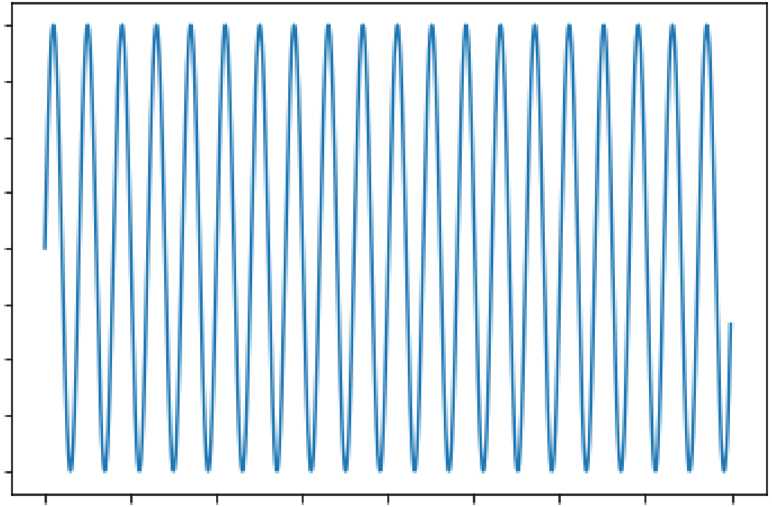
# Audio Classification

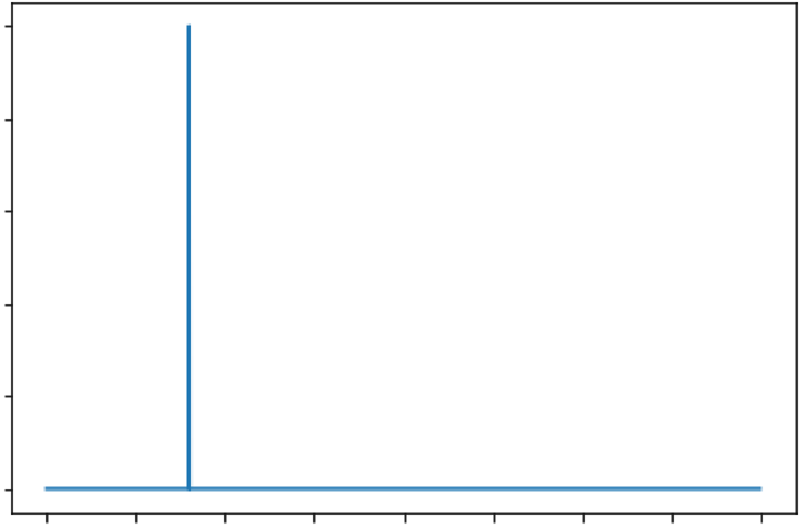
Colin Jemmott
DSC 96

# Audio processing

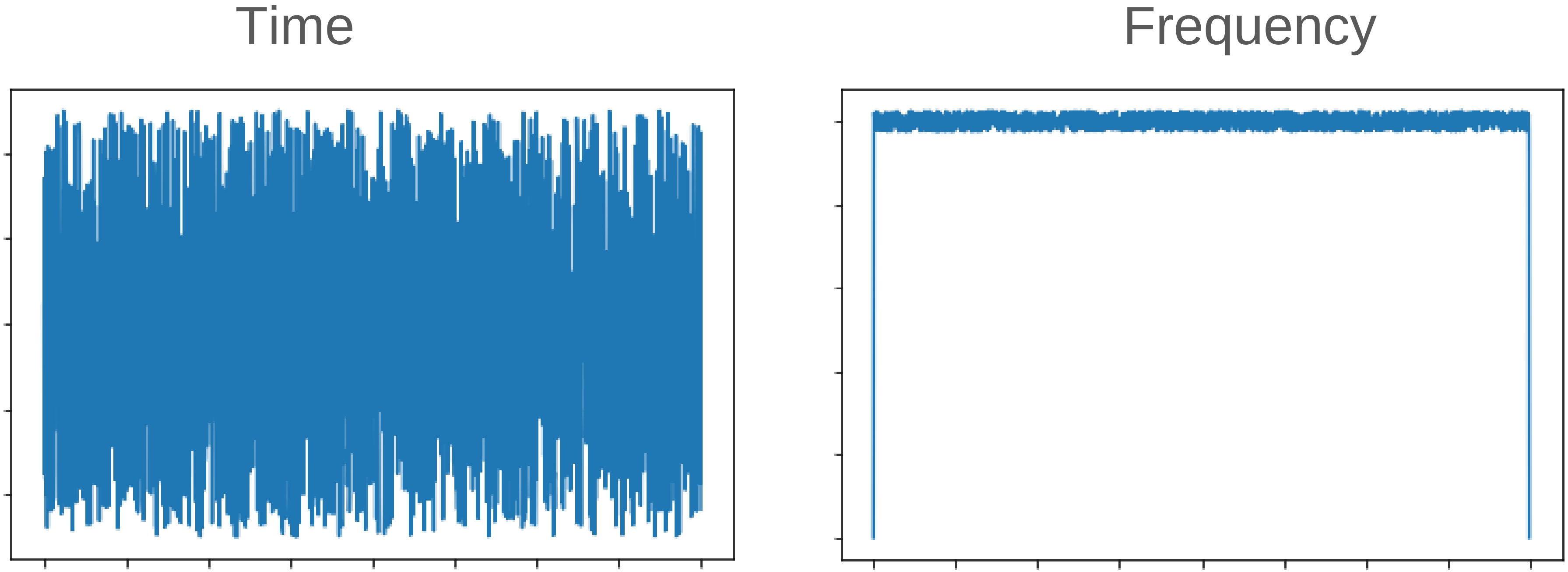
Time



Frequency



# Audio processing



# Digital Signal Processing

DSP is most strongly associated with linear transformations such as:

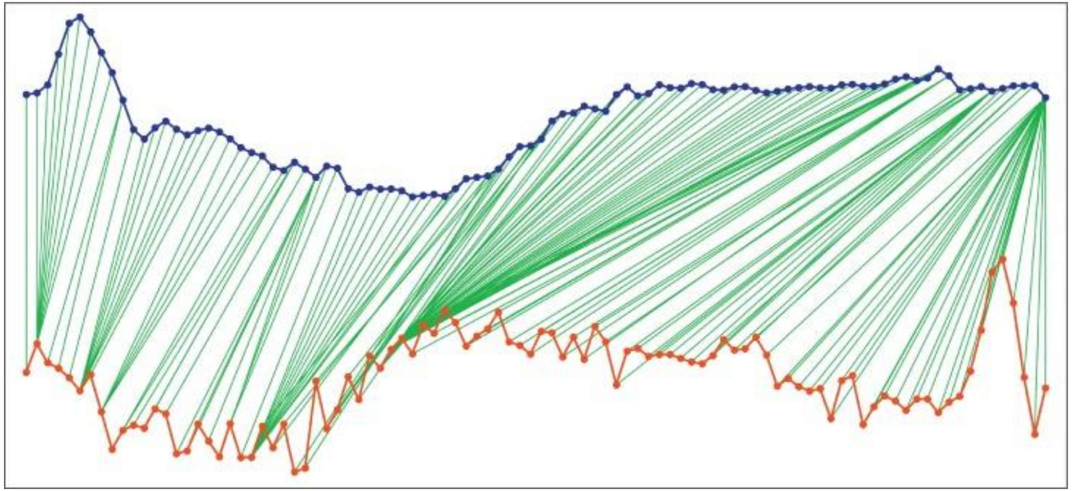
- **Filtering** (changing the relative magnitude of different frequencies)
- **Reverberation** (adding time delayed copies)

Other common DSP techniques for audio signals include:

- **Amplitude compression** (making everything loud)
- **Lossy data compression** (making files smaller and less accurate)
- Manipulations to make speech more intelligible or music sound better

# Speech Recognition History

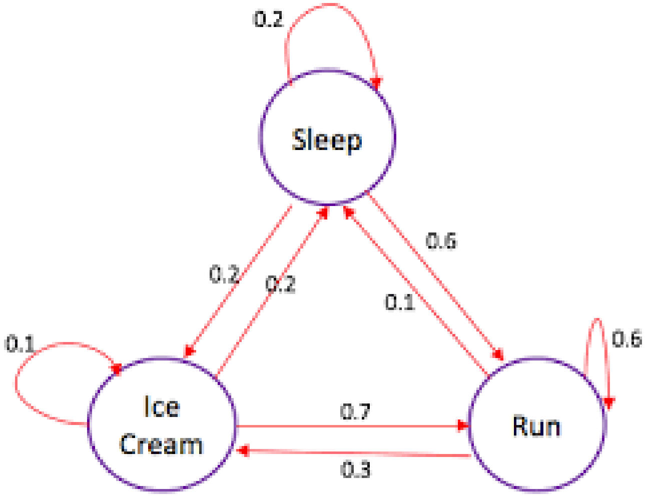
1960s Dynamic Time Warping (frames). 200 words (with isolated words)



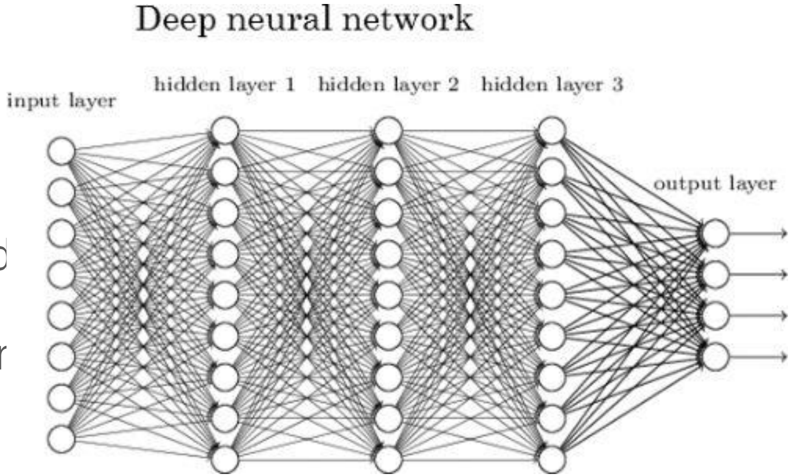
# Speech Recognition History

1960s Dynamic Time Warping (frames).  200 words (with isolated words)

1970s Markov Chains to encode language and syntax.  Continuous speech at one hour of processing per minute.



# Speech Recognition History



1960s Dynamic Time Warping (frames).  200 word

1970s Markov Chains to encode language and syn
of processing per minute.

1980s n-grams and faster computers: 20,000 words, still not realtime.

1990s commercial success with limited vocabulary and special microprocessors

2000s labeled data gathering and the NSA.
Poor accuracy prevented widespread use.

2010s Deep learning, Siri, Alexa, Cortana, Google.

# Why is speech recognition so hard?

Large vocabulary (proper nouns and jargon)

Speaker variability (accents!)

Word confusability

Context-dependency

Noise / reverberation

Conversational versus human-machine speech

# Assignment Today

Listen to a sound file

Read it into Python as an array

Add some noise, save, download and listen

Remove the long pauses from the speech (speaking recognition)

Use a pre-built audio classification API, and explore the effect of noise