Hidden Markov models

- The acoustic model computes the probability of feature vector sequences under the assumption that particular word sequence had produced these vectors.

- Variation in a word or phone's pronounciation manifests in: 1. duration 2. spectral content (freq.)

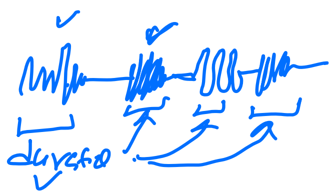
  1+2 = acoustic observation.

- $$p(w|x) = \frac{p(x|w) * p(w)}{p(x)}$$

  language model

  $p(x)$ dropped, being common denominator

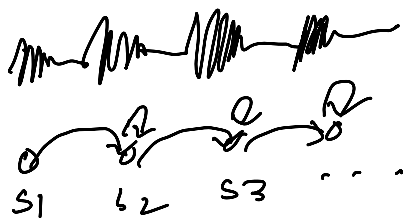
  acoustic model



Exercise: Today, in your phone, record the voice & see wave forms.

So in speech recognition (acoustic model) there are two processes:

$\underline{X}$ - time varies (temporal variability)



← no so easy

$\underline{Y}$ - spectral variability (freq.) - can be easily detected

$X \rightarrow$ is chain of sequences $s_0, s_1, \dots \rightarrow$ ∴ process is called Markov chain, it is hidden.

Y - is observable process.

in X: $p(S_{i+1} | s_0, s_1, \dots, S_i) = p(S_{i+1} | S_i)$

Y also depends on X. ?

क ख $= X$

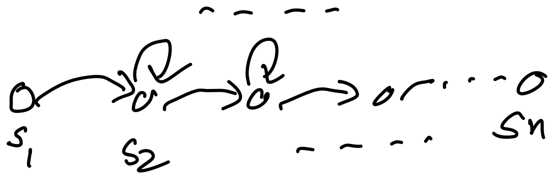
Freq $= Y$

By Freq. also we can estimate क, ख, etc. (i.e. X)

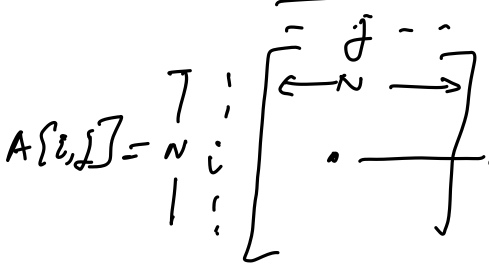
X & Y both are probabilistic ∴

this model is double probabilistic i.e. double stochastic model

& HMM is chosen since it is double stochastic model



$P(s_{i+1} | s_i)$, or $p(s_j | s_i)$ ∴ i, j prob.

✓ 

a_{ij} = prob. of transition from S_i to S_j

✓ $A = \{a_{ij} | i, j \in X\}$, $S = \{s_1, \ldots, s_n\}$

o/p. prob. of o/p $B = \{b_{i,j} | i, j \in X\}$ o/p distribution

prob. of initial state s_i, ∴ $P(s_0), P(s_1) \ldots P(s_n)$

$\Pi = \{\pi_i \mid i \in X\}$ initial prob. of states

$\therefore$ HMM. $\lambda(S, M, A, B, \Pi)$. $\quad p(s_{i+1}|s_i)$

M = total no. of unique o/p symbols per state

$a_{ij} = P(T_{i+1} = s_j \mid T_t = s_i)$

$\sigma_1 \quad \sigma_2 \ldots$

V = set of all possible o/p symbols $= \{v_1, v_2, \ldots, v_m\}$ (phonemes)

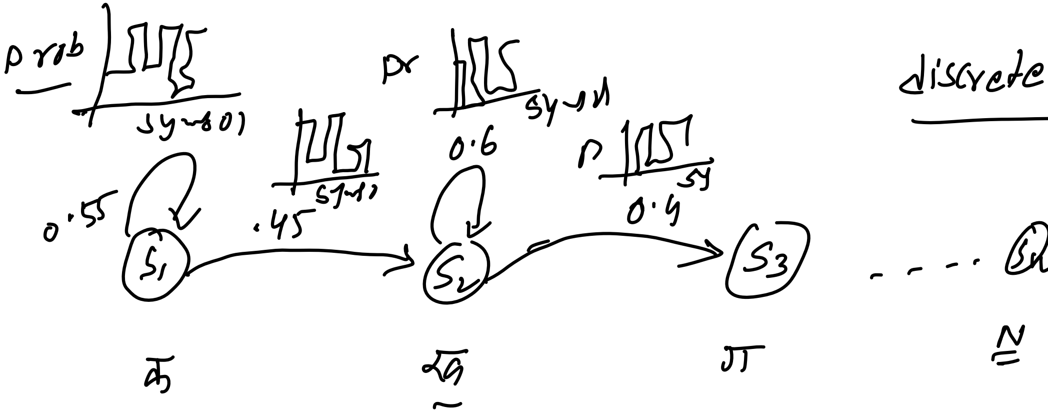
$O = (O_1, O_2, \ldots, O_T)$ at different times

$$P_\lambda(O \mid \Pi, A, B) = \sum_{q}^{T} a_{q_{t-1} q_t}\, b_{q_t}(O_t)$$

Expected duration of each state $= \boxed{1/(1-a_{ii})}$

$$\begin{array}{cc} & \cdots j \cdot \\ A(i,j) = i & \begin{bmatrix} \ddots & \\ & \end{bmatrix} \end{array}$$

$ii = 1,1$
$2,2$
$3,3$



discrete HMM

N no. of symbols (phonemes) are predecided

IPA