Simple Speech Recognition
Larry Rudolph
Speechbuilder +/-

- Speechbuilder is a great tool, but
- Grammar is limited
- Forces use of “action” and “attribute”
- “Domain” has lots of components
  - complicated compile of domain pieces - vocabulary, natural language, scoring
- Must start server with domain framework
Simpler than S-Builder

- Simple:
  - send audio to server
  - receive string in the domain
  - but what about domain?
  - open a connection with server w/ domain
  - stream audio over connection; get string when done
Streaming Audio

- Long process:
  - each step can be seconds.
- Streaming can speed things up (a lot)
Streaming Audio

- Start sending audio before user is done speaking
- Start recognition before audio done arriving
- Still must wait for return string
Streaming recognition

- Recall, recognition proceedings in several stages:
  - waveform to phonemes
  - phonemes to words
  - words to sentences
  - Natural language filtering
  - Grammar Parse

Much of this can be pipelined

NL requires whole sentence
Speed up process

• Do not do natural language filtering
• Do not do very limited vocabulary
• (what should one do with extraneous words -- fastest is to recognize them since ignoring them slows down parsing)
more speed

• pipeline whole process.
• what about grammar parse?
• do a Virturbi search
• get back confidence levels