LECTURE 14: EXAM NO. 1

The first exam can be found here.

Return to Main

ECE 8463: FUNDAMENTALS OF SPEECH RECOGNITION

Professor Joseph Picone
Department of Electrical and Computer Engineering
Mississippi State University

email: picone@isip.msstate.edu
phone/fax: 601-325-3149; office: 413 Simrall
URL: http://www.isip.msstate.edu/resources/courses/ece_8463

Modern speech understanding systems merge interdisciplinary technologies from Signal Processing, Pattern Recognition, Natural Language, and Linguistics into a unified statistical framework. These systems, which have applications in a wide range of signal processing problems, represent a revolution in Digital Signal Processing (DSP). Once a field dominated by vector-oriented processors and linear algebra-based mathematics, the current generation of DSP-based systems rely on sophisticated statistical models implemented using a complex software paradigm. Such systems are now capable of understanding continuous speech input for vocabularies of hundreds of thousands of words in operational environments.

In this course, we will explore the core components of modern statistically-based speech recognition systems. We will view speech recognition problem in terms of three tasks: signal modeling, network searching, and language understanding. We will conclude our discussion with an overview of state-of-the-art systems, and a review of available resources to support further research and technology development.

Tar files containing a compilation of all the notes are available. However, these files are large and will require a substantial amount of time to download. A tar file of the html version of the notes is available here. These were generated using wget:

```
wget -np -k -m http://www.isip.msstate.edu/publications/courses/ece_8463/lectures/current
```

A pdf file containing the entire set of lecture notes is available here. These were generated using Adobe Acrobat.

Questions or comments about the material presented here can be directed to help@isip.msstate.edu.

LECTURE 14: EXAM NO. 1

The first exam can be found here.

Index of /publications/courses/ece_8463/exams

Name	Last modified	Size	Description
Parent Directory	09-Feb-2002 08:26	-	
2002_spring/	08-Feb-2002 15:53	-	
current/	08-Feb-2002 15:53	-	

Index of /publications/courses/ece_8463/exams/current

Name	Last modified	Size	Description
Parent Directory	08-Feb-2002 16:03	-	
sp02_01.fm	08-Feb-2002 15:53	23k	
sp02_01.pdf	08-Feb-2002 15:53	26k	

Index of /publications/courses/ece_8463/lectures/current/lecture_14/exams

Name	Last modified	Size	Description
Parent Directory	08-Feb-2002 16:03	-	
sp02_01.fm	08-Feb-2002 15:53	23k	
sp02_01.pdf	08-Feb-2002 15:53	26k	

Number:

Problem	Points	Score
1 (a)	10	
1 (b)	10	
1 (c)	10	
1 (d)	10	
2 (a)	10	
2 (b)	10	
2 (c)	10	
3 (a)	10	
3 (b)	10	
3 (c)	10	
Total	100	

Notes:

1. The exam is closed books and notes. You are allowed one 8 1/2” x 11” double-sided sheet of notes.

2. Please indicate clearly your answer to the problem by some form of highlighting (underlining).

3. Your solutions must be legible and easy to follow. If I can’t read it or understand it, it is wrong. Random scribbling will not receive credit.

4. Please show ALL work. Answers with no supporting explanations or work will be given no credit.

5. Several problems on this exam are fairly open-ended. Since the evaluation of your answers is obviously a subjective process, we will use a market place strategy in determining the grade. Papers will be rank-ordered in terms of the quality of the solutions, and grades distributed accordingly.

1. Deep sea divers breath a mixture of air and helium called Heliox to avoid several problems associated with breathing compressed air under water. Heliox is lighter than air — its density is 75% lower than air. Speech produced while breathing Heliox sounds distorted (a classroom demonstration is provided).

(a) Predict the effect breathing Heliox has on the formant frequencies. Justify this answer using our linear acoustics model.

(b) Does the excitation signal change? Explain.

(c) For what value of the density (relative to air) would the speech become unintelligible to a human listener?

(d) Design a system that would descramble the diver's speech and produce a normal sounding speech signal. Is such a system physically realizable?

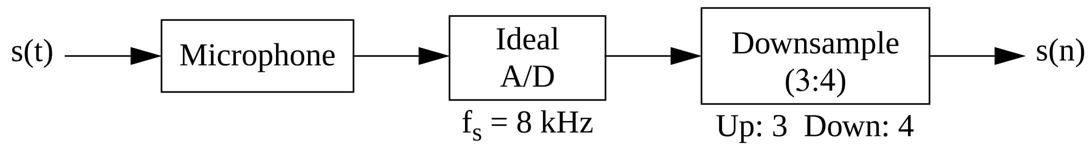
2. Consider a language which has a phoneme set that only contains four English consonants: *b, p, d, t*.

(a) Describe the similarities and differences between these sounds in as much linguistic detail as possible.

(b) Do the assumptions we make to justify frame-based processing in speech recognition of spoken English hold for this type of signal? Explain.

(c) Consider a voiced sound in this language produced with a fundamental frequency of 100 Hz. Would a listener perceive the 5th and 6th harmonics to be closer in frequency than the 20th and 21st harmonics? Explain.

3. Consider the system shown below:



(a) Suppose the microphone can be modeled by this equation: $y(t) = \alpha x(t) + \beta x^2(t)$. Typically, $\alpha > \beta$. Sketch the spectrum of the output signal, s(n), over the frequency range [0, 8 kHz] for an input that consists of a sinewave at 1 kHz.

(b) Sketch the spectrum of the output signal for a white noise input (flat spectrum). Assume both the A/D and the downsampler use ideal low pass filters.

(c) Suppose the input signal is the sum of a 1 kHz and 1.5 kHz sinewave. Sketch the spectrum of the output signal. Explain the influence of the microphone on this result. What aspect of the microphone would you improve? Why?