Characterizing and Processing Robot-Directed Speech

Paulina Varchavskaia • Paul Fitzpatrick • Cynthia Breazeal
MIT AI Lab Humanoid Robotics Group

Characterizing and Processing Robot-Directed Speech

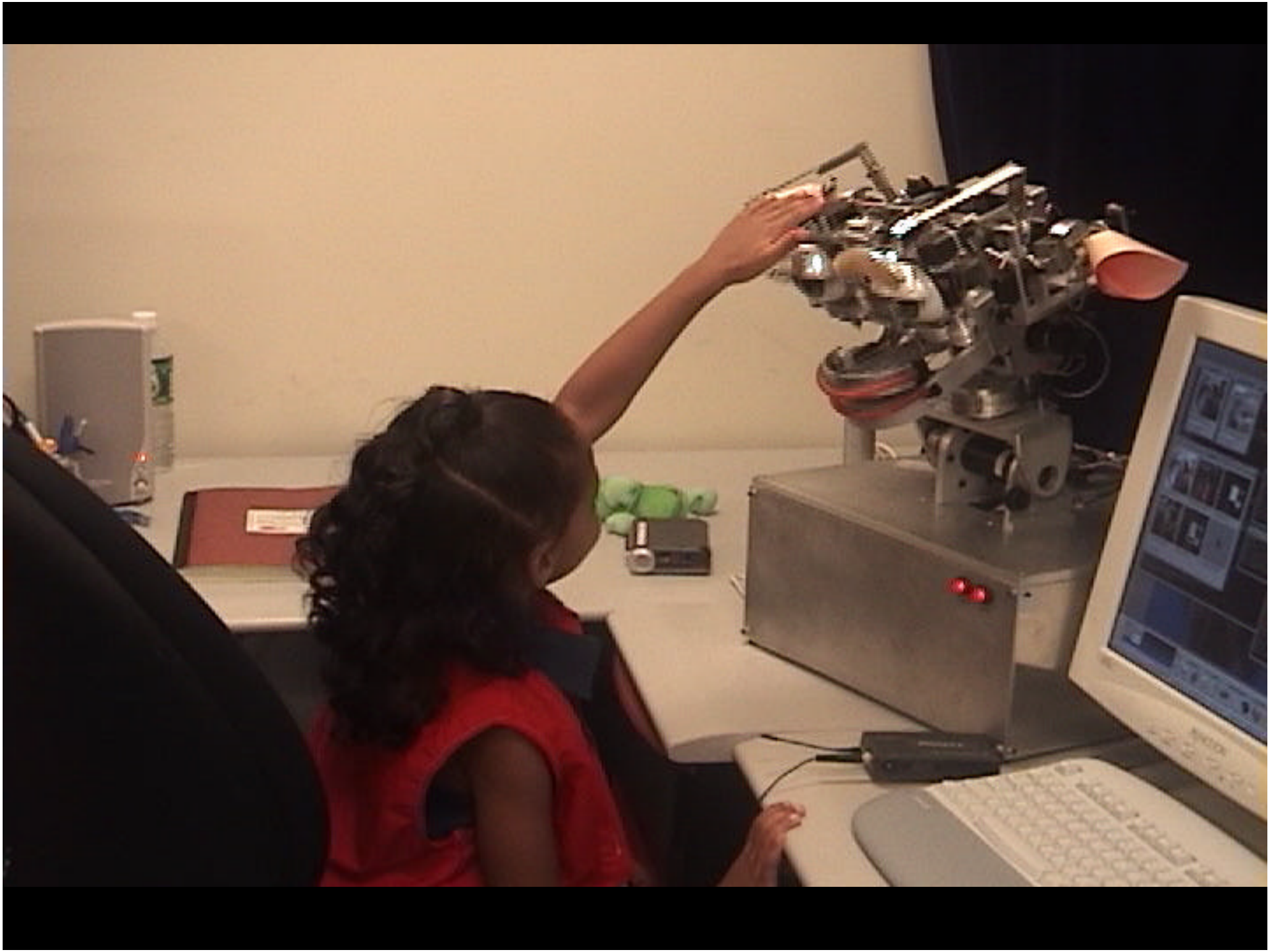
Paulina Varchavskaia • Paul Fitzpatrick • Cynthia Breazeal
MIT AI Lab Humanoid Robotics Group

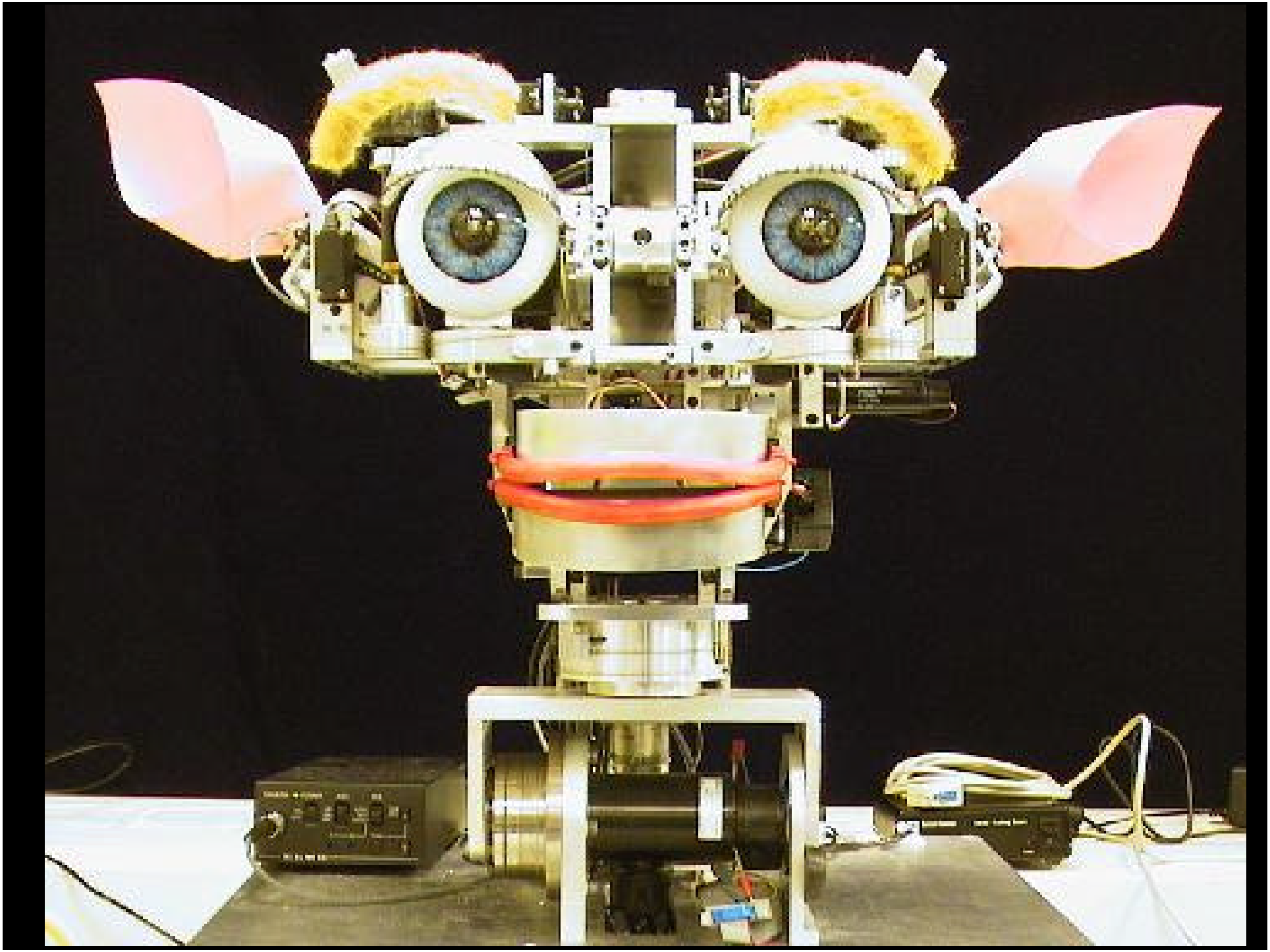
Characterizing and Processing Robot-Directed Speech

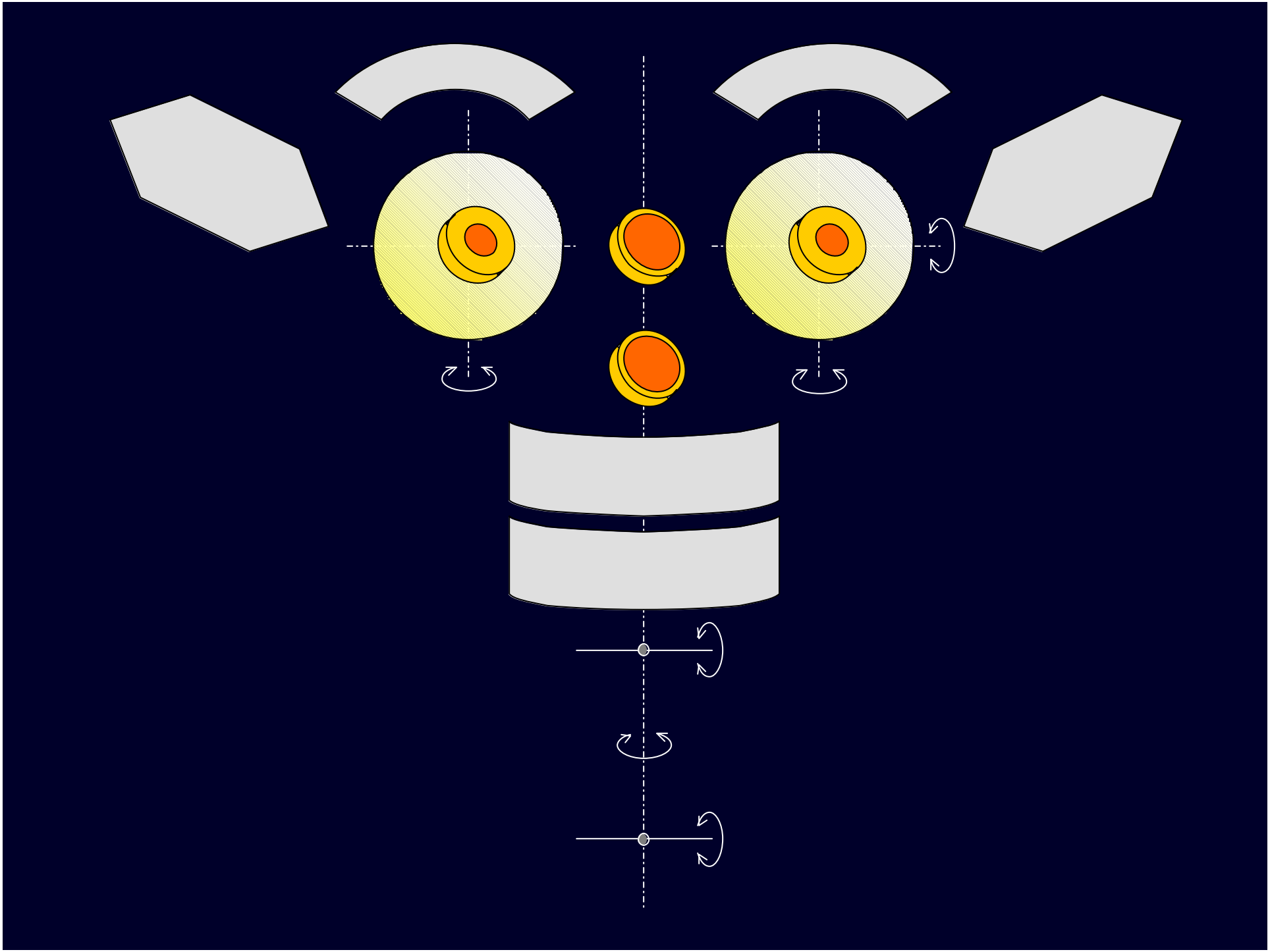
Paulina Varchavskaia • Paul Fitzpatrick • Cynthia Breazeal
MIT AI Lab Humanoid Robotics Group

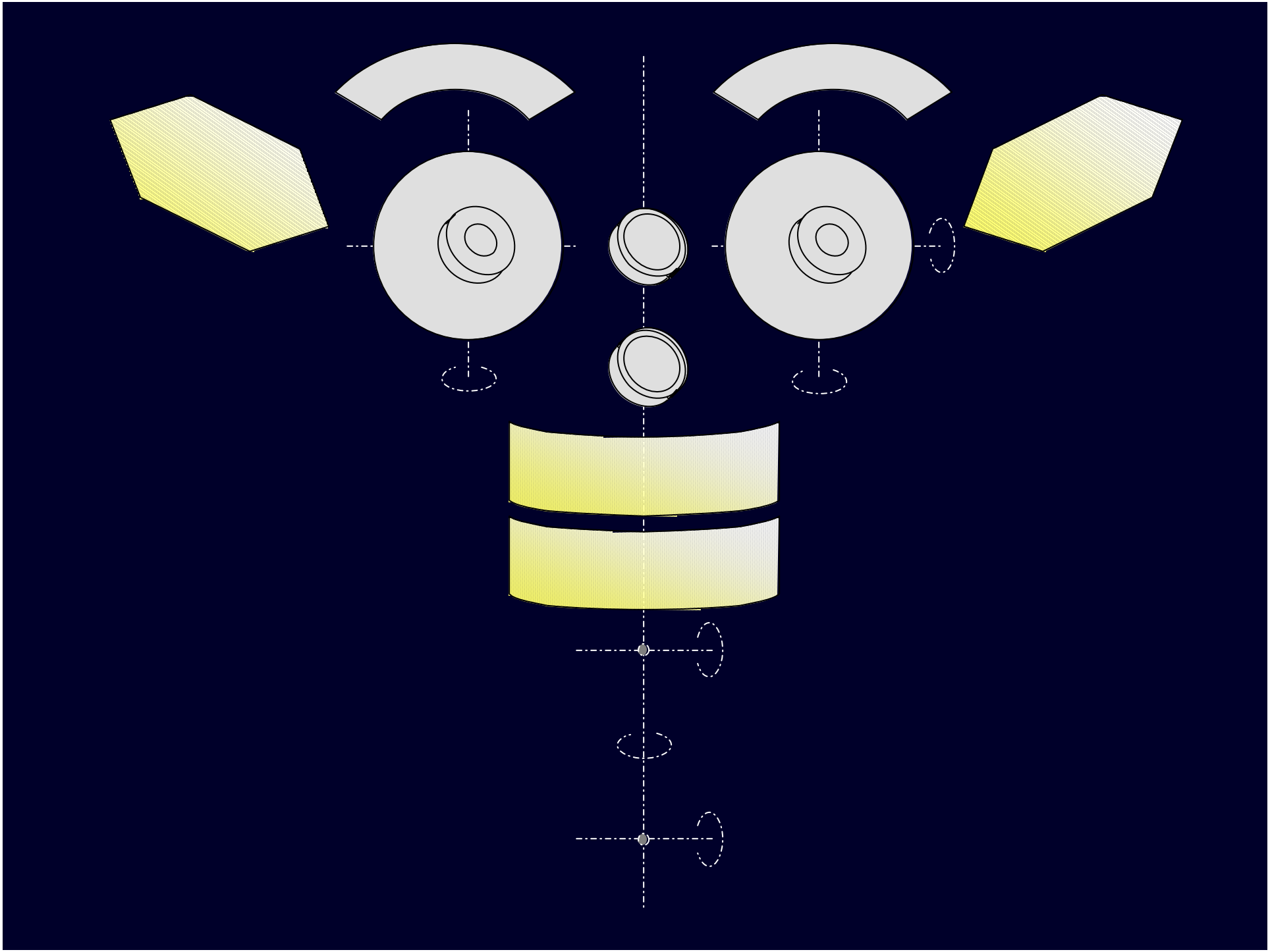
Characterizing and Processing Robot-Directed Speech

Paulina Varchavskaia • Paul Fitzpatrick • Cynthia Breazeal
MIT AI Lab Humanoid Robotics Group





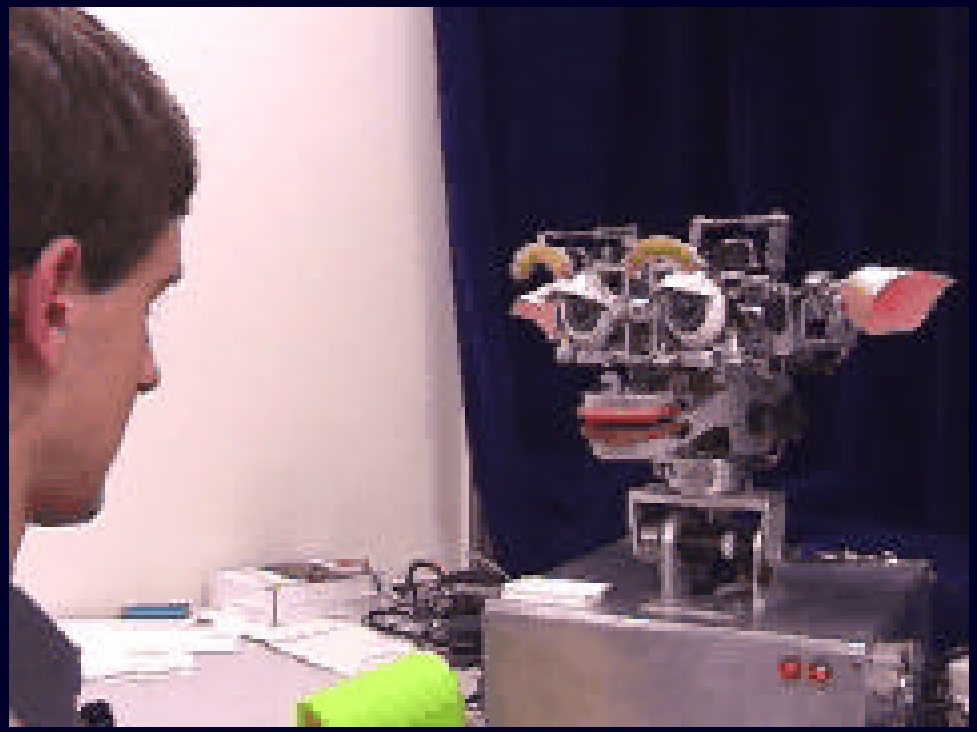


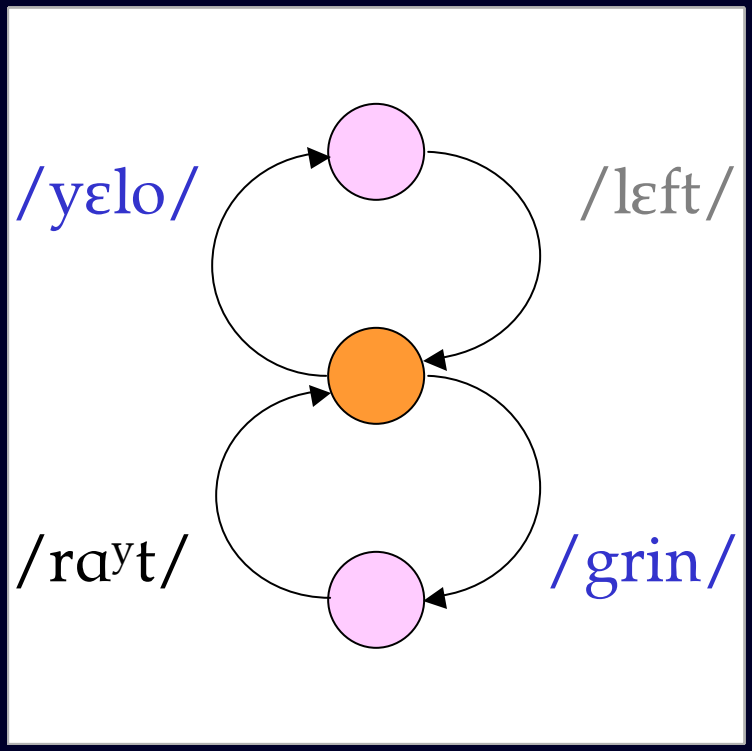


Speech Recognition

- What speech does the robot evoke?
- How can we process that speech?
  - Support a growing vocabulary
  - Without hurting recognition accuracy

Baseline System



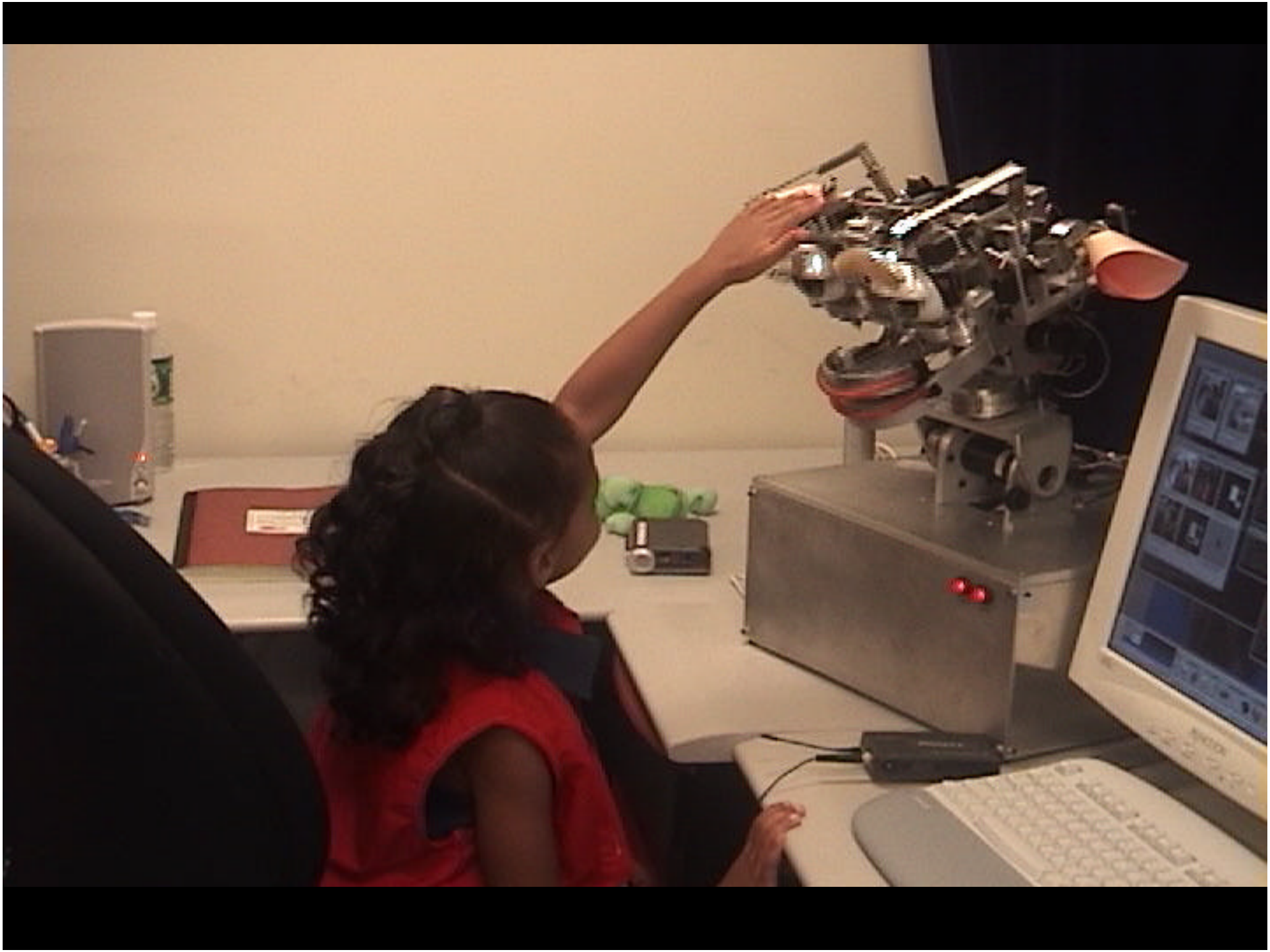


Baseline System

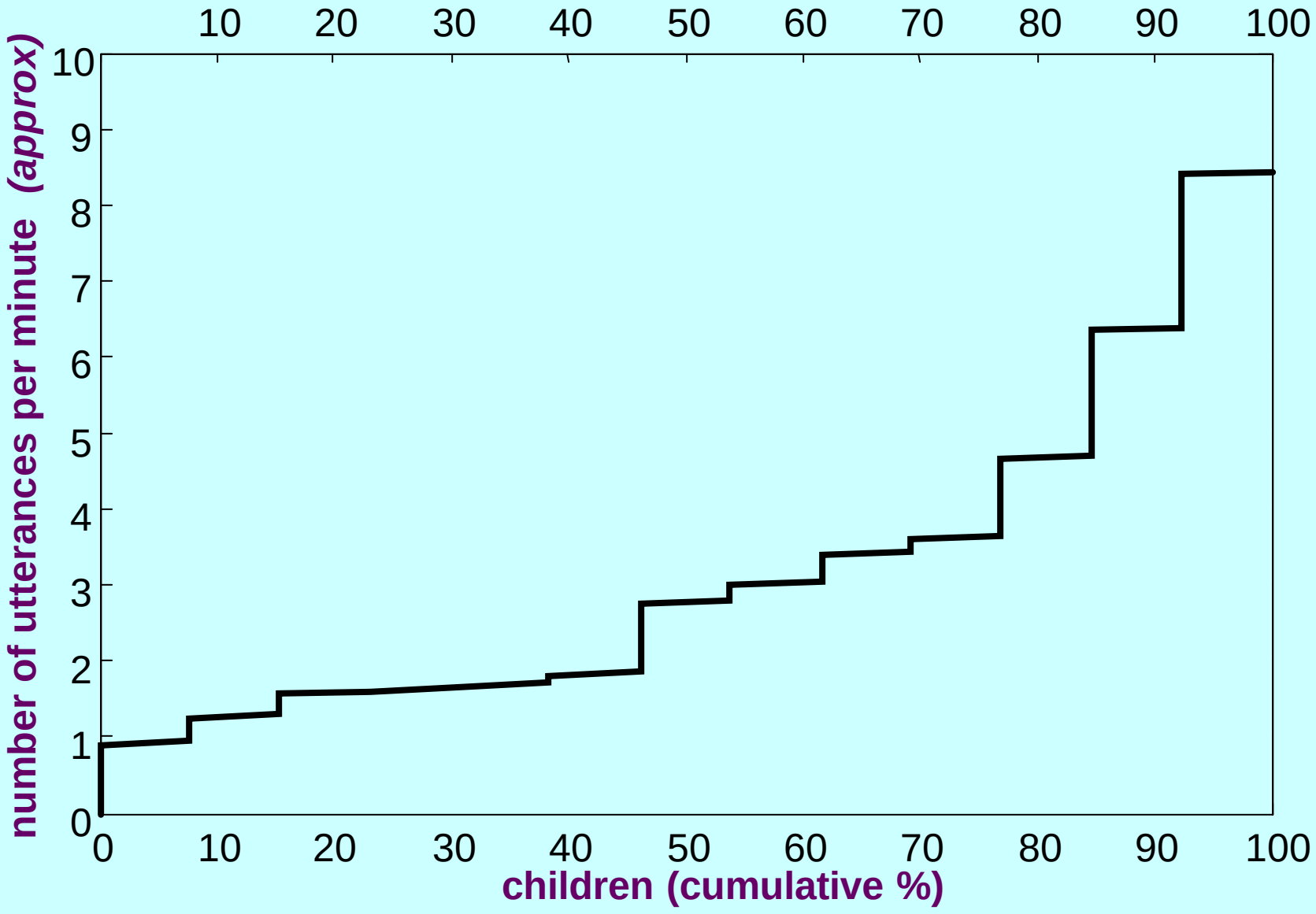
- “Magic word” to introduce vocabulary
- Separation of introduction from use
- Robot confused by unknown/filler words

Study: Characterizing Speech

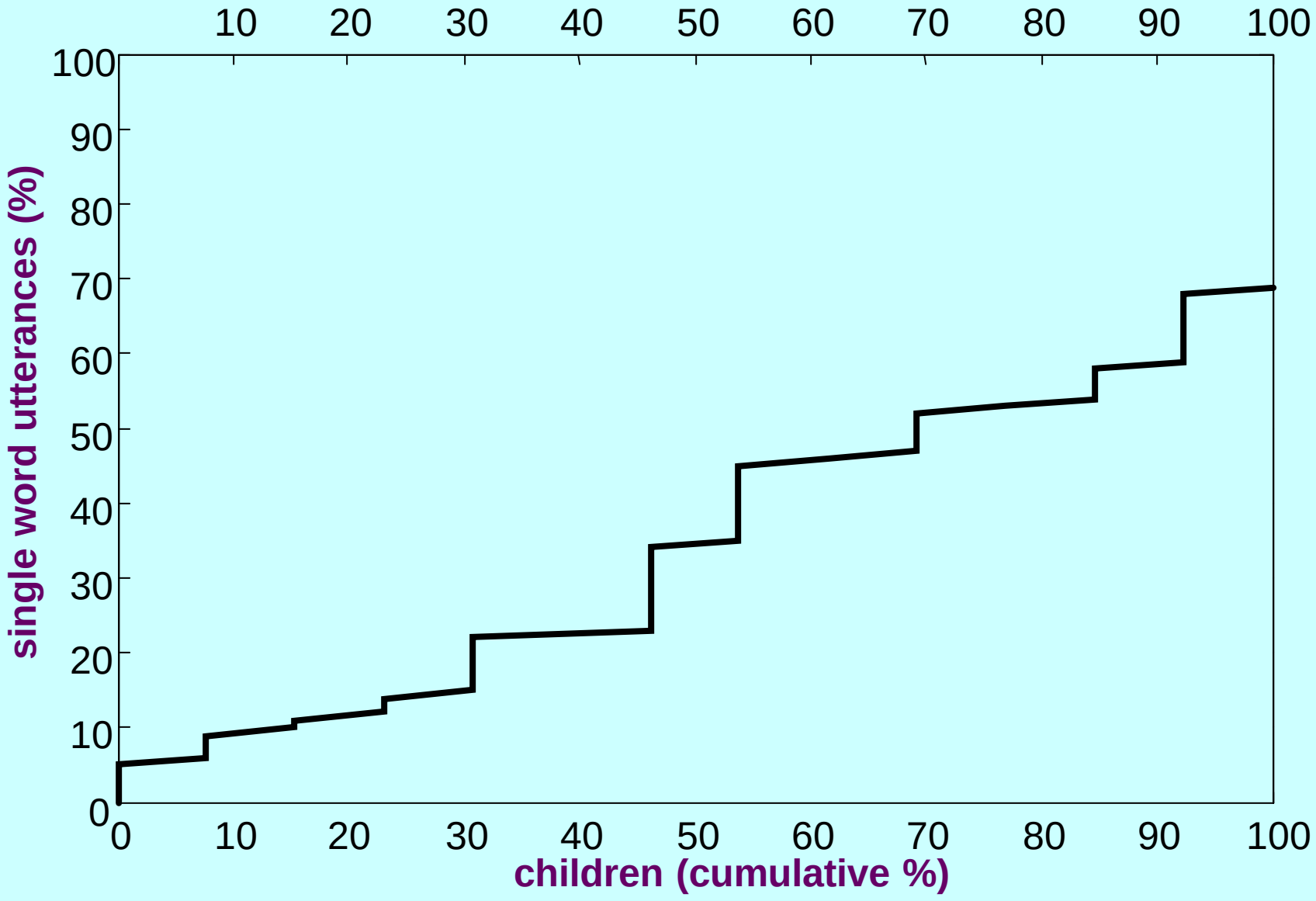
- 13 children
- 5-10 years old
- 20 minute sessions
- Some prompting *(Turkle et al)*



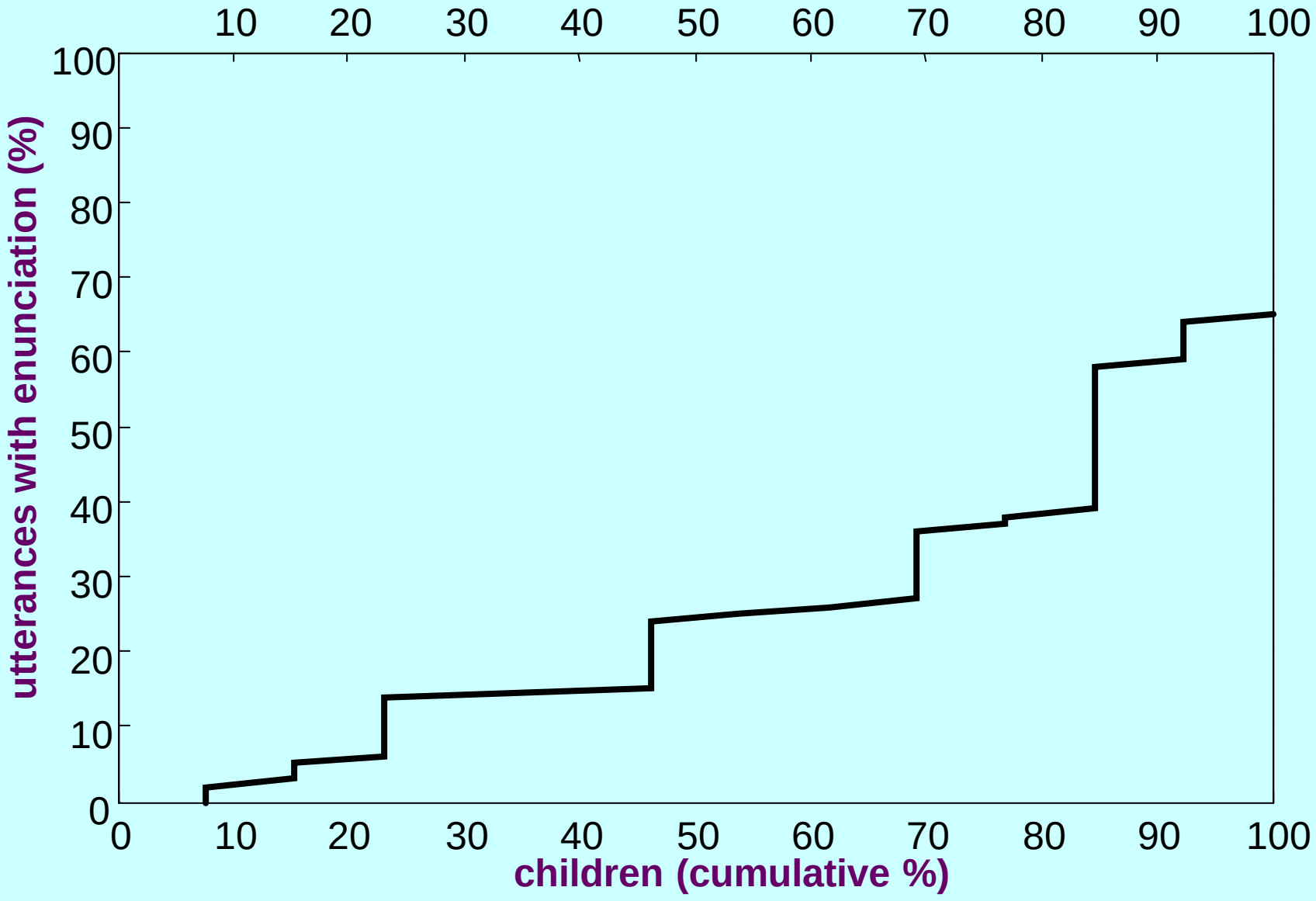
Number of Utterances



Single Word Utterances



Utterances with Enunciation



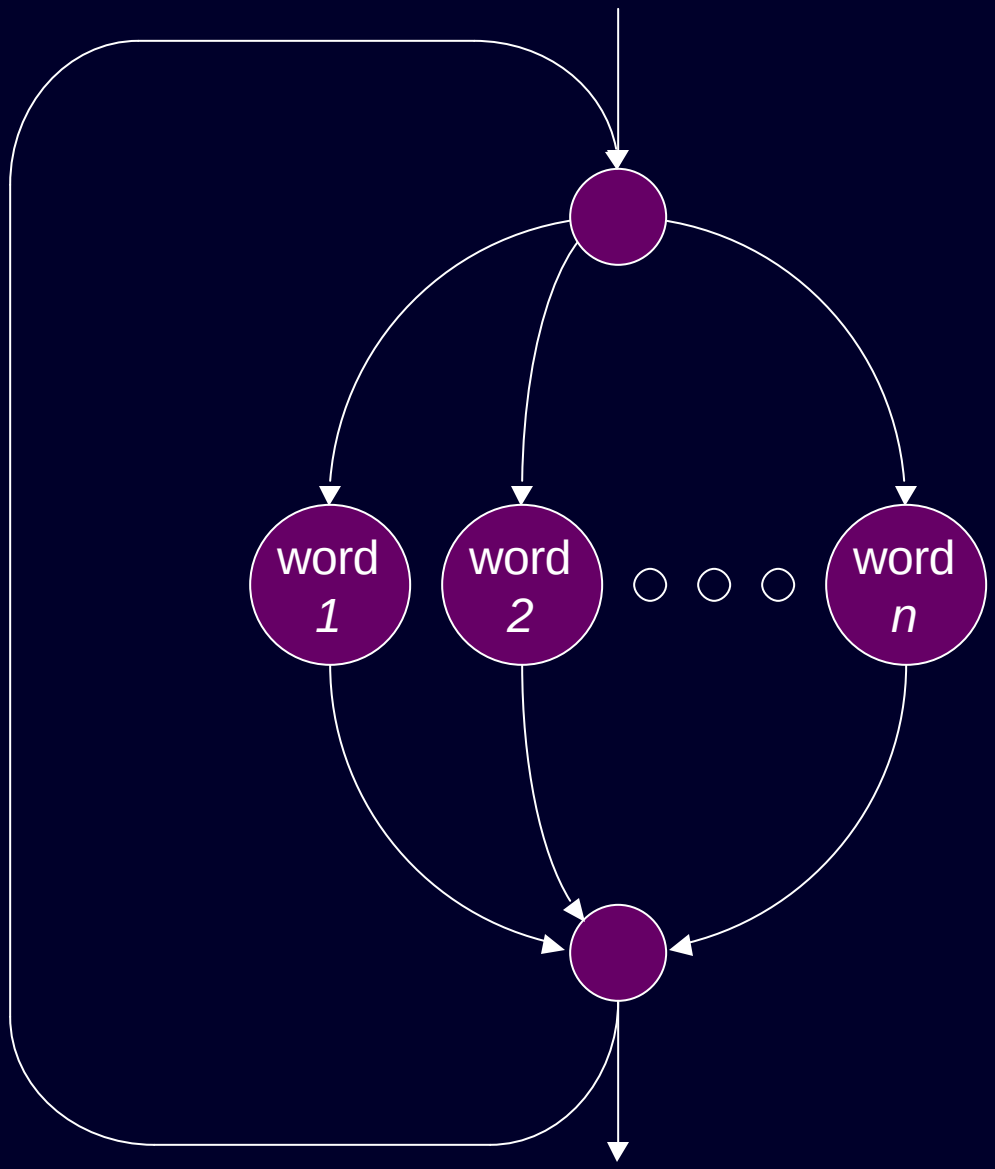
Cooperative Speech Exists

- Some clean word-learning scenarios
  - "Tiffany"
- Can we use them as leverage?
  - "My name is Tiffany"
- What recognition rates are possible?

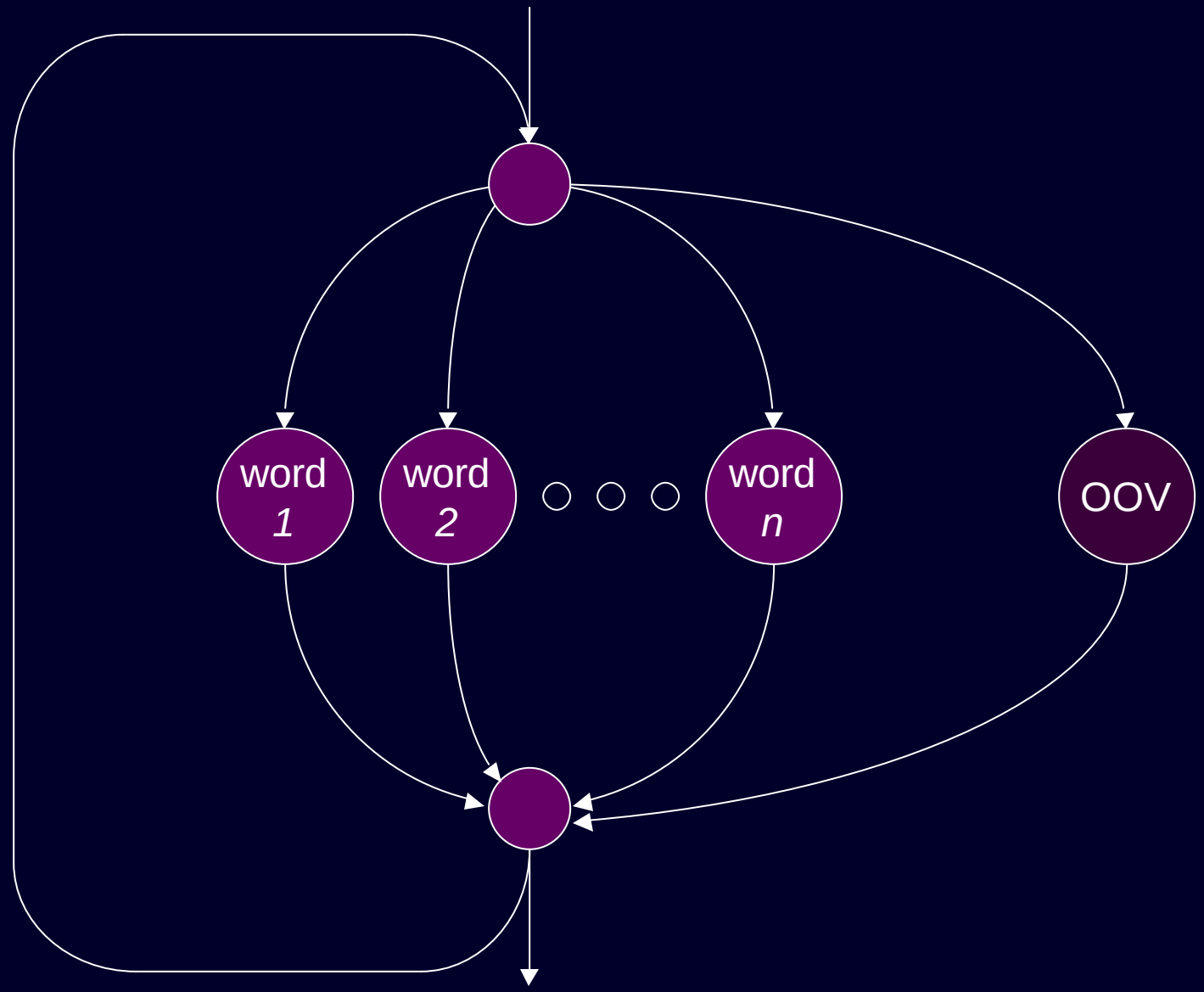
Study: Processing Speech

- "LCSINFO" domain *(Glass, Weinstein)*
- Provide some initial vocabulary
- Build language model without transcripts

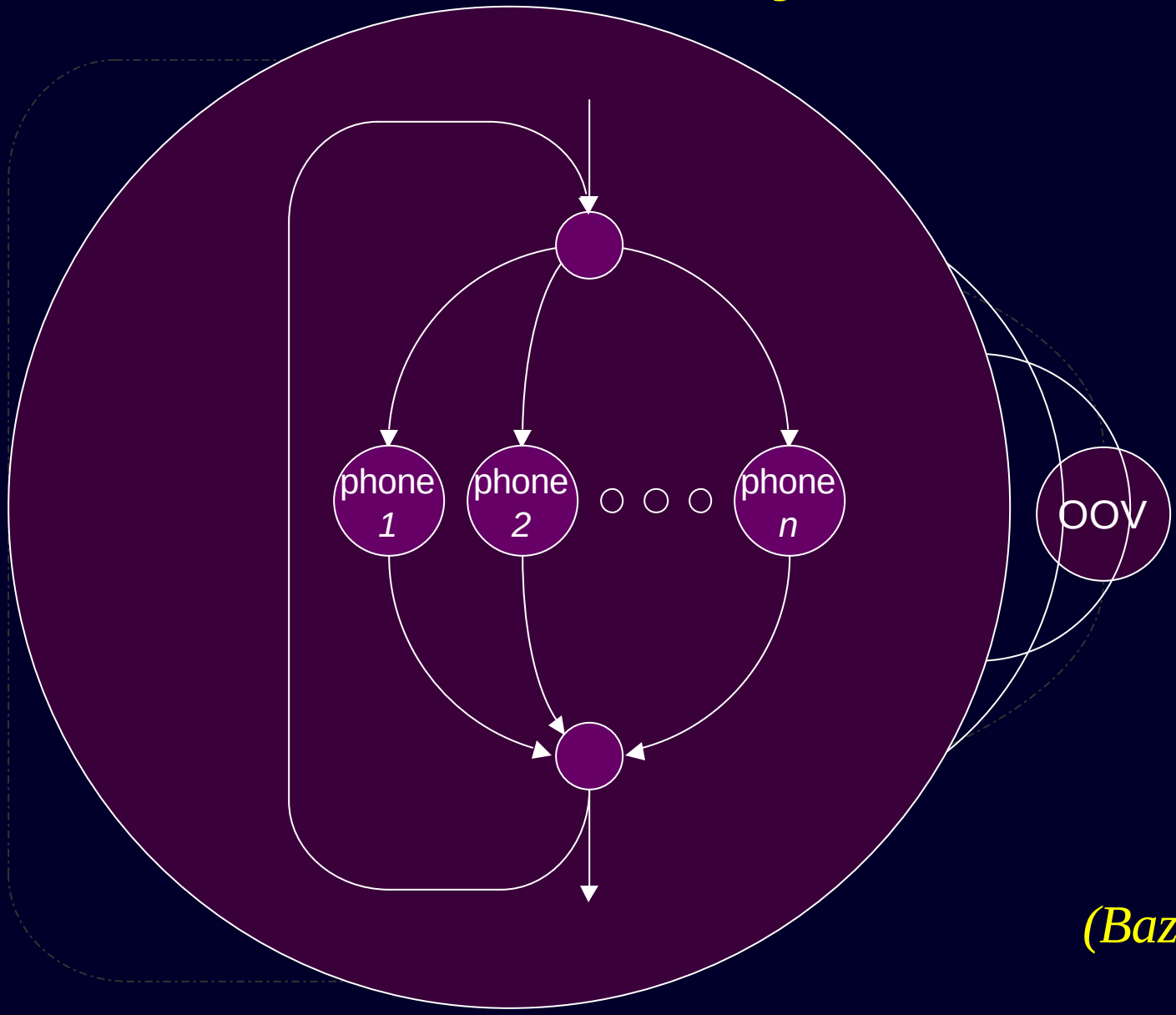
Language Model



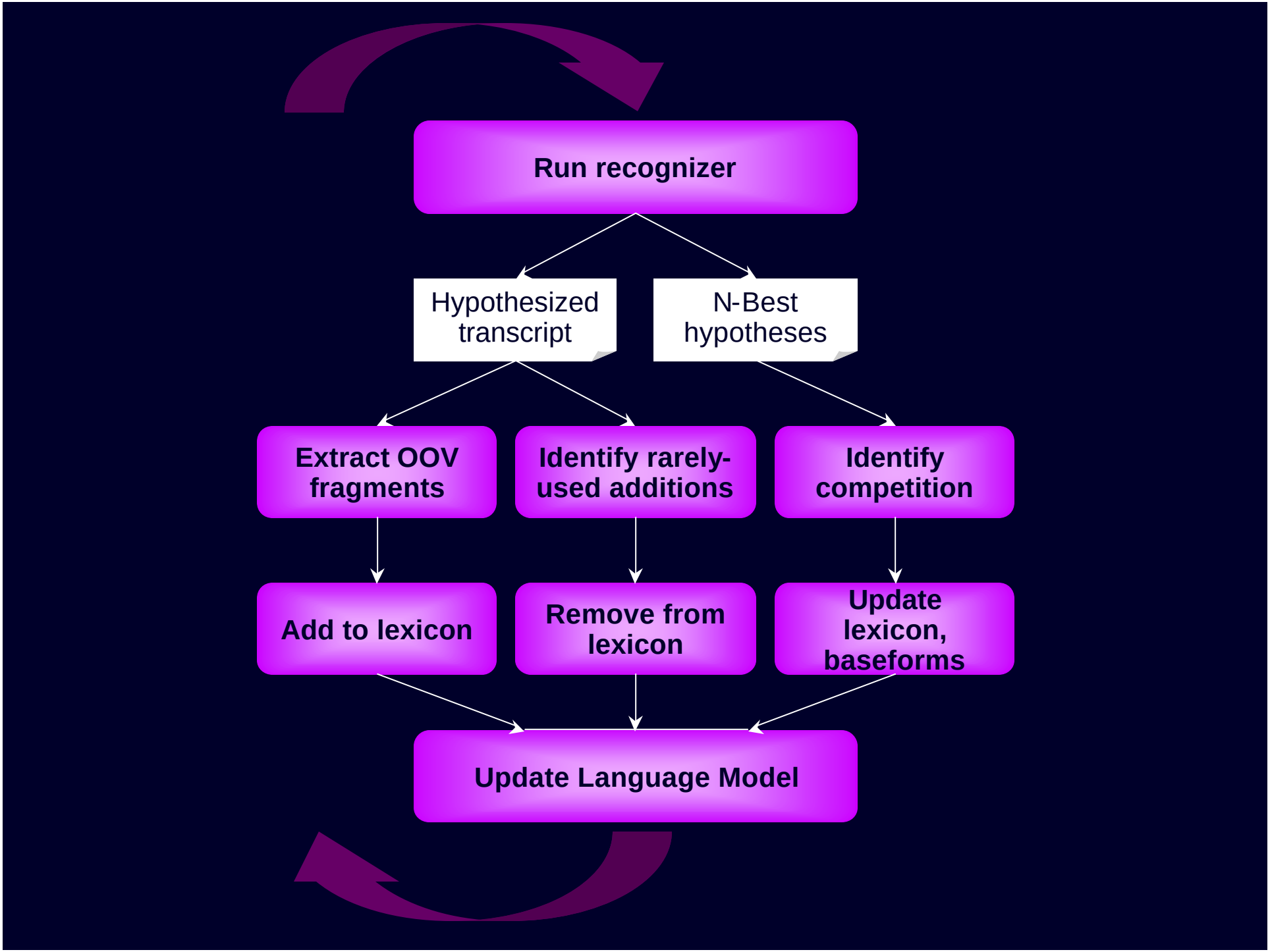
Out Of Vocabulary Model

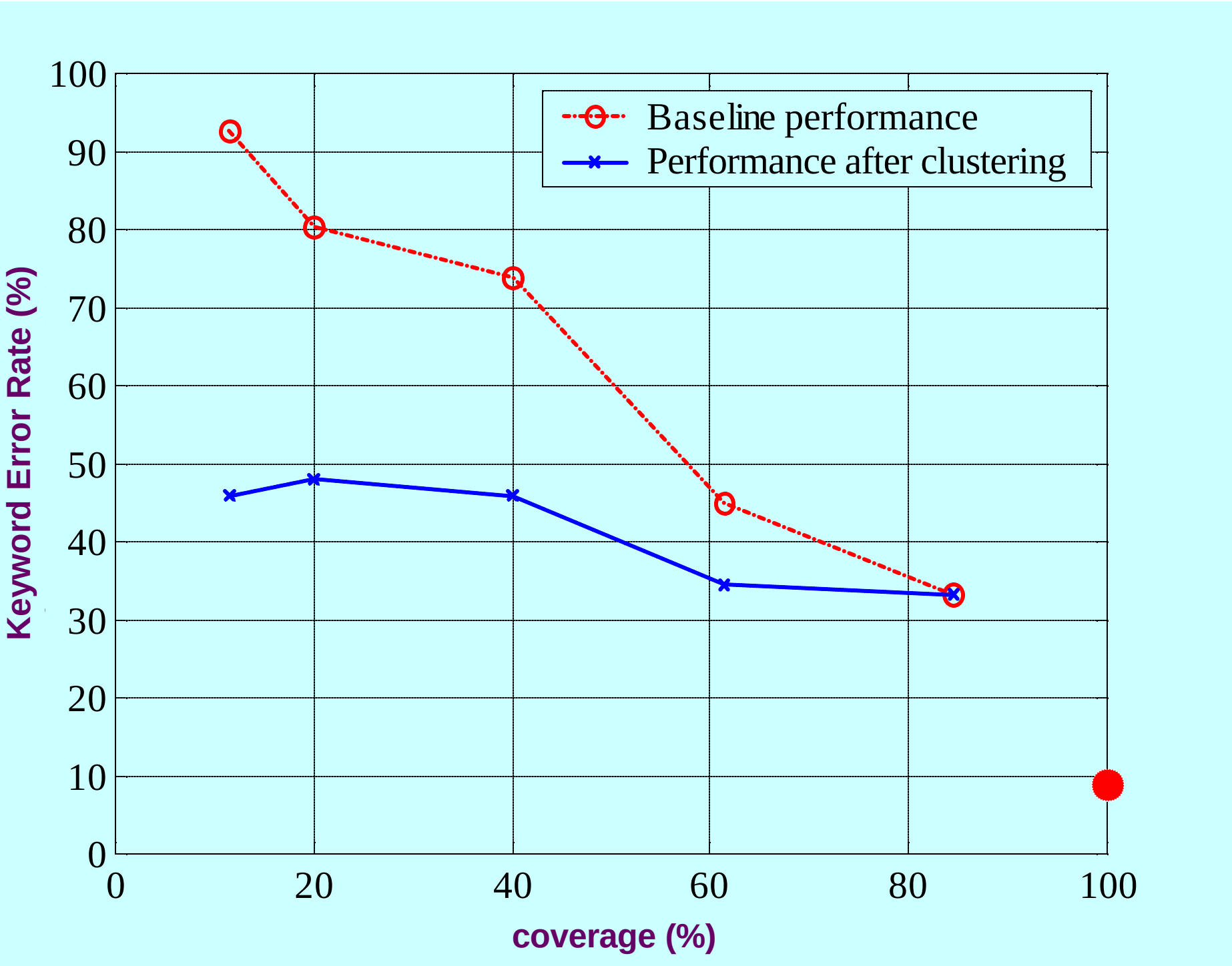


Out Of Vocabulary Model



(Bazzi)


Run recognizer
Hypothesized transcript
N-Best hypotheses
Extract OOV fragments
Identify rarely-used additions
Identify competition
Add to lexicon
Remove from lexicon
Update lexicon, baseforms
Update Language Model


Baseline performance
Performance after clustering
Keyword Error Rate (%)
coverage (%)
0
10
20
30
40
50
60
70
80
90
100
0
20
40
60
80
100

Qualitative Results

Given:

1600 Utterances

“email, phone, room, office, address”

Finds:

1. n ah m b er
2. w eh r ih z
3. w ah t ih z
4. t eh l m iy
5. k ix n y uw
6. p l iy z
7. ae ng k y uw
8. n ow
9. hh aw ax b aw
10. g r uw p

Conclusions

- What about acoustic models?
- Idiosyncratic vocabulary is important
- Pick a good name for your robot!

Acknowledgements

- MIT Initiative on Technology and Self
- MIT Spoken Language Systems group
- DARPA
- NTT