Signal Processing Techniques for Speech Recognition - Video course

COURSE OUTLINE

I. The Speech Production mechanism.

   a. Physiological and Mathematical Model.

   b. Relating the physiological and mathematical model.

   c. Categorization of Speech Sounds based on the source-system and the articulatory model.

II. Basic Speech Signal Processing Concepts.

   a. Discrete time speech signals, relevant properties of the fast Fourier transform and Z-transform for speech recognition, convolution, linear and non linear filter banks.

   b. Spectral estimation of speech using the Discrete Fourier transform.

   c. Pole-zero modeling of speech and linear prediction (LP) analysis of speech.

   d. Homomorphic speech signal de convolution, real and complex cepstrum, application of cepstral analysis to speech signals.

III. The Speech Recognition Front End.

   a. Feature extraction for speech recognition, Static and dynamic features for speech recognition, robustness issues, discrimination in the feature space, feature selection.

   b. Mel frequency cepstral co-efficients (MFCC), Linear prediction cepstral coefficients (LPCC), Perceptual LPCC.

IV. Distance measures for comparing speech patterns.

   a. Log spectral distance, cepstral distances, weighted cepstral distances, distances for linear and warped scales.

   b. Dynamic Time Warping for Isolated Word Recognition.

V. Statistical models for speech recognition.

   a. Vector quantization models and applications in speaker recognition.

   b. Gaussian mixture modeling for speaker and speech recognition.

   c. Discrete and Continuous Hidden Markov modeling for isolated word and continuous speech recognition.

NP-TEL

NPTEL

http://nptel.iitm.ac.in

Electronics & Communication Engineering

Pre-requisites:

1. There are no mandatory pre-requisites for the course.
2. It is designed to be self contained. However a course on Signals and Systems and Digital Signal Processing will come in handy for appreciating the course.

Additional Reading:

1. IEEE Transactions on audio, speech and language processing, (formerly speech and audio processing, ASSP), Available on IEEE Xplore accessible from inside all IITs and INDEST members.
   a. J. Makhoul, Linear Prediction: A Tutorial Review.
   b. JW Picone, Signal Modeling Techniques in Speech Recognition.
   c. SB Davis and P Mermelstein, Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences.
   d. H Hermansky and N Morgan, RASTA Processing of Speech.
   e. DA Reynolds and RC Rose, Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models.
   f. LR Rabiner and BH Juang, An Introduction to Hidden Markov Models.
   g. LR Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.

Hyperlinks:

1. The HTK toolkit for speech recognition http://htk.eng.cam.ac.uk/.

VI. Using the HTK toolkit for building a simple speech recognition system.

COURSE DETAIL

Sl. No.	Lecture Topic	No.of Hours
1	The Speech Production mechanism. a. Physiological and Mathematical Model. b. Relating the physiological and mathematical model. c. Categorization of Speech Sounds based on the source-system and the articulatory model.	05
2	Speech Signal Processing Concepts. a. Discrete time speech signals, relevant properties of the fast Fourier transform and Z-transform for speech recognition, convolution, linear and non linear filter banks. b. Spectral estimation of speech using the Discrete Fourier transform. c. Pole-zero modeling of speech and linear prediction (LP) analysis of speech. d. Homomorphic speech signal de convolution, real and complex cepstrum, application of cepstral analysis to speech signals.	10
3	The Speech Recognition Front End. a. Feature extraction for speech recognition, Static and dynamic features for speech recognition, robustness issues, discrimination in the feature space, feature selection. b. Mel frequency cepstral co-efficients (MFCC), Linear prediction cepstral coefficients (LPCC), Perceptual LPCC.	06
4	Distance measures for comparing speech patterns : Log spectral distance, cepstral distances, weighted cepstral distances, distances for linear and warped scales.	03
5	Dynamic Time Warping for Isolated Word Recognition.	03
6	Statistical models for speech recognition.	10

2. The Sphinx toolkit for speech recognition

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

Coordinators:

Dr. Rajesh M. Hegde
Department of Electrical EngineeringIIT Kanpur

	a. Vector quantization models and applications in speaker recognition. b. Gaussian mixture modeling for speaker and speech recognition. c. Discrete and Continuous Hidden Markov modeling for isolated word and continuous speech recognition.	
7	Using the HTK toolkit for building a simple speech recognition system.	03
	Total	40

References:

1. **Discrete-Time Speech Signal Processing: Principles and Practice**, Thomas F. Quatieri, Cloth, 816 pp. ISBN: 013242942X Published: OCT 29, 2001.
2. **Fundamentals of Speech Recognition,** L. Rabiner and B. Juang, Prentice-Hall SignalProcessing Series, Pages: 507, Year of Publication: 1993, ISBN:0-13-015157-2.
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.
4. **Corpus-Based Methods in Language and Speech Processing**, Steve Young et. al editors, 234 pages, Kluwer, ISBN 0-7923-4463-4.
5. **Discrete Time Processing of Speech Signals**, JR Deller, JG Proakis, JH Hansen, Year of Publication: 1993, ISBN:0023283017.
6. **Hidden Markov Models for Speech Recognition**, XD Huang, Y Ariki, MA Jack, Edinburgh University Press.
7. **Digital Processing of Speech Signals**, LR Rabiner and RW Schafer, Pearson Education.