

SGN-4106 Speech Recognition

Exam 6.11.2006.

You may answer in Finnish or English.

General note: to reduce hand computation, in a question which requires a numerical answer, you can give the algorithm that would give the correct answer and calculate the first few iterations of that algorithm to show how it works.

Problem 1. Explain the meaning of the following terms (1 point/term): 1) phon, 2) lexicon, 3) Mel scale, 4) cepstrum, 5) speaker adaptation, 6) Gaussian mixture model.

Problem 2. Draw a diagram of the standard front-end of a speech recognizer, explain how it operates and what properties of speech it relies on. (6 points)

Problem 3. Explain in general terms how an isolated word speech recognition system based on HMMs works. Specifically, what are the steps required to build such a system and how is it used for recognition? (6 points)

Problem 4. A 3-state HMM for 1-dimensional output vectors has the following parameters:

$$P(q_1|q_1) = 0.8, P(q_2|q_1) = 0.2, P(q_3|q_1) = 0.0,$$

$$P(q_1|q_2) = 0.0, P(q_2|q_2) = 0.5, P(q_3|q_2) = 0.5,$$

$$P(q_1|q_3) = 0.0, P(q_2|q_3) = 0.0, P(q_3|q_3) = 1.0,$$

$$\mu_1 = 0, \mu_2 = 1, \mu_3 = 2,$$

$$\sigma_1^2 = 1, \sigma_2^2 = 1.5, \sigma_3^2 = 2.$$

Calculate

- The total probability $P(o_1, o_2, o_3, o_4, o_5)$ of the observation sequence 0.5, 1.7, 0.4, 2.3, 2.8. (3 points)
- The most likely state sequence for this observation sequence beginning in state 1 and ending in state 3. (2 points)
- To reduce the risk of physical and mental injury during the exam, the following table lists the state-conditional output probabilities $P(o_i|q_i)$. Explain how to calculate the entries in this table. (1 point)

$P(o_i q_i)$	o_1	o_2	o_3	o_4	o_5
q_1	0.3521	0.0940	0.3683	0.0283	0.0079
q_2	0.2516	0.2385	0.2455	0.1827	0.1295
q_3	0.1506	0.1972	0.1448	0.1972	0.1841