Sound 2

Sound: the technicalities

Physical characteristics of sound (I)

Sound (to a physicist)

- is a pressure wave which travels in air at about 331m/s
  - (at 0 degrees: at 343m/s at 20 degrees C)
- with a frequency between 20 and 20,000 Hz (variations/second)

To a Psychologist…

- Sound is a perceptual effect caused by a pressure wave of between 20 and 20,000Hz being detected at the ear.

Physical characteristics of sound (II)

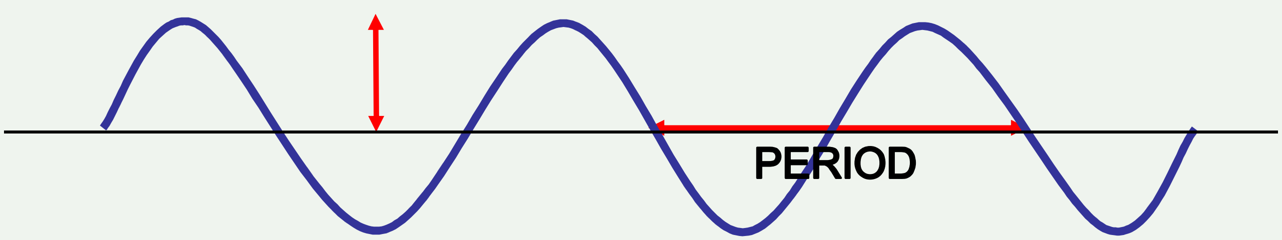
The pressure wave has two physical characteristics:

Amplitude

- the size of the pressure wave – strength of the rarefactions and compressions in the pressure wave

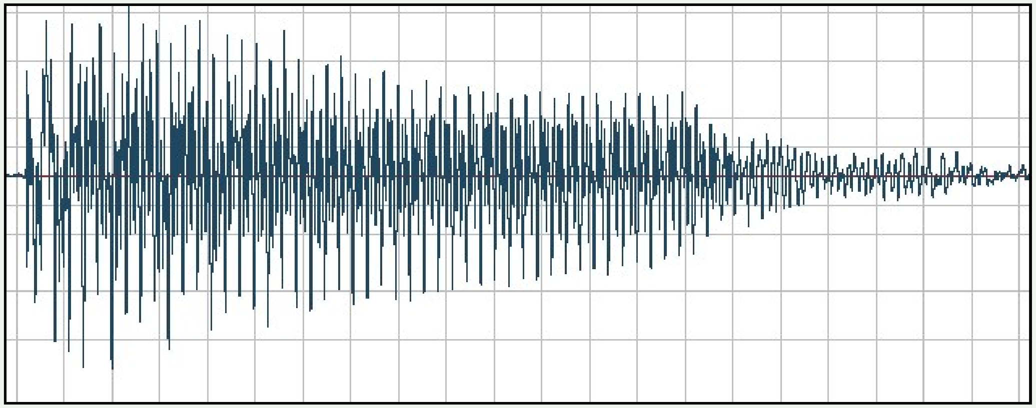
Frequency

- the number of compressions (or rarefactions) per second
- related to
  - the **period** of the sound, 1/frequency, and
  - the wavelength of the sound, speed/frequency

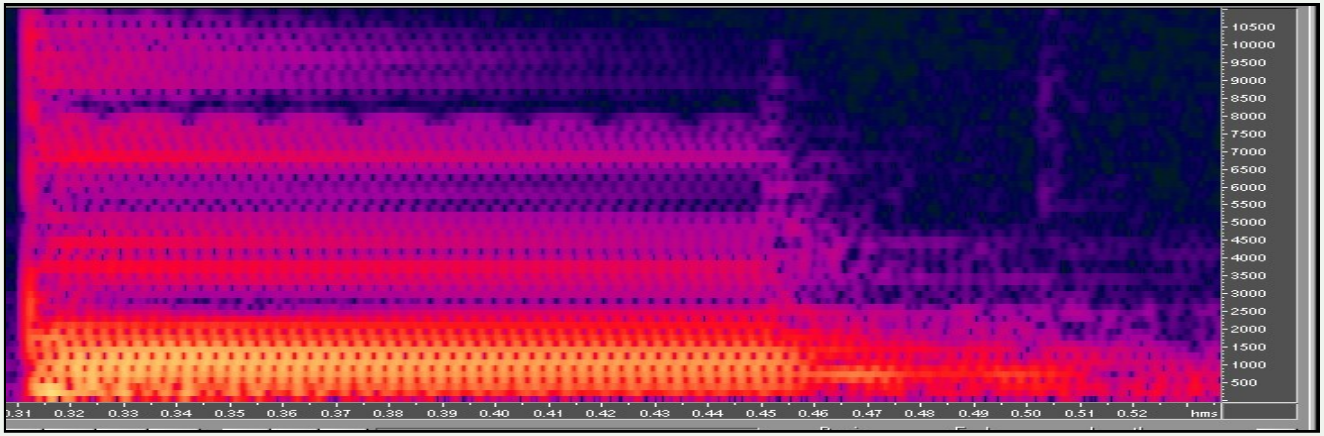


Characteristics of real sounds

Sound waveform: plucked guitar



Frequency spectrum



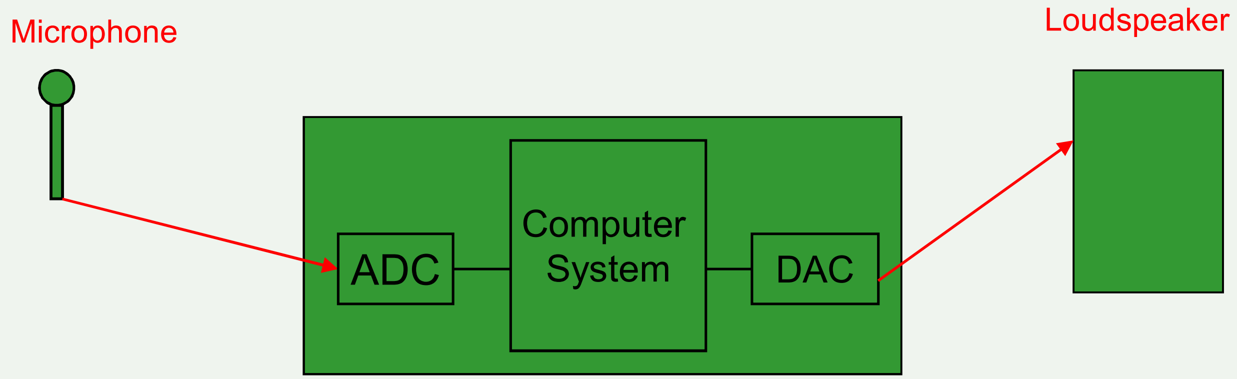
Devices for sound generation and transduction

For input to a computer, the pressure wave is

- converted to an analogue electrical signal (transduced)
- converted to a digital signal (digitised)

For output from a computer, the digitised signal is

- converted to an analogue signal
- converted to a pressure wave



Psychological characteristics of sound

From the perspective of sound being what we hear, sound has three defining characteristics:

- loudness: how intense the sound is perceived
- pitch: the sense of the sound having a tone
- timbre: the nature of the sound

As befits psychological descriptions, these are inexact.

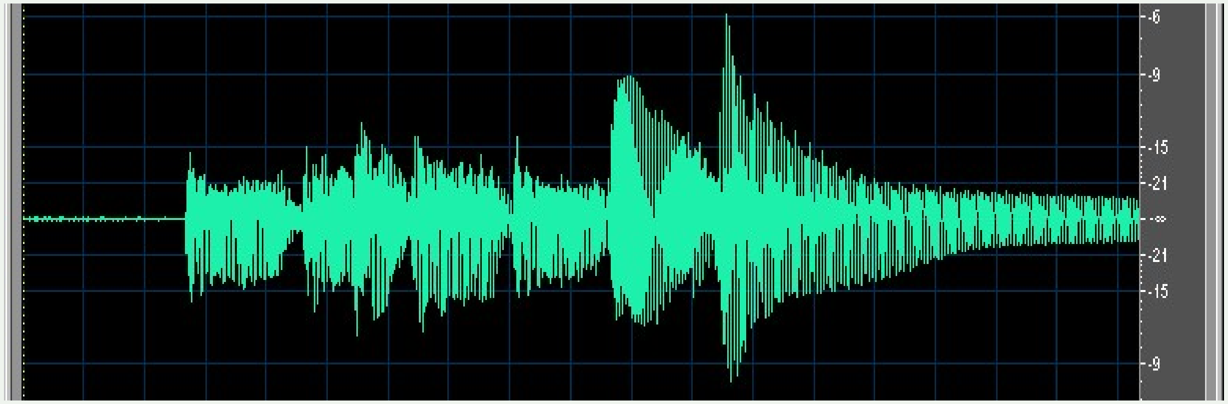
All sounds have a ***loudness***,

- but many have no ***pitch***

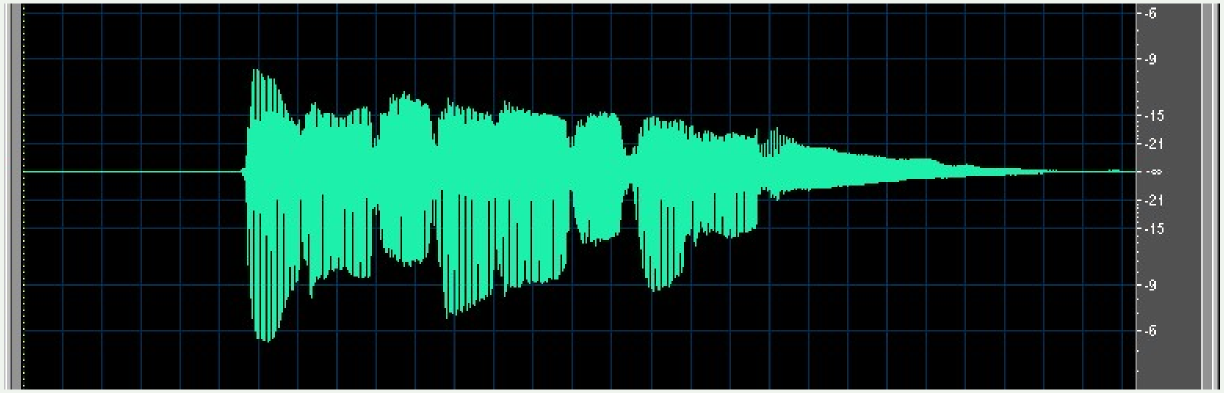
Timbre is often used as a catch-all term to describe those aspects of the sound not captured by loudness and pitch.

Some more real sounds

Spanish guitar



Saxophone



Pitch and Loudness

Pitch perception is complex

- Complex tones (many frequency components) often have a lower pitch than a pure tone of the same mean frequency
- Indeed, a low pitched tone may consist entirely of energy at high frequencies.

Apparent loudness of a sound depends on the frequency as well as the amplitude of the sound

- human ear responds differently to different frequencies
- young people can often hear higher frequencies than older people.

Measuring Loudness

Our ears have (essentially) a logarithmic response

- loudness depends on power:
  - proportional to (amplitude * amplitude)
- doubling the power of a sound does not make it twice as loud
- actually, (real, perceptual) loudness is difficult to compute

Decibels

- ratio of the power of two signals is measured in decibels (dB)
- this is a logarithmic scale
- if signal 1 has power P1, and signal 2 has power P2, then
- P2 is $\mathbf{10 \log_{10}(P2/P1)}$ dB louder than P1
- e.g. If P2 has 100 times the power of P1, it is 20dB louder

Measuring Loudness

Again: **10 $\log_{10}$(P2/P1) dB**

0 dB is threshold for a human to hear a sound of 1000Hz (**P1**)

20dB whisper

90dB loud music

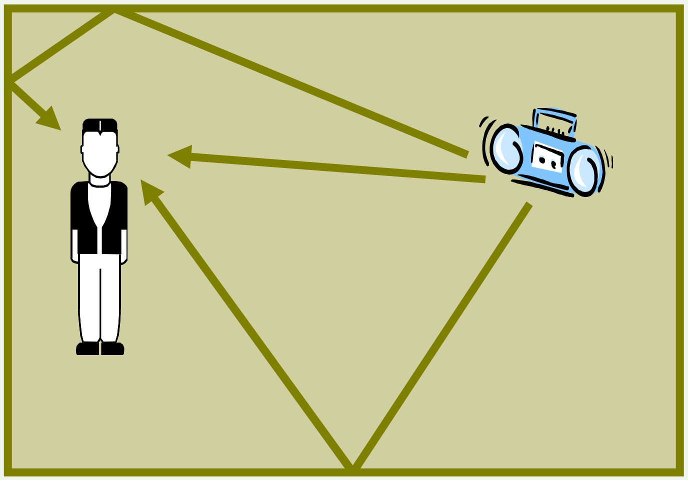
100dB risking damage

140dB aeroplane engine at close range

Perceived Sound Source Direction

Sound from a single source appears to come from that source

- whether it's a musical instrument or a single loudspeaker
- even although it gets reflected off walls etc.



The real sound field comes from many sources

- but human auditory scene analysis allows us to detect multiple sources, and to concentrate on one of them at a time

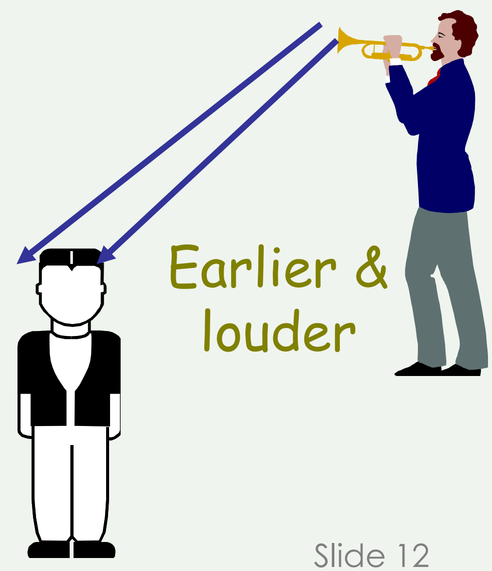
How does this happen?

We appear to use

- information in the fine time structure of monaural sounds to group sounds together
- the differences in timing and spectral intensity between the two ears to allow the listener to analyse the **auditory scene**

Physical correlates of direction are

- for horizontal direction
  - IID: interaural intensity difference
  - ITD: interaural time difference
- for front/back, elevation: spectral shape
- for distance: spectral shape, reflections



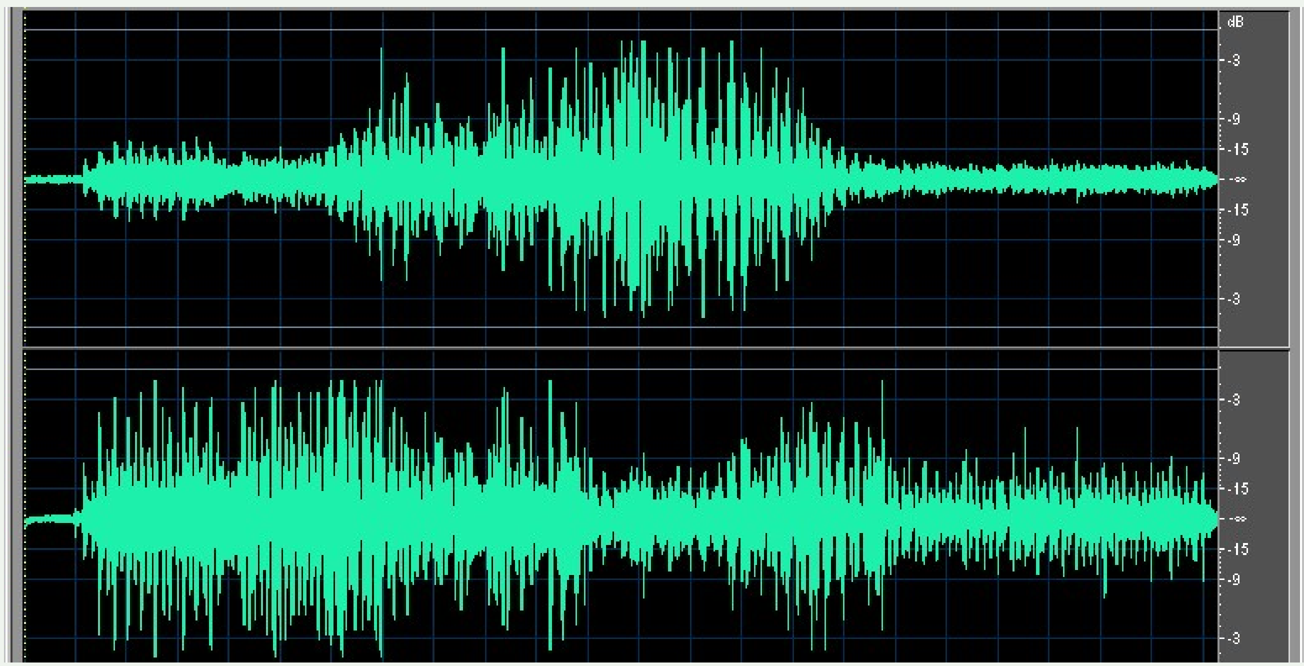
Synthesising Directional Sounds

Synthesised sound can be made to appear to come from particular directions

- original sound is modified, and different sounds played to each ear to create this illusion.
- The way in which a particular sound gets modified by the head on its way to the ear (or eardrum) is called the **Head-related Transfer Function** …and there's one for each ear

Binaural Sound

Sound recorded using a synthetic head



Head related transfer functions (HRTF)

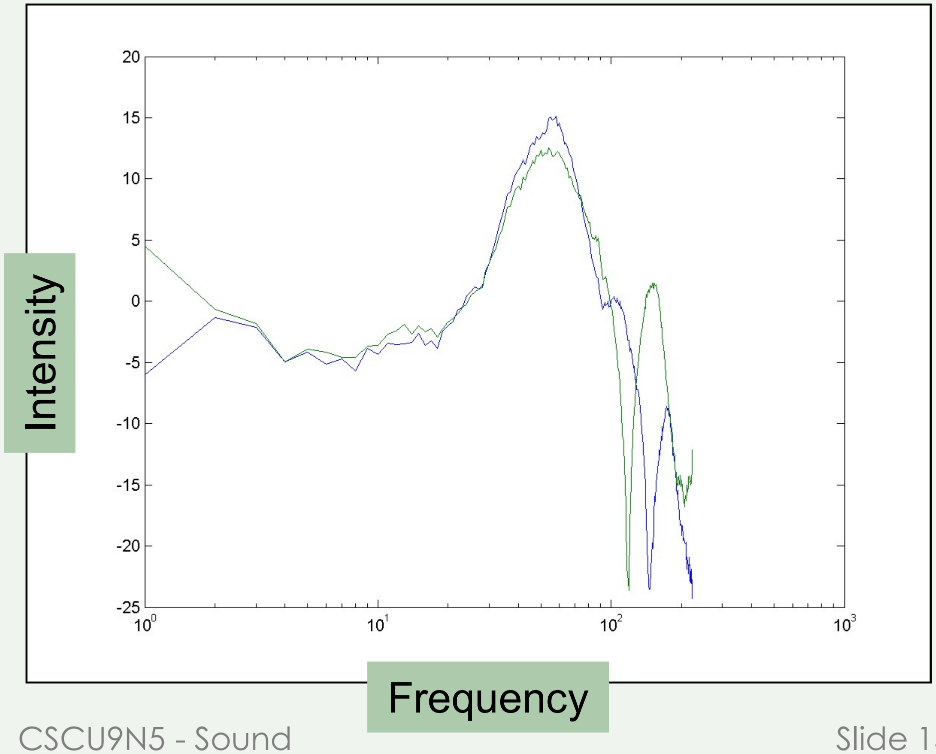
When a sound is played, the stimulus received depends on the angle of the stimulus …and it is different at each ear

HRTF for each ear:

Stimulus straight ahead

X-axis is frequency

Y axis intensity



Head related transfer functions (2)

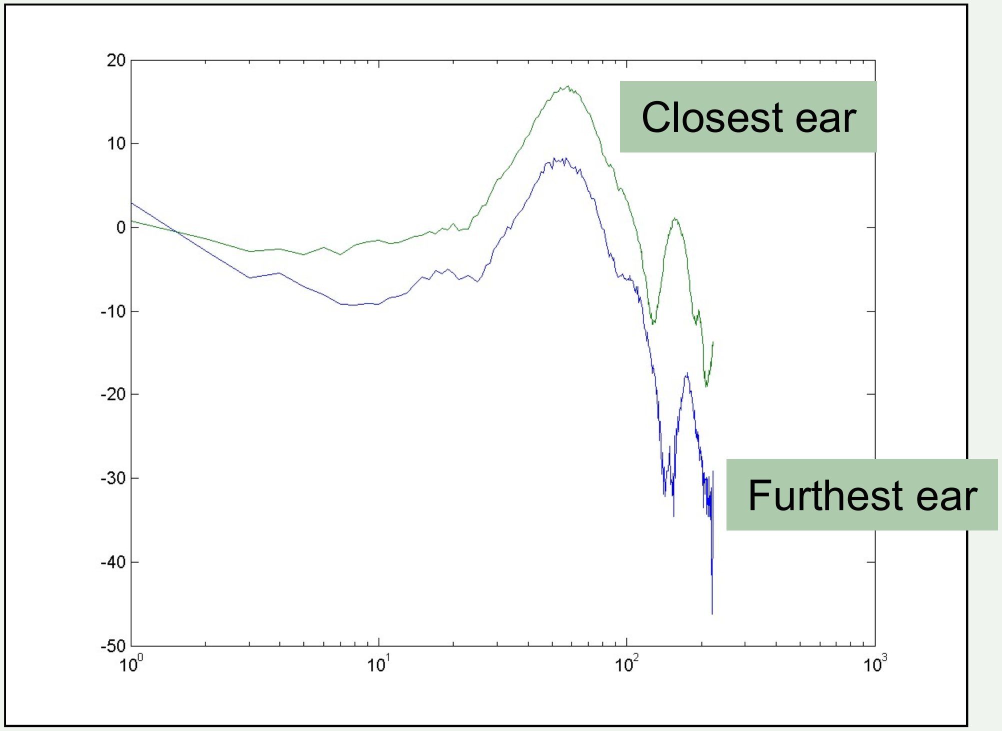
HRTF for each ear:

Stimulus 30 degrees

X-axis is frequency

Y axis intensity

Head shadow reduces intensity and can alter frequency spectrum



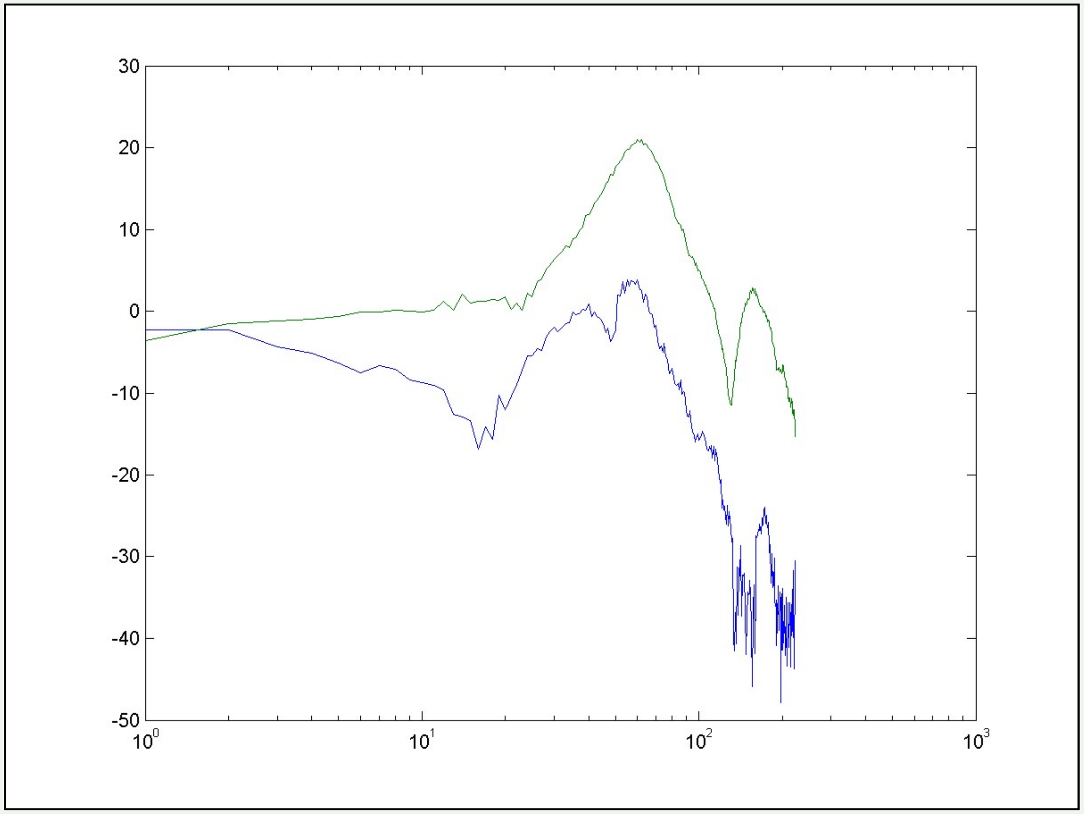
Head related transfer functions (3)

HRTF for each ear:

Stimulus 60 degrees

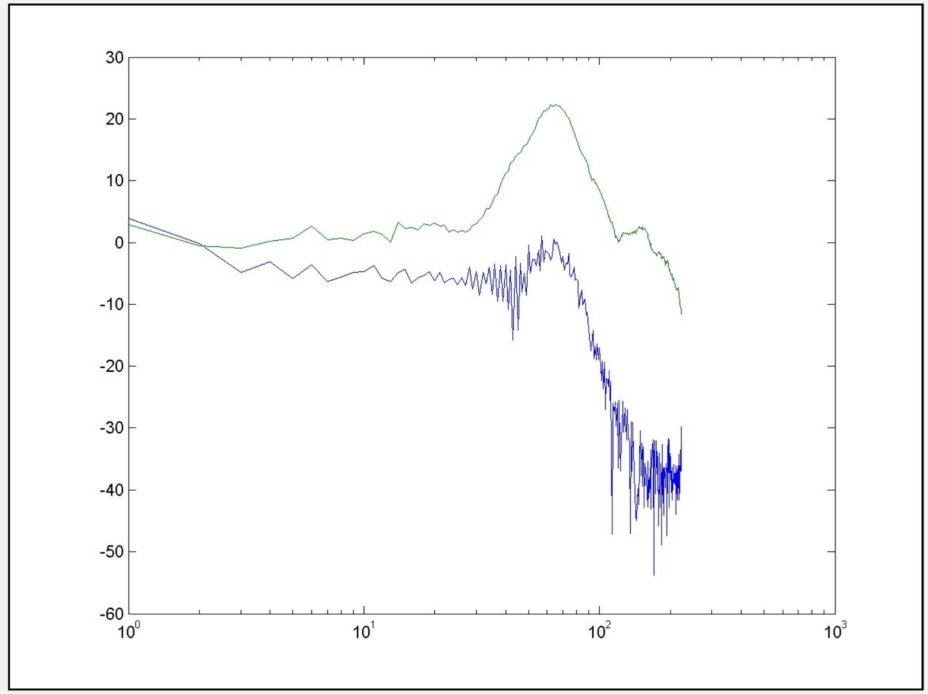
X-axis is frequency

Y axis intensity



Head related transfer functions (4)

HRTF for each ear:
Stimulus 90degrees
X-axis is frequency
Y axis intensity.
Note how the main difference is above 1000 Hz.



By modifying the original sound to mimic the HRTF, a sound can be made to appear to come from a particular direction.

Sound Transduction

Whenever sound is transduced, digitised, or reconverted to analogue, the original signal is altered in some way. When high quality reproduction is required, we need to keep this alteration to a minimum.

Transduction:

- Microphones and loudspeakers have a limited frequency response
  - they are more sensitive to sounds with certain frequencies
  - we would like a flat frequency response from 20 to 20KHz
- They also have a limited dynamic range
  - they cannot deal with sounds from the quietest up to the loudest
  - the range in energy of everyday sounds is huge

For some applications, we may sacrifice quality

- e.g. telephony: we care really only about comprehensibility

Digitising Sound

- Sound is digitised using an analogue to digital converter (ADC)
- Sound is converted back to analogue using a digital to analogue converter (DAC)
- Both forms of conversion can introduce alterations in the sound
  - but the ADC is the more problematic.

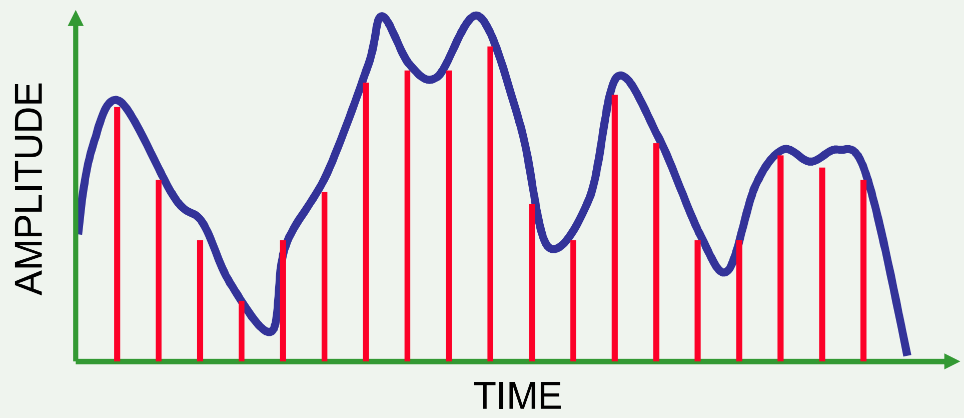
Analogue to digital conversion has two parameters:

  - **sampling rate**
  - **sample size**

Sampling Rate

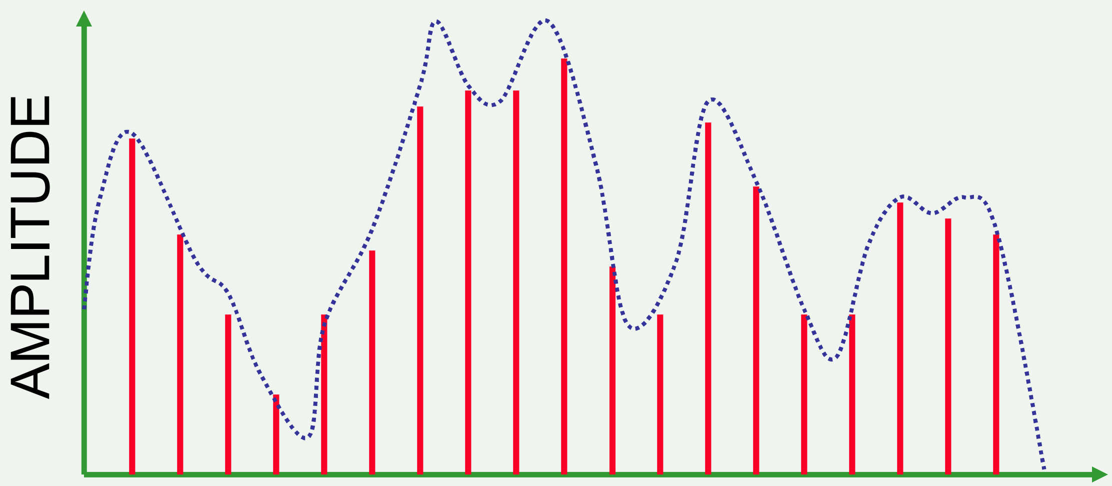
Sampling rate describes how frequently the analogue signal is converted

- Normally measured in samples/second
  - conversion is done regularly, at a fixed number of samples/second
- sampling rate must be at least twice the highest frequency of interest (Nyquist sampling theorem) otherwise aliasing can occur



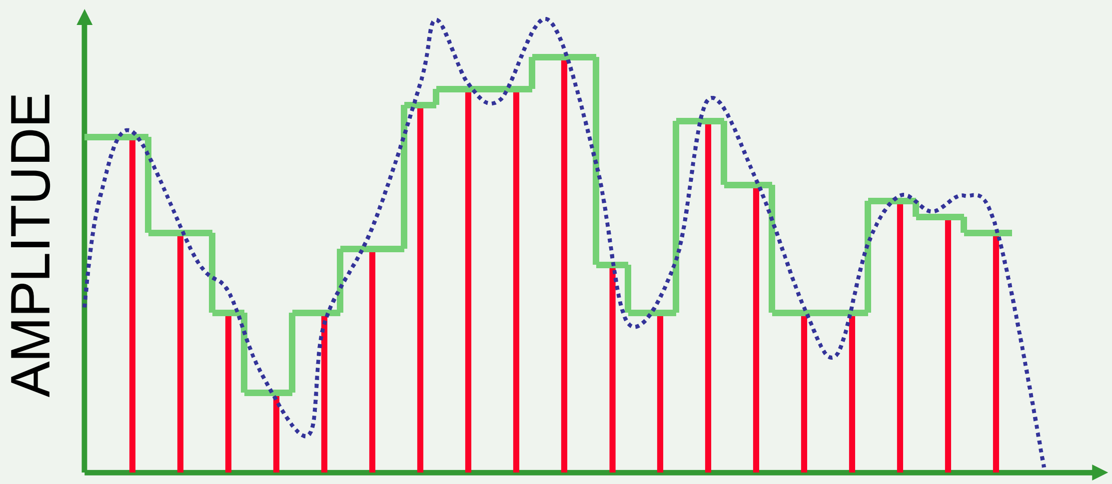
Signal Reconstruction

Quantization



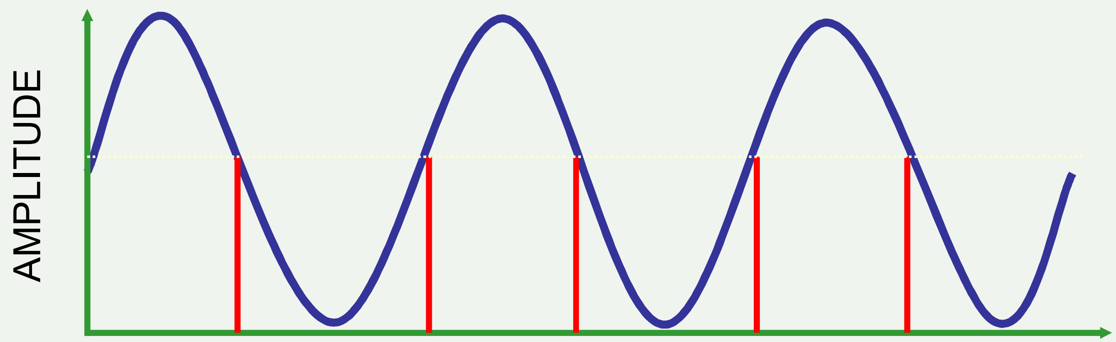
Signal Reconstruction

Sample and hold reconstruction

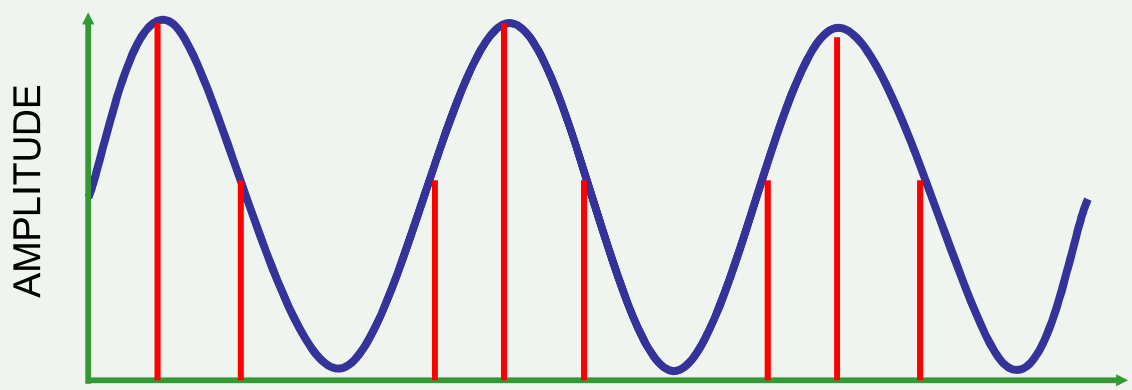


Aliasing

Aliasing occurs if a sound is sampled too slowly



Better...

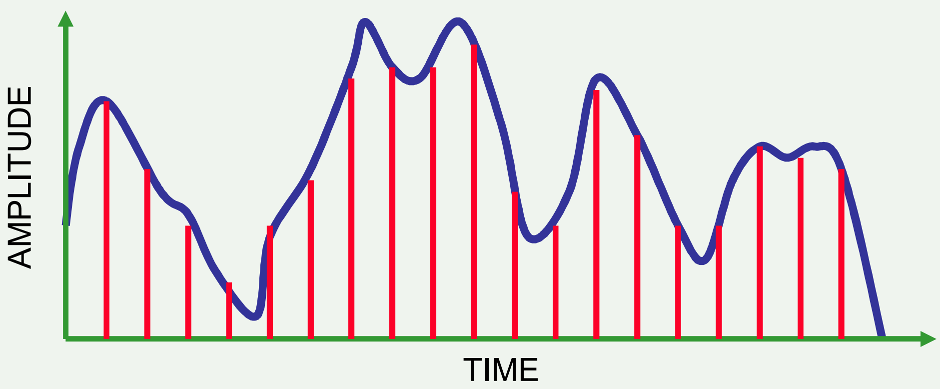


Sample size (I)

Sample size refers to the characteristics of the sample value taken each sample time, e.g. amplitude

Samples have a fixed length

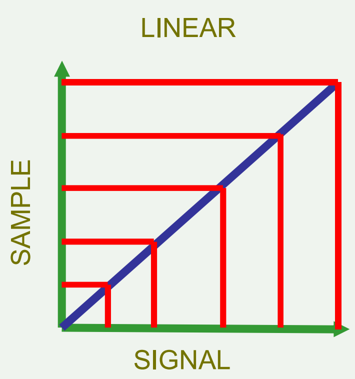
- 8-bit, (16-bit or 32-bit) which means each sample is a 2's complement 8-bit (16-bit or 32-bit) integer
- e.g. range -128 to +127 for 8-bit; -32768 to +32767 for 16-bit

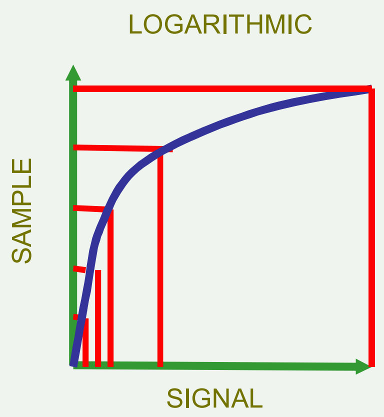


Sample Size (II)

Sampling may be linear or logarithmic

- **linear**: for sample value x, actual value is (x/maximum)* K for some K
- **logarithmic**: provides more resolution at lower levels
  - mu-law (μ-law) or A-law
  - a form of data compression





Sample Size (III)

Major concern for storage of a sampled sound is the total amount of data collected. Data length is proportional to sample rate * sample size

- 1 second of sound sampled at 44,100 16 bit samples/second uses

    44,100 * 2 = 88,200 bytes/second

- that is just 1 channel: stereo takes 176,400 bytes/second
    - about 10.5Mbytes/minute
- this is CD-audio quality

Data can be compressed

- but decompression must take place in real time
- more on sound data compression in the next sound lecture.

Power and Loudness: Dynamic Range

Dynamic range

- loudest measurable signal compared to quietest signal

Measured using decibels

- if signal has power P1, and signal 2 has power P2, then
- P2 is 10 log10(P2/P1) dB louder than P1

For example: 16-bit linear sampling

- Maximum amplitude approx 32000 (and so power is 32000*32000)
- Minimum amplitude is 1 (and so power is 1*1)
- Dynamic range is 10 log10(32000*32000) = 90dB

End of Lecture