SGN-4010 Puheenkäsittelyn Menetelmät
Speech Processing Methods
exam 31.1.2006

You may answer either in Finnish or English.

Problem 1. Explain briefly what the following terms mean (1 point/term): **a)** time-frequency resolution (aika-taajuusresoluutio), **b)** overlap-add, **c)** prosody (prosodia), **d)** lattice filter (ristikkorakenne), **e)** ARMA-model, **f)** Toeplitz matrix.

Problem 2.

a) When speaking after inhaling helium from a balloon, speech gets a 'Mickey Mouse'-quality due to shifted formant frequencies. Assuming that a person's vocal tract length is a uniform tube of length 16 cm and the speed of sound in helium is 970 m/s, what are her formant frequencies after inhaling helium? (2 points)

b) You are a world-famous engineer at a company which produces diving equipment. The mixture of air and helium which divers breathe causes the formants of speech to be shifted by the factor γ, i.e. with the mixture the formants are at frequencies $F_i' = \gamma F_i$, where F_i' is the i^{th} formant with the mixture and F_i is the i^{th} formant with normal air.

Your job is to devise a simple, practical signal processing algorithm which will convert the divers' speech into normal-sounding speech. Describe an algorithm that will do this. (4 points)

Problem 3. The autocorrelation function $r(k)$ of a frame of speech is

k	0	1	2	3	4	5	6
$r(k)$	12	11	10	8	6	4	2

The optimal 3^{rd}-order prediction-error filter $A_3(z)$ for this frame is

$$A_3(z) = 1 - 0.9318z^{-1} - 0.5000z^{-2} + 0.5682z^{-3}.$$

a) What is the optimal 3^{rd}-order prediction error energy E_3? (1 point)

b) What are the coefficients of the filter $p_0 + p_1 z^{-1} + p_2 z^{-2}$ which predicts the next sample of the frame as well as possible? (2 points)

c) What is the optimal 4^{th}-order prediction-error filter $A_4(z)$ for this frame? (3 points)

Problem 4. Explain the main principles of how rule-based and concatenative speech synthesizers work and what the main differences between them are. (6 points)