ELEN E6820:

Speech and Audio Processing and Recognition

Columbia University Dept. of Electrical Engineering
Spring 2002

Professor: Dan Ellis  <dpwe@ee.columbia.edu>

Web site:
http://www.ee.columbia.edu/~dpwe/e6820/
General information

• **Overview:**
  - fundamentals of acoustics & hearing
  - digital audio processing
  - speech recognition

• **Audience:**
  - engineers seeking a foundation for work in speech and audio

• **Course structure:**
  - weekly assignments (25%)
  - midterm exam (25%)
  - final project (50%)

• **Text:**
  *Speech and Audio Signal Processing*
Course outline

• **Fundamentals (4 weeks)**
  - DSP
  - Acoustics
  - Pattern recognition
  - Auditory perception

• **Audio processing (4 weeks)**
  - source and signal modeling
  - music analysis and synthesis
  - psychoacoustic-based audio compression

• **Speech recognition (4 weeks)**
  - acoustic feature vectors
  - modeling temporal structure
  - structure & evaluation of speech recognizers
Assignments & Projects

- **Weekly assignments**
  - research papers
  - MATLAB practical
  - written questions

- **Final project**
  - practical investigation into sound processing (MATLAB?)
  - developed in conjunction with professor
  - website/presentation/report
Examples of past projects

- Detecting sound events in basketball video archive
  - classifying ‘cheers’ in sport video soundtrack

- The S-Matrix: A novel approach to music segment detection
  - finding breaks between verse, chorus etc.
Psychoacoustic-based audio compression

- Exemplified by MPEG-Audio layer 3 (‘MP3’)

→ From CD rate \((44100 \times 2 \times 16 = 1.4 \text{ Mb/s})\) to 128 kb/s or less \((\leq 1.5 \text{ bits/sample})\)
Automatic Speech Recognition (ASR)

- **Standard speech recognition structure:**
  
  ![Diagram of ASR process]

  - Feature calculation
  - Acoustic model parameters
  - Acoustic classifier
  - Language model
  - HMM decoder
  - Understanding/application...

- **‘State of the art’ word-error rates (WERs):**
  - 2% (dictation) - 30% (telephone conversations)