Deep Learning for Speech Processing

CSCI-B 659 14328

TUESDAY, AUGUST 27, 2019

LECTURE #1

About me

- **Background:**
  - 4th year, Assistant Professor, Computer Science
  - Director of the Audio, SPeech, and Information REtreival (ASPIRE) Research Group
- **Main Research Interests:**
  - Speech/Signal processing, Machine/deep learning
  - Projects on: noise reduction, human-level speech quality assessment, privacy-preserving speech recognition
- **Teaching Experience**
  - Machine learning graduate (3x) and undergraduate (1x) levels
  - Deep learning for Speech Processing (1x prior)
  - Machine Audition and Perception (1x)
- **Industry Experience**
  - 3 years as software/systems engineer @ Lockheed Martin in Moorestown, NJ
  - Summer graduate intern @ Audience, Inc. (Knowles) in Mountain View, CA

About you?

- Name
- Degree and program (e.g. M.S. in Data science)
- Year in program
- Where are you from?
- What are your research interests?
- What do you hope to learn from this course?

Today's agenda

- Introductions
- What is speech processing?
- Why deep learning?
- Key topics, Course objectives and requirements

What is Speech Processing?

- **How would you define speech processing?**
- **Historical Definition**
  - "Manipulation of sampled speech signals … to obtain a new signal with some desired properties" by T. F. Quatieri
  - Example: Change a person's speech rate (e.g. talk faster or slower)
- **This definition does NOT cover all aspects of speech processing**

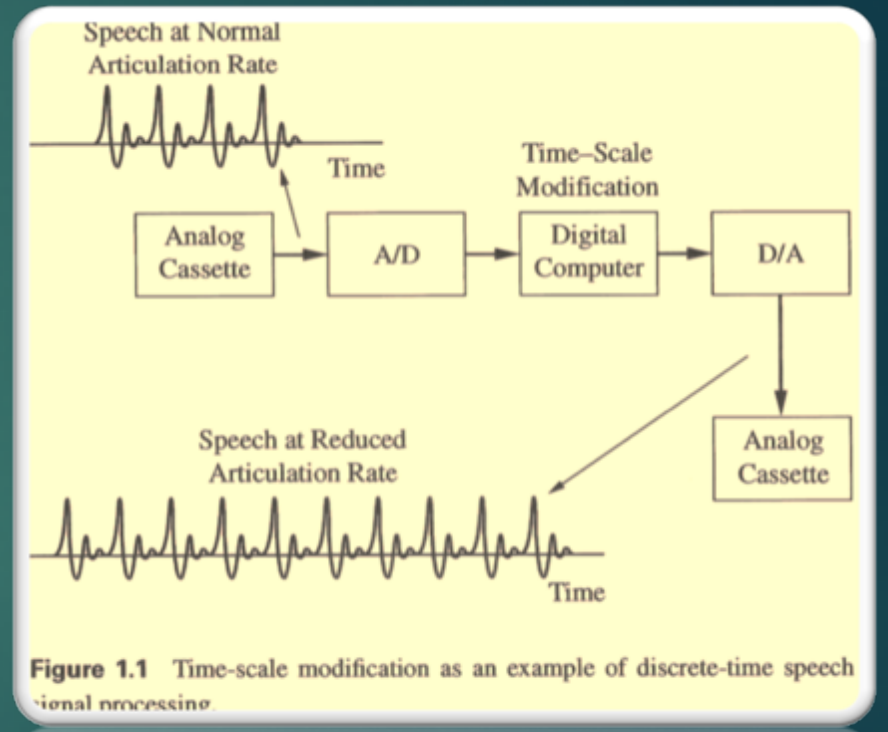
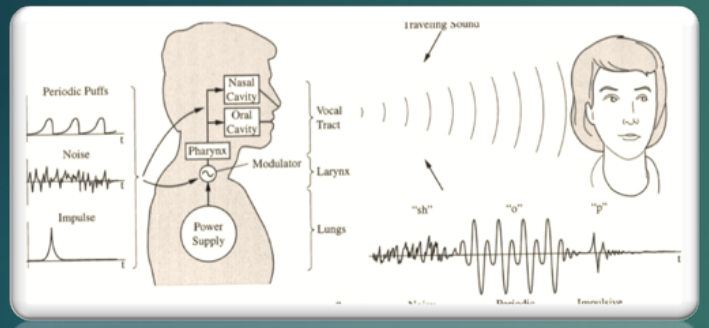


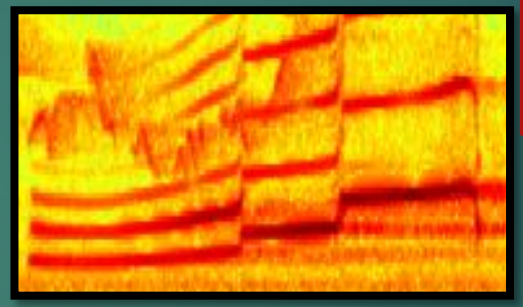
Figure 1.1 Time-scale modification as an example of discrete-time speech signal processing.

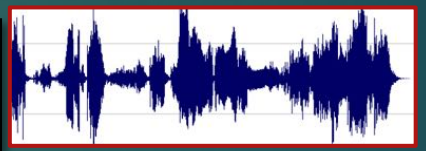
Stages of Speech Processing

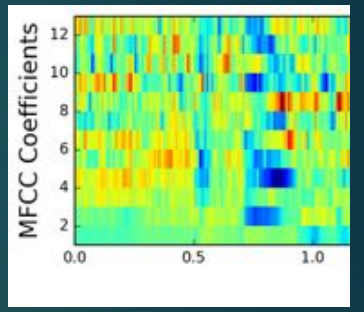
- **Pre-process stage**
  - Speech production and perception
  - Sound recording and characteristics



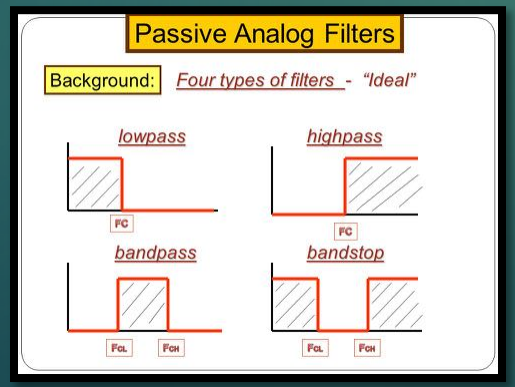
- **Feature extraction and Signal representation**

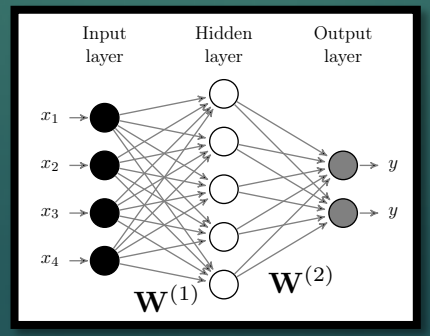


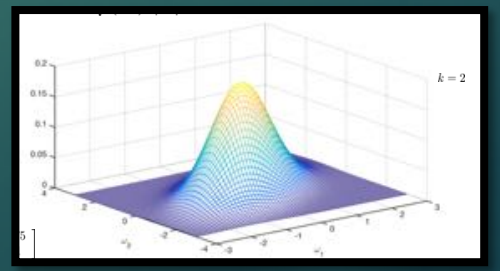




- **Processing and Data Manipulation**







Stages of Speech Processing

- **Example Applications**

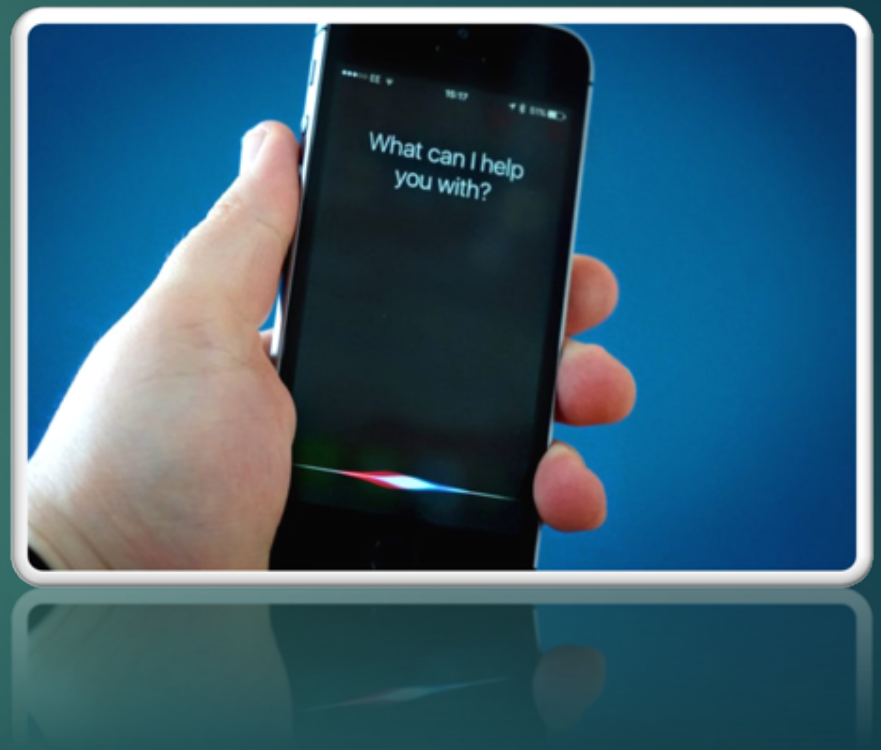
Speech enhancement

Automatic Speech Recognition

- **Performance Assessment:**
  - How well does the approach work?
  - In which environments? For which speakers?

What is needed to make Siri work?

- Demo: Siri
  - Me: “Hey Siri… What is today’s date?”
  - Siri: “It is Tuesday, August…”
- What does Siri need to do/understand, to answer this question?
  - How to capture the signal?
  - How to process and represent the signal?
  - Recognize what I’m saying?
  - Recognize who is speaking? (Demo)
  - How to reply? What I expect to hear?

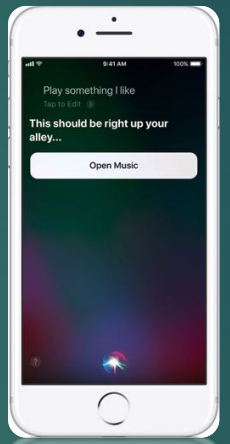


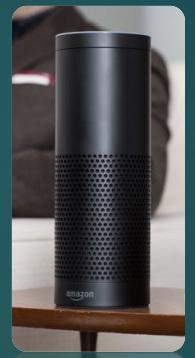
The Prevalence of Speech Processing!

Traditional

Health

Consumer





Why is this important?

- Speech is the primary form of communication between humans

skype™

- Now, speech is often (or will be) used to interface with machines

Why is this important?

- **A lot of information is contained in speech**
  - Speaker? age? gender? nationality?
  - Language? Number of speakers?
  - Location? left? right? close? Inside? Outside?
- **More R&D is needed**
  - Current devices really aren‘t that smart!
  - There are many things that they cannot do or do not do well

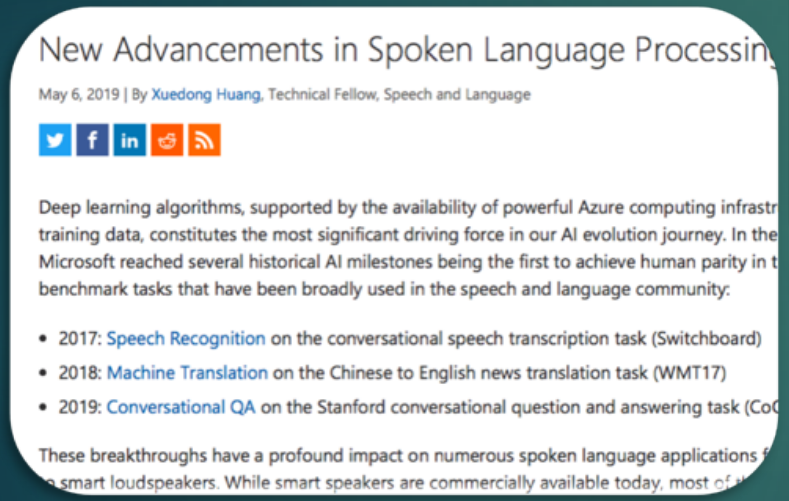
Commercial Interest

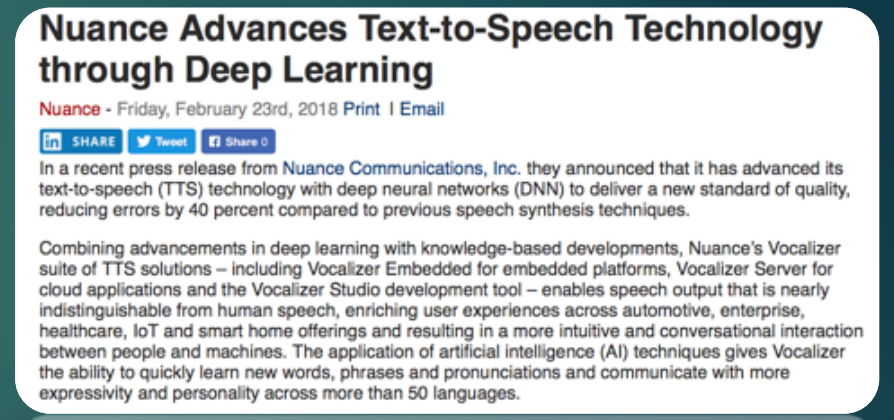
Today's agenda

- Introductions
- What is speech processing?
- Why deep learning?
- Key topics, Course objectives and requirements

Why deep learning? It Works!

- Advancements in speech applications



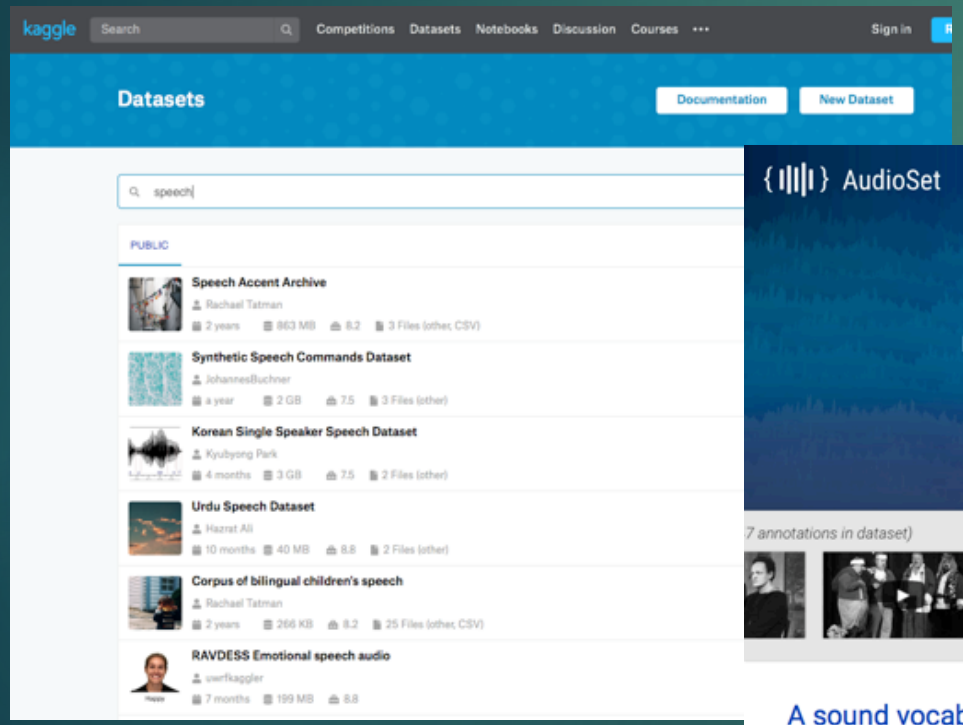


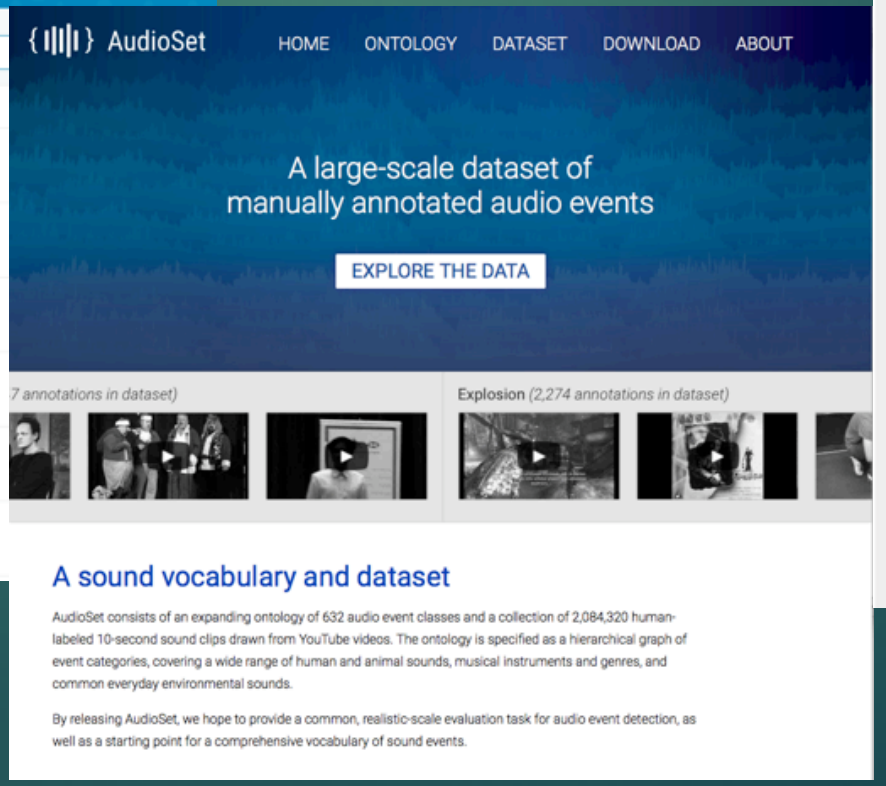
Why deep learning? It's Available

- There are many deep (machine) learning libraries

Why deep learning? It's Available

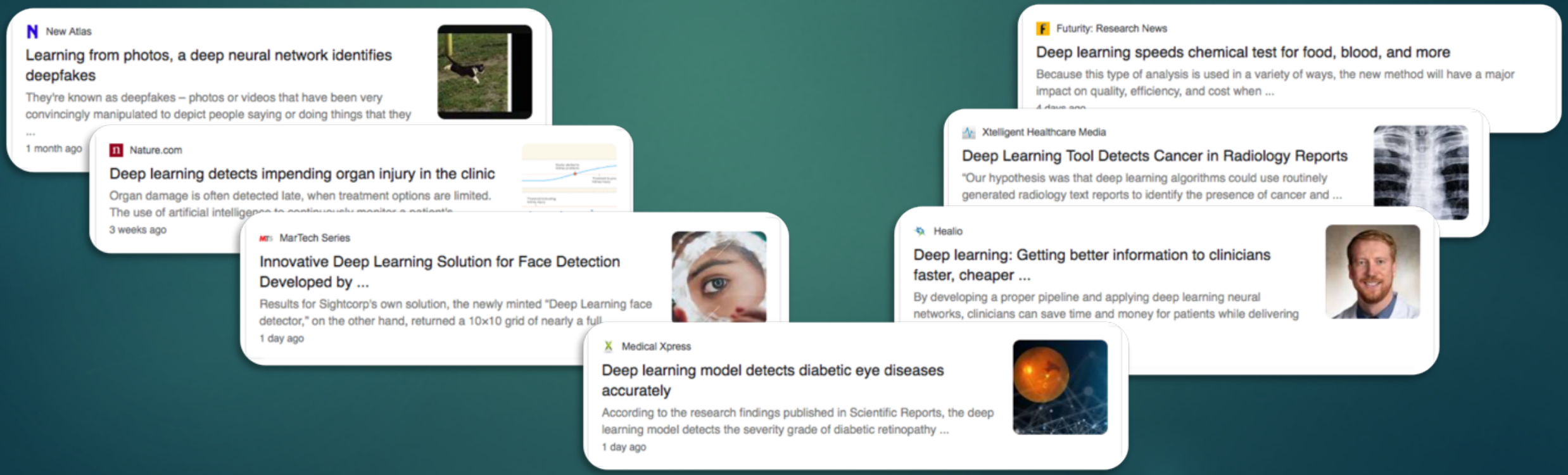
- There is an abundance of data!





Why deep learning? It's Useful

- Many problems are being successfully addressed with deep learning

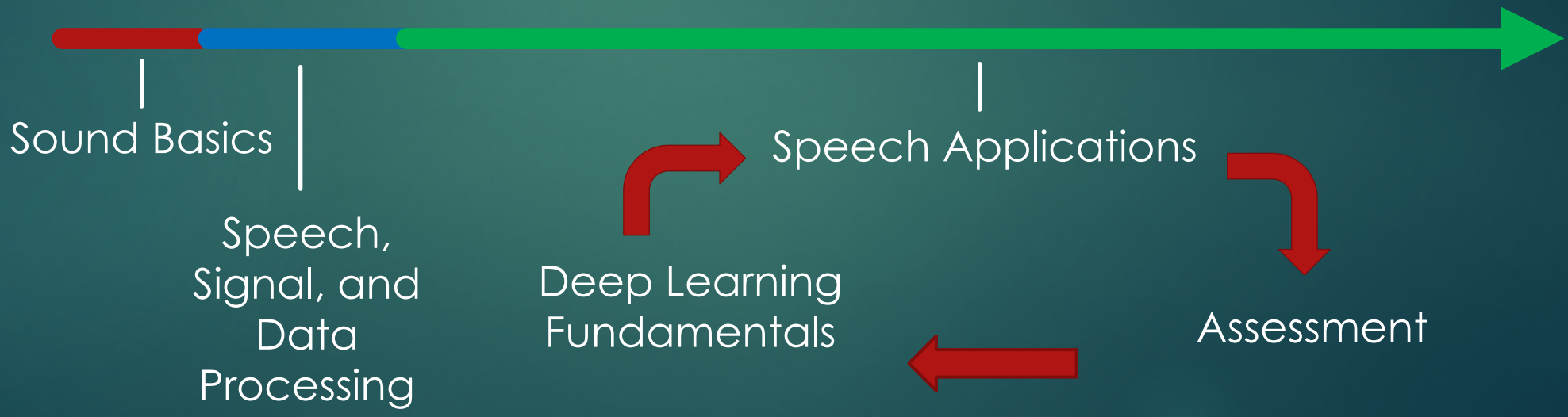


Today's agenda

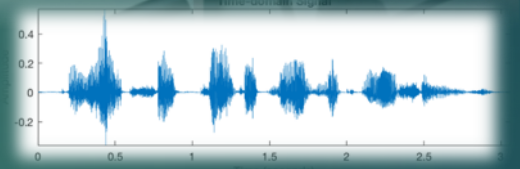
- Introductions
- What is speech processing?
- Why deep learning?
- Key topics, Course objectives and requirements

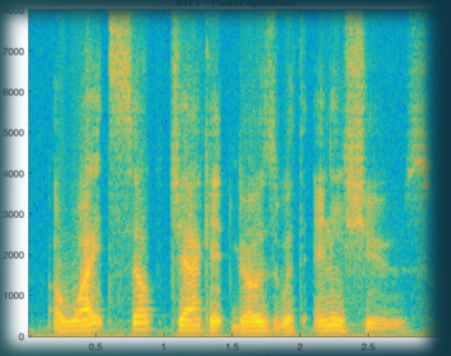
A Wholistic Viewpoint

- This class will introduce you to all stages of deep learning and speech processing research
- Programming assignments (in Python) will reinforce concepts at every stage
- Classes will toggle between lectures and practical experience

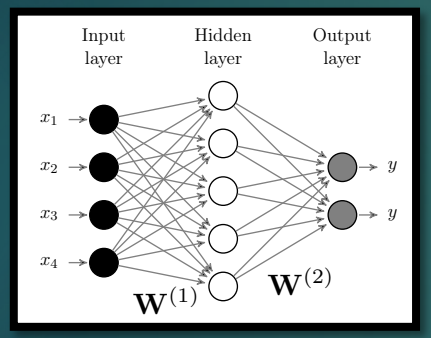


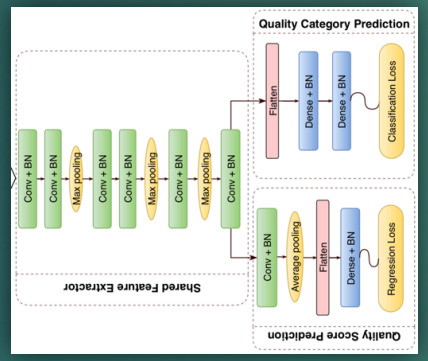
Key Topics





- **Sound and Speech Processing**
  - What is sound? How do we capture speech? How is it represented?
  - Signal power, energy, intensity; Sound level





- **Deep Learning**
  - Types of networks? DNN? RNN? CNN? LSTM?
  - How are networks optimized?
  - How much and what types of data?
  - Deep learning in Python

- **Deep Learning for Speech Processing**
  - How is deep learning used for different applications?
  - How do they perform? Assessment?
  - What work still needs to be done?

Course Description, Objectives,…

- See syllabus

My Expectations of You

- You are hard working
  - You will read course material and review class notes
  - You will spend at least 9 hours per week outside of class on this course
  - You will complete all post-class assignments
- You will participate in class (Ask questions)
- You are motivated to learn and succeed in class

Your Expectations of Me

- I will be transparent in grading
- I will be prepared for each class
- I am here to help you learn and answer your questions
- I will treat you with respect
- I am available, outside of class, to answer questions

Closing Items

- Course requirements
  - Come to class!
  - Participate!
  - Read material before class!
  - Ask questions!
- Next class:
  - Topic: Basics of sound and audio
  - Complete Homework #0 before next class
  - **Bring headphones**