Welcome to
Computer Audition
(ECE 277/477, AME 277/477, CSC 264/464, TEE 477)

Zhiyao Duan

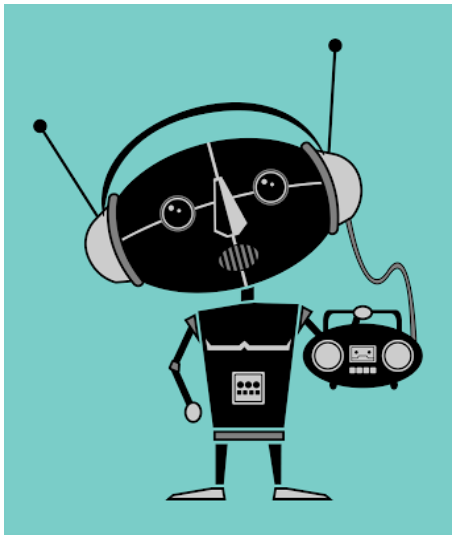
Assistant Professor of ECE and CS

University of Rochester

Human Audition

- Understanding the environment
- Communication
- Entertainment

Computer Audition



- Understanding the environment
- Communication
- Entertainment – entertain human

Some Key Problems

- Sound source identification


- Source localization
- Content understanding
  - Speech, event, melody, rhythm


- Source separation

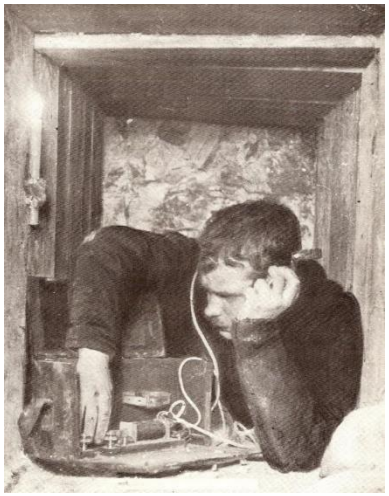
Tools for Sound Interaction

Create: Bone Flutes (7000 B.C.)

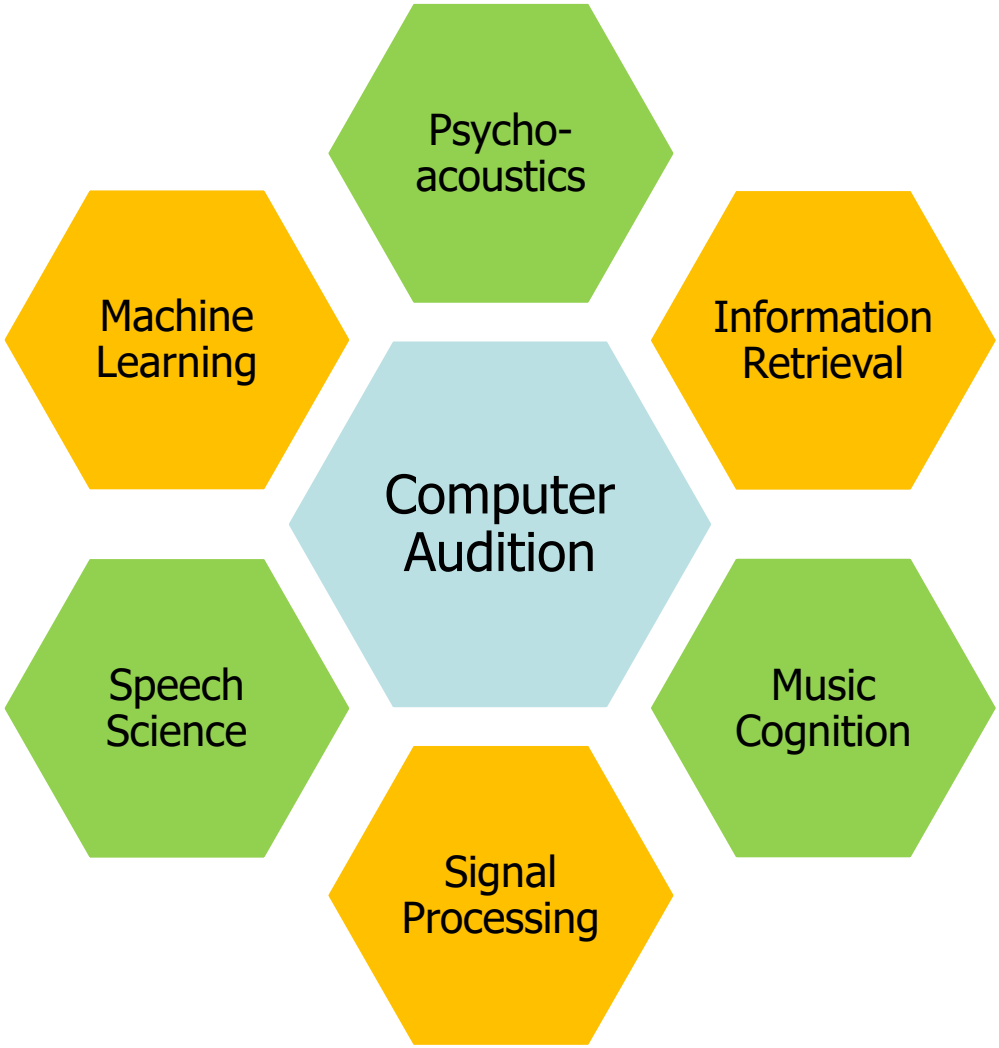
Modify: Delphi Theater (300 B.C.)

Record: Cylinder Phonograph (1899)

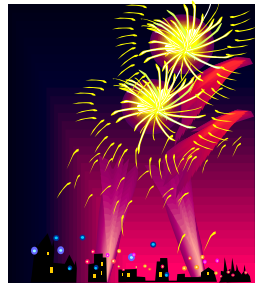
Transmit:
Crystal Radio
(1914)



Impact on Many Fields



Many Applications



Some Demos

- Automatic music accompaniment
  - http://www.music.informatics.indiana.edu/~craphael/music_plus_one/movies/movies.html
- Multimedia synchronization
  - https://www.audiolabs-erlangen.de/fau/professor/mueller/demos

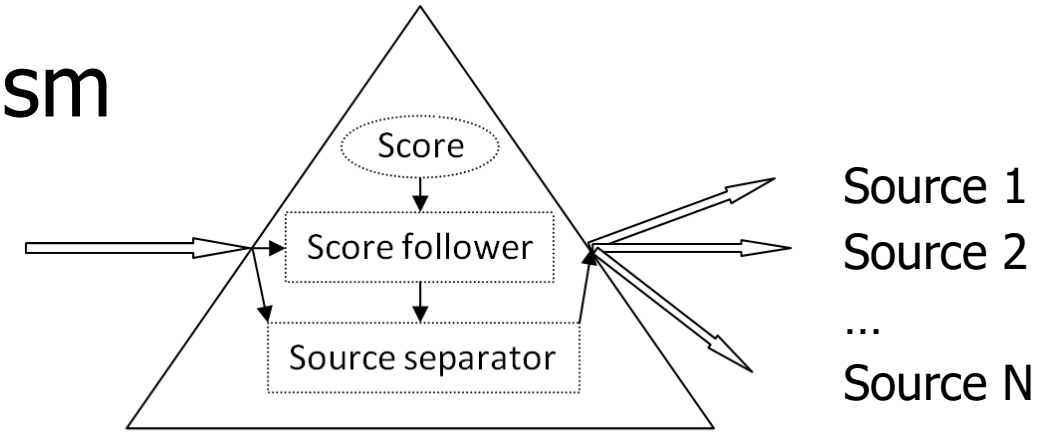
Some Demos

- Source Separation
  - Pop music separation [Takahashi, 2018]
    - https://sisec18.unmix.app/#/unmix/AM%20Contra%20-%20Heart%20Peripheral/TAU1
  - Violin/piano separation [Li, 2019]
    - Mixture:  violin:  piano: 
  - Speech/noise separation (speech enhancement) [Eskimez, 2018]
    - Mixture:  enhanced speech: 
  - Speech separation [Hershey, 2016]
    - Mixture:  female #1:  female #2: 
  - Audio-visual speech separation [Afouras, 2018]
    - http://www.robots.ox.ac.uk/~vgg/demo/theconversation/demos/vox/0/demo.html

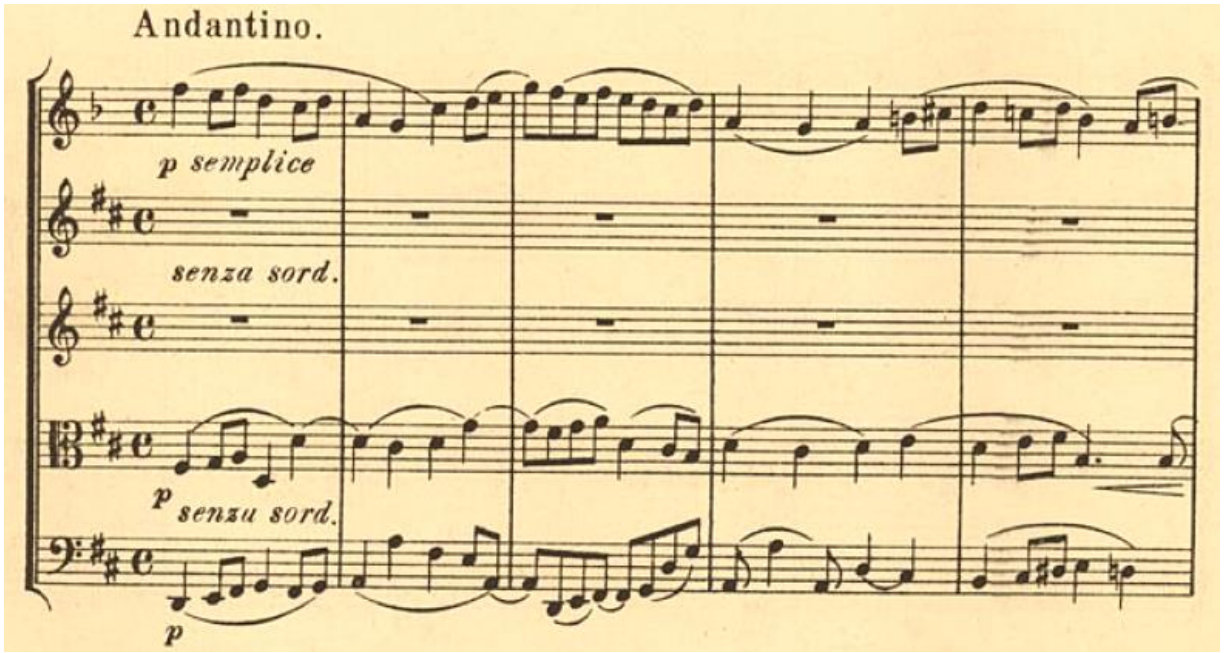
Some Demos

- Soundprism

Single-channel polyphonic music

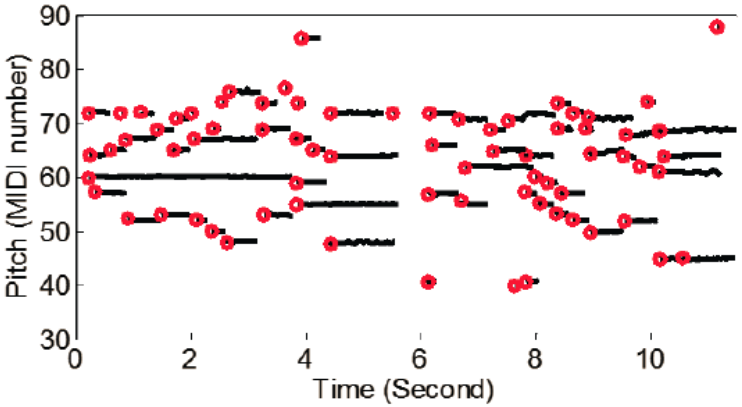


J. Brahms, Clarinet Quintet in B minor, op.115. 3rd movement

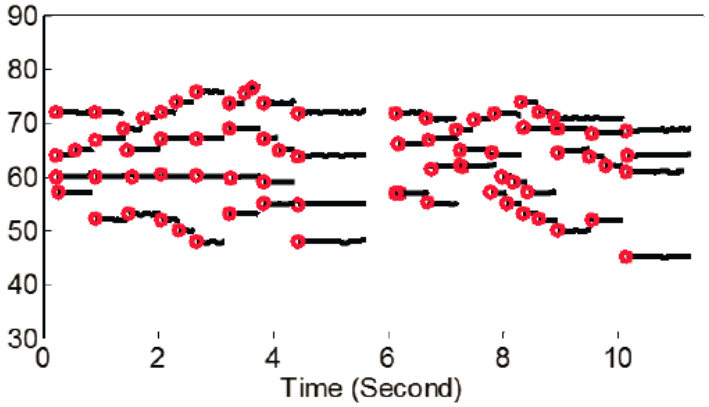


Some Demos

- Automatic music transcription
  - Multi-instrument transcription



Algorithm Transcription



Ground-truth Transcription

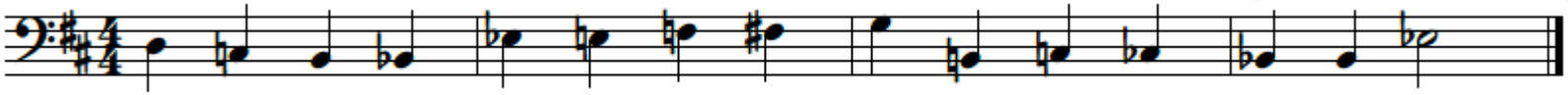
  - Context-dependent piano transcription [Cogliati, 2017]
    - Input: transcribed:
  - Deep learning based: "Onsets and Frames" [Hawthorne, 2017]
    - Input: transcribed:

Some Demos

- Acoustic event detection and localization
  - https://www.youtube.com/watch?v=iImkV6oKG_8
- Voice conversion
  - https://www.youtube.com/watch?v=RB7upq8nzIU
- Audio morphing
  - https://www.audiolabs-erlangen.de/resources/MIR/2015-ISMIR-LetItBee

Some Demos

- Automatic song writing
  - http://www.youtube.com/watch?v=3oGFogwcx-E
- Music Generation
  - https://www.youtube.com/watch?v=BfrNiqvKbLQ
- Music harmonization [Yan, 2018]



- Music generation [Yan, under review]
  - String trio:

Course Topics

- Fundamentals of human audition
- Auditory models
- Audio features (pitch, timbre, ect.)
- Audio modeling techniques
- State-of-the-art research topics
  - Polyphonic pitch analysis
  - Source separation
  - Sound identification
  - ……

Course Objectives

- General understanding of the field
- Deep understanding and hands-on research experience in a sub-field

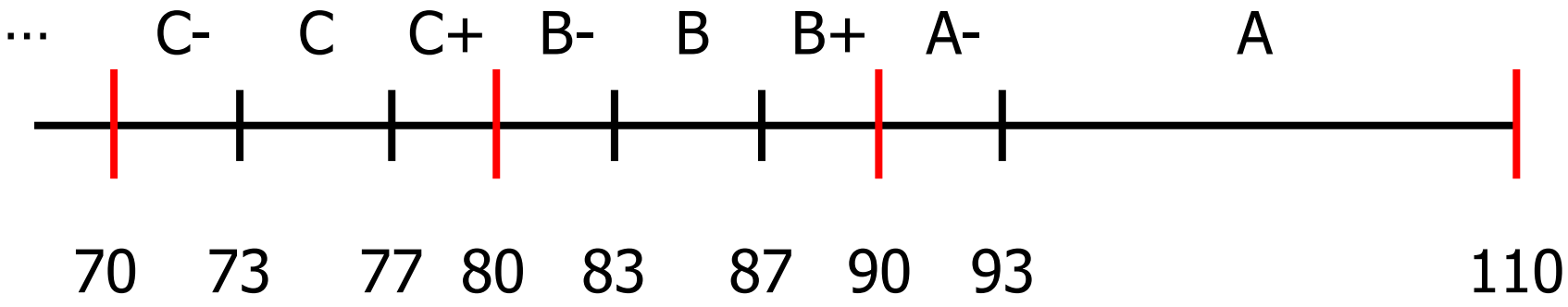
- Gain experience of the full cycle of research
- Able to think critically
- Improve presentation and writing skills

Assignments

- Total (110 points)
  - Homework (50 points)
  - Class paper review (14 points)
  - Presentation of research (10 points)
  - Course project (30 points)
  - Peer feedback (6 points)

- No exams

Grading



- No extra credit
- No curve
- 200-level students get 10 points boost

Important Policies

- Late homework penalty
  - 20% deduction each day
- Do your own work
  - Discussions are encouraged
  - No exchange of code
  - No copying of five or more consecutive words
  - Cite external sources
- Attendance is not taken, but class discussions are very important for learning

Prerequisites

- Signal Processing
  - ECE 246/446 or ECE 272/472 or equivalent
- Matlab or Python programming

- Preferred but not required
  - Machine learning such as SVM, Markov models, neural networks, clustering, etc.

Three Websites

- Course website
  - All materials (lecture notes, readings, assignments, etc.)
  - http://www.ece.rochester.edu/~zduan/teaching/ece477
- Blackboard:
  - Only for announcements and homework submissions
- Piazza
  - Only for discussions