Automatic Speech Processing
Dr. Mathew Magimai Doss
2023/2024 Autumn semester exam
20 January 2024

EPFL

First name:
Family name:
Section:
SCIPER:

Answer 4 out 5 questions

Question 1: *Components of Speech [60 points]*

Describe and illustrate, what is (a) an analysis window: definition, typical length, for what etc., (b) a power spectrum, and (c) a spectrogram? On power spectrum as well as on spectrogram, what are the typical properties of the speech signal that can be observed? *(30 points)*

In an analysis window, through speech signal analysis: (a) how to determine whether the speech signal is voiced or unvoiced? (b) how to estimate pitch frequency? and (c) how to extract formant related information? Concisely describe the key steps. *(30 points)*

Automatic Speech Processing
Dr. Mathew Magimai Doss
2023/2024 Autumn semester exam
20 January 2024

First name:
Family name:
Section:
SCIPER:

Question 2: Speech Analysis-Synthesis *[60 points]*

In the course, we have learned that the speech signal can be decomposed into source and system components, which can be put back together to get the speech signal.

a) **What are the two main methods to achieve that? Explain concisely.** *(10 points)*
b) **How is this understanding applied in speech coding to reduce the bit rate? Describe concisely what happens on the transmitter side and what happens on the receiver side. Illustrate with an example calculation how the bit rate is reduced when compared to sample-by-sample coding and transmission of the speech waveform.** *(20 points)*
c) **How is this understanding applied to build statistical parametric (i.e., hidden Markov model-based) text-to-speech system? Describe concisely how we go from input text to speech.** *(30 points)*

Automatic Speech Processing
Dr. Mathew Magimai Doss
2023/2024 Autumn semester exam
20 January 2024

EPFL

First name:
Family name:
Section:
SCIPER:

Question 3: *Statistical Automatic Speech Recognition [60 points]*

Starting from the speech waveform, clearly describe the different processes (building blocks) involved in statistical continuous speech recognition and what type of prior information and/or models is being used.

- **Clear block diagram of the processing steps, including inputs and outputs.** *(5 points)*
- **How are the lexical constraints being modeled and exploited? Type of model? Training (if any)?** *(5 points)*
- **How are the syntactic/grammatical constraints being modeled and exploited? Type of model? Training?** *(10 points)*
- **How is the acoustic information modeled? Type of model? Training?** *(10 points)*
- **Concisely explain how decoder puts together the different information together to output text. Optimization criteria?** *(10 points)*
- **Given a speech utterance and the trained HMM of all the phones, how can we infer/recognize the phonetic sequence in the speech utterance? How can we integrate phonotactic constraints to improve the inference? Clearly explain the steps with the justification for the choice of Markov model topology. What kind of errors can occur in the inferred/output phonetic sequence?** *(20 points)*

Automatic Speech Processing
Dr. Mathew Magimai Doss
2023/2024 Autumn semester exam
20 January 2024

EPFL

First name:
Family name:
Section:
SCIPER:

Question 4: *Automatic Speaker Recognition [60 points]*

- **Why is speech suitable for biometric person recognition? Which acoustic features can be used for speaker recognition? Justify your answer.** *(10 points)*
- **What is a speaker verification system?** *(20 points)*
    a. **What is the input? What is the output?**
    b. **What is the verification/decision criterion?**
    c. **What is the enrollment procedure?**
    d. **What hyperparameter do we also need to estimate and how does it relate to the decision criterion?**
- **What is a speaker identification system? Hint: input, output, decision criterion. How can it be determined if speaker of the input test speech signal is known or unknown to the speaker identification system?** *(15 points)*
- **What is presentation attack? How can speaker verification/recognition systems be protected against presentation attacks?** *(15 points)*

Automatic Speech Processing
Dr. Mathew Magimai Doss
2023/2024 Autumn semester exam
20 January 2024

EPFL

First name:
Family name:
Section:
SCIPER:

Question 5: Evaluation of Speech Processing Systems *[60 points]*

- **How to evaluate text-to-speech systems?** *(15 points)*
- **How to evaluate automatic speaker verification systems?** *(15 points)*
- **How to evaluate automatic speech recognition systems?** *(15 points)*
- **How to evaluate automatic speaker verification systems?** *(15 points)*

For each of these systems, describe concisely the types of errors, the performance measure, the method to measure performance and the resources needed to evaluate the systems.