

L1: Course introduction

Course introduction

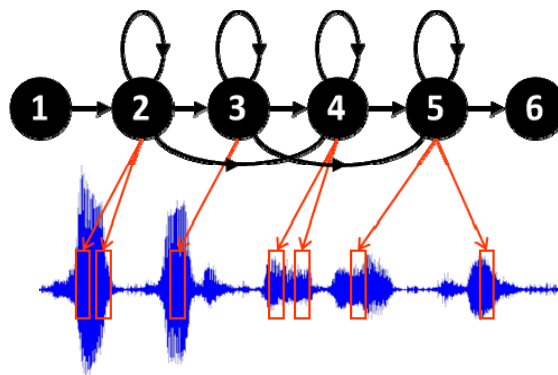
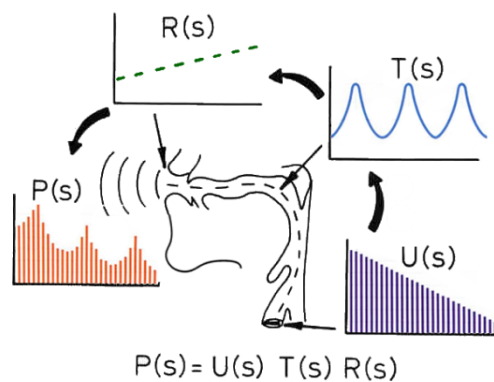
Course logistics

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[ɪ,ɪ]d ʌn 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/phonology
- Human computer interaction
- Perceptual psychology

The course objectives 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 outcomes, 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 1:50-2:40pm
- HEB 137

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/csce630_f17/

Recommended references

- B. Gold, N. Morgan and D. Ellis, *Speech and Audio Signal Processing: Processing and perception of speech and music*, 2nd Ed., Wiley, 2011
- 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
- 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 Kaldi 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	8/28	Course introduction	
	8/30	Speech production and perception	
	9/1	Organization of speech sounds	
2	9/4	Signals and transforms	
	9/6	Signals and transforms	
	9/8	Digital filters	
3	9/11	Digital filters	
	9/13	Short-time Fourier analysis and synthesis	
	9/15	Short-time Fourier analysis and synthesis	HW1 assigned
4	9/18	Linear prediction of speech	
	9/20	Linear prediction of speech	
	9/22	Source estimation	
5	9/25	Source estimation	
	9/27	Cepstral analysis	
	9/29	Cepstral analysis	HW1 due
6	10/2	Probability, statistics, and estimation theory	HW2 assigned
	10/4	Probability, statistics, and estimation theory	
	10/6	Pattern recognition principles	
7	10/9	Pattern recognition principles	
	10/11	Template matching	
	10/13	Hidden Markov models	
8	10/16	Hidden Markov models	
	10/18	Review/catch-up day	HW2 due
	10/20	Midterm exam	
9	10/23	Refinements for HMMs	
	10/25	Refinements for HMMs	
	10/27	Kaldi speech recognition system	HW3 assigned
10	10/30	Kaldi speech recognition system	
	11/1	Large vocabulary continuous speech recognition	
	11/3	Large vocabulary continuous speech recognition	
11	11/6	Speaker recognition	
	11/8	Speaker recognition	
	11/10	Speech synthesis	
12	11/13	Speech synthesis	HW3 due
	11/15	Speech synthesis	
	11/17	Speech synthesis	
13	11/20	Proposal presentations	Project proposal
	11/22	Proposal presentations	
	11/24	Thanksgiving holiday	
14	11/27	Speech modification	
	11/29	Speech modification	
	12/1	Speech modification	
15	12/4	Review/catch-up day	
	12/6	Final exam	
	12/8	No class	
16	12/12	Project presentations: 3:30–5:30 pm	Project report