

© 2019 The University of Sheffield

# COM3502/4502/6502 SPEECH PROCESSING

## Lecture 15 The Fourier Transform

The University Of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 15, slide 1

1

© 2019 The University of Sheffield

## Phase in Spectral Analysis

- In the last lecture it was shown how the spectrum could be computed by 'cosine correlation' (Pd Example 14-6)
- However, this only worked because the target signals all had 'zero phase'
- The correlation between two cosines varies according to the phase difference between them
  - in-phase → maximum correlation
  - 90° phase difference → zero correlation
  - 180° phase difference → maximum negative correlation
- In fact the correlation between two cosines varies as a cosine!

The University Of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 15, slide 2

2

© 2019 The University of Sheffield

## Phase in Spectral Analysis

**Multiplication of two cosines** →  $\cos(a) \cdot \cos(b) = \frac{1}{2}(\cos(a-b) + \cos(a+b))$

**Multiplication of two cosine waves** →  $\cos(\omega t) \cdot \cos(\omega t) = \frac{1}{2}(\cos(0) + \cos(2\omega t))$   
 $= \frac{1}{2}(1 + \cos(2\omega t))$  → Constant + cosine at twice the frequency

**Multiplication of two cosine waves differing in phase** →  $\cos(\omega t) \cdot \cos(\omega t - \phi) = \frac{1}{2}(\cos(\phi) + \cos(2\omega t - \phi))$   
 → Cosine of the phase difference + cosine at twice the frequency

The University Of Sheffield. COM3502-4502-6502 Speech Processing: Lecture 15, slide 3

3

© 2019 The University of Sheffield

## Phase in Spectral Analysis

**Correlation of two cosine waves differing in phase** →  $q = \sum_{i=0}^{N-1} \cos(\omega n T) \cdot A \cos(\omega n T - \phi)$   
 $= \sum_{i=0}^{N-1} \left( \frac{A}{2} (\cos(\phi) + \cos(2\omega t - \phi)) \right)$   
 $= \sum_{i=0}^{N-1} \left( \frac{A}{2} (\cos(\phi)) \right) + \sum_{i=0}^{N-1} \left( \frac{A}{2} (\cos(2\omega t - \phi)) \right)$   
 $= \sum_{i=0}^{N-1} \left( \frac{A}{2} (\cos(\phi)) \right)$   
 $= \frac{A \cdot N}{2} \cos(\phi)$  → Cosine of the phase difference  
 $= \alpha \cos(\phi)$

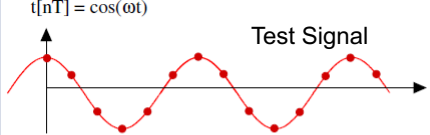
The University Of Sheffield. COM3502-4502-6502 Speech Processing: Lecture 15, slide 4

4

© 2019 The University of Sheffield

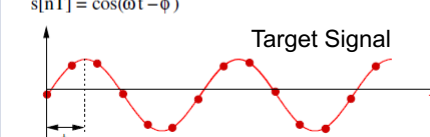
## Phase in Spectral Analysis

$t[nT] = \cos(\omega t)$



Test Signal

$s[nT] = \cos(\omega t - \phi)$



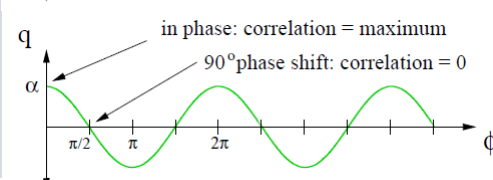
Target Signal

$\phi$

$Q$

in phase: correlation = maximum


$90^\circ$  phase shift: correlation = 0



$\alpha$

$\pi/2$     $\pi$     $2\pi$     $\phi$

This means that cosine correlation cannot detect a  $\pi/2$  ( $90^\circ$ ) phase shift in a target signal!



The University Of Sheffield.

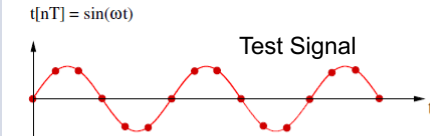
COM3502-4502-6502 Speech Processing: Lecture 15, slide 5

5

© 2019 The University of Sheffield

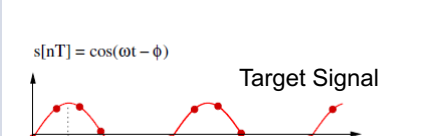
## Sine Correlation

$t[nT] = \sin(\omega t)$



Test Signal

$s[nT] = \cos(\omega t - \phi)$



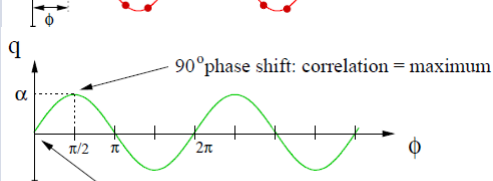
Target Signal

$\phi$

$Q$

$90^\circ$  phase shift: correlation = maximum


in phase: correlation = 0



$\alpha$

$\pi/2$     $\pi$     $2\pi$     $\phi$

The correlation between a cosine and a sine is a sine!  
(and it varies according to the phase difference between them)



The University Of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 15, slide 6

6


© 2019 The University of Sheffield

## Sine Correlation

Correlation of sine wave with cosine wave

$$\begin{aligned}
 q &= \sum_{i=0}^{N-1} \sin(\omega nT) \cdot A \cos(\omega nT - \phi) \\
 &= \sum_{i=0}^{N-1} \cos(\omega nT - \pi/2) \cdot A \cos(\omega nT - \phi) \\
 &= \sum_{i=0}^{N-1} \left( \frac{A}{2} (\cos(\phi - \pi/2) + \cos(2\omega nT - \phi - \pi/2)) \right) \\
 &= \sum_{i=0}^{N-1} \left( \frac{A}{2} (\sin(\phi) + \sin(2\omega nT - \phi)) \right) \\
 &= \sum_{i=0}^{N-1} \left( \frac{A}{2} (\sin(\phi)) \right) + \sum_{i=0}^{N-1} \left( \frac{A}{2} (\sin(2\omega nT - \phi)) \right) \\
 &= \sum_{i=0}^{N-1} \left( \frac{A}{2} (\sin(\phi)) \right) \\
 &= \alpha \sin(\phi)
 \end{aligned}$$

Sine of the phase lag



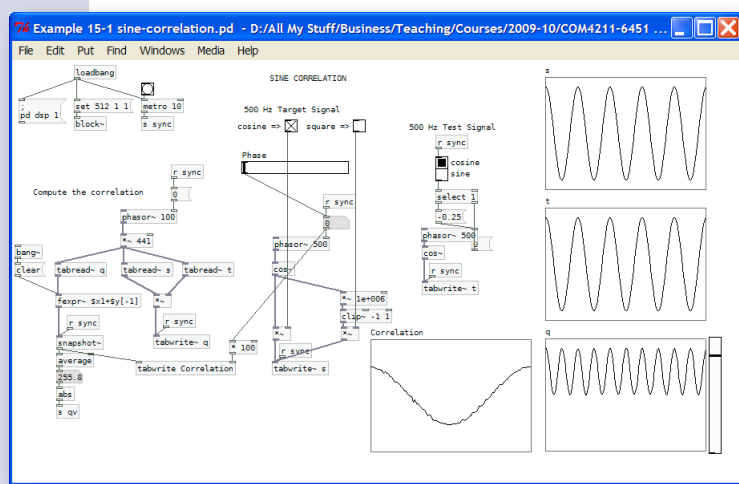
The University of Sheffield.


COM3502-4502-6502 Speech Processing: Lecture 15, slide 7

7

© 2019 The University of Sheffield

## Sine Correlation





The University of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 15, slide 8

8

© 2019 The University of Sheffield

## Sine and Cosine Correlation

- Cosine correlation with cosine is a cosine
- Sine correlation with cosine is a sine
- Sine and cosine correlations are  $90^\circ$  ( $\pi/2$ ) out of phase
- So can cosine and sine correlation be combined in some useful way?

The University Of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 15, slide 9

9

© 2019 The University of Sheffield

## Amplitude and Phase

- Correlate a signal with *both* sines *and* cosines ...
 
$$\begin{aligned} \text{cosinecorrelation} &= \alpha \cos(\phi) \\ \text{sinecorrelation} &= \alpha \sin(\phi) \end{aligned}$$
- Note ...
 
$$(\alpha \cos(\phi))^2 + (\alpha \sin(\phi))^2 = \alpha^2$$
- Hence the amplitude of the sinusoidal component independent of phase is given by ...
 
$$\alpha = \sqrt{(\text{cosinecorrelation})^2 + (\text{sinecorrelation})^2}$$
- The phase of this component is given by ...
 
$$\tan(\phi) = \text{sinecorrelation} / \text{cosinecorrelation}$$

$$\therefore \phi = \tan^{-1}(\text{sinecorrelation} / \text{cosinecorrelation})$$

The University Of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 15, slide 10

10

© 2019 The University of Sheffield

# General Spectral Analysis Algorithm

- Recall that any periodic signal can be expressed as the sum of a fundamental sinusoid and its harmonics (*Fourier*)
- The individual components at a frequency  $\Omega = p\omega$  can be found by correlating  $s(nT)$  with  $\cos(\Omega nT)$  and  $\sin(\Omega nT)$
- Let  $c(\Omega)$  be the cosine correlation and  $s(\Omega)$  the sine correlation ...

$$c(\Omega) = \sum_{n=0}^{N-1} s(nT) \cdot \cos\left(\frac{2\pi np}{N}\right) \quad p = 0, 1, \dots, N-1$$

$$s(\Omega) = \sum_{n=0}^{N-1} s(nT) \cdot \sin\left(\frac{2\pi np}{N}\right) \quad p = 0, 1, \dots, N-1$$

$$a_p = \sqrt{c(\Omega)^2 + s(\Omega)^2}$$

$$\phi_p = \tan^{-1}\left(\frac{s(\Omega)}{c(\Omega)}\right)$$

- This is the **Discrete Fourier Transform** (*DFT*)

COM3502-4502-6502 **Speech Processing**: Lecture 15, slide 11

11

© 2019 The University of Sheffield

# Discrete Fourier Transform

COM3502-4502-6502 **Speech Processing**: Lecture 15, slide 12

12

© 2019 The University of Sheffield

## Complex Numbers: A *Reminder*

Imaginary axis

Real axis

unit circle

$z = e^{j\theta}$

$y = \sin(\theta)$

$x = \cos(\theta)$

$\theta$

$$z = x + jy$$

$$\text{Magnitude } |z| = \sqrt{x^2 + y^2}$$

$$\text{Phase } \theta = \tan^{-1}\left(\frac{y}{x}\right)$$

$$e^{j\theta} = \cos(\theta) + j \sin(\theta)$$

The University Of Sheffield.
COM3502-4502-6502 **Speech Processing**: Lecture 15, slide 13

13

© 2019 The University of Sheffield

## Complex Formulation of the DFT

- The DFT is often expressed using '**complex number notation**'
- The cosine and sine correlations are associated with the real and imaginary parts of a complex number ...

$$\begin{aligned}
 S_p &= \sum_{n=0}^{N-1} s(nT) \cdot \cos\left(\frac{2\pi np}{N}\right) - j \sum_{n=0}^{N-1} s(nT) \cdot \sin\left(\frac{2\pi np}{N}\right) \\
 &= \sum_{n=0}^{N-1} s(nT) \left[ \cos\left(\frac{2\pi np}{N}\right) - j \sin\left(\frac{2\pi np}{N}\right) \right] \\
 &= \sum_{n=0}^{N-1} s(nT) \cdot e^{-j\left(\frac{2\pi np}{N}\right)}
 \end{aligned}$$

The University Of Sheffield.
COM3502-4502-6502 **Speech Processing**: Lecture 15, slide 14

14

© 2019 The University of Sheffield

## Complex Formulation of the DFT

- Hence the DFT can be expressed as ...

$$S_p = \sum_{n=0}^{N-1} s(nT) \cdot e^{-j\left(\frac{2\pi np}{N}\right)} \quad p = 0, 1, \dots, N-1$$

... where  $S_p$  is a complex number whose magnitude and phase correspond to that of the spectrum of  $s(nT)$  at a frequency  $p/NT$

- Note also the 'inverse DFT' ...

$$s(nT) = \frac{1}{N} \sum_{p=0}^{N-1} S_p \cdot e^{j\left(\frac{2\pi np}{N}\right)} \quad n = 0, 1, \dots, N-1$$

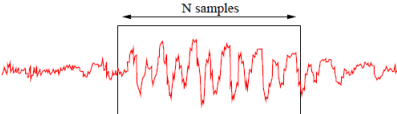
The University of Sheffield. COM3502-4502-6502 **Speech Processing: Lecture 15, slide 15**

15

© 2019 The University of Sheffield

## Using the DFT

**Signal:**

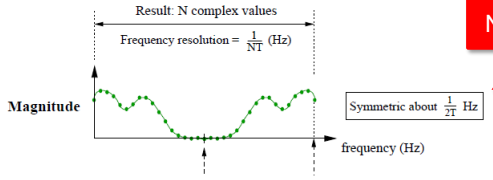


↓ DFT

**Spectrum:**

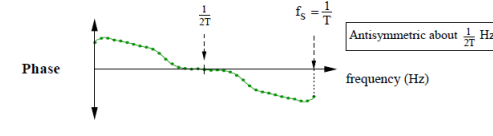
Result: N complex values  
Frequency resolution =  $\frac{1}{NT}$  (Hz)

**Magnitude**



Symmetric about  $\frac{1}{2T}$  Hz

**Phase**



Antisymmetric about  $\frac{1}{2T}$  Hz

This is the explanation for Nyquist's theorem

The University of Sheffield. COM3502-4502-6502 **Speech Processing: Lecture 15, slide 16**

16




© 2019 The University of Sheffield

## Time vs. Frequency Resolution

Frequency resolution =  $1/NT$  (Hz)

- Increasing the analysis frame  $N$ 
  - decreases the spacing between the spectral components
  - reduces the ability to respond to changes in the signal
- Hence *large*  $N$  leads to **'narrowband analysis'**
  - good spectral resolution
  - poor time resolution
- Decreasing the analysis frame  $N$ 
  - increases the spacing between the spectral components
  - increases the ability to respond to changes in the signal
- Hence *small*  $N$  leads to **'wideband analysis'**
  - good time resolution
  - poor spectral resolution

This is the *time-frequency trade-off* we saw in Lecture 4

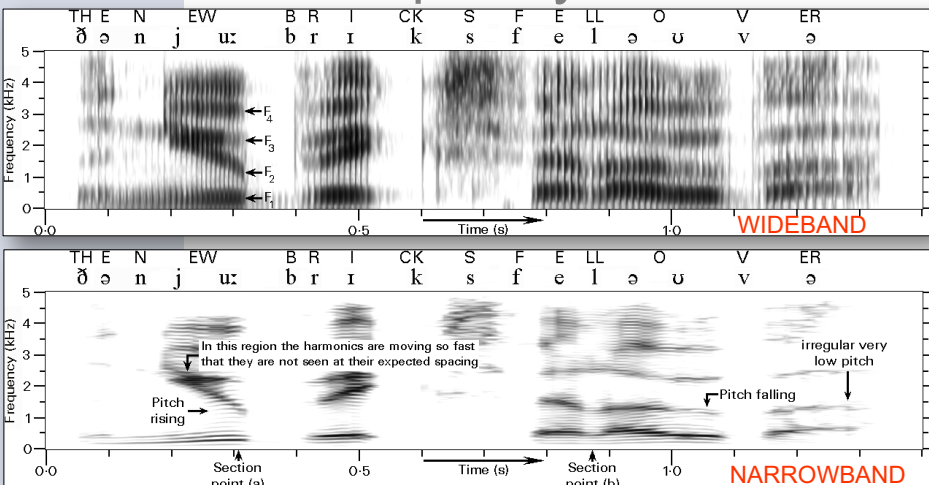
 The University of Sheffield.


COM3502-4502-6502 **Speech Processing: Lecture 15, slide 17**

17

© 2019 The University of Sheffield

## Time vs. Frequency Resolution



 The University of Sheffield.

Taken from: Holmes, J. N., & Holmes, W. (2002). *Speech Synthesis and Recognition*: Taylor & Francis.

COM3502-4502-6502 **Speech Processing: Lecture 15, slide 18**

18

© 2019 The University of Sheffield

## Time vs. Frequency Resolution

In `[wsprobe~]` ...

- narrowband analysis uses a frame/block size of 8192 samples (~190 msec)
- wideband analysis uses a frame/block size of 512 samples (~12 msec)

University Of Sheffield. COM3502-4502-6502 Speech Processing: Lecture 15, slide 19

19

© 2019 The University of Sheffield

## Implicit Periodicity in the DFT

- The DFT computes the spectrum at  $N$  evenly spaced discrete frequencies
- The algorithm assumes periodicity outside the analysis frame (*with a period equal to the frame length*)
- This means that discontinuities will arise for ...
  - periodic signals with a *non-integer* number of cycles in the analysis frame
  - all aperiodic signals
  - all stochastic signals
- Such discontinuities give rise to unwanted spectral components

University Of Sheffield. COM3502-4502-6502 Speech Processing: Lecture 15, slide 20

20

© 2019 The University of Sheffield

## Implicit Periodicity in the DFT

The diagram illustrates the process of the DFT on a signal with an integer number of cycles. At the top, a continuous sine wave is shown with a dashed box indicating an "Integer number of cycles". Below this, an "Analysis frame (N samples)" is shown, containing exactly one full cycle of the sine wave. This frame is then repeated to show "Implicit periodicity". The signal is then processed by the "DFT", resulting in a "Spectrum of assumed signal (as desired)" which shows a single sharp peak at the correct frequency on a plot of Amplitude vs. Frequency.

COM3502-4502-6502 **Speech Processing: Lecture 15, slide 21**

21

© 2019 The University of Sheffield

## Implicit Periodicity in the DFT

The diagram illustrates the process of the DFT on a signal with a noninteger number of cycles. At the top, a continuous sine wave is shown with a dashed box indicating a "Noninteger number of cycles". Below this, an "Analysis frame (N samples)" is shown, containing a partial cycle of the sine wave. This frame is then repeated to show "Implicit periodicity", which creates a "Discontinuity" at the boundaries of the frame. The signal is then processed by the "DFT", resulting in a "Spectrum of assumed signal (has unwanted components)" which shows a main peak at the desired frequency and several smaller, "Undesirable components due to discontinuity" at other frequencies on a plot of Amplitude vs. Frequency.

COM3502-4502-6502 **Speech Processing: Lecture 15, slide 22**


22

© 2019 The University of Sheffield

## Windowing

- The discontinuities arising from segmenting the signal into frames distorts the spectrum
- The distortion can be reduced by multiplying each signal frame with a **'window function'**
- The most common window function is the **'Hamming Window'** (proposed by Richard W. Hamming) ...

$$w(nT) = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi n}{N-1}\right)$$

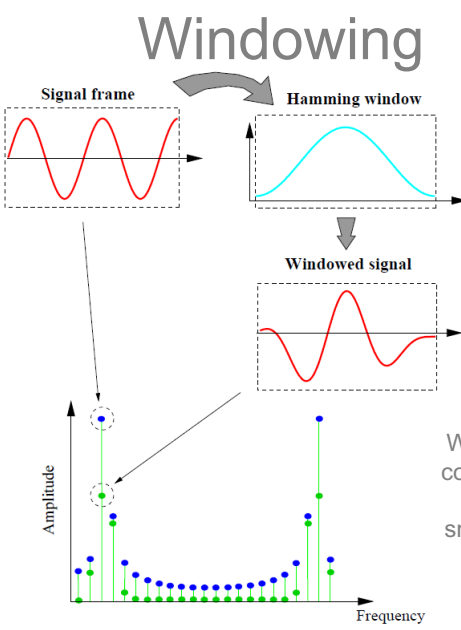

 The University Of Sheffield.

COM3502-4502-6502 **Speech Processing: Lecture 15, slide 23**


23

© 2019 The University of Sheffield

## Windowing



Windowing attenuates the components caused by the discontinuity, but also smears the spectral peaks


 The University Of Sheffield.

COM3502-4502-6502 **Speech Processing: Lecture 15, slide 24**

24

© 2019 The University of Sheffield

## Windowing

Example 15-3 windowing.pd - D:/All My Stuff/Business/Teaching/Cours...

File Edit Put Find Windows Media Help

WINDOWING

Signal\_Frequency

pd dft~

Frame Hanning wFrame

DFT wDFT

Sharper peak  
Higher floor

Broader peak  
Lower floor

The University of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 15, slide 25

25

© 2019 The University of Sheffield

## The Fast Fourier Transform (FFT)

- Implementation of the DFT requires the order of  $N^2$  multiply-add operations
- By exploiting symmetry, it is possible to devise an algorithm that requires only  $N \log_2 N$  multiply-add operations
  - e.g. for  $N=2048$ , the result is  $\sim 100x$  faster
- This more efficient algorithm is the ...  
**'Fast Fourier Transform'** (FFT)
- The FFT requires that the window/frame should be a power of 2 in size
- This can be achieved by ...
  - choosing the appropriate analysis frame size, and/or
  - zero-padding a frame to the nearest power of 2


The University of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 15, slide 26


26

© 2019 The University of Sheffield

## The FFT in Pure Data



- There are two key FFT objects available in Pd ...
  - `[rfft~]` computes the forward transform
  - `[ifft~]` computes the inverse transform
- We saw both `[rfft~]` and `[ifft~]` in action in Example 12-3
- The magnitude spectrum can be computed by squaring the real and imaginary outputs, then taking the square root `[sqrt~]`
- Alternatively, the Pd object `[framp~]` outputs frequencies and amplitudes directly
- `[rfft~]` followed by `[framp~]` is used in `[wsprobe~]`



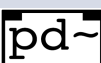
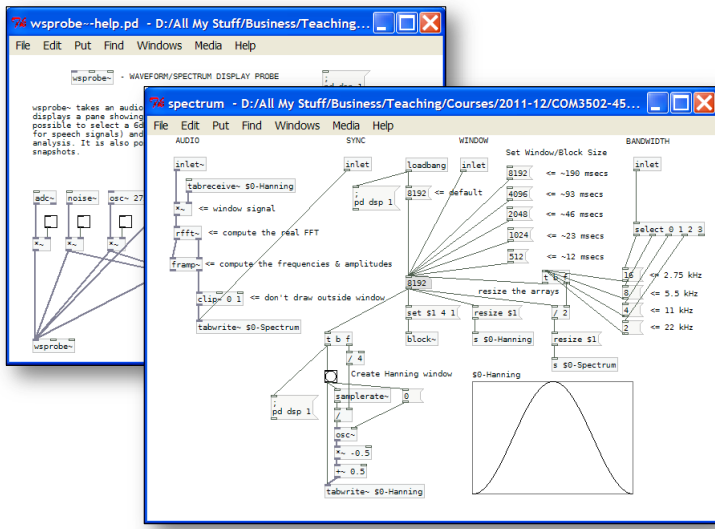
The University of Sheffield.

COM3502-4502-6502 **Speech Processing: Lecture 15, slide 27**


27

© 2019 The University of Sheffield

## The FFT in Pure Data

The screenshot shows a Pure Data patch titled 'wsprobe~ - WAVEFORM/SPECTRUM DISPLAY PROBE'. It includes an 'inlet~' for audio, followed by a 'tabreceive~ s0-Hanning' window, 'rfft~' for the real FFT, and 'framp~' for frequency and amplitude computation. The output is processed by 'clip~ 0 1' and 'tabwrite~ s0-Spectrum'. A 'pd dsp 1' object is used for display. A separate window titled 'spectrum' shows a graph of the magnitude spectrum with a Hanning window overlaid. The graph has a frequency axis from 0 to 22 kHz and an amplitude axis from -120 to 120. The Hanning window is a smooth curve peaking at 0 dB.



The University of Sheffield.

COM3502-4502-6502 **Speech Processing: Lecture 15, slide 28**

28

© 2019 The University of Sheffield

## This lecture has covered ...

- Phase in spectral analysis
- Sine and cosine correlation
- The discrete Fourier transform (*DFT*)
- Time versus frequency resolution
- Windowing
- The fast Fourier transform (*FFT*)


The University Of Sheffield.

COM3502-4502-6502 **Speech Processing: Lecture 15, slide 29**

29

© 2019 The University of Sheffield

## Any Questions ?



The University Of Sheffield.


COM3502-4502-6502 **Speech Processing: Lecture 15, slide 30**

30

© 2019 The University of Sheffield

Next time ...

The 'Z' Transform



The University Of Sheffield.

COM3502-4502-6502 **Speech Processing:** *Lecture 15, slide 31*