Automatic Speech Processing
Dr. Mathew Magimai Doss
2022/2023 Autumn semester exam
23 January 2023

EPFL

First name:
Family name:
Section:
SCIPER:

Answer 4 out 5 questions

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

How is speech signal typically modeled?
(a) Speech signal is the result (convolution) of two phenomena: what are they? Describe concisely a method to deconvolve the two phenomena. ***(20 points)***

(b) What are formants? What kind of information can they convey? Justify what kind of spectrogram is best suited for observing formants. Based on the deconvolution method described in answer to Part (a), concisely explain how formant related information can be extracted? ***(20 points)***

(c) What is pitch frequency? What kind of information can they convey? Justify what kind of spectrogram is best suited for observing pitch frequency. Based on the deconvolution method described in answer to Part (a), concisely explain how pitch frequency information can be extracted? ***(20 points)***

Automatic Speech Processing
Dr. Mathew Magimai Doss
2022/2023 Autumn semester exam
23 January 2023

First name:
Family name:
Section:
SCIPER:

Question 2: Speech and Speaker *[60 points]*

Given "only" two speech utterances recordings (no additional information is available),

a) **how to automatically determine if those two speech utterances represent the same lexical item, i.e., word? Clearly explain the steps, the features, and the matching algorithm. Discuss concisely the decision errors.** ***(30 points)***

b) **how to automatically determine if those utterances were uttered by the same speaker? Clearly explain the steps, the features, and the matching algorithm, Discuss concisely the decision errors.** ***(30 points)***

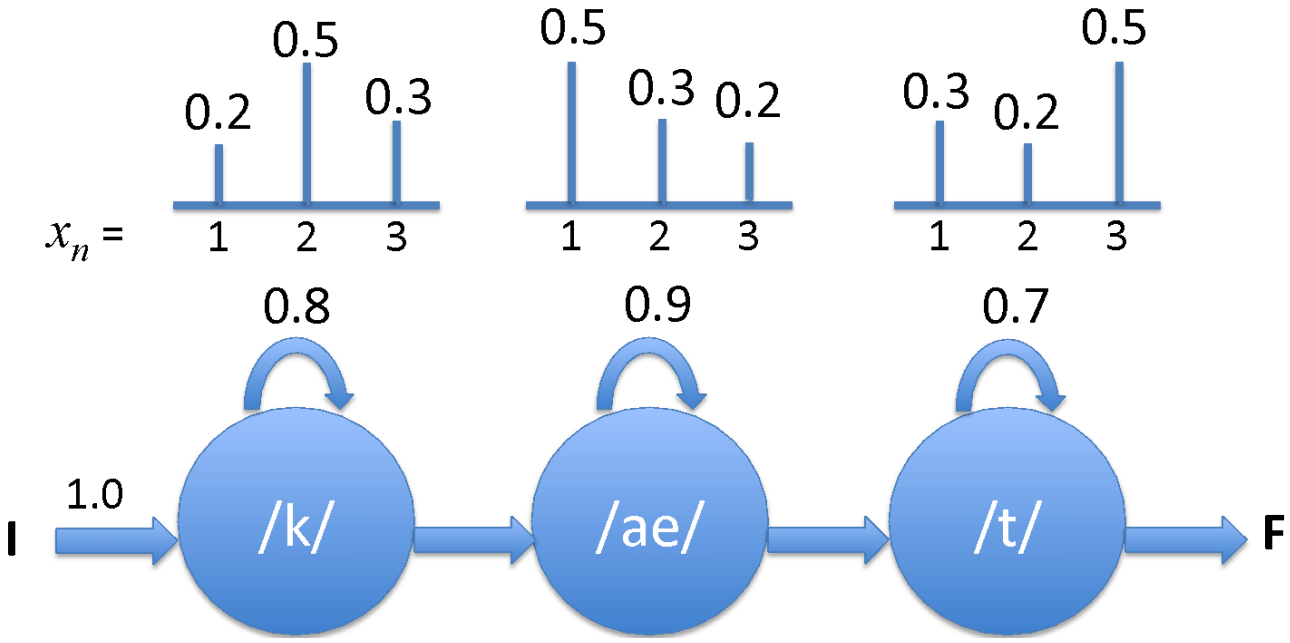
Automatic Speech Processing
Dr. Mathew Magimai Doss
2022/2023 Autumn semester exam
23 January 2023

EPFL

First name:
Family name:
Section:
SCIPER:

Question 3: *Sequence modeling using Markov Models* [60 points]

- **Define concisely discrete Markov models (DMM). What are the parameters of a DMM? How is DMM employed in statistical automatic speech recognition systems? How are the DMM parameters estimated?** ***[20 points]***
- **Define concisely hidden Markov models (HMM). What are the parameters of an HMM? How is HMM employed in statistical automatic speech recognition systems? How are the HMM parameters estimated?** ***[20 points]***
- **Let x_n denote a discrete (acoustic) feature observation at time frame n which belongs to the set of discrete observations {1, 2, 3}.**
  Given
  1) A feature observation sequence $X = [x_1, x_2, .. x_n \ldots x_N] = [2, 2, 3, 1, 1, 1, 2, 1, 3, 3]$
  2) The HMM *M (see figure below)* of word *CAT* with its parameters



Estimate *P(X|M). [20 points]*

Automatic Speech Processing
Dr. Mathew Magimai Doss
2022/2023 Autumn semester exam
23 January 2023

EPFL

First name:
Family name:
Section:
SCIPER:

Question 4: *Lexical constraints [60 points]*

- **Illustrate the following hidden Markov model (HMM) topologies**
  - **i. a K-state left-to-right HMM**
  - **ii. a K-state fully connected ergodic HMM**

  For illustration, choose a value of K of your choice. ***(10 points)***
- **In automatic speech recognition systems, what resource is needed to integrate lexical constraints and how is it obtained? How can the lexical constraints for new words (e.g., a new name, a new place) be obtained in an automatic manner? Which of the HMM topology in (a) is best suited to integrate lexical constraints in ASR systems? Justify concisely.** ***(20 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? Clearly explain the steps with the justification for the choice of HMM topology. How can we integrate phonotactic constraints to improve the inference? What kind of errors can occur in the inferred phonetic sequence?** ***(30 points)***

Automatic Speech Processing
Dr. Mathew Magimai Doss
2022/2023 Autumn semester exam
23 January 2023

EPFL

First name:
Family name:
Section:
SCIPER:

Question 5: Text-to-Speech Synthesis *[60 points]*

Describe text-to-speech synthesis system

- **What is input? What is output?** ***(6 points)***
- **What are the two major building blocks of a text-to-speech (TTS) system and how are they put together to synthesize speech?**
  - **What is the goal of the natural language processing block?**
  - **What is the goal of the digital signal processing/speech processing block?**

  Illustrate the synthesis process (or steps) at a broad level for an example input phrase: "*Dr. Mary had a 10 Kg little lamb*". *(24 points)*
- **Describe concisely the basic principle of unit selection concatenative speech synthesis. What is the cost function for synthesizing speech?** ***(10 points)***
- **What resources could be shared for development of TTS and automatic speech recognition systems? Justify concisely.** ***(10 points)***
- **How are text-to-speech synthesis systems evaluated?** ***(10 points)***