© 2019 The University of Sheffield

COM3502/4502/6502
SPEECH PROCESSING

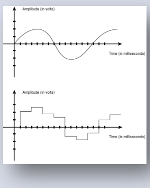
Lecture 13
Digital Signals

The University Of Sheffield.

1

© 2019 The University of Sheffield

Analogue vs. Digital Signals



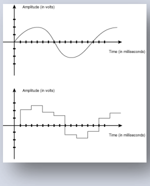
- All natural signals (*including speech*) are 'analogue' …
  - they can assume an infinity of possible values
  - they exist continuously in time
- Perfect processing of analogue signals thus requires infinite resolution and/or infinite storage
- Any practical system (*computers or living organisms*) must sacrifice fidelity in order to process signals using finite resources
- Signals must therefore be 'quantised' in each dimension

The University Of Sheffield.

2

© 2019 The University of Sheffield

Analogue vs. Digital Signals

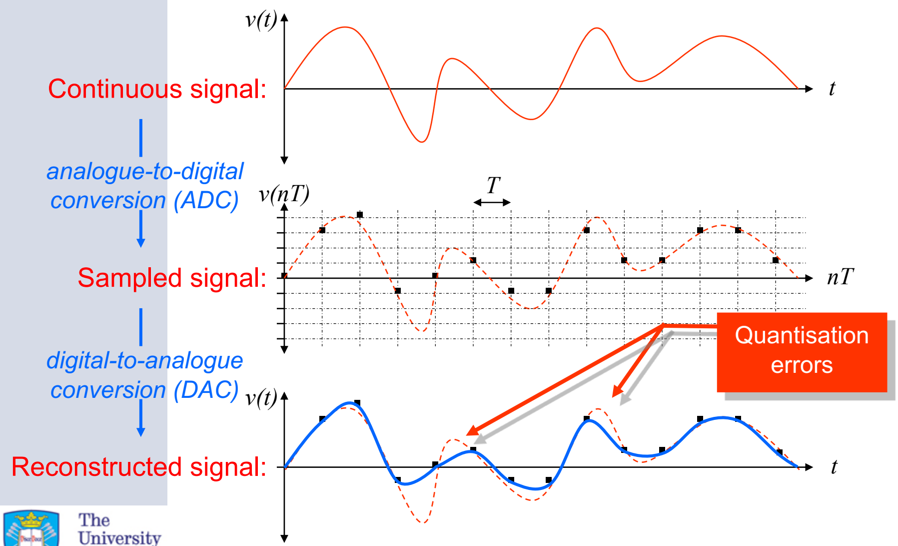


- Quantisation of a signal sequence is achieved by '**sampling**'
- Speech signals (*e.g. from a microphone*) are typically …
  - quantised in amplitude
  - sampled in time
- The process of signal quantisation and sampling is known as:
  '**pulse code modulation**' (PCM)

The University Of Sheffield.

© 2019 The University of Sheffield

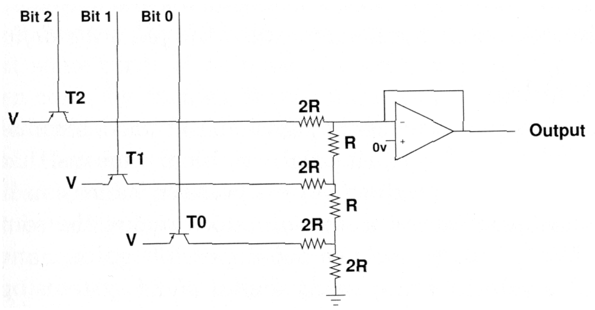
Quantisation and Sampling



The University Of Sheffield.

© 2019 The University of Sheffield

Digital-to-Analogue Conversion

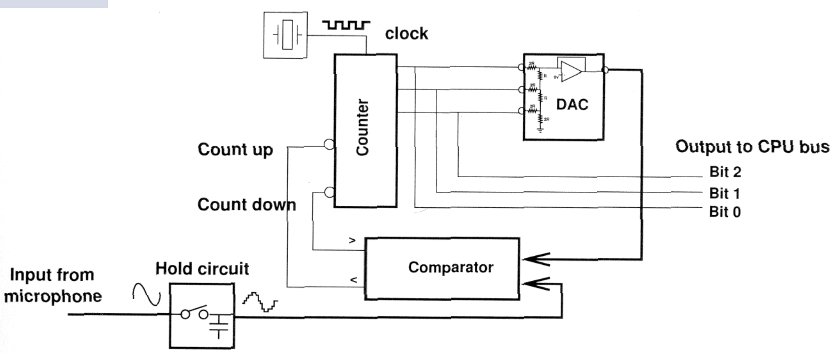


Taken from: Farnell, A. (2008). ***Designing Sound.*** London: Applied Scientific Press Limited.

5

© 2019 The University of Sheffield

Analogue-to-Digital Conversion



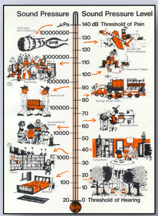
Taken from: Farnell, A. (2008). ***Designing Sound.*** London: Applied Scientific Press Limited.

6

© 2019 The University of Sheffield

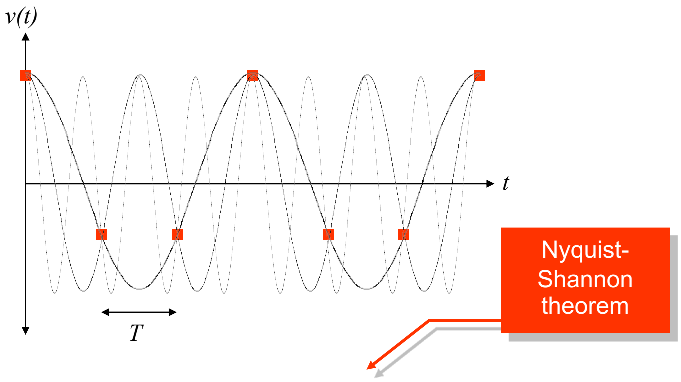
Amplitude Quantisation

- The range of numbers that can be used to represent a signal's amplitude defines the '**dynamic range**' of the system
- If n bits are used to store each amplitude value, then 2^n-1 possible values can be represented
- The dynamic range is $20.log_{10}(2^n-1)$ dB
- E.g. *16* bits of amplitude quantisation gives …
  - 65535 possible values
  - 96 dB dynamic range



© 2019 The University of Sheffield

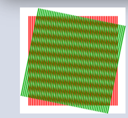
Sampling



To determine the frequency of a sampled sinusoid, you need *at least* two samples per period

© 2019 The University of Sheffield

Sample Rate



- The 'Nyquist-Shannon sampling theorem' states that we need at least *twice* the number of sample points per second as the highest frequency in a signal …

  sampling rate > 2 x *max signal frequency*

- Sampling a signal lower than the Nyquist rate leads to 'aliasing' (*energy at frequencies higher than the sampling rate are reflected back into the lower frequencies*)

- E.g. aliasing can occur in digital images if the spatial frequencies are higher than the pixel resolution (*the result is a 'Moiré pattern'*)

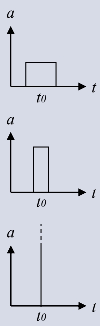
- It is usual to low-pass filter a signal *before* sampling in order to avoid aliasing …

  $f_c < 0.5 f_s$

  where f_c is the filter 'cut-off frequency'
  and f_s is the 'sampling frequency'

© 2019 The University of Sheffield

Sampling Theory



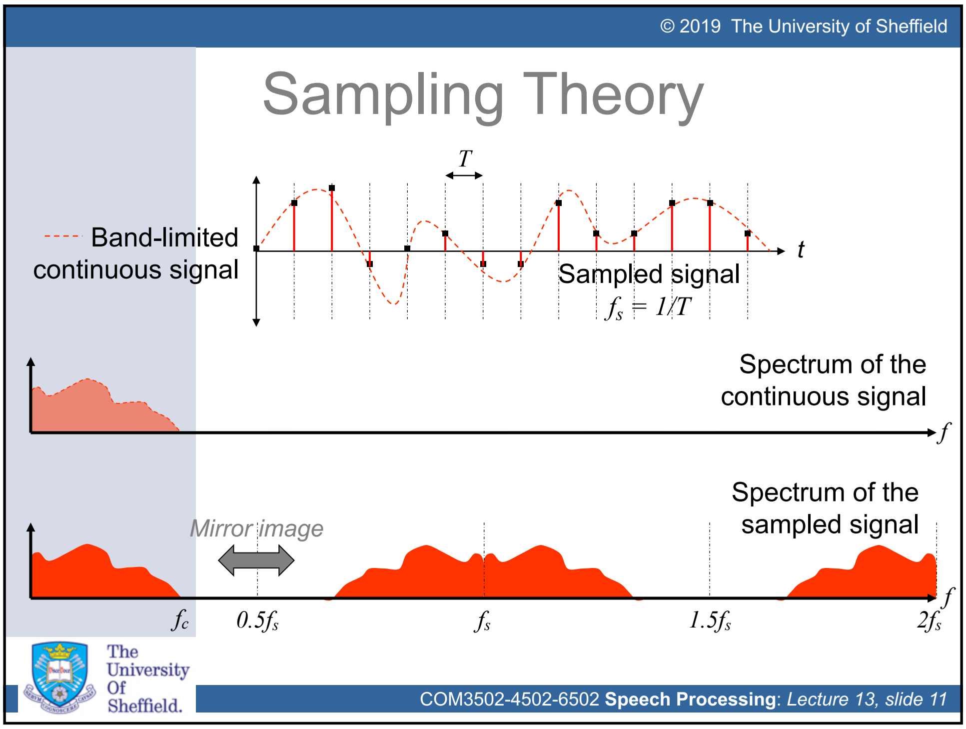
- Sampling a continuous signal $s(t)$ is equivalent to multiplying it by a 'Dirac delta function'

- This impulse is a rectangle of *unit area* centred around t_0 whose width tends to 0

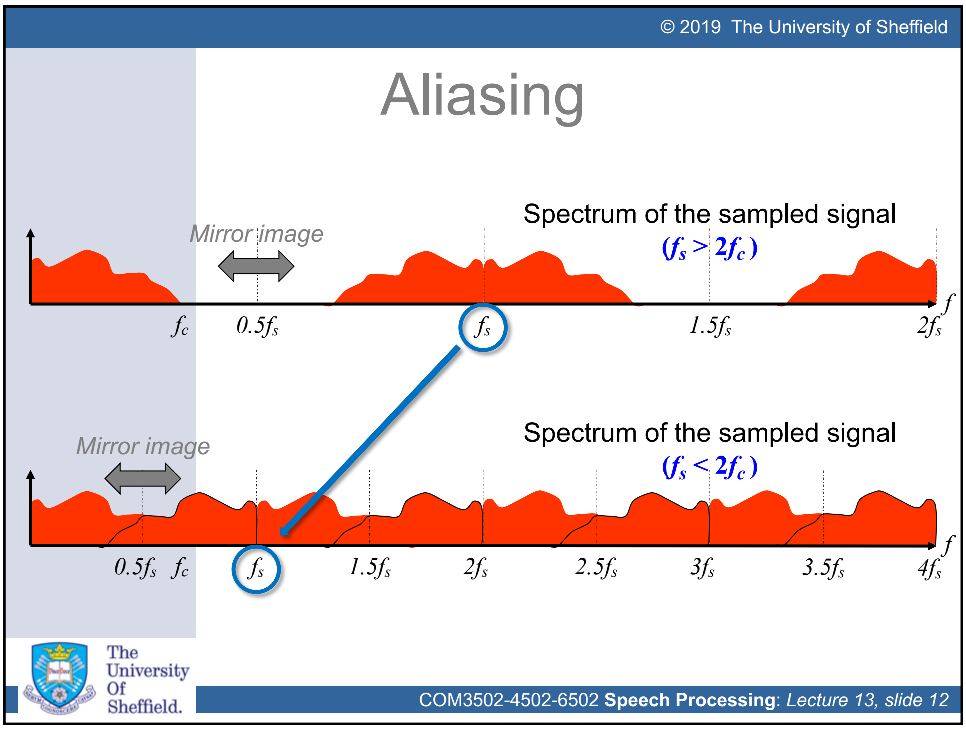
$$\delta(t - t_0) = \begin{cases} \infty, & if\ t = t_0 \\ 0, & if\ t \neq t_0 \end{cases}$$

$$\int_{-\infty}^{\infty} \delta(t)dt = 1$$

$$\int_{-\infty}^{\infty} s(t)\delta(t - t_0)dt = s(t_0)$$


© 2019 The University of Sheffield
Sampling Theory
Band-limited continuous signal
T
t
Sampled signal
$f_s = 1/T$
Spectrum of the continuous signal
f
Spectrum of the sampled signal
Mirror image
f_c
$0.5f_s$
f_s
$1.5f_s$
$2f_s$
f
The University Of Sheffield.
COM3502-4502-6502 **Speech Processing**: *Lecture 13, slide 11*

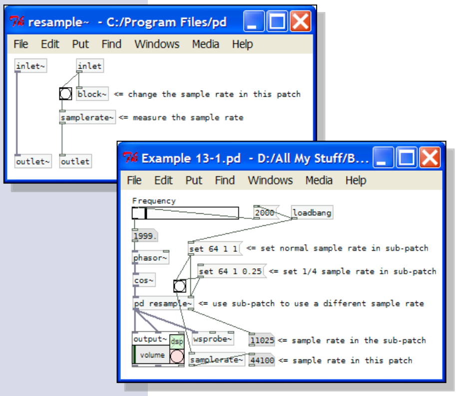
11


© 2019 The University of Sheffield
Aliasing
Spectrum of the sampled signal
$(f_s > 2f_c)$
Mirror image
f_c
$0.5f_s$
f_s
$1.5f_s$
$2f_s$
f
Spectrum of the sampled signal
$(f_s < 2f_c)$
Mirror image
$0.5f_s$
f_c
f_s
$1.5f_s$
$2f_s$
$2.5f_s$
$3f_s$
$3.5f_s$
$4f_s$
f
The University Of Sheffield.
COM3502-4502-6502 **Speech Processing**: *Lecture 13, slide 12*

12

© 2019 The University of Sheffield

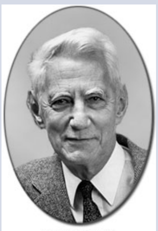
Aliasing



The University Of Sheffield.

© 2019 The University of Sheffield

Coding Theory



Claude Shannon (1916-2001)

- Derived from 'information theory' (*founded in 1948 by Claude Shannon*)
- Based on probability theory and statistics
- The most important quantities of information are …
  - 'entropy' (*the information in a random variable*)
  - 'mutual information' (*the amount of information in common between two random variables*)
- Information is usually expressed in bits
  - entropy indicates how easily data can be 'compressed' (*due to redundancy*)
  - mutual information can be used to find the communication rate through a channel

The University Of Sheffield.

© 2019 The University of Sheffield

Coding Theory

Entropy …

$$H(X) = -\sum_{x \in X} p(x) \log p(x)$$

Mutual information …

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

The University Of Sheffield.

© 2019 The University of Sheffield

Digital Coding: *Images*

- Example image:
  - 360 pixels wide
  - 480 pixels high
  - 3 colours (*RGB*)
  - 1 byte/colour
- Raw data (*bitmap*):
  - 518.4 kilobytes (*360 x 480 x 3*)
- 'Lossless' coding (*LZW-TIF*)
  - 378 kilobytes
- 'Lossy' coding (*jpeg*)
  - 31 kilobytes

The University Of Sheffield.

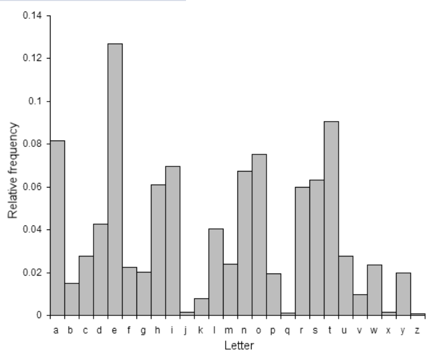
© 2019 The University of Sheffield

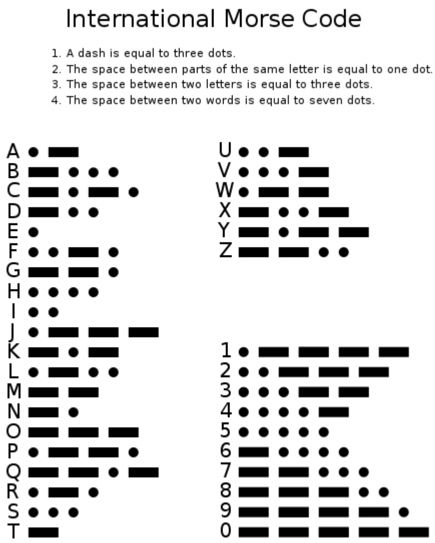
Digital Coding: *Images*

17

© 2019 The University of Sheffield

Digital Coding: *Language*





18

© 2019 The University of Sheffield

Digital Coding: *Signals*

- The amount of information needed to 'encode' a signal is defined by …
  - amplitude quantisation (*in bits/sample*)
  - sampling rate (*in samples/second*)
- Digital signals are thus characterised in terms of their 'data rate' (*in bits/second - bps*)
  - ethernet LAN = 10 Gbps
  - wireless LAN = 600 Mbps
  - ADSL modem = 24 Mbps
  - 4G mobile data = ~10 Mbps
  - 3G mobile data = ~4 Mbps

Note
b = bits
B = bytes

The University Of Sheffield.

19

© 2019 The University of Sheffield

Digital Coding: *Speech*

- Speech has …
  - a bandwidth of ~10 kHz
  - a dynamic range of ~50 dB
- Hence, the minimum quantisation and sampling requirements would seem to be …
  - 20 kHz sampling rate
  - 8 bit quantisation
  - i.e. **160 kbps**
- However, it is possible to reduce both the bandwidth and dynamic range significantly before suffering a major drop in speech intelligibility
  - e.g. a telephone has a bandwidth of ~300 Hz to ~3.5 kHz (*which is a problem for sounds such as* [f] *and* [s])
- Digital speech 'codecs' make good use of lossy compression schemes (*by exploiting the 'source-filter' model of speech*)

Note
b = bits
B = bytes

The University Of Sheffield.

20

© 2019 The University of Sheffield

Audio/Speech Codecs

- DVD audio (*24 bit 192 kHz PCM*) = 4,608 kbps
- CD audio (*16 bit 44.1 kHz PCM*) = 705.6 kbps
- 16 bit 16 kHz PCM = 256 kbps
- telephone (*8 bit 8 kHz ADPCM*) = 64 kbps
- MP3 of these lectures = 24 kbps
- mobile phone (*GSM CELP*) = 13 kbps
- VOIP (*low rate*) = 8 kbps
- NATO vocoder (*channel*) = 2400 bps
- NATO vocoder (*LPC10e*) = 2400 bps
- NATO vocoder (*MELP*) = 1200 bps

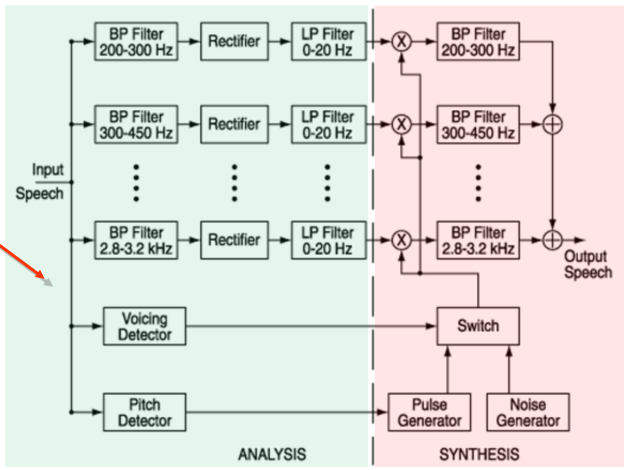
Note
b = bits
B = bytes

"vocoder" = "voice coder"

The University Of Sheffield.

© 2019 The University of Sheffield

Channel Vocoding



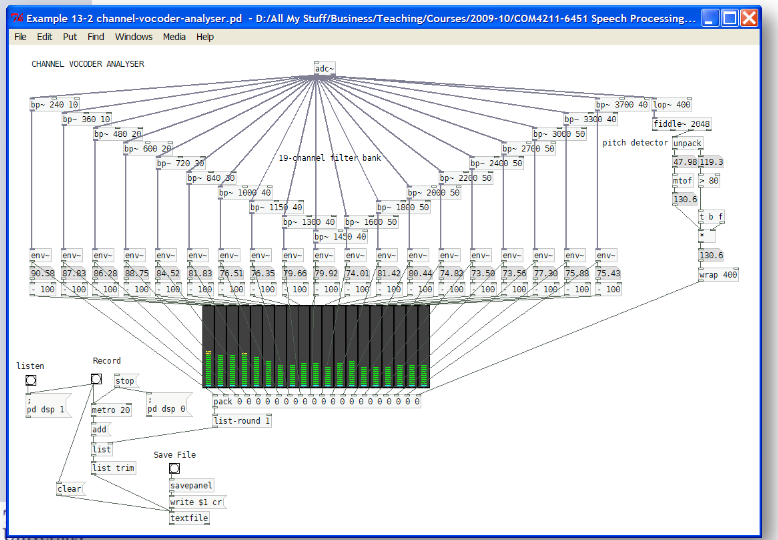
Exploits 'source-filter' separation

Flanagan, J. L., .Schroeder, M., Atal, B., Crochiere, R., Jayant, N., & Tribolet, J. (1979). Speech coding. *IEEE Trans. Communications*.

The University Of Sheffield.

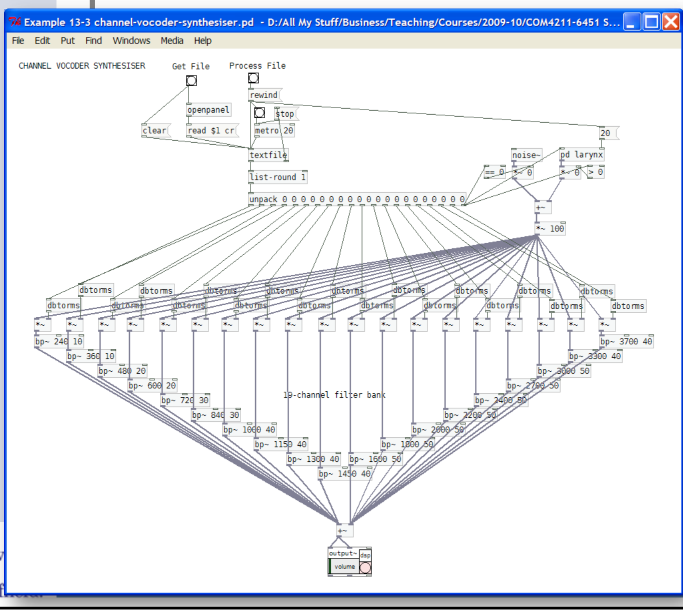
© 2019 The University of Sheffield

A 6 kbps Channel Vocoder in Pd



© 2019 The University of Sheffield

A 6 kbps Channel Vocoder in Pd



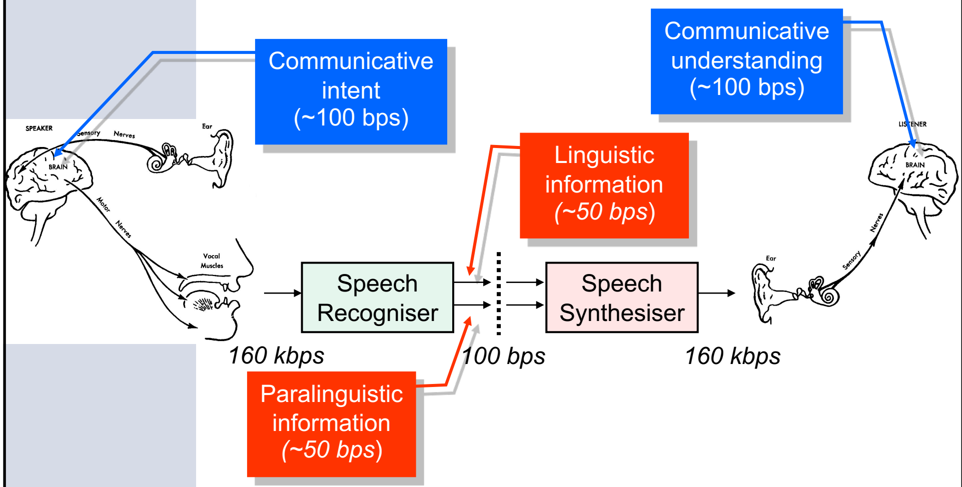
© 2019 The University of Sheffield

Digital Speech Coding

- The ‘information rate’ in speech is estimated to be only ~100 bps!
  - linguistic information = ~50 bps
  - paralinguistic information = ~50 bps
- So why do we need kbps vocoders?
- The way to code signals at lower rates is to exploit any ‘redundancies’ in the signal
- For speech, this is achieved using predictive models (*more on this in later lectures*)
- The ultimate predictive model for speech is ‘speech recognition’ + ‘speech synthesis’

© 2019 The University of Sheffield

Very Low Rate Speech Coding



© 2019 The University of Sheffield

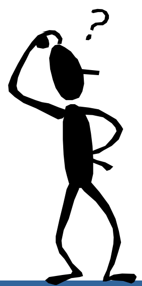
This lecture has covered …

- Analogue vs. digital
- Quantisation and sampling
- A-to-D and D-to-A conversion
- Dynamic range
- Sampling theory
- Aliasing
- Coding theory
- Audio and speech coding

The University Of Sheffield.

© 2019 The University of Sheffield

Any Questions ?



The University Of Sheffield.

© 2019 The University of Sheffield

Next time …

Waveform Processing