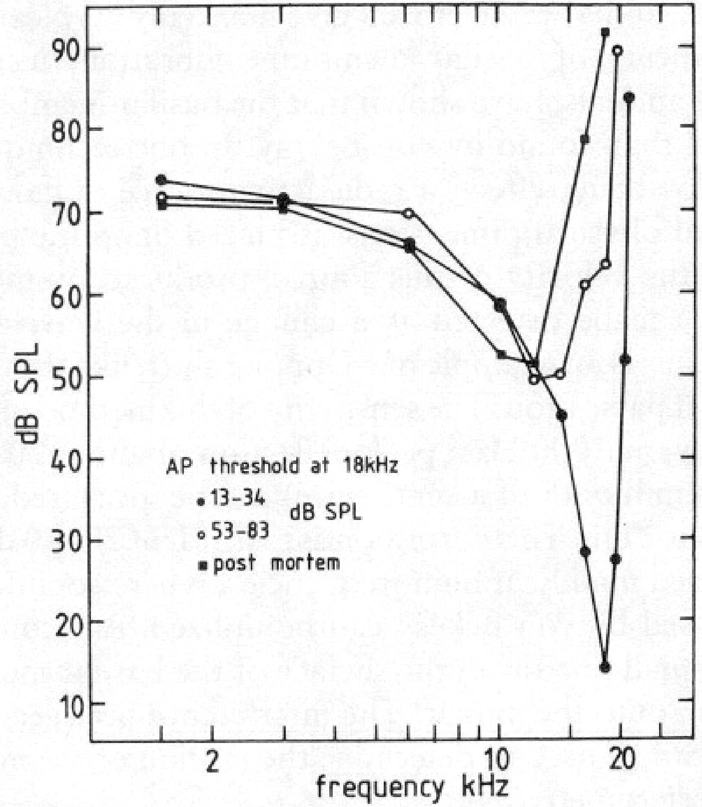
17. Patterns of vibration on the basilar membrane

1. We all know that if we have a length of string and we sharply flick one end of it in a vertical direction, then this causes a wave to travel down the string.
2. Exactly the same thing happens in the basilar membrane. The displacement in the basilar membrane caused by the inward motion of the stapes is transmitted along the basilar membrane just like the impulse is transmitted along a string.
3. Now, the linear density of a string or rope is roughly constant throughout its length. So when the disturbance is transmitted along it, the wavelength of the disturbance remains roughly the same and the amplitude diminishes gradually as the energy is dissipated.
4. However, the basilar membrane does not have a homogeneous structure like a string. Instead, it is rather narrow and stiff at the basal end and gets wider and more flexible the nearer you get to the apical end.
5. So when the cochlea responds to a sinusoidal stimulus, the wavelength and amplitude of the wave in the basilar membrane change as it travels along it. The wavelength gets shorter and the amplitude gradually increases until it reaches a certain point on the basilar membrane, after which it rapidly decrases.
6. This diagram shows the shape of the basilar membrane at two different instants in time when it is stimulated by a 200Hz sinusoid. This dotted curved line shows the track or locus of the points of maximum displacement in the wave. As you can see, the amplitude increases gradually up to a point and then falls rapidly after that point. This dotted line is called the *amplitude envelope* of the basilar membrane response.
7. Because the basilar membrane gets wider and more flexible, the nearer you get to the apex, the point on the basilar membrane at which the maximum amplitude occurs (i.e., the peak of the amplitude envelope) depends on the frequency of the stimulus.
8. When the stimulus has a high frequency, the peak of the amplitude envelope is near to the base where the basilar membrane is narrow and stiff. When the stimulus has a low frequency, the peak of the amplitude envelope is nearer to the apex where the basilar membrane is wide and flexible.

9. This diagram here from von Békésy (1960) shows the amplitude envelope of the basilar membrane response for sinusoidal stimuli of various frequencies.
10. Note that, whatever the stimulus frequency, the disturbance on the basilar membrane does not travel much further after the amplitude has peaked.
11. Since the point on the basilar membrane at which the maximum displacement occurs varies with frequency, the basilar membrane effectively separates out the sinusoidal components in the stimulus. In other words, it performs a crude form of Fourier analysis.
12. When the stimulus is a steady sinusoidal simple tone, every point on the basilar membrane responds by vibrating up and down in an approximately sinusoidal manner at the same frequency as the stimulus.
13. However, points on the basilar membrane further away from the oval window will show a phase lag behind those points nearer the oval window.
14. Also, because the width and flexibility of the basilar membrane increase the nearer you get to the apex, each point on the membrane will have its own unique resonant frequency. The closer this resonant frequency is to that of the stimulus, the greater the amplitude with which that point will vibrate.

18. Frequency resolution of the basilar membrane





- *Relative bandwidth* of a bandpass filter is the ratio of the bandwidth to the centre frequency.
- Q is a measure of the sharpness of the tuning of a bandpass filter. It is the reciprocal of the relative bandwidth (i.e., ratio of centre frequency to bandwidth).

18. Frequency resolution of the basilar membrane

1. We've seen that the basilar membrane performs a sort of Fourier analysis on the stimulus. That is, it separates out the sinusoidal components in the stimulus. But what is the *frequency resolution* of the basilar membrane? That is, how close must two simple tones be in frequency before the basilar membrane is unable to distinguish between them?

2. We measure this by assuming that each point on the basilar membrane behaves a bit like a bandpass filter, with a particular centre frequency, a bandwidth and a sloping transition band at the upper and lower cutoff frequencies.

3. It is often hard to measure the -3 dB bandwidth accurately for a point on the basilar membrane, so the -10 dB bandwidth is usually used instead in this case. The -10 dB bandwidth is the difference between the two frequencies at which the power of the output of a filter is 10 dB less than the power of the input.

4. The -10 dB bandwidth is not the same for all points on the basilar membrane. However, the ratio of the bandwidth to the centre frequency *is* roughly the same for most points on the basilar membrane.

5. The ratio of the bandwidth of a bandpass filter to its centre frequency is called its *relative bandwidth* and the reciprocal of the relative bandwidth—that is, the ratio of the centre frequency to the bandwidth—gives a measure of how 'sharply' tuned the filter is to its centre frequency. This measure of sharpness of tuning is called Q when the bandwidth is the -3 dB bandwidth. However, in the case of the basilar membrance, the bandwidth measured is the -10 dB bandwidth and, in this case, the measure of sharpness is denoted by Q_{10dB}.

6. Most of the early work on the response of the basilar membrane was done by von Békésy (1928, 1942, 1960). He measured the response of the basilar membrane to very loud sinusoids (about 140 dB SPL) at fairly low frequencies in recently deceased humans, using a microscope with stroboscopic illumination to measure the vibration amplitude.

7. For the range of frequencies that von Békésy measured, the relative bandwidth was about 0.6. That is, the bandwidth at each point was about 60% of the centre frequency. This relative bandwidth is far too wide to explain our ability to 'hear out' the partials in a complex tone. It is also too wide to explain the fact that individual neurons in the auditory nerve only respond to a very narrow range of frequencies—that is, they are very sharply tuned.

8. Moore (1989, p. 21) suggests that this discrepancy is due to two factors. First, the basilar membrane is now believed to be non-linear. This means that one cannot necessarily predict how it will respond to quiet sounds by examining how it responds to very loud ones like those that von Békésy used in his experiments. Second, in more recent experiments on animals, it has been shown that the response of the basilar membrane is *physiologically vulnerable*—that is, it changes depending on how healthy the subject is. So the response of the basilar membrane in human cadavers is going to be significantly different from its response in healthy living subjects.

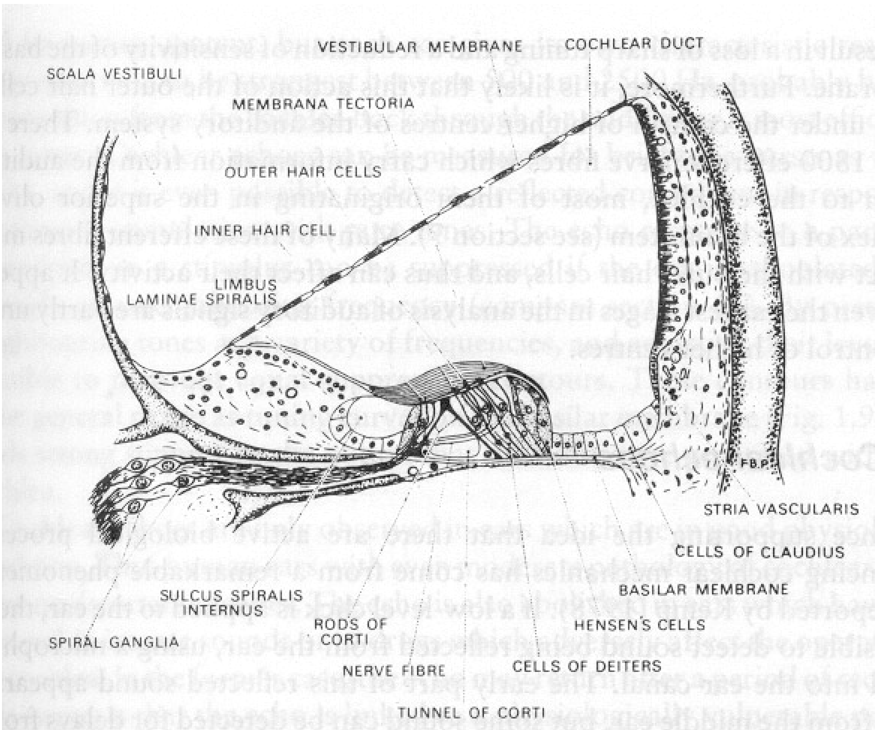
9. More recent experiments carried out on living animals have shown that the living basilar membrane is tuned much more sharply than suggested by von Békésy's experiments.

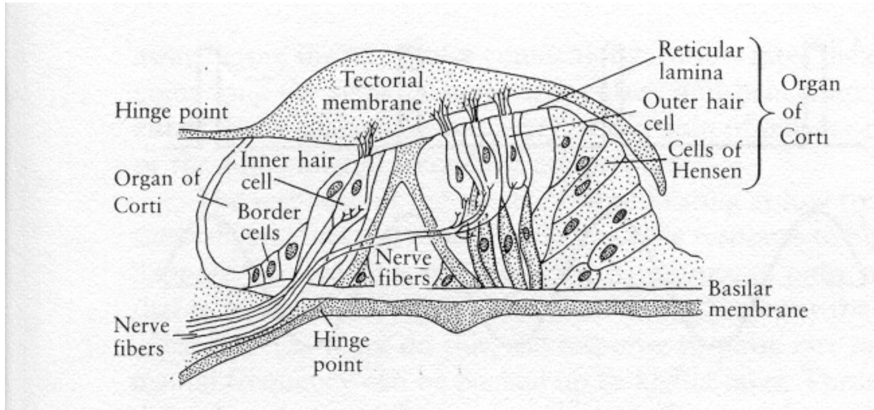
10. For example, this graph here shows the least sound level required to cause a constant velocity of motion at a particular point on the basilar membrane as a function of frequency. As you can see, at the start of the experiment when the animal was still healthy, only a very low sound level was required to cause a response at the best frequency. However, as the condition of the animal deteriorated, the tuning becomes less and less sharp.

11. So recent experiments seem to show that in normal healthy ears, each point on the basilar membrane is sharply tuned and behaves like a narrow bandpass filter with a value of Q_{10dB} in the range 3–10 (i.e., with a bandwidth between 1/10th and 1/3 of the centre frequency).

12. It also seems that this sharpness of tuning is the result of *active* processes—that is, the cochlea does not just send signals *to* the brain, it also receives signals *from* higher brain centres that affect the way that it responds to sounds.

19. The transduction process and the hair cells





19. The transduction process and the hair cells

1. We've seen that the basilar membrane acts as a sort of Fourier analyser with a limited resolving power.
2. In the last few minutes I'm just going to talk briefly about how the spectral information obtained by the basilar membrane is conveyed from the cochlea into the auditory nerve and then to the brain.
3. This diagram here is a more detailed cross section of the cochlea.
4. As you can see, just above the basilar membrane is a gelatinous structure called the *tectorial membrane*. And between the tectorial membrane and the basilar membrane there are some special cells called *hair cells*. The hair cells and tectorial membrane form part of a structure called the *organ of Corti*.
5. This diagram shows the organ of Corti in more detail.
6. The hair cells are divided into two groups by an arch called the *tunnel of Corti*. The group of hair cells nearer the outside of the cochlea are called the *outer hair cells* and those on the inside of the tunnel of Corti are called the *inner hair cells*.
7. In humans, the outer hair cells are organized into five rows, running along the basilar membrane and the inner hair cells are arranged as a single row. In total there are about 25000 outer hair cells, each with about 140 hairs protruding from it and about 3500 inner hair cells, each with about 40 hairs.
8. The hairs on the outer hair cells actually make contact with the tectorial membrane but the inner hair cells probably do not.
9. The tectorial membrane is effectively fixed or 'hinged' at its inner edge so that when the basilar membrane moves up and down, the tectorial membrane slides over it with a shearing motion, causing the hairs on the hair cells to be displaced.
10. This causes the inner hair cells to fire and send signals up the auditory nerve to the brain.

11. It seems that most of the afferent neurons that connect to the cochlea (that is, the ones that carry signals to the brain) are connected to the *inner* hair cells, each hair cell being contacted by about 20 neurons. So it seems to be the *inner* hair cells that are concerned with conveying information about the incoming sound to the brain.

12. Most of the 1800 or so efferent neurons that connect to the cochlea (that is, the ones that carry information from the brain to the cochlea) connect to the outer hair cells and it seems that signals carried in these neurons can cause the outer hair cells to change their length and shape, consequently affecting the way that the basilar membrane responds to sounds (see Pickles, 1988).

13. In experiments where subjects have been dosed with drugs that only affect the outer hair cells, it has been shown that this leads to a reduction of sensitivity in the basilar membrane—that is, each point on the basilar membrane becomes less sharply tuned and responds to a wider bandwidth of frequencies.

14. This shows that active processes involving signals being sent from the brain down to the outer hair cells contribute to the sharpness of the response of the basilar membrane.

References

Campbell, M. and Greated, C. (1987). *The Musician's Guide to Acoustics*. J. M. Dent & Sons Ltd., Melbourne.

Dowling, W. J. and Harwood, D. L. (1986). *Music Cognition*. Series in Cognition and Perception. Academic Press, Orlando, FL.

Moore, B. C. J. (1989). *An Introduction to the Psychology of Hearing*. Academic Press, London and San Diego, third edition.

Morgan, D. E. and Dirks, D. D. (1975). Influence of middle ear contraction on pure-tone suprathreshold loudness judgements. *Journal of the Acoustical Society of America*, **57**, 411–420.

Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing*. Academic Press, San Diego, second edition.

Pierce, J. R. (1992). *The Science of Musical Sound (Revised Edition)*. W. H. Freeman and Co., New York.

Roads, C. (1996). *The Computer Music Tutorial*. MIT Press, Cambridge, MA. With contributions from John Strawn, Curtis Abbott, John Gordon and Philip Greenspun.

von Békésy, G. (1928). Zur Theorie des Hörens: die Schwingungsform der Basilar membran. *Phys. Z.*, **29**, 793–810.

von Békésy, G. (1942). Über die Schwingungen der Schneckentrennwand beim Präparat und Ohrenmodell. *Akust. Z.*, **7**, 173–186.

von Békésy, G. (1960). *Experiments in Hearing*. McGraw-Hill, New York. Translated and edited by E. G. Wever.

Whelan, P. M. and Hodgson, M. J. (1978). *Essential Principles of Physics*. John Murray, London.