

# **An Introduction to Large Vocabulary Continuous Speech Recognition**

## **Abstract**

The goal of Large Vocabulary Continuous Speech Recognition system (LVCSR) is to develop all aspects of speech recognition in the domain of spontaneous, conversational speech as opposed to planned or read speech. In this tutorial, we will cover all theoretical and practical aspects involved in building a state-of-the-art LVCSR system. The tutorial will discuss in detail the various blocks in a LVCSR including:

1. Feature extraction: Mel-frequency cepstral coefficients and Perceptual Linear prediction (PLP) cepstral coefficients
2. Discriminant Analysis and Linear Transformations.
3. Basics of Hidden Markov Modeling
4. Model training: Forward-backward algorithm, Baum-Welch re-estimation
5. Triphone Modeling and Decision-tree state clustering
6. N-gram language models
7. Recognition and Decoding
  
8. Adaptation and Normalization for robustness to channel and speaker variations.

Finally, as an example, we will discuss an LVCSR system that was built for the evaluation of American Broadcast News as a part of the U.S. DARPA/NIST speech-to-text evaluation.

## **Bio**

S. Umesh completed his PhD from the University of Rhode Island in 1993, and was a PostDoctoral Fellow at the City University of New York until 1996. From 1996 to July 2009, he was with the Department of Electrical Engineering at IIT-Kanpur, first as Assistant Professor and finally as Professor. He is currently Professor of Electrical Engineering at IIT-Madras. He has also been a visiting researcher at AT&T Research Laboratories, USA; at Machine Intelligence Laboratory Cambridge University Engineering Department, UK and the Department of Computer Science, RWTH-Aachen, Germany. He is a recipient of the AICTE Career Award for Young Teachers in 1997 and the Alexander von Humboldt Research Fellowship in 2004. His recent research interests have been mainly in the area of speaker-normalization and acoustic modeling and their application in large vocabulary continuous speech recognition systems. During his stint at Cambridge University in 2004, he was part of the U.S. DARPA's Effective, Affordable Reusable Speech-to-text (EARS) programme. The aim of the project is to very significantly advance the state-of-the-art while tackling the hardest speech recognition challenges including the transcription of broadcast news and telephone conversations. Similarly in 2005 he was part of the RWTH-Aachen's TC-STAR project for transcription of speech from European Parliament's Plenary Sessions.