

# Welcome to Computer Audition

(ECE 277/477, AME 277, CSC 264/464, TEE 477)

Zhiyao Duan

Associate Professor of ECE and CS

University of Rochester

# Human Audition

---

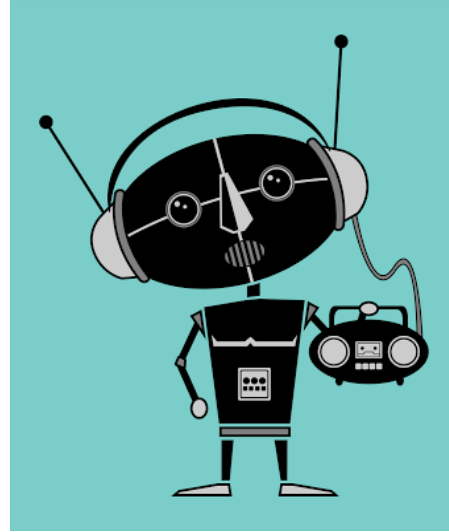


- Understanding the environment
- Communication
- Entertainment



# 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

---



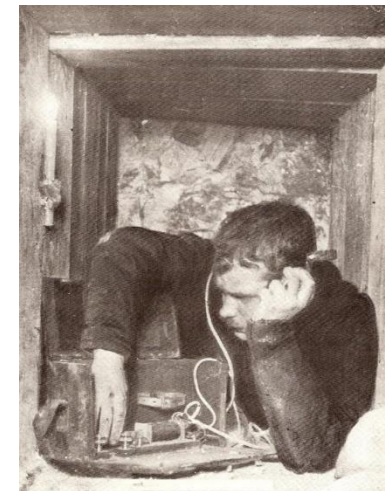
**Create:** Bone Flutes (7000 B.C.)



**Modify:** Delphi Theater (300 B.C.)



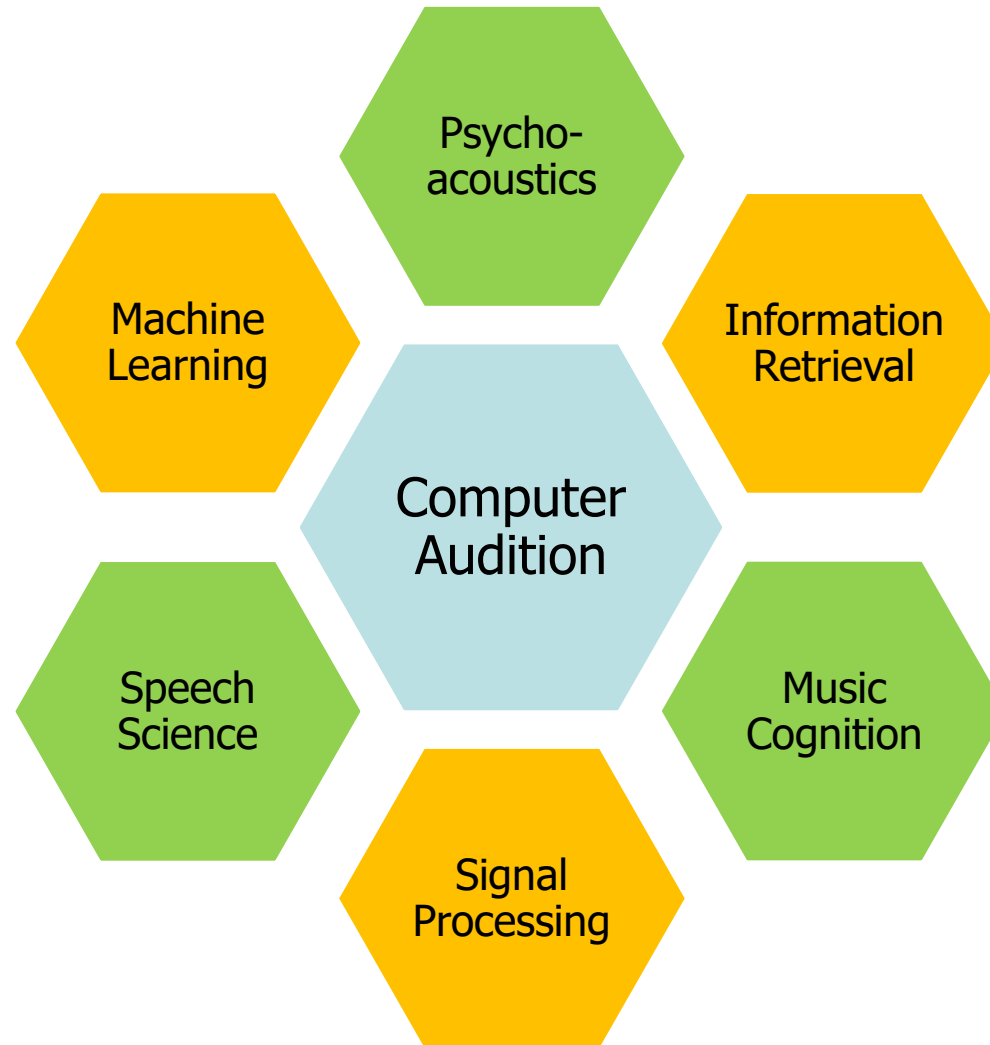
**Record:** Cylinder Phonograph (1899)



**Transmit:**  
Crystal Radio  
(1914)

# Impact on Many Fields

---

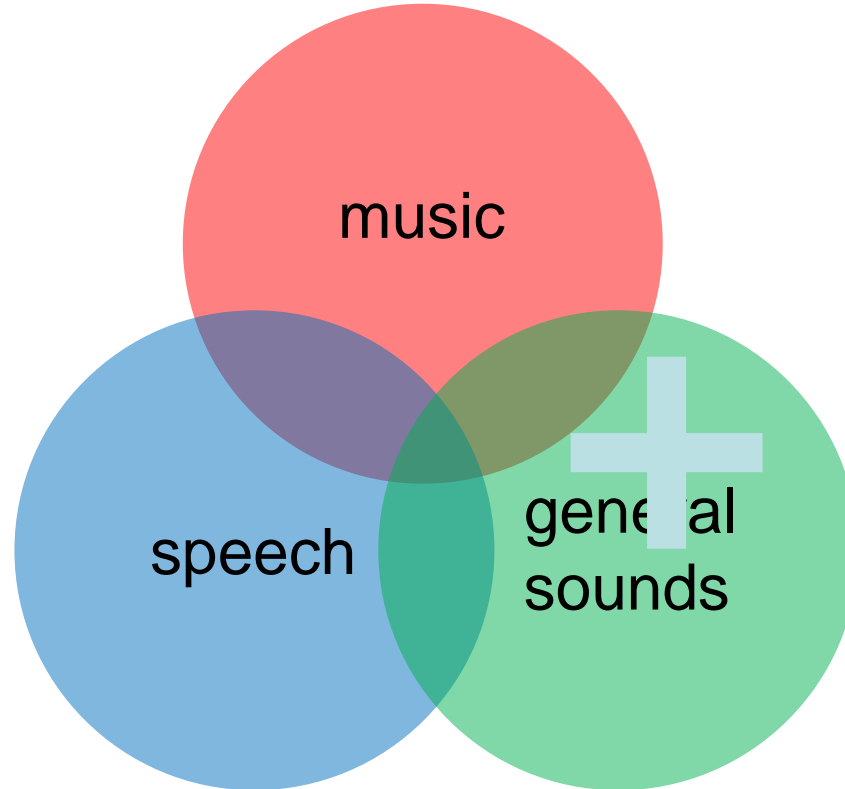


# Many Applications



# Research Areas

---



vision

text

EEG

other modalities



# Some Demos

---

- Automatic Music Transcription

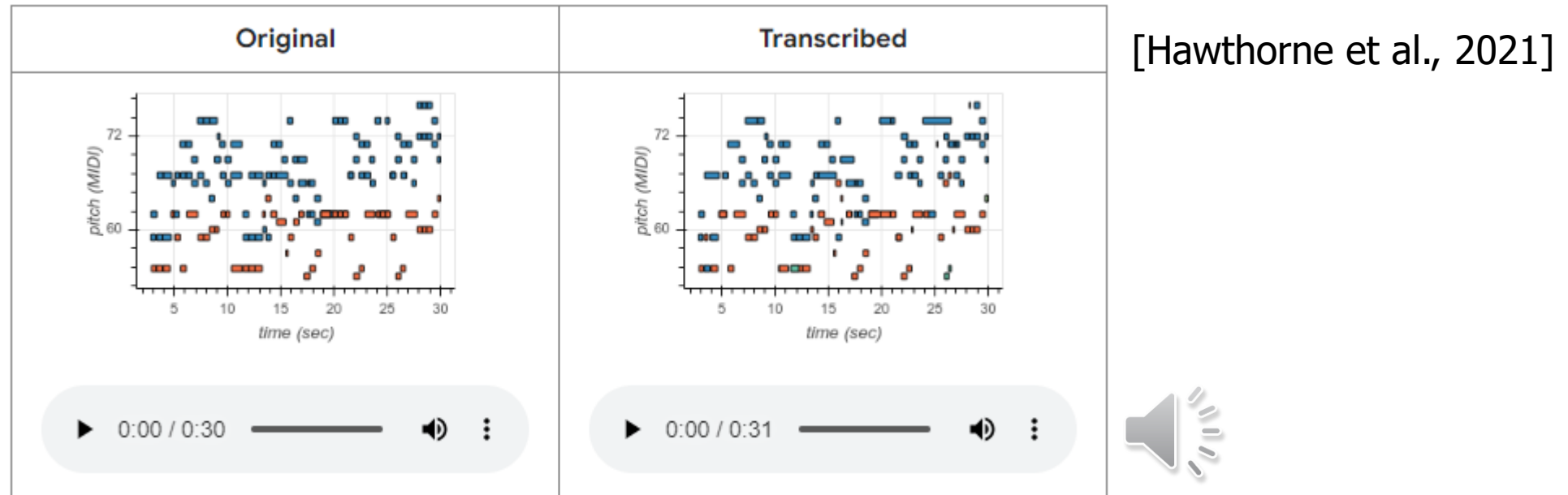
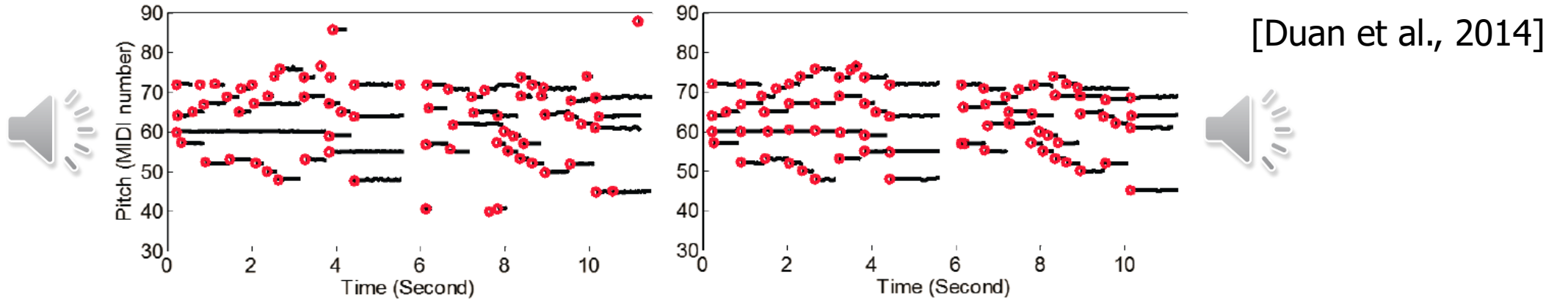
[Kong et al., 2020]

## **GiantMIDI-Piano: A MIDI dataset for classical piano music compositions**

ByteDance AI Lab

- Transcribed piano solo MIDI files.
- 2,784 composers
- 10,848 compositions
- 1,237 hours

# Some Demos



<https://magenta.tensorflow.org/transcription-with-transformers>

# Some Demos

---

- Pop music separation [Takahashi et al., 2018]
  - <https://sisec18.unmix.app/#/unmix/AM%20Contra%20-%20Heart%20Peripheral/TAU1>
- Violin/piano separation [Li, 2019]
  - Mixture:  violin:  piano: 
- Speech separation [Hershey et al., 2016]
  - Mixture:  female #1:  female #2: 
- More demos about speech separation
  - <https://cslikai.cn/project/Pure-Audio/>

# Some Demos

---

- Automatic Music Accompaniment
  - Music Plus One



[Raphael, 2001]

[https://music.informatics.indiana.edu/~craphael/music\\_plus\\_one/movies/movies.html](https://music.informatics.indiana.edu/~craphael/music_plus_one/movies/movies.html)

# Some Demos

---

- Symbolic music generation

- Music harmonization [Yan, 2018]



- Generation from scratch

- String trio:



# Some Demos

---

- Music audio generation
  - OpenAI's Jukebox: generating songs given lyrics, genre and artist
    - <https://openai.com/research/jukebox>
  - Google's MusicLM: generating music from text
    - <https://google-research.github.io/seanet/musiclm/examples/>

# Some Demos

---

- Text-to-Speech
  - IBM Watson
  - <https://www.ibm.com/demos/live/tts-demo/self-service/home>
  
- Voice conversion
  - ControlVC: [Chen & Duan, 2023]
  - <https://melissachen15.notion.site/melissachen15/ControlVC-Audio-Demo-dd0ea58c5b7f434a81af9cbcd67f56f6>

# Some Demos

---

- Voice editing

## VoCo: Text-based Insertion and Replacement in Audio Narration

Zeyu Jin\*, Gautham J. Mysore†, Stephen DiVerdi†, Jingwan Lu† and Adam Finkelstein\*

\* Princeton University † Adobe Research

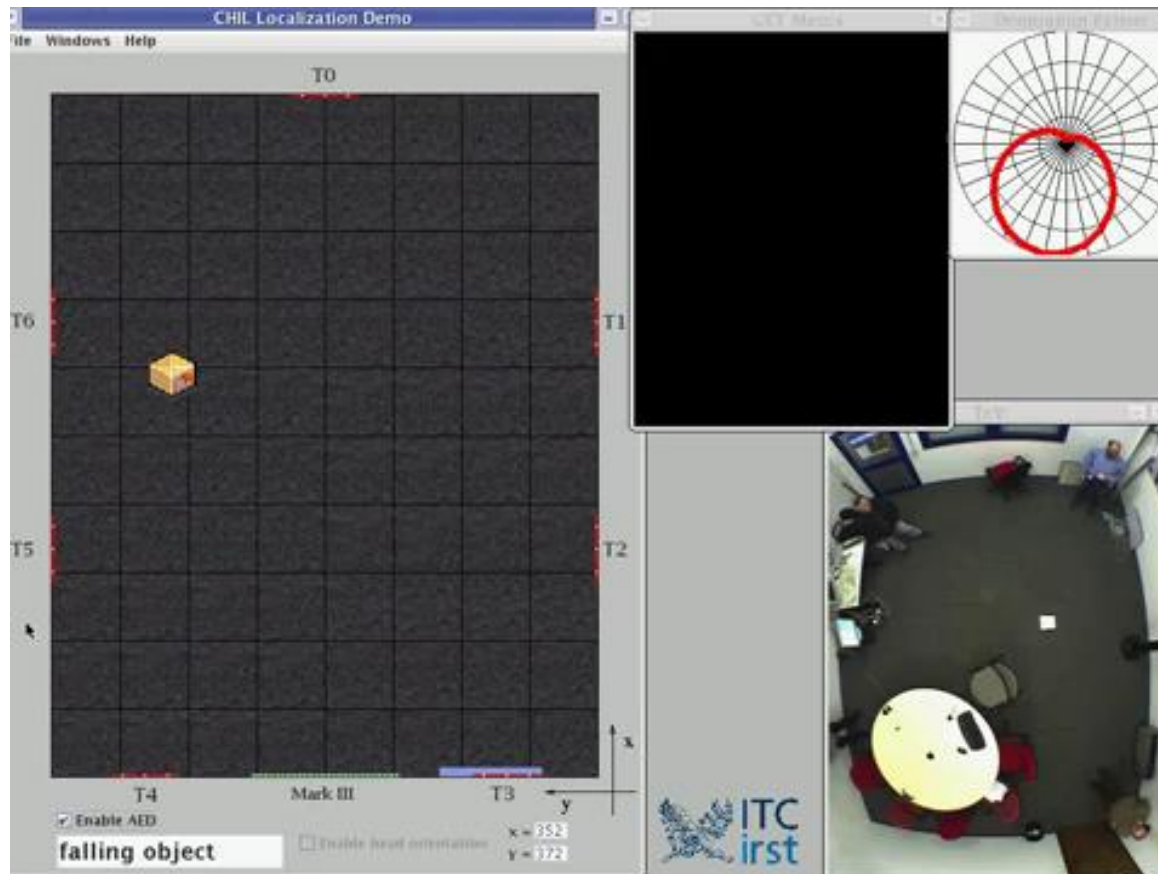
SIGGRAPH 2017

<https://www.youtube.com/watch?v=RB7upq8nzIU>



# Some Demos

- Acoustic event detection and localization



[https://www.youtube.com/watch?v=iImkV6oKG\\_8](https://www.youtube.com/watch?v=iImkV6oKG_8)

# Some Demos

---

- Audio-visual speech separation [Afouras, 2018]
  - <http://www.robots.ox.ac.uk/~vgg/demo/theconversation/demos/vox/0/demo.html>
- Speech-driven talking face generation [Eskimez et al., 2020]



# 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

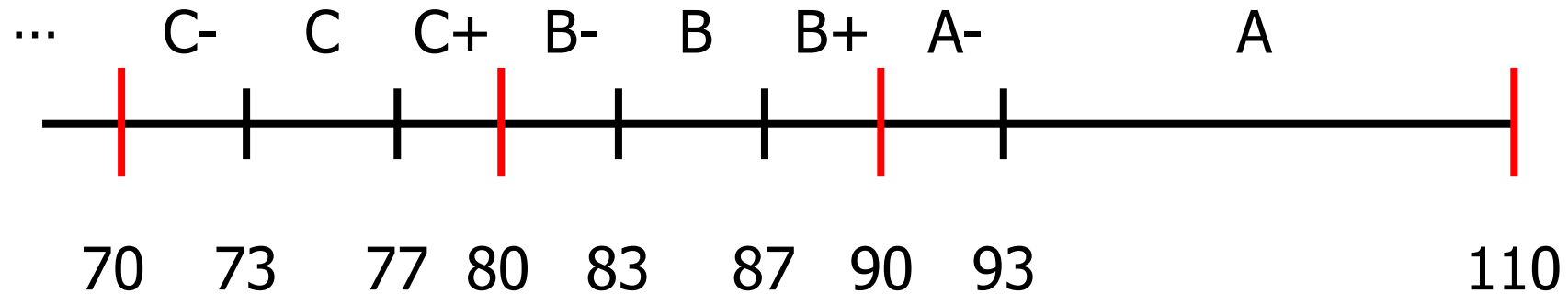
# Assignments

---

- Total (110 points)
  - Homework (60 points)
    - HW1-HW4: Python/Matlab programming
    - HW5-HW6: Python programming for deep learning
  - Class paper review (20 points)
  - Course project (30 points)
    - Proposal (5 points)
    - Status update (5 points)
    - Peer review (5 points)
    - Final report (5 points)
    - Presentation/demo (10 points)
- No exams

# Grading

---



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

# Important Policies

---

- Late homework penalty
  - 20% deduction of full grade each day
- Do your own work
  - Discussions are encouraged
  - No exchange of code
  - No copying of five or more consecutive words
  - Cite external sources
  - AI tools (e.g., ChatGPT) can only be used to polish text but not to generate content
- Attendance is not taken, but class discussions are very important for learning

# Prerequisites

---

- Signal Processing
  - ECE 246/446 or ECE 272/472 or equivalent
- Python or Matlab programming
- Preferred but not required
  - Machine learning such as SVM, Markov models, neural networks, clustering, etc.



# Two Websites

---

- Course website
  - All materials (lecture notes, readings, assignments, etc.)
  - <http://www.ece.rochester.edu/~zduan/teaching/ece477>
- Blackboard:
  - For announcements, homework submissions, and Q&A