Semi-complete list of questions for the Speech Processing Exam

Tom Bäckström

2020

Format: Open-book, online exam using mycourses. Answers returned as text-file through MyCourses.
Allowed:

- Any material provided on course as well as internet sources, search enginges and tools.
- Any software such as Matlab, Mathematica, Python, Word Processing, Turbo Pascal etc.

Not allowed:

- Collaboration with anyone. All answers must be 100% personal.
- Copy-pasting. Every word and formula must be written personnally.

With great freedom comes great responsibility. Even the smallest indications of cheating will be reported and investigated. We will use both cheating-detection tools such as turnitin and manual checking to detect cheating.

Question categories

There are always four questions and the questions are:

1. detail questions (6x1p)
2. basic question (speech production, speech production modelling, 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

The following questions are examples for each category:

1. (Detail question) Explain the main idea of following words and concepts (1p each): *(This is not an exhaustive list, but we will choose 6 words of this type)*
   (a) Phoneme
   (b) Phonation
   (c) Phone
   (d) Vocal tract
   (e) Formant
   (f) Coarticulation
   (g) Onset and offset ($\frac{1}{2}$p+$\frac{1}{2}$p)
   (h) Intonation
   (i) Voiced and unvoiced signals
   (j) Source-filter model
   (k) Objective and subjective quality evaluation ($\frac{1}{2}$p+$\frac{1}{2}$p)
   (l) Fundamental frequency and pitch ($\frac{1}{2}$p+$\frac{1}{2}$p)
   (m) Speaker recognition, verification and diarization (two correct=$\frac{1}{2}$p, three correct=1p)
   (n) Concatenative synthesis and synthesis by speech production modelling ($\frac{1}{2}$p+$\frac{1}{2}$p)
   (o) Features (in the context of, for example, speech or speaker recognition, or voice activity detection)
   (p) Expert and naïve listeners ($\frac{1}{2}$p+$\frac{1}{2}$p)
   (q) ADPCM
   (r) Differential privacy
   (s) Privacy by design
   (t) Federated learning
   (u) Noise gate
   (v) etc.
2. (Basic questions) *(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).
- 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 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)?
- etc.

3. (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,*
   - *algebraic codebook,*
   - *short-time Fourier transform (STFT),*
   - *overlap-add,*
   - *entropy coding,*
   - *voice activity detection (VAD),*
   - *fundamental frequency estimation,*
   - *signal-to-noise ratio, noise reduction 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 μ-law quantization,*
- *feature extraction,*
- *etc.*

4. (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 of speech processing methods during their entire life-cycle (during algorithm and product development, during standardization and marketing of the product, as well as during the time product is in use)? Discuss the strengths and weaknesses of available approaches. (6p)
   - Suppose you are hired by an online magazine to write an article which 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 your 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)
   - Describe 3 different application scenarios for speaker recognition (3p). How do they differ in the training-phase (1p) and evaluation (2p)?