Course Handout

Instructors: Rajesh M. Hegde
TAs: Priyadarshini Dwivedi, Deepali Kushwaha

Course Content

Overview of speech recognition, Modeling the speech production mechanism, Source-system model of speech, Physiological and Mathematical categorization of speech sounds. Discrete time processing of speech signals, Relevance of the DFT, the ZT, convolution. Filter banks, and analytical pole-zero modeling in speech recognition. Short time Fourier Analysis and Spectral estimation models for Speech – DTFT DFT. Pole zero modeling and All pole modeling of speech, LPC model for speech. Basics of Speech Coding. Homomorphic speech signal deconvolution, cepstral analysis, Features for speech recognition: MFCC. Vector Quantization, Pattern Recognition. GMMs for speaker and Language Identification.

Grading Policy

- Exams:

    - Mid Term - 30 %
    - End Term - 30 %

- Assignments - 20 %
- Term Presentation - 20 %

Cheating Policy: There will be zero tolerance when cheating is detected. It may result in F grades. We will follow IITK rules wherever they apply.

Assignments

- Assignments are continuous (every week-one assignment)
- All the assignments contain equal weightage.
- Announcements regarding the assignments will be posted on the assignment page.
- The assignments will be uploaded in the course website on https://hello.iitk.ac.in every week.
- Assignments must be submitted in the due time.
- There will be zero-tolerance if plagiarism is found.

Term Presentation

- Term presentation will be conducted in groups/batches. Each group will contain 4-5 students.
- They have to prepare a presentation on the given topic.
- The presentation will be around last week of February, in virtual mode using zoom/mookit followed by questionnaire.
- The details regarding the groups and the presentation topic will be intimated in the announcements or through mail.

Examination

- The exams will be conducted as per the dates mentioned in the academic calendar

    - Mid-term Examination: Feb 21-26, 2022
    - End-term Examination: Apr 25-May 4, 2022

- The exams will be online using https://hello.iitk.ac.in.

- The exam question will be available on time on the course webpage and the scanned copy of the answer sheets has to be submitted by the student on the due time.

- The students will me monitored by the instructor and the TAs on the video conferencing using zoom/mookIt.

- The students will be live when writing the exams.

- There will be zero-tolerance if any cheating will be detected.

References

- Discrete-Time Speech Signal Processing: Principles and Practice, Thomas F. Quatieri, ISBN:013242942X Published: OCT 29, 2001
Chapters 2, 3, 5, 6, 7, 13 (CMS and SS only), 14 (14.2, 14.3 only)

- Fundamentals of Speech Recognition, L. Rabiner and B. Juang, Prentice-Hall Signal Processing Series, Pages: 507, Year of Publication: 1993, ISBN:0-13-015157-2
Chapters 1, 2, 3, 6, 8

- Speech and Audio Signal Processing: Processing and perception of speech and music B. Gold and N. Morgan, Wiley 2000, ISBN: 0-471-35154-7
Chapters 5 ,6, 7, 8, 9, 19, 20, 21, 22, 23, 24, 25, 26, 28 (overview only)

- Corpus-Based Methods in Language and Speech Processing, Steve Young et. al editors, 234 pages, Kluwer, ISBN 0-7923-4463-4
Chapters 2, 3

- Discrete Time Processing of Speech Signals, JR Deller, JG Proakis, JH Hansen, Year of Publication: 1993, ISBN:0023283017
Chapters 1, 2, 4, 6, 10, 11, 12

- Hidden Markov Models for Speech Recognition, XD Huang, Y Ariki, MA Jack, Edinburgh University Press
Chapters 2,3,4,5,6,8

- Digital Processing of Speech Signals, LR Rabiner and RW Schafer, Pearson Education
Chapters 3, 4, 6, 7, 8.

- Automatic Speech Recognition: A Deep Learning Approach, D. Yu and L. Deng, Springer, 2016

- https://kaldi-asr.org/doc/

- The HTK book