



Speech Recognition for Dialects & Spoken Tutorials

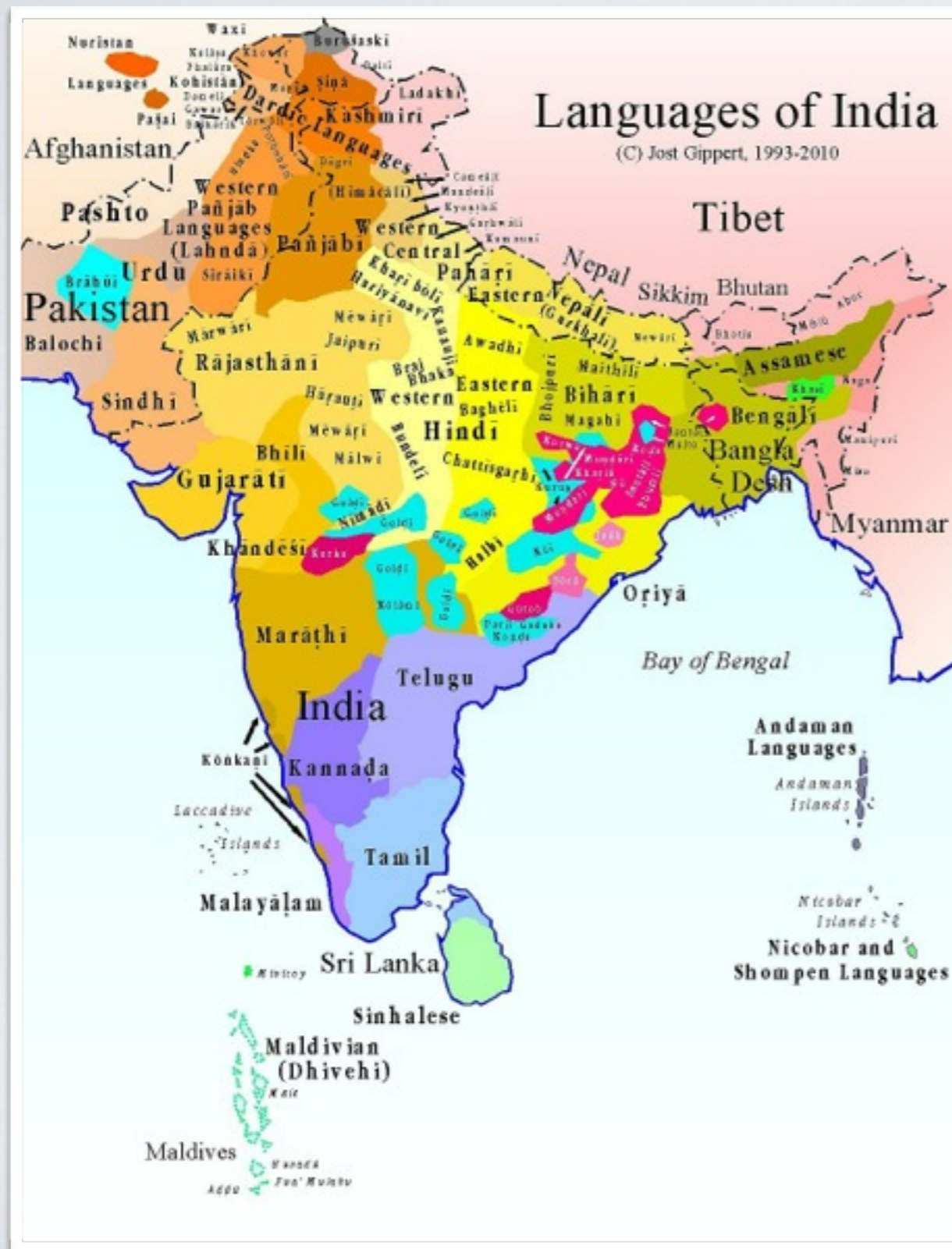
M.Tech. 1 Seminar Topics

Preethi Jyothi
Department of CSE, IIT Bombay

Automatic Speech Recognition

- Automatic Speech Recognition (ASR) is one of the oldest (early 1900s) and most complex sequence prediction tasks
- Modern ASR systems are dominated by statistical methods pioneered by [Jelenik '76]
 - Noisy channel model: Given an input speech utterance, what is the most likely text sequence?
- Current state-of-the-art in ASR involves a complex pipeline with many machine learning components

Languages in the Indian subcontinent



- 30 Indian languages spoken by >1M native speakers
- Hindi and Bengali among the world's most populous languages
- Despite this, Indian languages (barring Hindi) considered to be low-resource for ASR

Challenges

- Rich language diversity (more than 150 languages and more than 1500 dialects!)
- Morphological Complexity
 - Dravidian languages pose extra challenge, being *agglutinative*
 - Lack of standard lexicons/morphological analysers
- Syntactic Complexity
 - E.g. free word order
- Limited prior work
 - Lack of diversity in ASR tasks
 - Lack of annotated corpora in many Indian languages

Seminar Topics

- Speech recognition of Indian dialects
 - Topic 1: Acoustic model adaptation using dialectal speech
 - Topic 2: Discriminative pronunciation and language modelling for dialectal speech
- Automatic transcription of spoken tutorials in Indian languages
 - Topic 3: Leveraging side information for automatic transcription of spoken tutorials

Seminar Topics

- Speech recognition of Indian dialects
 - Topic 1: Acoustic model adaptation using dialectal speech
 - Topic 2: Discriminative pronunciation and language modelling for dialectal speech
- Automatic transcription of spoken tutorials in Indian languages
 - Topic 3: Leveraging side information for automatic transcription of spoken tutorials

Acoustic model adaptation using dialect speech

- To handle dialects, either build a) an ensemble of dialect-specific recognizers or b) a common language-specific recognizer.
 - E.g.: Strategy adopted by Google VoiceSearch: Route spoken query to a specific dialectal recognizer based on location information.
- Potential for large improvements in current strategies and this is a largely unexplored area for dialects of Indian languages.
- *Reading:* Papers on acoustic model adaptation using both Hidden Markov Models and Deep Neural Network based systems.

Seminar Topics

- Speech recognition of Indian dialects
 - Topic 1: Acoustic model adaptation using dialectal speech
 - Topic 2: Discriminative pronunciation and language modelling for dialectal speech
- Automatic transcription of spoken tutorials in Indian languages
 - Topic 3: Leveraging side information for automatic transcription of spoken tutorials

Pronunciation/language modelling of dialectal speech

- To extend an ASR system to a new dialect, pronunciation/language models are enhanced by adding pronunciation variants/new words.
 - Can we automatically learn phonological rules governing pronunciation differences in dialect speech (compared to the standard dialect)?
 - How to devise good discriminative models to learn weights for these rules? How about for language models?
- *Reading:* Papers on discriminative pronunciation and language modelling.

Seminar Topics

- Speech recognition of Indian dialects
 - Topic 1: Acoustic model adaptation using dialectal speech
 - Topic 2: Discriminative pronunciation and language modelling for dialectal speech
- Automatic transcription of spoken tutorials in Indian languages
 - Topic 3: Leveraging side information for automatic transcription of spoken tutorials

Transcribing Lectures using S(l)ide Information

- Transcribing lectures and leveraging information on slides to build contextual language models
 - Will be using data from spokentutorial.org
- Produce sub-titles. Could we add sentence markers? Will require an augmented language model and detecting informative cues in the speech signal.
- *Reading:* Papers on language modelling and prosodic analysis of speech.

Interested?

- Requires a strong grasp of probability and statistics.
- Coding component – all seminar topics will require:
 - Building an ASR system for an Indian language of your choice, using the open-source ASR toolkit, Kaldi.
 - Subsequent MTPs: Develop new techniques addressing your research problem and incorporate them into the above ASR system