So...

• What’s so special about talking to an actual robot?

• Basically a voice-enabled chatbot with arms, right?
Well...

Talking to Robots: Making it Work!

Pepper is @ A10 booth at #RSAC19 - love this humanoid!
Well...

Talking to Robots: Making it Work!
What’s different?

Text bot  Voice bot  Avatar  Physical robot

Mirroring/similarity  Shared goals  Engagement  Conversion rate
What’s different?

• Physical robots/avatars result in a high level of user engagement than text or audio-only bots
• Robots carry greater sense of trust and credibility, feeling of shared goals between bot and user (Cynthia Breazeal)
• Sounds great! BUT that means that users also have a higher expectation of engagement from the robot.
How do we get more engagement?

• Best-of-breed AI solutions from all possible sources, both 3rd party and home-grown
• Real-time, multi-model, human-in-the-loop, when AI is unsure
Best-of-breed AI: Components

Talking to Robots: Making it Work!

We Make Robots Smarter™
Best-of-breed AI: STT harder when...

- Higher perplexity language model
- Talker further away from device
- Higher noise
- Higher reverberation ($T_{60}$)
- Multiple talkers
Best-of-breed AI: STT very good for...

- 3rd party APIs, devices like Echo, Google Home, HomePod, Siri, Cortana
- Medium-high perplexity language
- <= 2m from device
- Low noise SNR >= 20 dB
- Low reverberation $T_{60} \sim 0.5s$
- Single talker
## Best-of-breed AI: STT with various factors

<table>
<thead>
<tr>
<th>Use case</th>
<th>Perplexity</th>
<th>Mic dist (m)</th>
<th>dB SNR</th>
<th>$T_{60}$</th>
<th>WER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexa/Google Home (2016)</td>
<td>Medium</td>
<td>2</td>
<td>20</td>
<td>0.5</td>
<td>20</td>
</tr>
<tr>
<td>CHiME-1: binaural living room (2011)</td>
<td>6.3 (V. low)</td>
<td>2</td>
<td>9-- -6</td>
<td>0.3</td>
<td>8.2 @ 6 dB</td>
</tr>
<tr>
<td>CHiME-2: CHiME-1 + more vocab + speaker movement (2013)</td>
<td>6.3,110</td>
<td>2</td>
<td>9-- -6</td>
<td>0.3</td>
<td>4,17 @ 6dB</td>
</tr>
<tr>
<td>REVERB: Single speaker in office room (2013)</td>
<td></td>
<td>1,2.5</td>
<td>20</td>
<td>0.7</td>
<td>30—50 (1—8 mics)</td>
</tr>
<tr>
<td>CHiME-3: Tablet w/ 6 mics outside (2015)</td>
<td>0.4</td>
<td>5</td>
<td>0</td>
<td>5.8</td>
<td></td>
</tr>
<tr>
<td>CHiME-4: CHiME-3 + 1,2,6 mics (2016)</td>
<td>0.4</td>
<td>5</td>
<td>0</td>
<td>2.2, 3.9, 9.2</td>
<td></td>
</tr>
<tr>
<td>CHiME-5: Dinner party (mult talkers) (2018)</td>
<td>2</td>
<td>High</td>
<td>0.5</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Pepper in trade show</td>
<td>Medium/low</td>
<td>1</td>
<td>15—5</td>
<td>1+</td>
<td>???</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Space</th>
<th>$T_{60}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home</td>
<td>0.55</td>
</tr>
<tr>
<td>Office</td>
<td>0.4—0.7</td>
</tr>
<tr>
<td>Mall</td>
<td>1.7—3.2</td>
</tr>
<tr>
<td>Airport</td>
<td>3+</td>
</tr>
</tbody>
</table>
Best-of-breed AI: STT in noise

CER v. SNR

Talking to Robots: Making it Work!
Human-in-the-loop

- STT, NLU pretty good
- DM, NLG still need work for natural speech

- Real-time human-in-the-loop
- Determine when AI is less confident
- Pull in human operator in real-time
- User gets right answer
- Collect most valuable training data
**HARI Architecture**
OUR INNOVATIVE PLATFORM

**HARI**
Combined Human and Machine Intelligence
Supervised by a Human Operator

**VBN NETWORK**
Isolated From Internet
Secure and High Performance
Global

**CLOUD ROBOTS**
Natural and Engaging Human-Robot Interaction
Focusing on: Recognition, Understanding, and Dialog

**CloudMinds AI**

**HARI Switch**

**Best-of-Breed AI**

**Smart Vision**

**Natural Language Processing**

**SLAM**

**Grasping**

**VBN**

**RCU**

**CLOUD ROBOTS**
We’re Excited!

• How to build great voice apps

• Tons of industry-seasoned examples

• Chapters 1–5 released

• Completion end 2019
THANK YOU