CS 441 – Automatic Speech Recognition
Course Syllabus and Calendar – Spring 2006

Professor Don Colton

Brigham Young University Hawaii

1 Overview

Automatic Speech Recognition (ASR) is a computerized process. It receives as its input a digitized speech recording. It produces as its output a transcription. In the computing world, ASR is an unsolved problem, which means that nobody has yet demonstrated the ability to do it as well as humans can.

This course introduces speech recognition. Various topics in speech recognition are explored, including: phonemes and phonetic transcription; digitization of speech together with sampling frequency (samples per second) and quantization error (bits per sample); Fourier transformation of the raw signal into the frequency domain; formants; speech recognition in the CSLU rapid application developer (RAD); and psychophysiology including critical bands, loudness, frequency response, masking.

The goal of the course is to give students a substantial introduction to a computer-related subject area where research is ongoing. Students may discover an interest in ASR that will carry them into a PhD program, or may simply discover an interest in research in general with some other particular focus.

Grading is largely based on participation. Assignments are designed to introduce material, and generally not for mastery of material. For example, students may be assigned to write a Haiku poem using as many different phonemes as possible. Students may be assigned to develop a simple speech-based application that conducts a dialog between the student and a computer, such as ordering pizza. Students with programming skill may be asked to write programs that manipulate and recognize speech, such as converting a phonetic transcription of numeric speech (such as a telephone number, a credit card number, or an address) into the underlying numbers.

Early in the semester, the instructor will introduce concepts. Later in the semester, students may build projects that will cause greater learning to take place.

This course may be appropriate for two types of student. First, it is intended for Computer Science students who have already achieved some skill in programming, generally CS 301 (Algorithms), and who want to explore the rich task domain of speech input problems, including spoken language translation. Second, it may be appropriate for Linguistics students who have already achieved some skill using phonetic transcription, and who want to explore the current realities and future promise of linguistic computerization as a labor-saving device. The goal of the course is to take these two types of students and any others interested in ASR and to move them forward in understanding and ability, bringing them closer to their own personal goals.

Fundamentally this course is experimental. It exists because the instructor, Brother Colton, did his PhD work in Computer Science and Engineering, in the area of spoken language systems, and this course gives him the chance to share his interests and excitement with students. The content will change somewhat from year to year as the instructor discovers things that work well, tries out things that seem promising, and discards things that did not work as well as planned. This means two things to you the student: the course will probably never be the same twice in a row, and if you have particular interests you would like to see addressed, speak up. It is possible that the calendar can be rearranged to fit in your area of interest. Some topics are best addressed in response to a question. Some topics that are incredibly interesting will never come up unless someone asks. There is simply too much material available in the field of spoken language systems to attend to all of it this semester. So please feel free to suggest topics that are of interest to you.

1.1 The Course

This course is a work in progress. The subject of this course is my research specialty from my PhD program. Unfortunately, I know a lot more than we can possibly cover in a two-credit course, and I am still experimenting with the exact mix of topics that works best.

Also, unfortunately I have not met the textbook in this field that I am satisfied with. Therefore I will provide handouts and directions as needed. Many things are linked to my course website. I am still experimenting with the exact mix of topics that works best.

Here is what I am reasonably sure I can say:

- **Course Number:** CS 441
- **Title:** Automatic Speech Recognition
- **Course Description:** Introduction to automatic speech recognition by computers, including digital sampling, Fourier transformation, phonemic classification by neural network, and Viterbi search. (Prerequisite: CS 440 or Linguistics background.)
- **Textbook:** none at this time.
- Class Time: M-T-Th-F 8:00–8:50 AM
  Final Exam: Thu 17 Jun, 8:00–8:50 AM
  Classroom: GCB 153

1.2 The Instructor

- **Instructor (me):** Don Colton
- **My email:** don@colton.byuh.edu
- **My Office:** GCB 130 B
- **Office Hours:** Daily 11:00 AM to 11:50 AM.

2 Subject to Change

It is possible that I will revise some aspects of the course as we go along. Any changes I make are likely to be to your advantage. If any of my changes seems unfair to you, let me know. I will try to correct it.

3 Course Calendar

I intend to address the following topics in about this order.

Although the paragraphs state "will learn" and "will do" there is probably too much on the plate, and will will do some subset of the following list of items.

3.1 Wave Files

Students will learn manner in which sounds are recorded digitally for use on computers. They will explain quantization. Students will explain why so many sound formats exist and how they are used. Students will identify the most popular formats for recording sound on computers, and will explain their advantages and disadvantages.

3.2 Phonetics

Students will become familiar with the international phonetic alphabet (IPA). They will use it to transcribe recorded spoken utterances into IPA. They will translate written English into IPA. They will translate IPA into written English.

We will introduce WorldBet, an IPA-compatible method for transcribing utterances using common computer characters (ASCII).

3.3 Spectrograms

Students will learn how to read a spectrogram and will demonstrate their ability by reading spectrograms in homework and on tests.

3.4 Neural Networks

Students will be introduced to neural networks as a way of estimating the likelihood that a given frame of speech comes from an attempt to articulate a specified phoneme. Students will explain how a neural network is trained and will discuss issues affecting recognition accuracy.

3.5 Project

Students will complete a project. CS students will write programs to translate text into IPA. Advanced students may write programs that translate IPA into text.

3.6 Ideas

Here are some ideas I am not sure how to use just yet.

* `(s|sh)(cons)(r|l|w)(vowel)(cons)` = syllable

4 Special Needs

Brigham Young University Hawaii is committed to providing a working and learning atmosphere, which reasonably accommodates qualified persons with disabilities. If you have any disability that may impair your ability to complete this course successfully, please contact the students with Special Need Coordinator, Leilani A'una at 293-3518. Reasonable academic accommodations are reviewed for all students who have qualified documented disabilities. If you need assistance or if you feel you have been unlawfully discriminated against on the basis of disability, you may seek resolution through established grievance policy and procedures. You should contact the Human Resource Services at 780-8875.

5 Preventing Sexual Harassment

Title IX of the education amendments of 1972 prohibits sex discrimination against any participant in an educational program or activity that receives federal funds, including Federal loans and grants. Title IX also covers student-to-student sexual harassment. If you encounter unlawful sexual harassment or gender-based discrimination, please contact the Human Resource Services at 780-8875 (24 hours).