

# 11-751: Speech Recognition and Understanding

Grading Information and Review

(12-03-2008)

# Important Dates

---

<b>Date</b>		<b>Where</b>	<b>Time</b>
Friday December 12 <sup>th</sup>	Final Examination	Newell-Simon Hall 3002	10:00 – 13:00
Monday December 15 <sup>th</sup>	Project Presentation	Doherty Hall 1209 (here)	16:30 –
Tuesday December 16 <sup>th</sup>	Project Report Due	My office (RM 221, 407 S.Craig street)	24:00

## Today:

- Details of term project (Presentation and Report)
- Details of final examination
- What you should review
- Review of homeworks (Wilson)

# 11-751: Grading

---

The final grade consists of:

40% Exam

30% Homework

30% Term project

# Term Project Grading

---

1. Quantity and quality of work
2. Report
3. Presentation
4. Task difficulty
5. Amount of support needed

# Term Project Reports

- **Due:** December 16<sup>th</sup> 24:00 (Hand to me or place at my office door)
- Detailed scientific report, up to 10 pages:
  - State the problem you tackled **1/2 page**
  - Literature survey, relevant work **1 page**
  - Describe the approach(es) you used **1 1/2 pages**
  - Experimental results **2 pages**
  - Analyze of Results **3 pages**
    - Reasons for success or failure.
    - Limitations of the approach, solution. What can be improved?
  - Lessons Learnt **1 page**
    - Changes in goals/timeline
    - What problems did you encounter?
  - Conclusions and future directions **1 page**

# Term Projects Presentations

- **When:** Monday December 15th, 16:30 -
- **Where:** Doherty Hall 1209
- **Research Presentation**
  - 20 min for single speaker (15min talk + 5min Q+A)
  - ~ 15 minutes talk **10-12 slides**
  - Introduction to the problem you solved **1 slide**
  - The approaches you used **2 slides**
  - Experimental results **1-3 slides**
  - Analysis of the results **1-2 slides**
  - Possibly improvements **1 slide**
  - Lessons Learnt **1 slide**
  - Conclusion **1 slide**
  - Future work (if you want to continue) **1 slide**

# Term Projects (What to keep in mind)

---

- What is the problem you are investigating?
- How have others attempted to solve this problem?
- What are the limitations of previous approaches?
- What is your solution?
  - How does your approach overcome these limitations?
- Describe the experimental setup in detail
  - Another student should be able to read your report and reproduce your results
  - Consistency of setup across experiments
- What are the results? Show comparative results in graphs, and describe results in text
  - There should be a reason for each result you include
- What do these results mean?
- Conclusions and future directions

# Final Exam (Closed Book)

---

- **DATE:** Friday, December 12, 2008
- **TIME:** 10:00 – 13:00am
- **PLACE:** Newell-Simon Hall 3002
- **WHAT YOU CAN BRING:**
  - Calculator, Ruler, Watch
  - Pens (please don't use pencil for the final filling)
- **BE ON TIME**
  - **Start 10:00am sharp:** If you are late - it's your time you are wasting
- **DURING EXAM**
  - **Keep an eye on the time:** will indicate every hour how much time is left

# How to Prepare For the Exam

---

- In order of importance

## 1) Revisit and understand lectures slides

- Digital Signal Processing and Representations of Speech Signals
- Hidden Markov Models
- Acoustic Modeling
- Language Modeling
- Dynamic Time Warping and Search
- Adaptation, Discriminative Training, Applications

## 2) Revisit homework questions and solutions

## 3) Re-read Chapters in “Spoken Language Processing”

- Chapter 6:                      Speech Signal Representation
- Chapters 8,9:                    HMMs and Acoustic modeling
- Chapter 11:                    Language Modeling
- Chapters 12,13:                Search

# Exam Structure

---

- **Section 1 (short answer problems)**
  - 20 multiple choice & 10 short questions
  - Answer them all!
- **Section 2 (long answer problems)**
  - From 6 questions, select and answer 4
  - If you answers to more than 4 problems, **CLEARLY INDICATE** which 4 problems you want us to mark  
(If not indicated we will mark first 4)
  - We will not consider more than 4 problems!
- Each question is tagged with the number of points you will get for the correct answer

# Section 1 (Example Questions)

- **Multiple choice questions (Select One)**

- A 4-gram language model determines the frequency with which the word  $w_i$  occurs given the history of  $w_{k-1}, w_{k-2}, \dots, w_{k-n}$ .  $n$  has the size of: **(a)** 1 **(b)** 2 **(c)** 3 **(b)** 4 **(c)** 5

- **Multiple choice questions (Select One or More)**

- The following statements are true for Word Error Rate.  
**(a)** WER can be negative; **(b)** WER can be larger than 100%; **(c)** WER is determined by the Levenshtein distance between the reference and hypothesis; **(d)** WER is always larger than the phone error rate; **(e)** WER is based on the sum of word insertion, deletion, and substitution errors

- **Questions requiring short answers or illustrations**

- Write down the fundamental equation of speech recognition.

OR

Give a block diagram of a speech recognition system. Label the input and output.

# Section 2 (Example Questions)

---

- **Answer 4 of the 6 problems**

- The problems are likely to cover the focus areas of this course

- **Search:**

Imagine you are asked to reduce the computational effort in an Large Vocabulary Speech Recognition System. Which methods could be applied to reduce the computational effort?

Name 5 methods and explain each in 1 or 2 sentences.

# Topics Covered in This Course

---

- Introduction to Speech Recognition
- Spoken Language: Speech Production & Perception
- Digital Signal Processing
- Classification
- Dynamic Programming and Single Word Recognizers
- Hidden Markov Models and their Recognizers
- Acoustic Modeling and Context Dependencies
- Language Modeling
- Search for Speech Recognition
- Adaptation, Discriminative Training
- Natural Language Processing and Understanding
- Spoken Language Translation

# Topics You Should Revise

---

- Introduction to Speech Recognition
- Spoken Language: Speech Production & Perception
- **Digital Signal Processing**
- Classification
- **Dynamic Programming and Single Word Recognizers**
- **Hidden Markov Models and their Recognizers**
- **Acoustic Modeling and Context Dependencies**
- **Language Modeling**
- **Search for Speech Recognition**
- **Adaptation, Discriminative Training**
- Natural Language Processing and Understanding
- Spoken Language Translation

# Digital Signal Processing

---

- Nature of Speech: Sound, Formants, Acoustic Features
- Describe and explain feature extraction:
  - Sampling, Sampling theorem, aliasing
  - Continuous-time and Discrete-time Fourier Transform
  - Short-time Fourier Analysis (conventional Fourier does not capture time-varying nature of speech signal, solved by windowing and applying DTFT)
  - Effect of windowing
- Z-Transform (generalization of DTFT)
  - Poles, zeros give insight to frequency response of linear system
  - Features for speech recognition
    - Cepstral coefficients
    - LPC-derived cepstral coefficients (LPCC)

# Dynamic Programming and Single Word Recognizers

---

- Comparing Complete Utterances: Problems
- Endpoint Detection and Speech Detection
- Approaches to Alignments of Vector Sequences
- Time Warping
- Distance Measure between two Utterances
- The Minimal Editing Distance Problem
- Dynamic Programming
- Utterance Comparison by Dynamic Time Warping
- Constraints for the DTW-Path
- The DTW Searchspace
- DTW with Beam Search
- The Principles of Building Speech Recognizers
- Isolated Word Recognition with Template Matching

# Dynamic Programming and Single Word Recognizers

---

- Comparing Complete Utterances: Problems
- Endpoint Detection and Speech Detection
- Approaches to Alignments of Vector Sequences
- Time Warping
- Distance Measure between two Utterances
- The Minimal Editing Distance Problem
- Dynamic Programming
- Utterance Comparison by Dynamic Time Warping
- Constraints for the DTW-Path
- The DTW Searchspace
- DTW with Beam Search
- The Principles of Building Speech Recognizers
- Isolated Word Recognition with Template Matching

# Hidden Markov Models and HMM-based Recognizers

---

- Problems with Pattern Matching
- Speech Production seen as Stochastic Process
- Examples for HMMs
- Typical HMM-Topologies
- Formal Definition of Hidden Markov Models
- Some Properties of Hidden Markov Models
- Three Main Problems of Hidden Markov Models
- The Evaluation Problem: The Forward Algorithm
- The Decoding Problem: The Viterbi Algorithm
- The Learning/Optimization Problem
- The Forward-Backward Algorithm

# Acoustic Modeling

---

- Discrete vs Continuous HMMs
- Parameter Tying
- Codebook Sizes
- Pronunciation Variants
- Context Dependent Acoustic Modeling
- Speech Units
- Clustering of Context
- Bottom-Up vs. Top-Down Clustering
- Distances Between Model Clusters
- Clustering with Decision Trees

# Language Modeling

---

- Language Modeling in Automatic Speech Recognition
- Probabilities of Word Sequences
- Interpolation of Language Model Parameters
- Parameter Smoothing
- Measuring the Quality of Language Models, Perplexity
- Different Kinds of Language Models
  - Cache, Trigger, Multilevel, Interleaved, Morpheme-Based, Context-Free Grammars, ...
- Practical Issues
  - Spontaneous Speech
  - Unknown Words
  - Different Languages

# Search for Speech Recognition

---

- The Search in Automatic Speech Recognition
- DTW → pattern based recognition
- Viterbi Decoding → model based recognition
- Continuous speech recognition
  - Reasons against predicting word boundaries
  - Two level DP
  - One stage DP, Search strategies, stack decoder
- Optimization: How to waste not too much Computation Time
  - Tree-Search, Pruning, Pruning with Beamsearch
- Search with LM / Grammar
- Multi-Pass Searches, Problems and Examples
- Producing more than one Hypothesis, Problems
- Speeding up the Search
- Search with Context-Dependent Models

# Adaptation and Discriminative Training

---

- Motivation of Adaptation Techniques
  - Variations in Speech
  - Estimating a Transformation
  - Different Types of Adaptation
  - Adaptation Methods
  - Vocal Tract Length Normalization
  - Normalization and Run-on Recognition
  - Maximum A Posteriori Adaptation (MAP)
  - Maximum Likelihood Linear Regression (MLLR)
  - Which Adaptation to Use under which Circumstances
- Discriminative Training
  - Motivation
  - MMIE vs. MPE

# Exam: Topics You Should Revise

---

- Introduction to Speech Recognition
- Spoken Language: Speech Production & Perception
- **Digital Signal Processing**
- Classification
- **Dynamic Programming and Single Word Recognizers**
- **Hidden Markov Models and their Recognizers**
- **Acoustic Modeling and Context Dependencies**
- **Language Modeling**
- **Search for Speech Recognition**
- **Adaptation, Discriminative Training**
- Natural Language Processing and Understanding
- Spoken Language Translation