EE 84?3 — Fundamentals of Speech Recognition

Effective date: Spring Semester 1999

New catalog listing:

EE 84?3: Fundamentals of Speech Recognition. (3). Three hour lecture. Speech Production and Perception; Acoustic Phonetics; Discrete-Time Models of Speech Production; Short-Term Spectral Measurements; Linear Prediction; Dynamic Programming; Hidden Markov Models; Statistical Language Modeling; Neural Networks; An Overview of Commercial Speech Recognition Systems.

1.     DETAILED COURSE OUTLINE

LECTURE
(5 hours)  1.  REVIEW OF DIGITAL SIGNAL PROCESSING/INFORMATION THEORY
        A.  Overview of a Speech Recognition Systems
        B.  Sampling, Transforms, Frequency Response
        C.  Probability and Random Processes
        D.  Probabilistic Distance Measures and Maximum Likelihood Classification
        E.  Entropy, Information, Divergence

(6 hours)  2.  Fundamentals of Speech Production and Perception
        A.  Speech Waveforms and Spectrograms
        B.  The Acoustic Theory of Speech Production
        C.  Phonemics and Phonetics
        D.  Classification of Vowels and Consonants
        E.  Sound Propagation
        F.  The Digital Speech Production Model

(6 hours)  3.  Short-Term Spectral and Temporal Features
        A.  Time-Domain Windowing
        B.  Recursive-in-Time Approaches
        C.  Autocorrelation and Covariance
        D.  Linear Prediction
        E.  Cepstral Techniques
        F.  Front-Ends in Speech Recognition

(5 hours)  4.  Dynamic Programming
        A.  The Principle of Optimality
        B.  Word Spotting Via Unconstrained Endpointing
        C.  Syntactic Constraints, Network Search, and Beam Search
        D.  The One-Stage DP Algorithm
        E.  The Pattern Recognition Paradigm

(6 hours)  5.  Hidden Markov Models
        A.  Examples of Markov Processes
        B.  Hidden Processes  and Doubly Stochastic Systems
        C.  The Baum-Welch Algorithm
        D.  The Viterbi Algorithm
        E.  Training of Speech Recognition Systems
        F.  Continuous Density Systems

(6 hours)  6.  Recognition Architectures
        A.  Acoustic Model Topologies
        B.  Scaling in HMMs
        C.  The Supervised Training Paradigm
        D.  The DTW Analogy

E.  Continuous Speech Recognition Using HMMs
                        F.  The Network Search Problem (Parsing Strategies)

(6 hours)   7.  Statistical Language Modeling
                        A.  Formal Language Theory
                        B.  Probabilistic Grammars
                        C.  N-Gram Language Models and Perplexity
                        D.  Training and Interpolation Strategies
                        E.  Decoding Strategies for Large Scale Systems
                        F.  State-of-the-Art HMM-Based Speech Recognition Systems

(5 hours)   8.  Neural Networks in Speech Processing
                        A.  The Artificial Neural Network
                        B.  Multilayer Perceptrons
                        C.  Hybrid Speech Recognition Systems
                        D.  Recurrent Neural Networks
                        E.  Comparisons of Various Approaches

## 2.      METHOD OF EVALUATION

| | |
|---|---|
| Project Demonstration | 15% |
| Project Paper | 10% |
| Project Presentation | 25% |
| Final Exam | 50% |

## 3.      JUSTIFICATION

This course is being added to the Electrical and Computer Engineering curriculum as the entry-level graduate course in speech recognition. It develops the basic notions of statistical pattern recognition and shows how statistical methods can be applied to the speech to text problem. Students are expected to be upper-level graduate students at the time they take this course, and be familiar with some aspects of the speech or natural language processing problem. In addition to developing the underlying theory, students are forced to turn theory to practice by implementing a portion of a speech recognition in software for their semester-long project. There are only a few courses of this type taught at major universities in the U.S. This course is intended to complement MS State's strong research profile in this area.

## 4.      SUPPORT

This course will be taught by existing personnel in the Department of Electrical and Computer Engineering. This new workload has been considered in the departmental staffing plan. Semester-long projects utilize existing computer facilities in the department. Current library holdings are adequate to meet the needs of this course.

## 5.      INSTRUCTOR OF RECORD

Joseph Picone

## 6.    GRADUATE STUDENT REQUIREMENTS

None.

## 7.    PLANNED FREQUENCY

Planned frequency is once every two years.

## 8.    EXAMINATION OF DUPLICATION

This new course will not duplicate any existing course in the MSU Bulletin.

## 9.    METHOD OF INSTRUCTION CODE

???

## 10.    PROPOSED C.I.P. NUMBER

???

## 11.    PROPOSED 20-CHARACTER ABBREVIATION

ECE SPEECH REC.

## 12.    PROPOSED SEMESTER EFFECTIVE

Spring 1999

## 13.    OTHER APPROPRIATE INFORMATION

Textbook:          J. Deller, et. al., *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., 1995, ISBN 0-02-328301-7.

## 14.    PROPOSAL CONTACT PERSON

Mike Nosser, 325-3912