The 1996 Mississippi State University Conference on

Speech Recognition

What: EE 8993 - 02 Project Presentations
Where: 432 Simrall, Mississippi State University
When: May 1, 1996 — 1:00 to 4:00 PM

SUMMARY

The Department of Electrical and Computer Engineering invites you to attend a mini-conference on Speech Recognition, being given by students in EE 8993 — Fundamentals of Speech Recognition. Papers will be presented on a wide range of topics including signal processing, Hidden Markov Models, search, and language modeling.

Students will present their semester-long projects at this conference. Each student will give a 12 minute presentation, followed by 3 minutes of discussion. After the talks, each student will be available for a live-input real-time demonstration of their project. These projects account for 100% of their course grade, so critical evaluations of the projects are welcome.

Session Overview

1:00 PM — 1:10 PM: J. Picone, Introduction

1:15 PM — 1:30 PM: **J. Trimble**, “Front-end of a Speech Recognizer”

This proposal describes a plan to design and implement the front-end of a speech recognition sytem. The front end must derive a smooth spectral estimate of a signal in order to produce feature vectors that are compatible with the acoustic models of the system. Linear prediction provides an efficient and simple means of computing these feature vectors. Its basic purpose is to as accurately predict currents values of a signal based on a weighted sum of the signal’s previous values. In addition, an even better spectral estimator, cepstral analysis, will be implemented.

1:30 PM — 1:45 PM: **L. Webster**, “Front End Modeling with Special Emphasis on FFTs, LPC, and Feature Selection”

This proposal describes a plan to design and implement the front-end of a speech recognition sytem. The front end must derive a smooth spectral estimate of a signal in order to produce feature vectors that are compatible with the acoustic models of the system. Linear prediction provides an efficient and simple means of computing these feature vectors. Its basic purpose is to as accurately predict currents values of a signal based on a weighted sum of the signal’s previous values. In addition, an even better spectral estimator, cepstral analysis, will be implemented.

1:45 PM — 2:00 PM: **R. Seelam**, “Implementation of Statistical Modeling Techniques and Channel Adaptation Techniques”

Implementation of various Statistical Modeling Techniques is necessary for the building of a Speech Recogniser. Statistical Modeling is done to learn the nature of the multi-variate random process generating the signal parameters. In this direction, pre-whitening transformations will be performed on the parameters to eliminate redundancy and to make the analysis easier. The transformations will be performed on the input vector to produce an uncorrelated Gaussian random vector, containing only “information-bearing” parameters. For some algorithmically complex computations such as the computation of the covariance matrix, existing software will be used. Evaluation will be done by comparing the performance of the software on different classes of data. Channel adaptation techniques will be implemented so as to make the parameters robust to changes in the acoustical environment. For this purpose, two particularly simple, but effective algorithms, Cepstral Mean Normalisation/Subtraction and RASTA have been chosen. The Codeword-Dependent Cepstral Normalisation method will also be studied. Evaluation will be done by comparing the performance of my software with existing public-domain software.

2:00 PM — 2:15 PM: **A. Ganapathiraju,** “Implementation of Viterbi Beam Search Algorithm”

Speech Recognition can be treated in a very general sense as a structured search problem. Correct recognition is defined as outputting the most likely word sequence given the language model, the acoustic model and the observed acoustic data. In this project I plan to implement a very commonly used search algorithm, Viterbi Beam Search. For this purpose continuous observation HMMs will be implemented to represent phonemes and will be trained on a large data set. The Viterbi reestimation algorithm will be used for the purpose of training the HMMs. The search algorithm will then be used to output the most likely phoneme sequence that matches the observed acoustic data. The code will be implemented in GNU C++ and the structure will be made very generic so as to allow for using other search algorithms at a later stage. The design will keep in mind the algorithm’s integration with various other modules like the language model, and the front-end signal processor, to form a simple continuous speech recognizer. For evaluating the performance of the search engine, results from other public-domain speech recognizers will be compared. One of the main aims of this project will be to study the recognition performance dependence on various aspects of the HMM parameters like the number of states per model.

2:15 PM — 2:30 PM: **N. Deshmukh**, "Efficient Search Algorithms for Large Vocabulary Continuous Speech Recognition"

Automatic speaker-independent speech recognition has made significant progress from the days of isolated word recognition. Today large-vocabulary continuous speech recognition (LVCSR) over complex domains such as news broadcasts and telephone conversations is a reality. A major component of this success is the result of recent advances in search techniques that support efficient, sub-optimal decoding over large search spaces. These evaluation strategies are capable of dynamically integrating information from a number of diverse knowledge sources to determine the correct word hypothesis. In this project we propose to implement two major classes of such decoding algorithms viz. multipass N-best search and search using models with a weighted mixture of Gaussian probabilities as density functions. The performance of this LVCSR system will be evaluated on speech data in the public domain and compared with that of other recognizers as a benchmark. We will also evaluate the search algorithm modules in isolation using statistical measures. The final software will be placed in the public domain at the Institute of Signal and Information Processing (ISIP).

2:30 PM — 2:45 PM: **O. LaGarde**, "Language Modeling and Grammar Construction for an HMM Continuous Speech Recognition System"

This project will consist of the implementation of Language Model (LM) objects for the construction of regular stochastic grammars based on Bigrams and Ngrams for both words in a training text and for phones in phonetic series equivalents for those words. An externally produced word-to-phone dictionary compliant with the Worldbet symbol set will be acquired as a supporting resource. A method of deriving Ngrams as the joint product of weighted Bigrams will be defined and investigated, Implementation of grammar construction and polling objects will target a Continuous Speech Recognition (CSR) system's search engine as the principle user; the objects will provide query response services for probabilities of current and proposed states in the search engine's domain as well as sequences of hypothesized next-states. The models will serve primarily to assist in formation of sentences from hypothesized word sequences through the implementation of a token-based grammar and Unigram/Bigram/Ngram generation scheme. Although the current CSR organization does not require phonetic-level support, such support will implicitly be provided in the token-based model organization. The LM will be tested using input perplexity as a benchmark and both alternate Ngram generation methods and independently produced models providing similar services as measures of performance. Deliverables will include object interface and implementation specifications, results of Ngram generation methodology experimentation, and a set of C++ coded language model construction and request servicing objects.

2:45 PM — 3:00 PM: **S. Given, "**Development of an N-Gram Based Language Model"

An essential element of any speech recognition system is the language model. A language model attempts to identify and make use of the regularities in natural language to better define language syntax for easier recognition. One major obstacle in speech recognition is variability and uncertainty of message content. This, coupled with inherent noise, distortion and losses that occur in speech, emphasize the need for a good language model. Several different types of language modelling techniques exist. This project will concern itself mainly with statistical language modelling. Statistical language modelling uses large amounts of text to automatically compute the model's parameters. This is called training. Language models can be compared using standard measures such as perplexity and recognition or word error rate. This project will use perplexity as a benchmark. A good language model will provide *a priori* probabilities for all possible queries that the search algorithm may request pertaining to the learned vocabulary. Hence, the complexity of the model is directly related to the size of the corpus upon which it is trained.

3:00 PM — 4:00 PM: Demonstrations in 414 Simrall