Exam

Instructions

Format

Allowed: Pen, paper, ID card and *handwritten notes*.
No other extra material is allowed (including no calculators, no printed or copied literature, no computers, no phones, etc.). You are allowed to take any amount of handwritten notes you like, but it is clear that from a larger pile it is harder to find anything.

Extent of answers

Good answers to "detail questions" (see below), corresponding to 1 point each, are typically 1-3 sentences. The length of good answers to the other questions scale according to amount of points. For example, good essay answers to "application questions" are typically one page in length.

Answering questions with how what and why

Structuring any text with the three questions, "how", "what", and "why", is a very effective way of describing things in general. In the example questions below, typically the question itself already states "what" is done, such that your task is to answer the two remaining questions of "how" and "why". Often, questions also explicitly ask for "definition" and "motivation".

To illustrate this type of answer, consider the following example question ``Describe AD-conversion of speech?''. The ``how'' part of the answer could be:

Speech is typically picked up by an analog microphone, where the acoustic pressure waveform is converted to an electric signal where the voltage corresponds to the pressure. Analog to digital (AD) conversion is then achieved by quantizing the voltage to steps, where each step is encoded into a digital number.

This answer does not reveal why such conversion is useful or important, nor does it explain what we can do with the digital signal. That is where ``why'' comes in:

Modern signal processing methods are, to a large extent, based on digital computations, such as filtering and time-frequency analysis. While many such operations are possible in the analog domain, they are much easier and cheaper to implement in digital form. Speech signals are therefore often converted to a digital format with an analog-to-digital converter, such that we can, for example, easily visualize the speech signal with a spectrogram or attenuate noise with speech enhancement methods.

Observe that in many of the example questions below, it is a requirement to answer both parts, "how" and "why" to receive a full score.

Exam structure

There are four questions:

1. Detail questions (6x1p).
2. Basic questions (speech production, speech production modeling, basic phonetics) (most likely 1x6p, but also 2x3p questions are possible).
3. Algorithmic questions (most likely 2x3p, but either 1x6p or 3x2p are also possible).
4. High-level (=abstraction-level) application question (1x6p).

Example Questions (UPDATED for 2023)

Detail question

Explain the main idea of the following words and concepts (1p each): *(This is not an exhaustive list, but we will choose 6 words of this type)*

- Phoneme
- Phonation
- Phone
- Vocal tract
- Formant
- Coarticulation
- Onset and offset (0.5p + 0.5p)
- Intonation
- Voiced and unvoiced signals
- Source-filter model
- Objective and subjective quality evaluation (0.5p + 0.5p)
- Fundamental frequency and pitch (0.5p + 0.5p)
- Voice activity detection and wake word detection (0.5p + 0.5p)
- Speaker recognition, verification and diarization (two correct=0.5p, three correct=1p)
- Features (in the context of, for example, speech or speaker recognition, or voice activity detection)
- Expert and naïve listeners (0.5p + 0.5p)
- ADPCM
- Differential privacy
- Privacy by design
- Federated learning
- Noise gate
- Jitter and shimmer (0.5p + 0.5p)
- etc.

Basic question

(Essay questions, but drawings are often helpful. Not an exhaustive list of topics.)

- Explain how humans produce speech (what physiological processes are involved 3p, what acoustic effect do these processes have 1p and what type of phonations are these related to 2p).
- Describe the source filter model of speech production (model description 3p, connection to speech production 2p, application in speech processing 1p).
- Describe how the quality of speech processing algorithms can be evaluated (main categories of quality measures 1p and short descriptions of those categories 1p).
- What categories of information are present in speech signals (2p) and how can you group them (1p)?
- Describe which types of private information can a speech signal contain? (An exhaustive list is not possible, but describe the range of information types. 3p)
- What are the basic issues in security and privacy of speech technology? (3p)
- What is the F0 of a speech signal (definition 1p)? In contrast, what are F1, F2, F3 ... (definition 1p)? What processes in speech production causes these effects (2p)? What types of information do these carry (2p)?
- Explain the systems design of a speech interface, such as a smart speaker in terms of a flow-diagram. Which blocks will it need and how are they connected? (blocks 2p, connections 2p, motivation 2p)

Algorithmic questions

A combination of following questions. Exact wording might vary. Observe that some combinations of keywords and questions are not applicable.

- What is *keyword* (definition, 1p)?
- What is *keyword* used for in speech processing (objective/motivation, 2p)?
- How is *keyword* applied in speech processing (application/algorithm, 2p)?
- How does *keyword* relate to speech production or perception (background, 2p)?

Here the *keyword* can be for example:
linear prediction,

- spectral subtraction,
- Wiener filtering (including derivation of the scalar estimator used in speech enhancement),
- beamforming,
- short-time Fourier transform (STFT),
- overlap-add,
- entropy coding,
- voice activity detection (VAD),
- wake-word detection,
- fundamental frequency estimation,
- signal-to-noise ratio, noise reduction factor (or noise attenuation factor) and speech distortion index,
- mel-frequency cepstral coefficients (MFCC),
- zero-crossing rate (ZCR),
- cepstrum,
- pulse-code modulation (PCM) and its differential and adaptive variants,
- uniform, log and mu-law (μ-law) quantization,
- feature extraction,
- jitter and shimmer,
- etc.

In addition, there can be simple math-questions:

- In the context of linear prediction, derive the optimal coefficients in the minimum mean square error (MMSE) sense. (mathematical definition of an LPC filter 1p, derivation 1p, final formula 1p)
- Describe how a linear classifier is defined, for a given vector of features (1p). Derive the optimal classifier (2p).
- Suppose a signal can attain values −1, 0, +1 and the probability of 0 is 50%, while ±1 both have a probability of 25%. How many bits does the optimal coding of this signal require on average per symbol (1p)? What would an optimal encoding be (give an example of the bitstring for each of the symbols) (1p)? Given that optimal encoding, how would you encode the sequence −1, 0, +1, 0 (1p)?

Application question

Essay questions, but drawings can sometimes be helpful. Wording may vary, but the topics will be one of the following.

- How is the output speech quality evaluated with speech processing methods during their entire life-cycle (during the algorithm and product development, during standardization and marketing of the product, as well as during the time product is in use)? What resources are needed? Discuss the strengths and weaknesses of available approaches. (6p)
- Suppose you are hired by an online magazine to write an article that compares available VoIP services. The target user group is business users, where people want to have the ability to talk both one-to-one, but also in teleconference scenarios where multiple persons attend the meeting simultaneously. You are given a limited budget for expenses.
  What types of performance evaluations would you use and why? (3p)
  How would you implement the test in practice? (3p)
- Describe the basic structure of speech codecs. (1p) Which are the main components and which features of the speech signal do they model? (3p) What approach is used within a codec to optimize parameters? (1p) How is output quality evaluated? (1p)
- Describe the most typical approaches in speech enhancement with both single (2p) and multichannel (2p) signals. Include a discussion about noise estimation. (1p) How is output quality evaluated? (1p)
- Suppose you are developing a new speech enhancement algorithm and you are in the process of evaluating performance. You have already measured the PESQ scores for 100 speech samples for a baseline method and your proposed new method (1 result per file per method = a 2×100 matrix). How do you analyze the results? How do you determine if your new proposed method is better than the baseline? (2p informal methods, 2p formal methods) To what extent is PESQ applicable and do you need other tests? (2p)
- How do you design a research and development project in speech enhancement? More specifically, suppose you have an information kiosk with a speech interface in a public space, and there is a need to make it work better in the presence of background noise. You are hired to lead a project to implement a speech enhancement module. What are the steps of this project? Description of steps (2p) as well as their motivation (2p). Are there environmental, social, or privacy questions that need to be taken into account? (2p)

Results from Exam on 16.10.2023 PDF document

→ Previous section
Exercises