Isolated Word Speech Recognition

- Speech recognition — Given a recording of an utterance, produce a text transcription of that utterance.

Isolated Word Speech Recognition

- Speech recognition — Given a recording of an utterance, produce a text transcription of that utterance.
- Continuous speech recognition is a hard problem!
  - Word boundary detection, elision, context . . .
- We make several simplifying assumptions:
  - Isolated

Overview of Speech Recognition

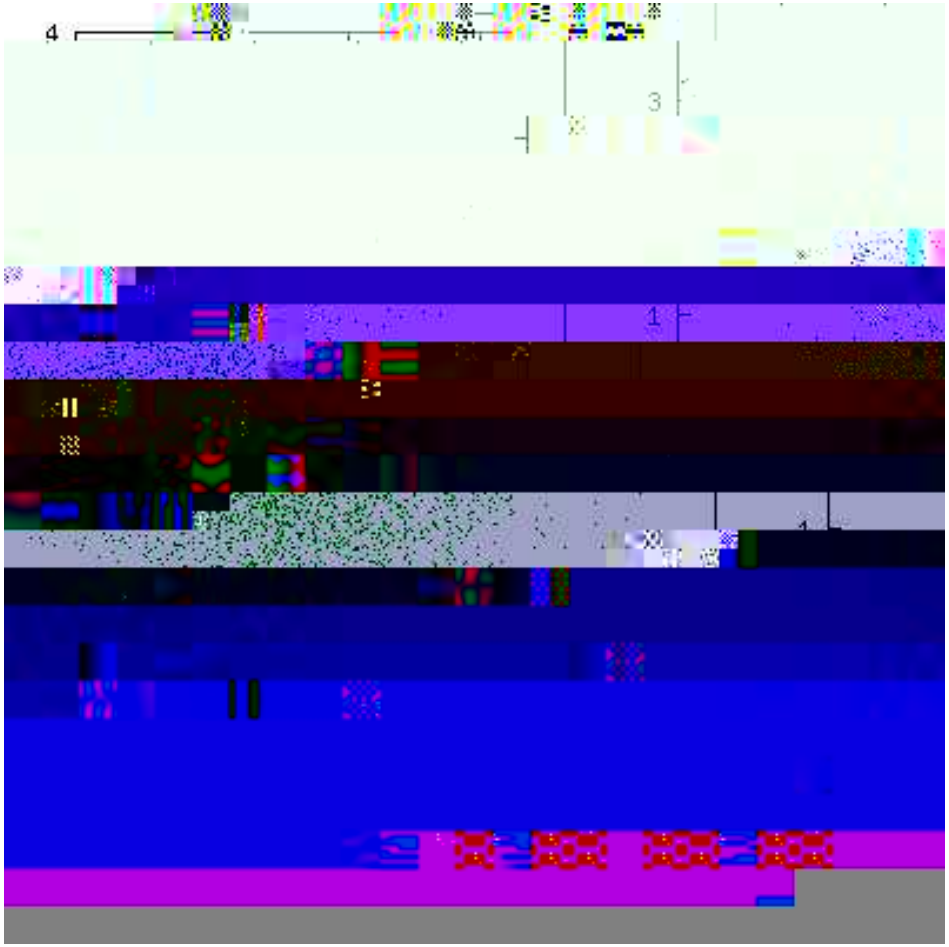
Preprocessing

Spectral coefficient extraction

In order to use the speech waveform for recognition, we convert our input from the time domain into Mel Frequency -1.14

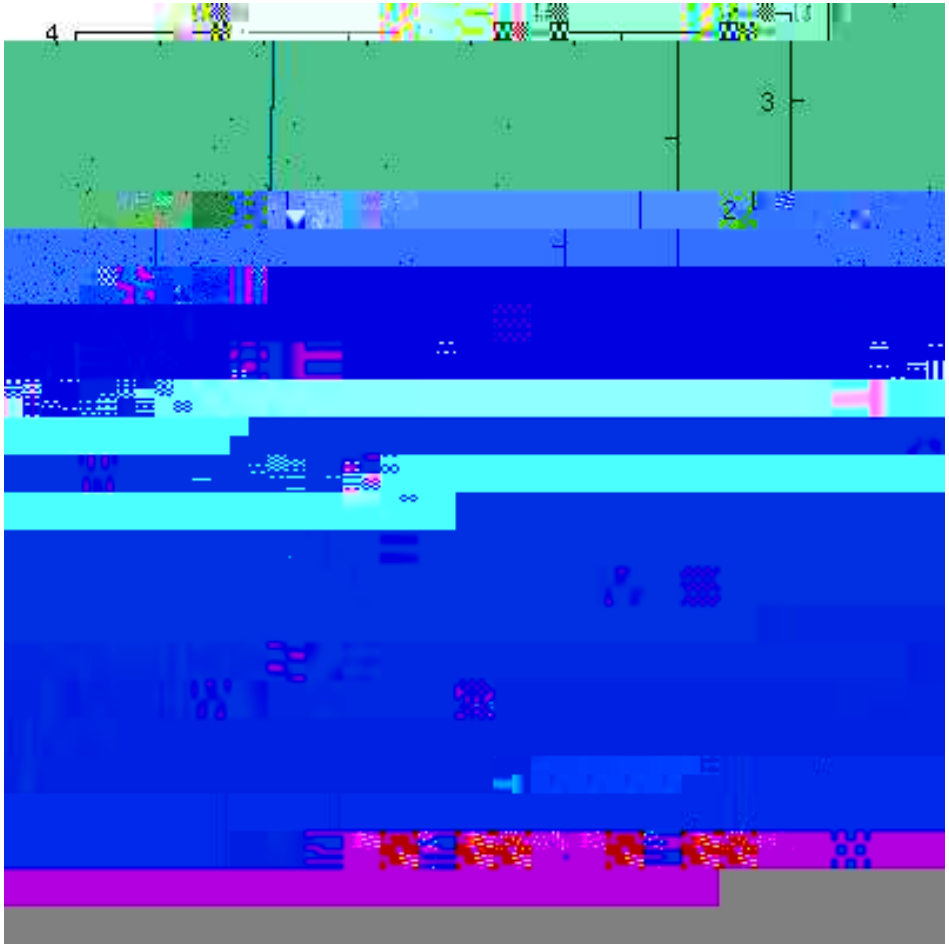
Feature Vectors and HMMs

How can we use a 12-dimensional



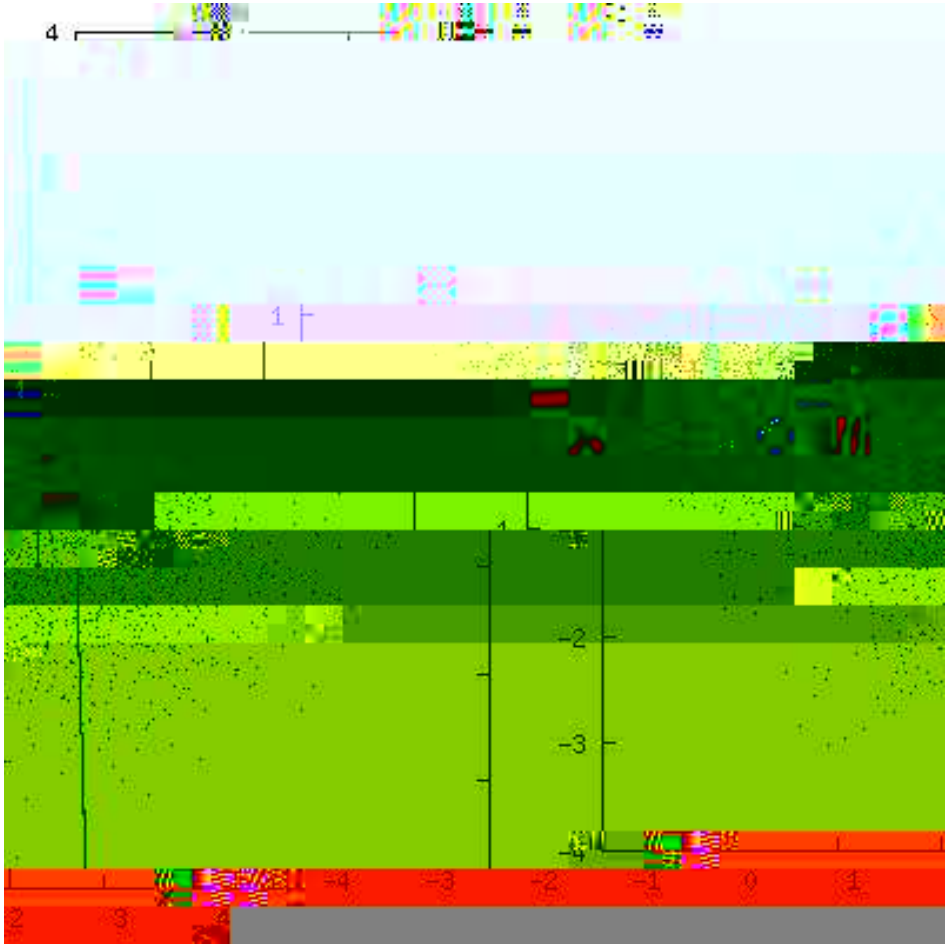
Vector Quantization Example

A 2D example (from `http://www.data-compression.com/vq.html`)



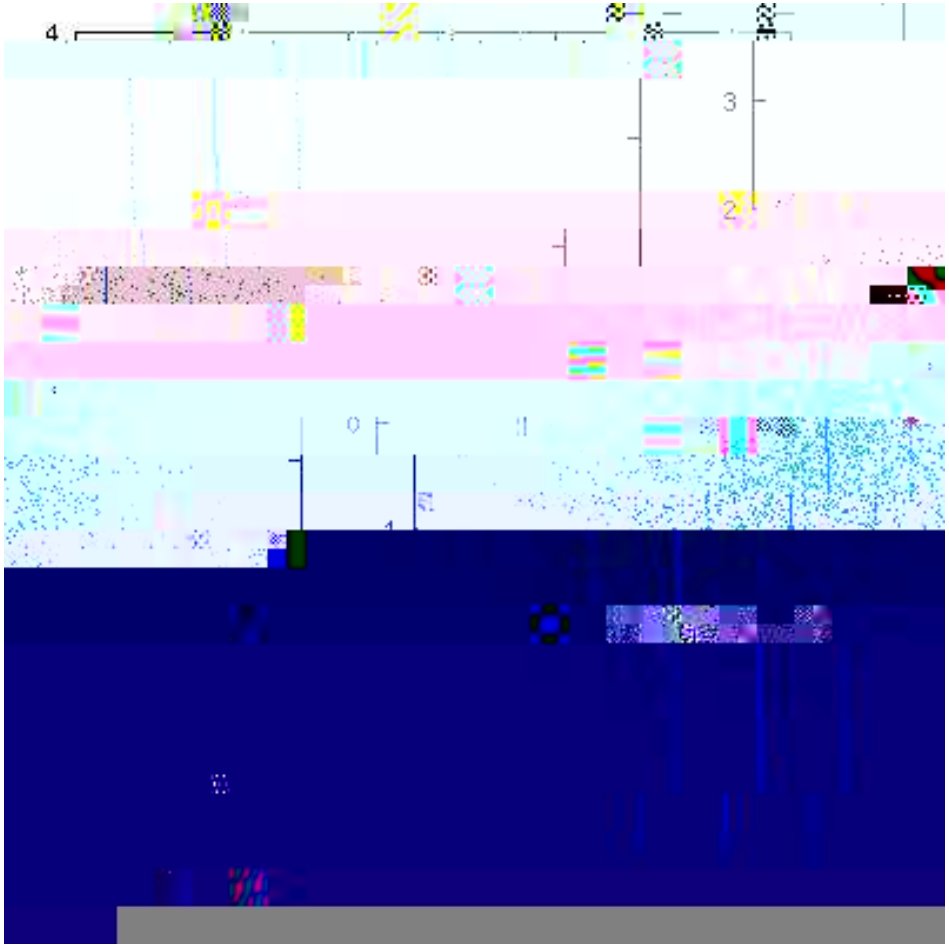
Vector Quantization Example

A 2D example (from
`http://www.data-compression.com/vq.html`)



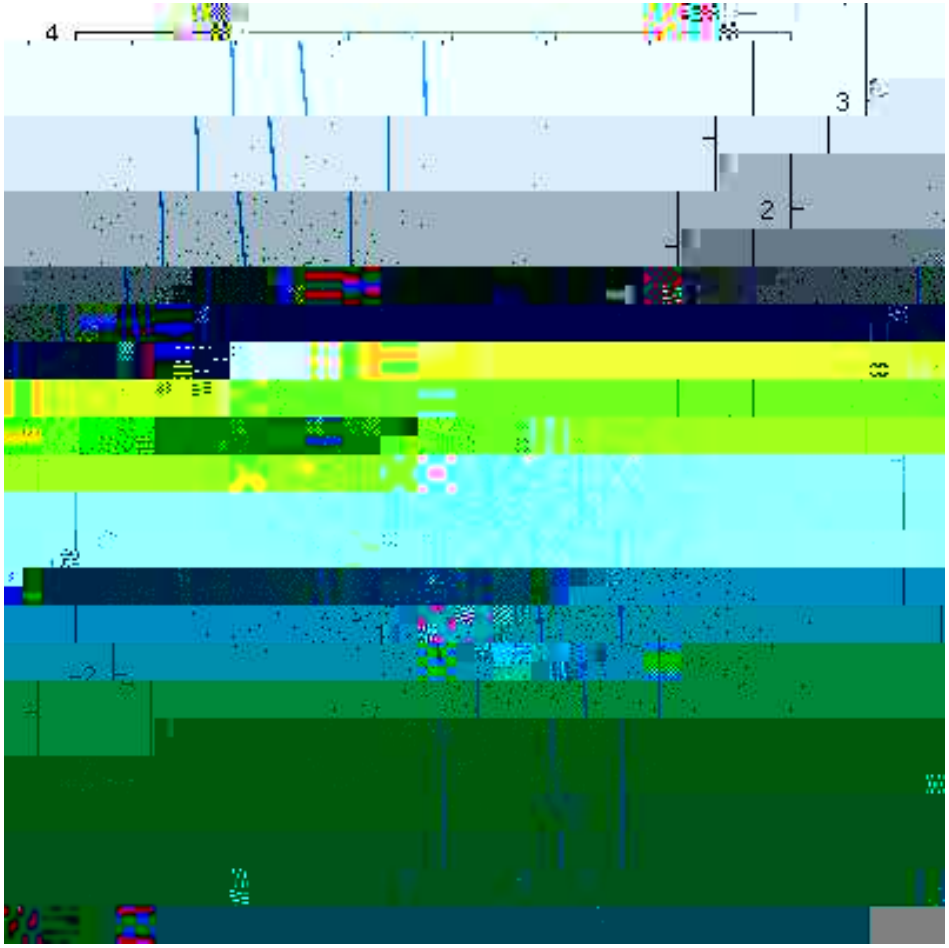
Vector Quantization Example

A 2D example (from `http://www.data-compression.com/vq.html`)



Vector Quantization Example

A 2D example (from
`http://www.data-compression.com/vq.html`)

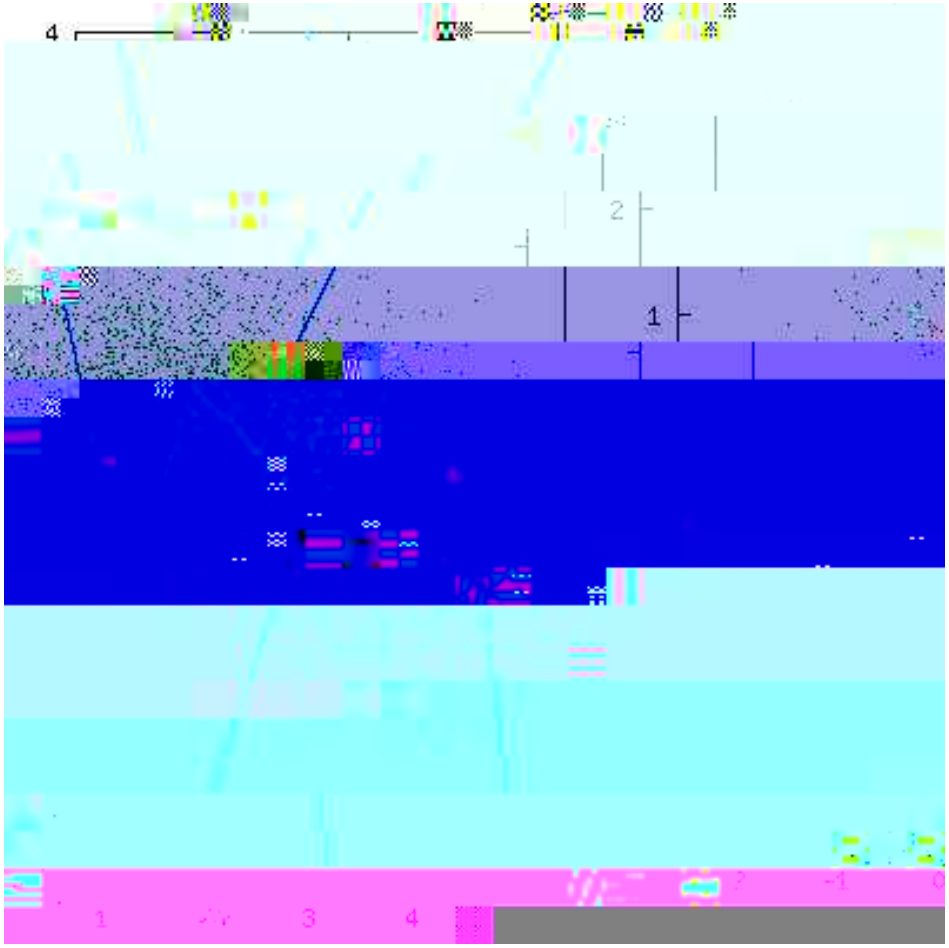


Vector Quantization Example

A 2D example (from
`http://www.data-compression.com/vq.html`)

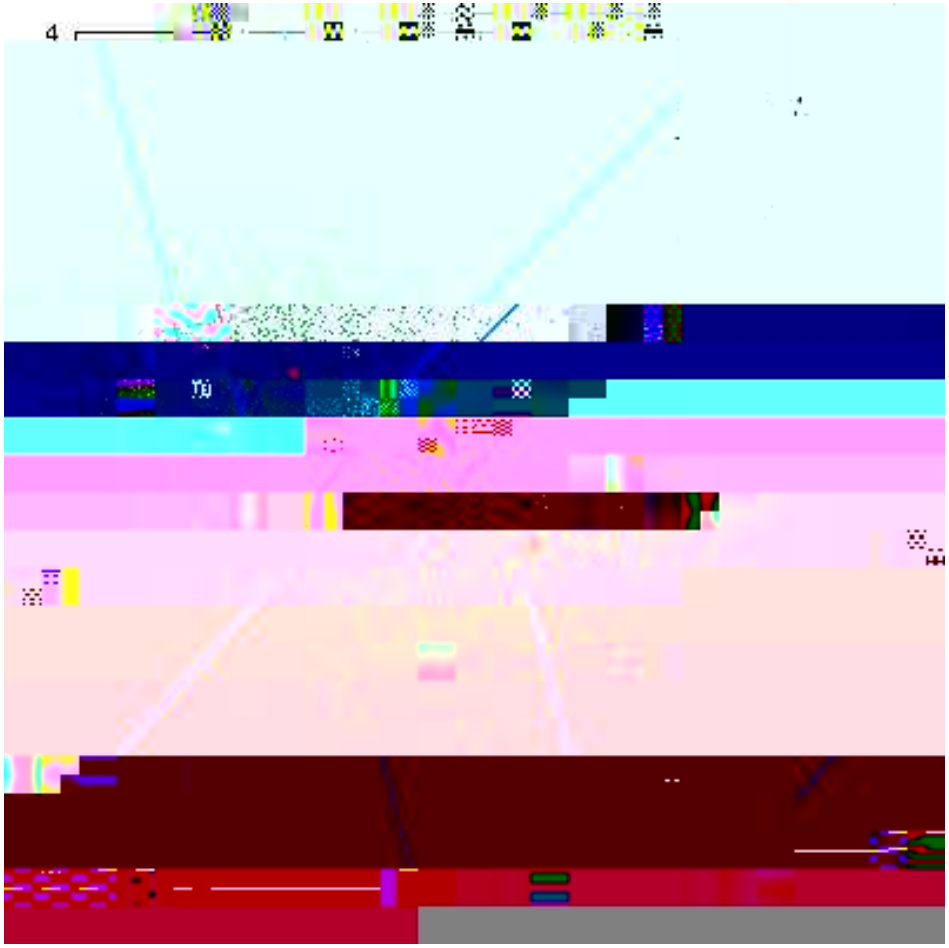
Vector Quantization Example

A 2D example (from `http://www.data-compression.com/vq.html`)



Vector Quantization Example

A 2D example (from `http://www.data-compression.com/vq.html`)



HMM Training

- We use the TI20 isolated-word speech corpus:
  - 16 speakers (8 male, 8 female)
  - Digits 0 – 9, commands such as ‘go’ and ‘enter’
  - 16 repetitions of each word from each speaker

HMM Training

- We use the TI20 isolated-word speech cor

Speech Recognition

To recognise an utterance:

- Extract the feature vectors of the utterance, and optionally quantiz

Experimental Results

- All 16 speakers used for training
  - Accuracy when using vector v

Experimental Results

- All 16 speakers used for training
  - Accuracy when using vector quantization depends on the codebook size

Conclusion

- Speech recognition in general is hars