

Tutorial 4
Speech Recognition

1. Why are a number of mixtures necessary to represent the MFCC parameters to be used in an HMM model that uses continuous pdf's?
2. An HMM system has A, B, and JI given by a_{ij} where i, j denote a state transition from i to j and the system has four states, and b_{ij} where i, j denote the i output in state j , and there are six outputs.

A	0.5	0.4	0.1	0		B	0.2	0.2	0.1	0.0
	0	0.2	0.6	0.2			0.1	0.1	0.1	0.2
	0	0	0.2	0.8			0.0	0.2	0.2	0.2
	0	0	0.2	0.8			0.3	0.2	0.0	0.1
							0.3	0.3	0.3	0.3
							0.1	0.0	0.3	0.2

The initial state is state 1.

- (a) Work out the forward probability for the following sequence

state	1	1	2	4	3	3	4
observation	5	1	2	4	5	3	5

- (b) For the observation sequence below, work out the most probable state sequence:

observation	5	1	3	4	2
-------------	---	---	---	---	---

3. A small speech recognition system uses six words – top, down, right, left, begin, end.

Figure 1, below shows the language state diagram model for a small speech recognition system using the words -

up, top, right, left, down and the state sequence of their use

In a recognition trial the phoneme HMM passes the following two sequences as the most probable utterances to the language model

left top right down with probability 0.75

up left right down with probability 0.7

What is the most probable choice if the language model is taken into account?

Language State Diagram

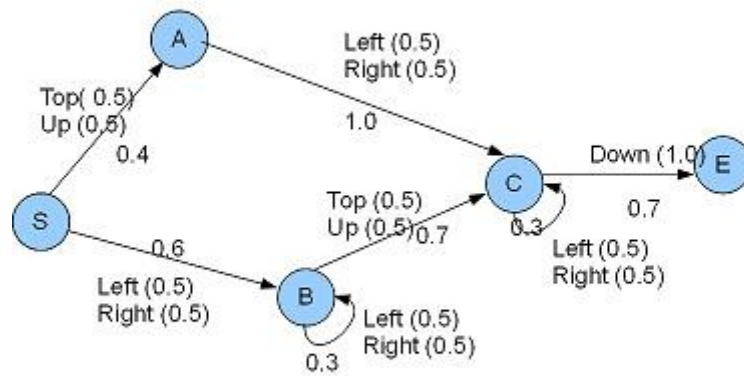


Figure 1

4. Describe briefly the 'top-down' and 'bottom-up' models for speech recognition highlighting the advantages and disadvantages of each type.
5. What is the meaning of embedded HMM's in a speech recognition system? Mention with reasons the type of speech recognition systems in which they can be used.