L1: Course introduction

Course introduction

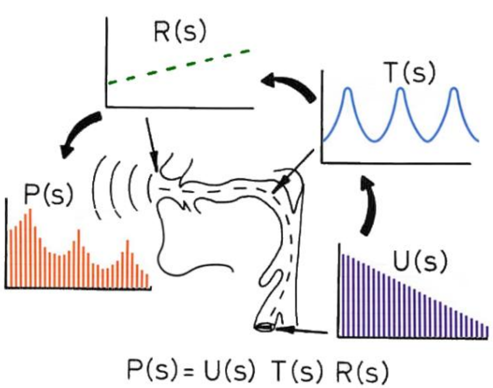
Course logistics

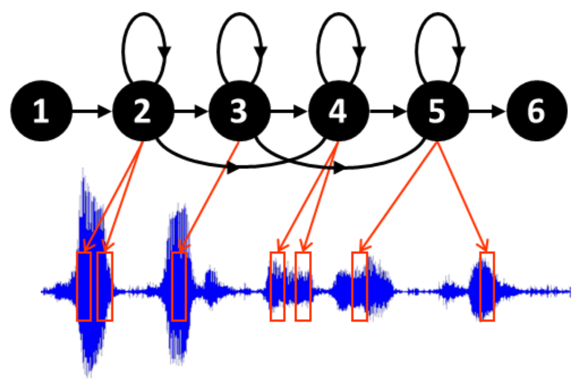
Course contents

Course introduction

What is speech processing?

- The study of speech signals and their processing methods
- Speech processing encompasses a number of related areas
  - **Speech recognition**: extracting the linguistic content of the speech signal
  - **Speaker recognition**: recognizing the identity of speakers by their voice
  - **Speech coding**: compression of speech signals for telecommunication
  - **Speech synthesis**: computer-generated speech (e.g., from text)
  - **Speech enhancement**: improving intelligibility or perceptual quality of speech signals





The music carried on until
ðə mju:zɪk kær[i,ɪ]d ɒn ʌntɪl

after midnight and then the
ɑ:ftə mɪdnaɪt[|,]ən[d] ðen[|,]ðə

drummers became tired and
drʌməz b[ɪ,ə]keɪm taɪəd[|,]ən[d]

the dancers became cold.
ðə dɑ:nsəz b[ɪ,ə]keɪm kəʊld|

Applications of speech processing

- Human computer interfaces (e.g., speech I/O, affective)
- Telecommunication (e.g., speech enhancement, translation)
- Assistive technologies (e.g., blindness/deafness, language learning)
- Audio mining (e.g., diarization, tagging)
- Security (e.g., biometrics, forensics)

Related disciplines

- Digital signal processing
- Natural language processing
- Machine learning
- Phonetics
- Human computer interaction
- Perceptual psychology

The course <u>objectives</u> are to familiarize students with

- Fundamental concepts of speech production and speech perception
- Mathematical foundations of signal processing and pattern recognition
- Computational methods for speech analysis, recognition, synthesis, and modification

As <u>outcomes</u>, students will be able to

- Manipulate, visualize, and analyze speech signals
- Perform various decompositions, codifications, and modifications of speech signals
- Build a complete speech recognition system using state of the art tools

Course logistics

Class meetings

– MWF 9:10-10:00am
– HRBB 126

Course prerequisites

– ECEN 314 or equivalent, or permission of the instructor
– Basic knowledge of signals and systems, linear algebra, and probability and statistics
– Programming experience in a high-level language is required

Textbook

– The course will not have an official textbook and instead will be based on lecture slides developed by the instructor from several sources
– Additional course materials may be found in the course website http://courses.cs.tamu.edu/rgutier/csce689_s11/

Recommended references

- J. Holmes & W. Holmes, *Speech Synthesis and Recognition*, 2nd Ed, CRC Press, 2001 (available online at TAMU libraries)
- P. Taylor, *Text-to-speech synthesis*, Cambridge University Press, 2009
- L. R. Rabiner and R. W. Schafer, *Introduction to Digital Speech Processing*, Foundations and Trends in Signal Processing 1(1–2), 2007
- B. Gold and N. Morgan, *Speech and Audio Signal Processing: Processing and perception of speech and music*, Wiley, 2000
- T. Dutoit and F. Marques, *Applied signal processing, a Matlab-based proof-of-concept*, Springer, 2009
- J. Benesty, M. M. Sondhi, and Y. Huang (Eds.), *Springer Handbook of Speech Processing*, 2008 (available online at TAMU libraries)
- X. Huang, A. Acero and H.-W. Hon, *Spoken Language Processing*, Prentice Hall, 2001

Grading

- Homework assignments
  - Three assignments, roughly every 2-3 weeks
  - Emphasis on implementation of material presented in class
  - Must be done individually
- Tests
  - Midterm and final exam
  - Closed-books, closed notes (cheat-sheet allowed)
- Project
  - Team-based, in groups of up to 3 people
  - Three types: application of existing tools, development of new tools, design of new algorithms

	Weight (%)
Homework	40
Project	30
Midterm	15
Final Exam	15

Course contents

Introduction (3 lectures)
- Course introduction
- Speech production and perception
- Organization of speech sounds

Mathematical foundations (4 lectures)
- Signals and transforms
- Digital filters
- Probability, statistics and estimation theory
- Pattern recognition principles

Speech analysis and coding (4 lectures)
- Short-time Fourier analysis and synthesis
- Linear prediction of speech
- Source estimation
- Cepstral analysis

Speech and speaker recognition (6 lectures)
- Template matching
- Hidden Markov models
- Refinements for HMMs
- Large vocabulary continuous speech recognition
- The HTK speech recognition system
- Speaker recognition

Speech synthesis and modification (4 lectures)
- Text-to-speech front-end
- Text-to-speech back-end
- Prosodic modification of speech
- Voice conversion

Tentative schedule*

Week	Date	Classroom meeting	Materials due
1	1/17	No class (MLK day)	
	1/19	Course introduction	
2	1/24	Speech production and perception	
	1/26	Organization of speech sounds	
3	1/31	Signals and transforms	HW1 assigned
	2/2	Digital filters	
4	2/7	Short-time Fourier analysis and synthesis	
	2/9	Linear prediction of speech	
5	2/14	Source estimation	
	2/16	Cepstral analysis	HW1 due
6	2/21	Probability, statistics, and estimation theory	HW2 assigned
	2/23	Pattern recognition principles	
7	2/28	Template matching	
	3/2	Hidden Markov models	
8	3/7	Review/catch-up day	HW2 due
	3/9	Midterm exam	
9	3/14	Spring Break	
	3/16	Spring Break	
10	3/21	Refinements for HMMs	HW3 assigned
	3/23	Large vocabulary continuous speech recognition	
11	3/28	HTK speech recognition system	
	3/30	Speaker recognition	
12	4/4	Speech synthesis (front-end)	
	4/6	Speech synthesis (back end)	HW3 due
13	4/11	Review/catch-up day	
	4/13	Proposal presentations	Project proposal
14	4/18	Prosodic modification of speech	
	4/20	Voice conversion	
15	4/25	Review/catch-up day	
	4/27	Final exam	
16	5/2	Prep day (no class)	
	5/4	Reading day (no class)	
17	5/9	Project presentations (8:00AM - 10:00PM)	Project report

*This timeline assumes MW meeting times