# Draft: Word Spotting without sense labels

PVS Rao and Sunil Kopparapu
Speech Group, Cognitive System Research Laboratory
Tata Infotech Limited, Mumbai

September 20, 2005

**Abstract**

## 1 Procedure

1. Words $\mathcal{W} \in W_1, W_2, \cdots, W_n$ has to be spotted (say) in continuous speech
2. Offline Processing (Continuous Speech)
   (a) Take a large amount of speech corpus
   (b) Use dendrogram[1] technique to segment the speech corpus - ideally each segment is different from the other (in some way).
   *Is it possible to segmented without providing a metric or measure? But there is a paper:* **A New Fast Algorithm for Automatic Segmentation of Continuous Speech**, *Iman Gholampour and Kambiz Nayebi, Electrical Engineering Department, Sharif University of Technology (Iran), ICSLP 1998 : In this paper a new method for automatic segmentation of continuous speech into phone-like units is addressed. Our method is based on a very fast presegmentation algorithm which uses a new statistical modeling of speech and searching in a multilevel structure, called Dendrogram, for decreasing insertion rate. Performance of algorithms have been tested over a large set of TIMIT sentences. According to these tests, our final segmentation algorithm is capable of detecting nearly 97% of segments with an average boundary position error of less than 7 msec and average insertion rate of less than 12.7%. In addition to acceptable precision, our overall segmentation scheme has very low computation cost and it can be implemented in real time on an average Pentium PC. The major advantage of presented algorithms is that no training or threshold estimation is needed in realizing them. Details of proposed algorithms and their performance results are included in the paper.*
   (c) Cluster all the speech segments into $M$ bins
   $M$ not very large, use k-means for segmentation

[1] By keeping track of the similarity score when new clusters are created, the dendrogram can often yield insights into the natural grouping of the data.

(d) Assign labels (arbitrary) to each bin

Note: These labels will not necessarily have sense in terms of speech units like phoneme

(e) Construct HMM for each bin using the speech samples in each bin.

3. Offline processing (Word $W_n$ to be spotted)

(a) Collect $N$ different utterances (different accent, environment) of the word $W_n$

(b) For each of the $N$ utterance get a trellis of labels using Viterbi (using the HMMs constructed with continuous speech).

(c) Combine all the $N$ labels ($N$ different utterances) of the word to form a STN (state transition network)

4. Word Spotting (test speech)

(a) Using the HMMs do speech recognition on the test speech.

(b) Construct a trellis

(c) Search through the trellis to spot for the word $\in \mathcal{W}$.