Questions A

A1. Explain the structure of the prediction error filter and of its inverse.

- ❑ What is the difference of the two filters $\boldsymbol{p}$ on slide 9?
- ❑ Which cost function is minimized?

A2. Cost function / Distances of spectral envelopes

- ❑ What are the cost function (distance function) requirements?
- ❑ Does the equation on slide 16 comply with these requirements?
- ❑ Describe the plots on pages 17 and 18 and compare them with regard to their influence on a subsequent pattern recognition.

A3. Cepstral coefficients

- ❑ What is the statement of Parseval's theorem used on slide 19?
- ❑ What is the influence of the cepstral coefficients c_i (slide 26) on the cost function (slide 16)?

A4. MFCCs: Logarithm and IDCT

- ❑ How is the logarithm in the MFCC calculation motivated?
- ❑ Explain the equation on slide 40. What meaning does the matrix $\boldsymbol{P}$ have?

Part 4. Feature extraction

Questions B

B1. What does speech recognition expect from feature extraction (slide 6)?

- ❑ Why are new features extracted every 10 to 20 ms?
- ❑ Why is it the goal to calculate as few coefficients as possible?

B2. Linear prediction

- ❑ Explain the equation $\boldsymbol{p}_{opt} = \boldsymbol{S}_{yy}^{-1}\, \boldsymbol{s}_{yy}(1)$.
- ❑ How is $\boldsymbol{s}_{yy}$ defined? How would the equation of a predictor with a delay of z^{-k} instead of z^{-1} look like? Where in the derivation does the delay appear?

B3. MFCCs: Overview

- ❑ Give an overview of the calculation of MFCCs.
- ❑ Which influence does the windowing have?

B4. MFCCs: Squared absolute, mel filtering

- ❑ Which information is lost in the process of calculating the absolute of the spectrum?
- ❑ Why is the mel filtering applied?
- ❑ Which influence does the mel filtering have on the data rate, on the spectral envelope, on the pitch and on the harmonics?

Answers B

B1. What does speech recognition expect from feature extraction (slide 6)?

- □ Speech signals are assumed to be approximately stationary in frames of 10 to 20 ms.
- □ The subsequent pattern recognition is most efficient if only few coefficients contain every relevant information.

B2. Linear prediction

- □ See slide 14.
- □ See definition of error signal on slide 11 and of $\boldsymbol{s}_{yy}$ on slide 14, $\boldsymbol{p}_{opt} = \boldsymbol{S}_{yy}^{-1}\, \boldsymbol{s}_{yy}(k)$.

B3. MFCCs: Overview

- □ See slide 28.
- □ See slide 31.

B4. MFCCs: Squared absolute, mel filtering

- □ The phase information is lost. It is assumed to be not relevant for speech recognition.
- □ The human perception of pitch motivates the Mel filtering. At high frequencies, a lower frequency resolution is sufficient.
- □ Influence of mel filtering: Reduction of the data rate, “smoothing” of the spectral envelope, neglect of the pitch and its harmonics.

Answers A

A1. Explain the structure of the prediction error filter and of its inverse.

- ❑ There is no difference. Only how the filter is connected (as FIR or IIR) makes the difference.
- ❑ Cost function: See slide 8.

A2. Cost function / Distances of spectral envelopes

- ❑ See slide 16.
- ❑ Yes. The requirement to be invariant to level/gain is ensured by the use of linear prediction.
- ❑ The cepstral distance relativizes the large distances in areas of high energy (low frequencies) and increases the influence of deviations in areas of low energy (high frequencies)

A3. Cepstral coefficients

- ❑ Interpretation of Parseval’s theorem: The power of a time-domain signal is equal to the power of its Fourier coefficients.
- ❑ See equation on slide 19, sum over the squared differences of the first $3N/2$ coefficients.

A4. MFCCs: Logarithm and IDCT

- ❑ The human perception of loudness motivates the use of the logarithm (slide 37).
- ❑ The matrix P is described in the last passage; it performs a reduction of dimensions.