Speech Processing 11-492/18-492

Sound ID
(slides from Bhiksha Raj)
What is in the audio scene

- Searching for specific things
  - Cars, talking, music

- Generically analyzing the audio
  - Find the “important” parts
Audio Alone
Summarizing Audio
Human judgment is non-standard

- Context matters
- Context includes a priori knowledge not represented in recording
- Two descriptions of a scene from a movie
  - "There were a series of beeps, and a bomb went off"
  - "A timer counted down, and then there was a big boom"
A Hierarchical Structure for Sound

Semantic Level

Intermediate Level

Audio Data

Lower-Level Units

Event Sequence

Event Dependencies

Crowd hum
Cheering (loud)
Talking (babble) + cheering (soft)

Metallic sound (bat hitting ball)

E₁ E₂ E₃ E₄
Audio Unit Detection

- **Low level acoustic units**
  - Similar mcep over time
- **Find repeated segments over time**
- **Find repeated patterns over time**