Phases of speech Recognition: (This is like compiler phases - ASR is complex process, hence divided into many stages for processing).

1. Signal processing (Input is speech signal & o/p is feature vector), phones = 46
2. phonetic Recognition (i/p is feature vector along with database of acoustic models that help in generation of phone lattice)
3. Word Recognition (i/p is phone lattice & a database of lexicon i.e. dictionary that comprises word models), o/p is the word
4. Task Recognition (it has i/p the word lattice, and the database of grammar

Computer

Audacity to record sound

word probability

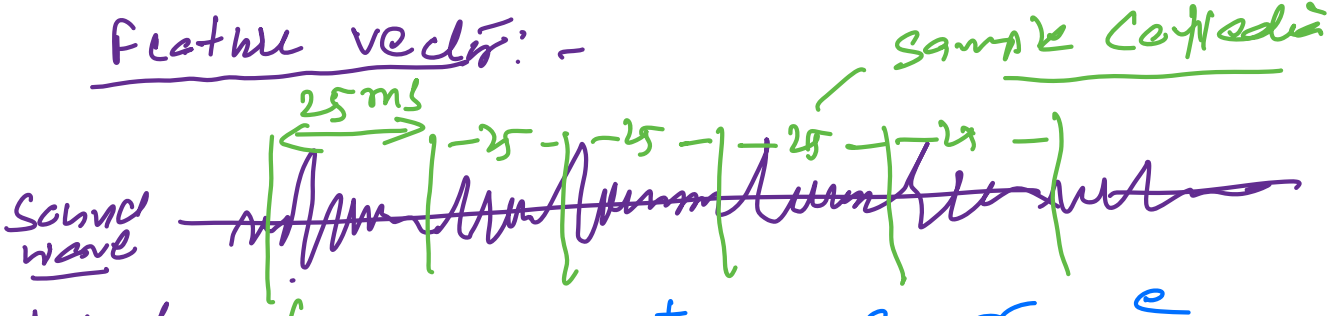
green .8

coin .9

so, sentence is → I have killed your coin. (so, correct is "green".

-- output of this last stage is final sentence.

Feature vector: -



- The speech signal is sampled every 25 msec
- its properties are sample (39 nos')
  - frequency
  - Amplitude
  - Power

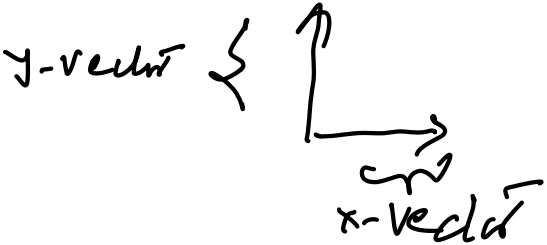
only these parameters go to next stages

These parameters are called feature vectors

e.g. $(x, y) = (5, 7)$ — two vectors x, y

$(x, y, z) = (5, 7, 10)$ — three vectors x, y, z.

→ 39 parameters means 39 vectors



The vectors take far less space, compared to the original signal, hence the vectors effectively provide compression (of data).

Search Algorithm :

- for phonemes : Easy search, as only 46 phonemes. search space is
- " Lexicon : medium search (1000s)
- " Sentence : difficult search as it requires to search out of millions of sentences

Let the sentence was:

"Recognize speech", is pronounced like,

like "wreck a nice beach" ✓

When this happens, we need to search a correct sentence. This is language part. ∴ we require a model of language

To recognize words individually, which is possible if words are sufficiently isolated (a word does not effect next word) we can simply search the words in lexicon

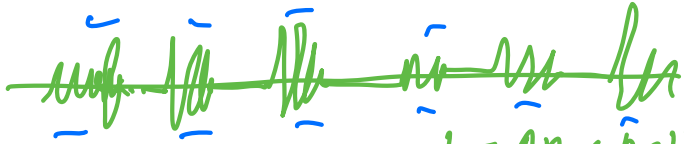
When words are isolated (slow speech) so that no word effect the pronunciation of any word the search space is less.

This, is called - IWR - Isolated word recognition.

- Its search algorithm is simply the lexicon search.

When entire sentence is spoken at speed, the situation becomes that of "Recognize speech" is recognized as "wreck a nice beach". This requires Language model based algorithm.

and the algorithm is called CWR - Continuous Word Recognition. This is time consuming also.



Which is flight to New Delhi

1 WR

Which is flight to New Delhi

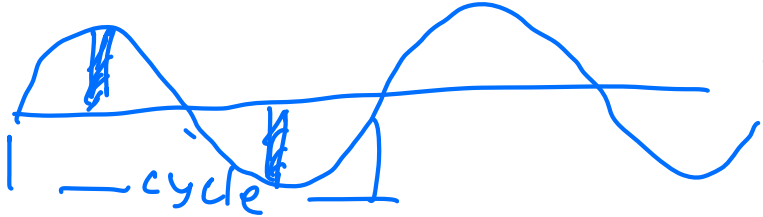
CWR

$\frac{1}{4 \times 10^{3}} = .25$ msec

Domestic Study: (India) 4 cycles/sec

2KHz

$\frac{1}{2 \times 10^{3} \times 2} =$



1 cycle

of sampling frequency is 2x signal frequency