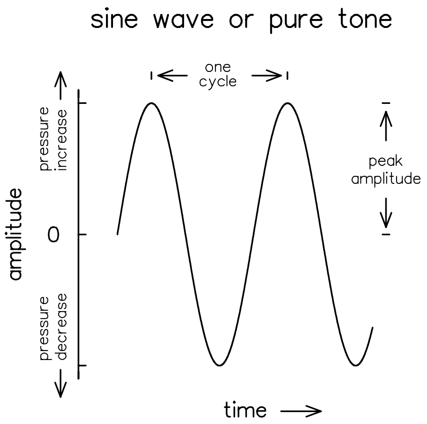
Sound waves and Fourier analysis

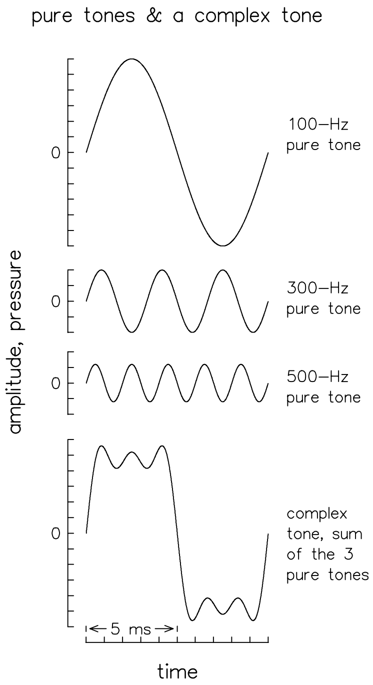
Waveforms. Auditory information tells us about small movements, which generate the sounds that our ears detect. Small movements produce sound by causing fluctuations of the air's pressure in their vicinity. These fluctuations follow the source's movements so that there are alternating increases and decreases in the air's pressure. This generates a sound wave as the fluctuations spread through the air. Sound informs us about the nature of the movement that produced it because the pressure fluctuations follow the source's movement. Thus, a graph of pressure against time, as a sound wave passes, mirrors the distant movements of the sound source. This graph is called a waveform.

Sine waves. A simple kind of movement is like the repeating pattern a pendulum's motion. Rapid movements of this kind produce sounds called pure tones, known also as sine waves, sine tones, sine-wave tones or sinusoids. Sounds of this precise type are not particularly common in our everyday environment, but they can be heard when a tuning fork is sounded, or when we whistle. The waveform of such a sound is shown below.



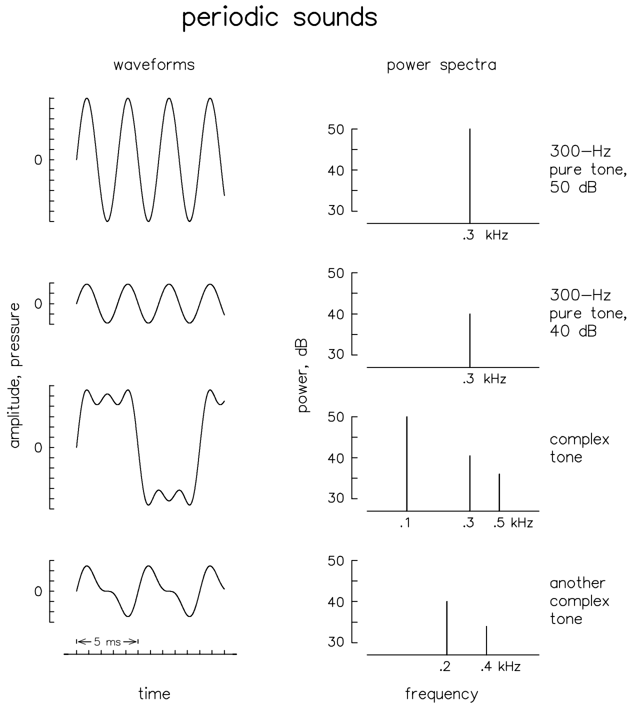
Each repetition of the wave's pattern is called a cycle. The rate at which cycles arrive is the rate that they were produced at the source. The number of cycles arriving in one second is known as the *frequency* of the sound, measured in Hertz and abbreviated Hz. 1000 Hz is a kilohertz, abbreviated kHz (see also appendix 1). Thus, slow vibrations give low frequency sounds and faster vibrations give high frequency sounds. Frequencies of pure tones that the ear can detect range from about 20 Hz up to nearly 20 kHz. The height of the wave's crest is called its peak amplitude. The *power* of a pure tone increases with its peak amplitude and is measured in decibels, abbreviated dB (see also appendix 2). This aspect of the waveform indicates the strength of the source's vibration. *Phase* is a term used to describe position within the wave's cycle, where there are 360 degrees (or 2π radians) in a full cycle. Thus, 180 degrees (or π radians) is half way through a cycle, 90 degrees (or $\pi/2$ radians) is one quarter of a cycle, and so on.

Complex tones. Typical everyday sounds are said to be complex because they contain a range of frequency components. This is true of individual speech sounds and individual notes in music as well as the clatters, clangs, bangs, bumps, rasps, rumbles, buzzes, beeps, fillips, fizzes, crunches, crackles, thumps, thuds, etc., that populate our auditory world. When components at higher frequencies are removed from such sounds they can seem muffled, or if lower frequencies are missing they can sound tinny. Individually, a single frequency component from a complex sound is just like a pure-tone with a particular frequency and dB. Therefore, different complex sounds can be obtained by *adding together pure tones* with different frequencies and dB. One complex sound is shown on the left, along with its 3 components. Adding together the pressures of each of the pure-tone components at each point in time forms the waveform of the complex sound.



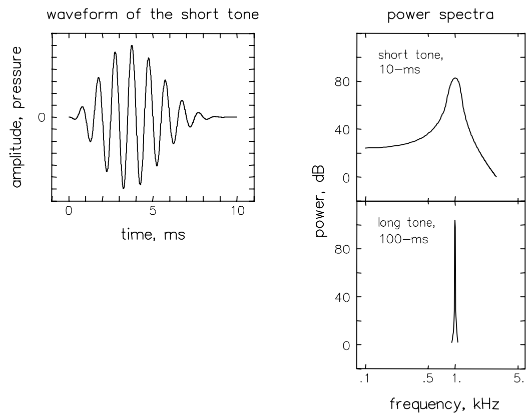
A complex sound can be analysed to find its frequency components. This is called *Fourier analysis* or frequency analysis. Ohm (1843) suggested that the ear performs a Fourier analysis in his 'acoustical law', and in broad terms he was correct. It is therefore common to represent sounds in terms of their frequency components, and such a representation is called the *power spectrum.*

Periodic sounds. Power spectra of simple and complex sounds are shown on the right along with their waveforms. These are called periodic sounds because their waveforms repeat. Their spectra are vertical lines whose left-right position represents the frequency of a component, while the line's height represents the power of a component in dB.



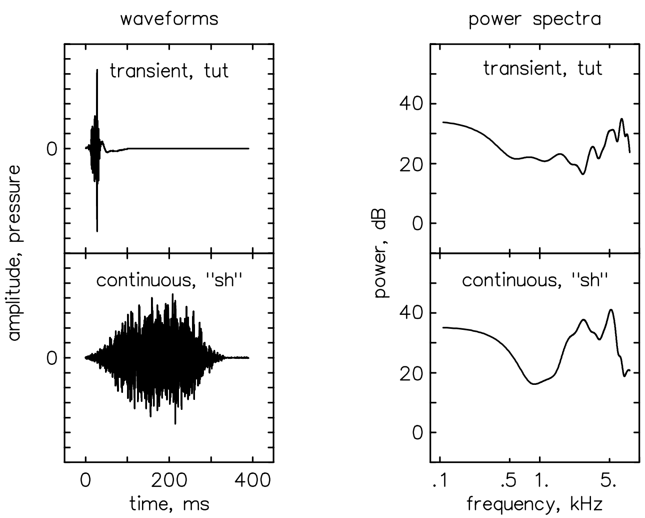
When a periodic sound is turned on or off, extra frequency components are added. The power spectra of such sounds contain numerous closely spaced frequency components, so

short & long periodic tones



in these cases the dB values are connected together by a continuous line. Spectra of a *short* and *a longer* tone are shown above. Notice that the longer tone's continuous spectrum approaches the 'ideal' of a line spectrum. This is increasingly the case as the tone gets longer.

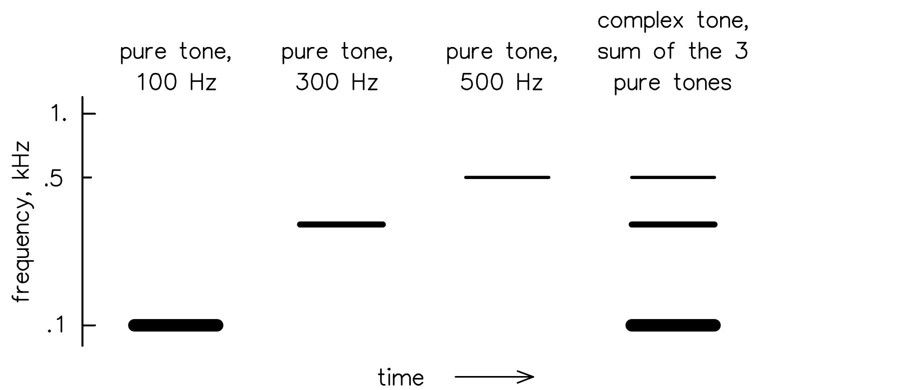
aperiodic sounds



Aperiodic sounds. Sounds with waveforms that do not repeat are called aperiodic sounds. Their power spectra are also continuous because they also contain numerous, closely spaced frequency components. *Transients* are sounds of this type, such as the tut (as in tut-tutting) whose waveform and power spectrum are shown on the right. The other aperiodic sounds are *continuous* and have amplitudes that fluctuate haphazardly from moment to moment. These sounds have a noise-like quality, such as the spoken "sh" that is also shown here.

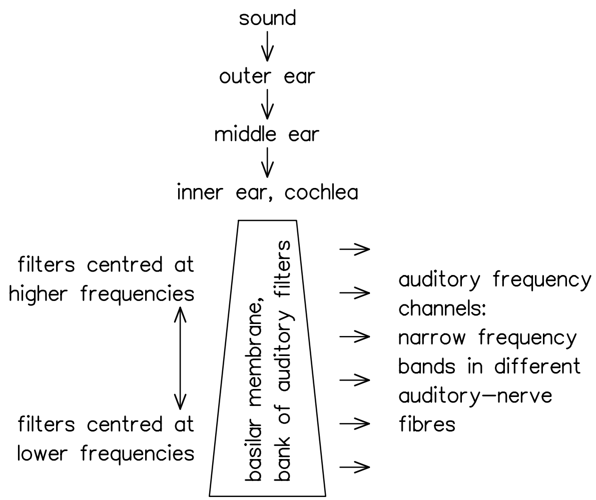
Auditory scenes. The spectra of many everyday sounds tend to change during the sound. To represent this, graphs called 'auditory scenes' or *'spectrographs'* are plotted. These show the frequencies of the sound's components on the vertical axis, and time on the horizontal axis. The darkness of the plotted points tries to represent the dB of components. This sort of display is shown below. The spectra of these particular sounds do not actually change with

spectrograph, or auditory scene



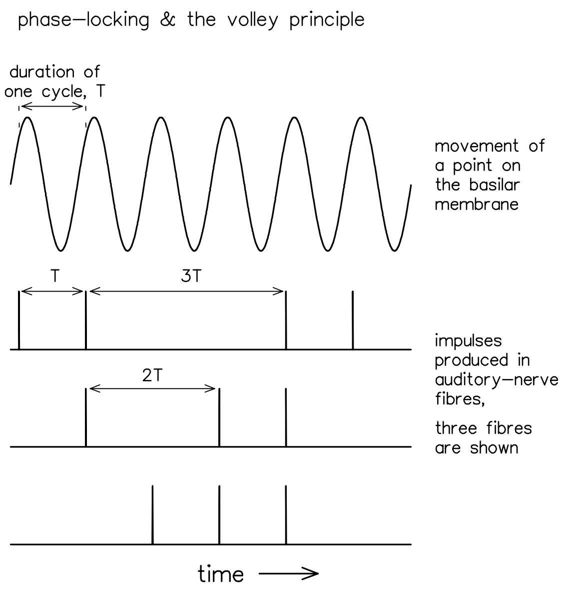
time. However, you can compare this type of display with the waveforms and power spectra of these same sounds that have been shown above.

Basilar membrane. The ear gives us a 'display' that resembles an auditory scene, broadly as shown below. The ear provides a Fourier *analysis* of the sound at each moment of time.



The basilar membrane acts like a *bank of 'auditory filters'*. Each filter picks out a different, narrow band of frequencies from the sound, and sends this information along one of the 25,000 fibres of the auditory nerve. This allocation of components to different fibres is a possible *'place code'* for their frequencies. The fibres are like *'frequency channels'* that are closely spaced and distributed over the entire range of audible frequencies. There is substantial overlap between the frequency ranges of neighbouring channels.

Auditory-nerve firing. The movement at a particular point on the basilar membrane follows the waveform of a frequency component, or a narrow range of frequency components, in the sound. This causes movement of hair cells which 'fire' if they are displaced by a sufficient amount. The firing of a hair cell causes an impulse to travel down the corresponding auditory-nerve fibre. The firing only happens at a particular point in the wave's cycle, as long as the component's frequency is lower than about 5kHz. This *'phase locking'* (Rose et al., 1968) gives rise to time intervals between neural impulses that are either the duration of the wave's cycle, or this duration multiplied by an integer (whole number), as shown below. These intervals are therefore closely related to the component's frequency, so there is here a possible *'time code'* for frequency.



Individually, the fibres shown do not fire on every cycle of the wave, but, taking their firings together as a group, there is an impulse from at least one of them on each cycle. This is called the *'volley principle'* as it allows a group of fibres to fire more rapidly than any one individual fibre.

Frequency resolution. The precision of frequency coding by the ear is called its frequency selectivity, frequency resolving power or frequency tuning. It can be measured in essentially two ways. One method tries to assess activity within a single channel in response to components with different frequencies. This gives a measurement known as the *shape of the auditory filter*. Narrow filter-shapes indicate good frequency resolution. The other method tries to assess activity across different frequency channels in response to a sound with a single, fixed frequency component. This gives what is called an *'excitation pattern'*; whose spread across channels is also narrow if frequency resolution is good. Both of these types of measurement can be made in living breathing people, albeit indirectly. The methods rely on the perceptual phenomenon of masking, which seems largely to be determined by characteristics of the early stages of auditory processing that have been described here. These and other related 'psychoacoustic' experiments will be considered in the next lecture.

Analysis & synthesis. Although sounds often have several frequency components, we do not generally experience the components as separate entities when the sound is played. This indicates that the ear's frequency analysis is accompanied by a synthesis of the products of analysis. Auditory phenomena that come about through this synthesis will also be considered in subsequent lectures.

Appendix 1 Frequency

In graphs representing sounds it is usual to plot frequency on an axis that has a log (logarithmic) scale. This gives equal frequency ratios an equal spatial separation. Thus, the spacing between 1kHz and .5kHz, a ratio of 2 to 1, is the same as the spacing between 10kHz and 5kHz. Similarly, the 10 to 1 ratio of 1kHz and .1kHz gives these two frequencies the same spacing as that between 10kHz and 1kHz.

It is said that a pair of notes in music is separated by a musical interval, and it is the ratio of the frequencies of the two notes that defines the interval. Thus, the ratio 2 to 1 is called an octave. This is divided into 12 equal intervals called semitones. Eight of these 12 steps are the notes of the diatonic scale (the white notes on a piano; doh, re, me, etc.). Four semitones are one third of an octave, which is an interval with a frequency ratio of 1.25 to 1.

Appendix 2 Decibels

Decibels are not plotted on a log scale on graphs representing sounds. However, the dB is found by taking the logarithm of a ratio of two sound-pressures. It is thus a logarithmic unit.

In determining the dB of a sound, the first step is to find what is called the 'root mean square pressure', p(rms), from the sound's waveform. This is found by considering amplitudes (pressures) at points on the waveform that are closely spaced in time. The square of each value is found, and then the mean (average) of these squared values is taken over a reasonably long interval of time. The square root of this mean is p(rms). The next step is to divide p(rms) by a 'reference pressure'. Different dB scales have different reference pressures so they are relative scales like those for temperature. Thus, different numbers from different dB scales can describe the same sound level. For sensation level, dB(SL), the reference pressure is the p(rms) of a sine wave when it is at threshold for a human listener. In the scientific measurement of sound the reference pressure is a standardised value, 0.00002 Newtons per square metre. The scale based on this universal standard is called sound pressure level, dB(SPL). Finally, p(rms) of the sound is divided by the reference pressure, the log (base 10) of this ratio is taken, and this logarithm is multiplied by 20 to give the dB value.

Further reading. There are several suggestions on the course reading list.

A. J. Watkins, April 2002