Digital Speech Processing

Professor Lawrence Rabiner
UCSB
Dept. of Electrical and Computer Engineering
Jan-March 2011
Course Description

This course covers the basic principles of digital speech processing:

- **Review of digital signal processing**
- **Fundamentals of speech production and perception**
- **Basic techniques for digital speech processing:**
  - short - time energy, magnitude, autocorrelation
  - short - time Fourier analysis
  - homomorphic methods
  - linear predictive methods
- **Speech estimation methods**
  - speech/non-speech detection
  - voiced/unvoiced/non-speech segmentation/classification
  - pitch detection
  - formant estimation
- **Applications of speech signal processing**
  - Speech coding
  - Speech synthesis
  - Speech recognition/natural language processing

A MATLAB-based **term project** will be required for all students taking this course for credit.
Course Information


- **Grading:**
  - Homework 20%
  - Term Project 20%
  - Mid-Term Exam 20%
  - Final Exam 40%

- **Prerequisites:** Basic Digital Signal Processing, good knowledge of MATLAB

- **Time and Location:** Tuesday, Thursday, 10:30 am to 11:50 am, Phelps 1437.

- **Course Website:** [www.ece.ucsb.edu/Faculty/Rabiner/ece259](http://www.ece.ucsb.edu/Faculty/Rabiner/ece259)

- **Office Hours:** Tuesday, 1:00-3:00 pm
Web Page for Speech Course

Click on 
Digital Speech Processing Course 
on left-side panel
Digital Speech Processing Course (Winter 2011)

No Cheating Policy:

No cheating declaration: (No Cheating Policy.pdf)

Monograph on Digital Speech Processing: Introduction to Digital Speech processing.pdf

Lectures:

- Introductory Material: (basic course material_winter_2011.pdf), 6-to-a-page: (basic course material_winter_2011_6tp.pdf)
- Lecture 1: Introduction to Digital Speech Processing: (Lecture 1 Fall 2010.pdf), 6-to-a-page: (Lecture 1 Fall 2010 6tp.pdf)
- Lecture 2: Review of DSP Fundamentals: (Lecture 2 Fall 2010.pdf), 6-to-a-page: (Lecture 2 Fall 2010 6tp.pdf)
- Lecture 3: Acoustic Theory of Speech Production: (Lecture 3 Fall 2010.pdf), 6-to-a-page: (Lecture 3 Fall 2010 6tp.pdf)
- Lecture 4: Speech Perception—Auditory Models, Sound Perception Models, MOS Methods: (Lecture 4 Fall 2010.pdf), 6-to-a-page: (Lecture 4 Fall 2010 6tp.pdf)
- Lectures 5-6: Sound Propagation in the Vocal Tract: (Lectures 5-6 Fall 2010.pdf), 6-to-a-page: (Lectures 5-6 Fall 2010 6tp.pdf)
- Lectures 7-8: Time Domain Methods in Speech Processing: (Lectures 7-8 Fall 2010.pdf), 6-to-a-page: (Lectures 7-8 Fall 2010 6tp.pdf)
- Lecture 10: Short Time Fourier Analysis Methods—Filter Bank Summation and Overlap Add: (Lecture 10 Fall 2010.pdf), 6-to-a-page: (Lecture 10 Fall 2010 6tp.pdf)
- Lecture 11: Speech Representations Based on STFT Analysis-Synthesis Methods: (Lecture 11 Fall 2010.pdf), 6-to-a-page: (Lecture 11 Fall 2010 6tp.pdf)
- Lecture 12: Homomorphic Speech Processing: (Lecture 12 Fall 2010.pdf), 6-to-a-page: (Lecture 12 Fall 2010 6tp.pdf)
- Lecture 13: Linear Predictive Coding (LPC) Methods: (Lecture 13 Fall 2010.pdf), 6-to-a-page: (Lecture 13 Fall 2010 6tp.pdf)
Web Page for Speech Course

Course lecture slides (6-to-page)
Web Page for Speech Course

Download homework assignments, speech files
Web Page for Speech Course

Download MATLAB (.m) files; Examine Project Suggestions

```
test_16k.wav; (test_16k.wav)
ah.wav; (ah.wav)
should.wav; (should.wav)
s1.wav; (s1.wav); pitch period contour for s1.wav; (pp1.mat)
s2.wav; (s2.wav); pitch period contour for s2.wav; (pp2.mat)
s3.wav; (s3.wav); pitch period contour for s3.wav; (pp3.mat)
s4.wav; (s4.wav); pitch period contour for s4.wav; (pp4.mat)
s5.wav; (s5.wav); pitch period contour for s5.wav; (pp5.mat)
s6.wav; (s6.wav); pitch period contour for s6.wav; (pp6.mat)
we were; (we were away a year ago.wav)
isolated digit training files; (digits_train.zip)
isolated digit testing files; (digits_test.zip)
isolated digit training files (raw-no endpoints marked); (digits_train_raw.zip)
isolated digit testing files (raw-no endpoints marked); (digits_test_raw.zip)
```

Matlab Files:

- loadwav.m: (loadwav.m)
- savewav.m: (savewav.m)
- loadraw.m: (loadraw.m)
- saveraw.m: (saveraw.m)
- grayscale.m: (grayscale.m)
- fxquant.m: (fxquant.m)
- psoct.m: (psoct.m)
- spectrom.m: (spectrom.m)
- LPC solutions: (cholesky_full.m), (durbin.m), (lattice.m)

Project Suggestions:

- General Project Suggestions: (Digital Speech Processing Projects.pdf)
- LPC Vocoder Project Details: (LPC Vocoder Project.pdf)
- Project Schedule (UCSB-2009):
- User Interface Example (Sound Spectrograms): (GUI_plot_spectrogram_ucsb.m), (select_dir.m)

Oded: Material: (oded_material.zip)
Course Readings

**Required Course Textbook:**

**Recommended Supplementary Textbook:**

**Matlab Exercises:**
Recommended References

References in Selected Areas of Speech Processing

Speech Coding:

References in Selected Areas of Speech Processing

Speech Synthesis:

• Y. Sagisaka, N. Campbell, and N. Higuchi, *Computing Prosody*, Springer Verlag, 1996
References in Selected Areas of Speech Processing

Speech Recognition:
• X. Huang, A. Acero and H-W Hon, *Spoken Language Processing*, Prentice Hall Inc, 2000
References in Digital Signal Processing


## The Speech Stack

<table>
<thead>
<tr>
<th>Speech Applications</th>
<th>Coding, synthesis, recognition, understanding, verification, language translation, speed-up/slow-down</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech Algorithms</td>
<td>Speech-silence (background), voiced-unvoiced, pitch detection, formant estimation</td>
</tr>
<tr>
<td>Speech Representations</td>
<td>Temporal, spectral, homomorphic, LPC</td>
</tr>
<tr>
<td>Fundamentals</td>
<td>Acoustics, linguistics, pragmatics, speech production/perception</td>
</tr>
</tbody>
</table>
Digital Speech Processing

Need to understand speech processing at all three levels

Practice

Concepts

Theory

Ability to implement theory and concepts in working code (MATLAB, C, C++)

Basic understanding of how theory is applied

Mathematics, derivations, signal processing
Course Outline – ECE 259A – Speech Processing

• Jan 4 - Lecture 1, Introduction to Digital Speech Processing
• Jan 6 - Lecture 2a, Review of DSP Fundamentals
• Jan 11 - Lecture 2b, Review of DSP Fundamentals
• Jan 13 - Lecture 3a, Acoustic Theory of Speech Production
• Jan 18 - Lecture 3b, Lecture 4, Speech Perception—Auditory Models, Sound Perception, MOS Methods
• Jan 20 - Lecture 5, Sound Propagation in the Vocal Tract—Fundamentals, Solutions of the Wave Equation
• Jan 25 - Lecture 6, Sound Propagation in the Vocal Tract—Lossless Tube Models, Digital Filters
• Jan 27 - Lecture 7, Time Domain Methods—Short - Time Energy, Magnitude, Zero Crossings, Autocorrelation
• Feb 1 - Lecture 8, Time Domain Methods—Short - Time Energy, Magnitude, Zero Crossings, Autocorrelation
• Feb 3 - Lecture 9, STFT Methods—Introduction, FBS, OLA, Modifications
• Feb 8 - Lecture 10-11, STFT Methods—Speech Representations Using Analysis-Synthesis Methods
• Feb 10 - Mid - Term Exam
• Feb 15 - Lecture 12a, Homomorphic Speech Processing—Analysis, Synthesis Methods
• Feb 17 - Lecture 12b, Homomorphic Speech Processing—Practical Implementations
• Feb 22 - Lecture 13, Linear Predictive Coding (LPC)—Introduction, Autocorrelation Method, Covariance Method
• Feb 24 - Lecture 14, LPC—Lattice Implementation, Frequency Domain Interpretations
• Mar 1 - Lecture_Algorithms—Speech Detection, V/U/S Classification, Pitch/Formant Estimation Algorithms
• Mar 3 - Lecture 15, Speech Waveform Coding—Uniform and Non-Uniform Quantization
• Mar 8 - Lecture 16, Speech Waveform Coding—Adaptive and Differential Quantization
• Mar 10 - Term Project Presentations (10-12 am)
• Mar 16 - Final Exam (8 am-11 am)
Other Potential Topics for Discussion/Term Projects

- Sinusoidal modeling of speech
- Speech modification and enhancement—slowing down and speeding up speech, noise reduction methods
- Speaker verification methods
- Music coding including MP3 and AAC standards-based methods
- Pitch detection methods
Term Project

• All registered students are required to do a term project. This term project, implemented *using Matlab*, must be a speech or audio processing system that accomplishes a simple or even a complex task—e.g., pitch detection, voiced-unvoiced detection, speech/silence classification, speech synthesis, speech recognition, speaker recognition, helium speech restoration, speech coding, MP3 audio coding, etc.

• Every student is also required to make a 10-minute Power Point *presentation* of their term project to the entire class. The presentation must include:
  – A short description of the project and its objectives
  – An explanation of the implemented algorithm and relevant theory
  – A demonstration of the **working** program – i.e., results obtained when running the program
Suggestions for Term Projects

1. Pitch detector – time domain, autocorrelation, cepstrum, LPC, etc.
2. Voiced/Unvoiced/Silence detector
3. Formant analyzer/tracker
4. Speech coders including ADPCM, LDM, CELP, Multipulse, etc.
5. N-channel spectral analyzer and synthesizer – phase vocoder, channel vocoder, homomorphic vocoder
6. Speech endpoint detector
7. Simple speech recognizer – e.g. isolated digits, speaker trained
8. Speech synthesizer – serial, parallel, direct, lattice
9. Helium speech restoration system
10. Audio/music coder
11. System to speed up and slow down speech by arbitrary factors
12. Speaker verification system
13. Sinusoidal speech coder
14. Speaker recognition system
15. Speech understanding system
16. Speech enhancement system (noise reduction, post filtering, spectral flattening)
The requirements for this project are a short description of the problem containing relevant mathematical theory and objectives of the project, a listing (with sufficient documentation and comments) of the program, and a demonstration that the program works properly.