

## Course Content

- I. Overview of the speech production mechanism
  - a) The physiological model of speech production
  - b) The mathematical (source-system) model of speech production
  - c) Relation between (a) and (b)
  - d) Physiological and mathematical basis of categorization of speech sounds
- II. Basic Signal processing techniques for speech recognition
  - a) Discrete time speech signals, relevant properties of the fast Fourier transform and Z-transform for speech processing. Convolution, filter banks, and analytical pole-zero modeling of the speech signal
  - b) Spectral estimation of speech using the Discrete Fourier transform
  - c) Pole-zero modeling, linear prediction (LP) analysis, perceptual linear prediction (PLP), Relative spectral filtering-PLP (RASTA-PLP) analysis of speech
  - d) Basics of speech coding LPC, CELP.
  - e) Homomorphic speech signal deconvolution, real and complex cepstrum, application of cepstral analysis to speech signals
- III. The speech recognition front end and pattern comparison techniques
  - a) Mel frequency cepstral co-efficients (MFCC), MVDR-MFCC, RASTA-PLP cepstral co-efficients.
  - b) Issues in feature vector extraction for speech recognition, Static and dynamic feature vectors for speech recognition, robustness issues, discrimination in the feature space, feature selection
  - c) Log spectral distance, cepstral distances, weighted cepstral distances, distances for linear and warped scales
- IV. Statistical models for speech recognition
  - a) Vector quantization models for speech and speaker recognition
  - b) Gaussian mixture modeling for speaker, language and speech recognition
  - c) Hidden Markov modeling for isolated word and continuous speech recognition
- V. Speech Recognition in practice
  - a) Using the open source HTK and Sphinx for speech recognition
  - b) Speech recognition for mobile devices

## Detailed Reference List:

\* Note that portions of these references will be used for teaching. While you are encouraged to read all the stuff in the book chapters mentioned below, it is not mandatory. You need to read only what is covered in the lectures which will be listed in the course notes link to the left of this page.

1. **Discrete-Time Speech Signal Processing: Principles and Practice**, Thomas F. Quatieri,  
Cloth, 816 pp. ISBN: 013242942X Published: OCT 29, 2001

Chapters 2, 3, 5, 6, 7, 13 (CMS & SS only), 14 (14.2, 14.3 only)

2. **Fundamentals of Speech Recognition**, L. Rabiner and B. Juang,  
Prentice-Hall SignalProcessing Series, Pages: 507, Year of Publication: 1993, ISBN:  
-13-015157-2

Chapters 1, 2, 3, 6, 8

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

Chapters 2, 3

5. **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

6. [IEEE Transactions on audio, speech and language processing](#), (formerly speech and audio processing, ASSP), Available on [IEEE Xplore](#) accessible from inside UCSD and on VPN from outside

[6.a J. Makhoul, Linear Prediction: A Tutorial Review](#)

[6.b JW Picone, Signal Modeling Techniques in Speech Recognition](#)

[6.c SB Davis and P Mermelstein, Comparison of Parametric Representations for Monosyllabic Word Recognition in](#)

[Continuously Spoken Sentences](#)

[6.d H Hermansky and N Morgan, RASTA Processing of Speech](#)

[6.e DA Reynolds and RC Rose, Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models](#)

[6.f LR Rabiner and BH Juang, An Introduction to Hidden Markov Models](#)

[6.g LR Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition](#)

7. The HTK toolkit for speech recognition

<http://htk.eng.cam.ac.uk/>

8. The Sphinx toolkit for speech recognition

<http://cmusphinx.sourceforge.net/html/cmusphinx.php>

9. Hidden Markov Models for Speech Recognition, XD Huang, Y Ariki, MA Jack,  
Edinburgh University Press

Chapters 2,3,4,5,6,8

9. Digital Processing of Speech Signals, LR Rabiner and RW Schafer, Pearson  
Education

Chapters 3, 4, 6, 7, 8.

[Rajesh Hegde](mailto:rhegde@iitk.ac.in)<rhegde@iitk.ac.in>

Dept. of Electrical Engg. IIT Kanpur