Speech interaction with personal assistive robots supporting aging-at-home for individuals with Alzheimer's disease

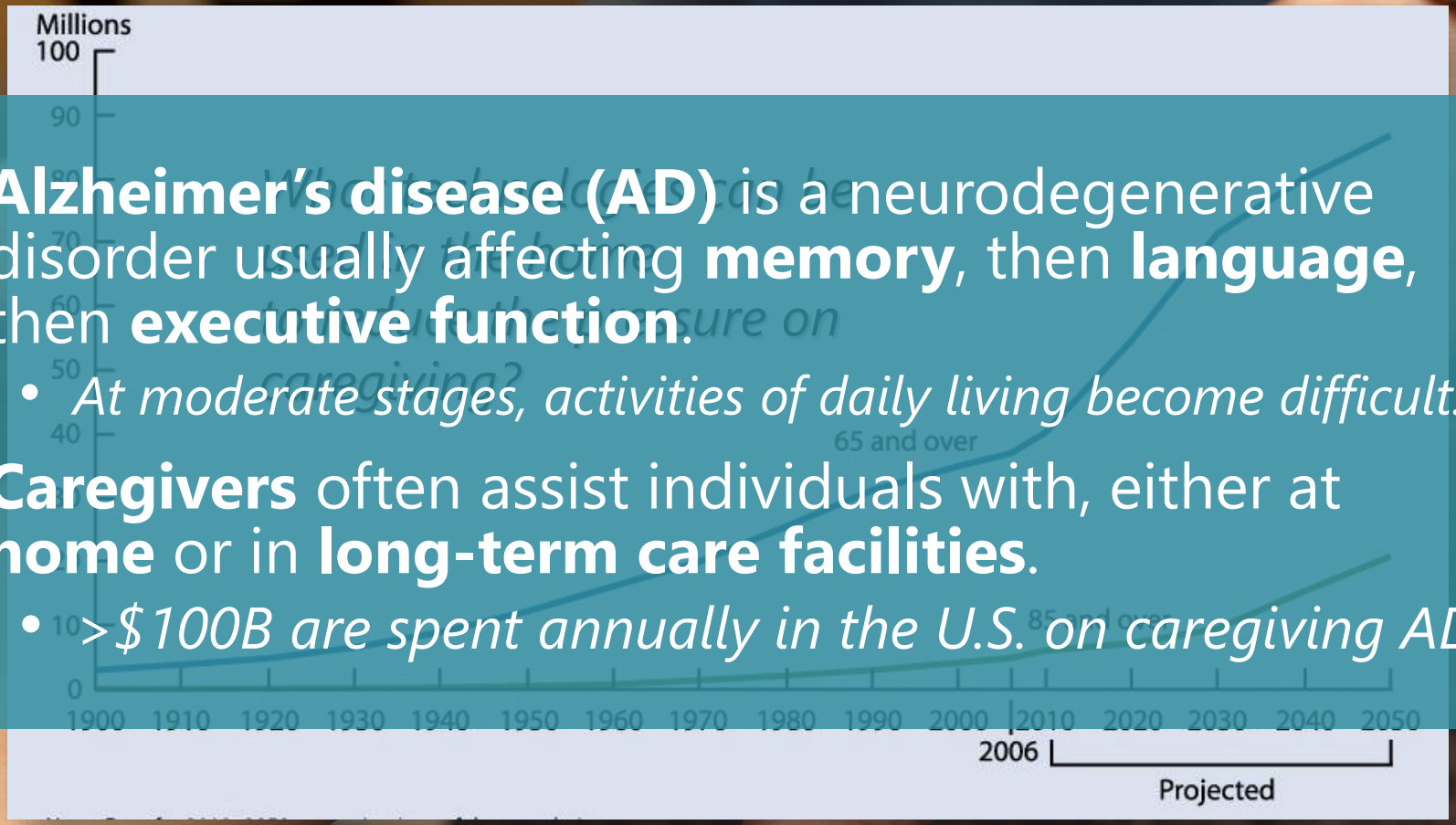
Frank Rudzicz[1,2],
Rosalie Wang[1,2], Momotaz Begum[3], Alex Mihailidis[1,2]

1 UHN Toronto Rehabilitation Institute

2 University of Toronto

3 UMass Lowell

Introduction



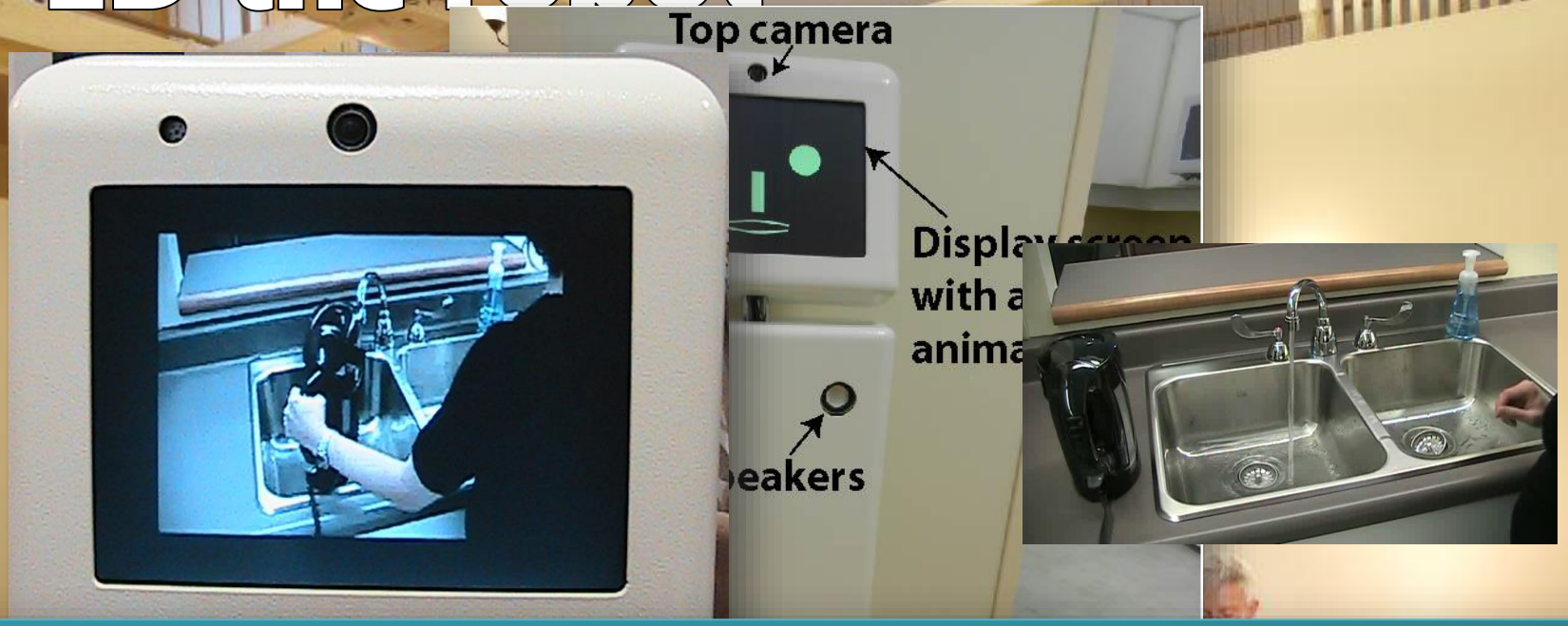
- **Alzheimer's disease (AD)** is a neurodegenerative disorder usually affecting **memory**, then **language**, then **executive function**.
  - *At moderate stages, activities of daily living become difficult.*
- **Caregivers** often assist individuals with, either at **home** or in **long-term care facilities**.
  - *>$100B are spent annually in the U.S. on caregiving AD.*

The HomeLab

- **'COACH'** automates support of daily activities.
  - E.g., hand-washing, tooth-brushing.
  - Uses partially-observable Markov decision processes (POMDPs) and **camera-only** input.

- *But what if the user does not want to spend their whole day in front of the sink?*

ED the robot



Early qualitative analysis indicated that **speech** is the most **desired** form of interaction with such a system.

Our **goal** is to implement two-way **spoken dialogue** that ***identifies*** and ***recovers*** from communication breakdowns.

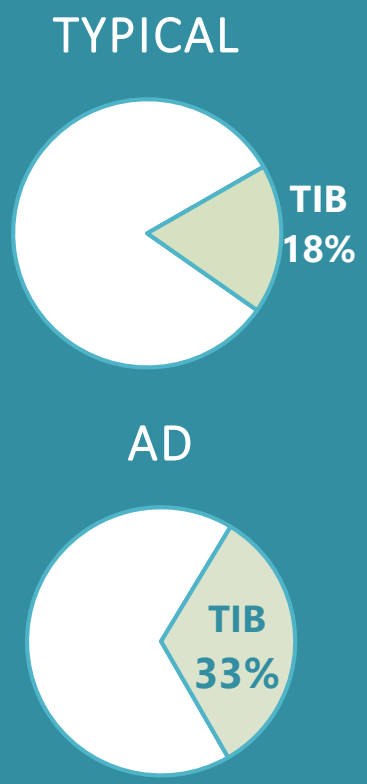
Related work

- There has been a *lot* of great work on supporting **older adults** with robots.

  - *However, speech interaction has been superficial.*

- We know a lot about how AD **affects language**.
  - Repetition, disfluency, paraphrasing (Guinn and Habash, 2012).
  - Can be used for **diagnosis** (Fraser, Meltzer, and Rudzicz., 2015).

Communication difficulties and Trouble-Indicating Behaviors



- In **dialogue**, people with AD have more **discourse-related difficulties**, including:
  - inattention,
  - Poor tracking of propositions and themes, and
  - deficits in working memory.

- **Trouble Indicating Behaviors (TIBs)** (Watson, 1999).
  - Difficulties can be **phonological**, morpho/**syntactic**, **semantic** (e.g., lexical access), or **discourse** (e.g., misunderstanding topic).
  - **Seniors with AD use TIBs significantly more** ($p<0.005$) than matched controls (Watson, 1999).

- ***What are these TIBs?***

Some common TIBs

1. **Neutral or non-specific requests for repetition (local)**.
   E.g., ***What? Huh?***

2. **Request for confirmation – repetition with reduction**.
   E.g., Speaker 1: *I went to the museum last night.*
   Speaker 2: ***Last night?***

…

Some common TIBs *(cont.)*

…

8. **Lack of uptake / lack of continuation**.
   Include i) *minimal feedback* indicating non-understanding,
   ii) *lack of contribution* to topic extension;
   iii) overriding/*interrupting*; and
   iv) abrupt *switch of topic*.

   E.g., Speaker 1: *Do you know what 'rhetorical' means?*
   Speaker 2: *Yes.*
   Speaker 1: *What?*
   Speaker 2: ***Oh, its a bit too hard, bit late too late to.***

Some common TIBs *(cont.)*

…

11. **Reprise / minimal dysfluency.**
    Reprises: partial or whole repetition/revision.
    Minimal dysfluencies: sound, syllable, or word repetition, pauses, and fillers.
        E.g., ***Eerrr, I want to we went to*** *the river.*

How do people avoid TIBs?

- ED should mimic **verbal strategies** of caregivers.
  - E.g., (Wilson et al., 2012) :
    1. **Speak slowly**.
    2. **Repeat** misunderstood prompts **verbatim**.
    3. Ask **closed-ended questions** (i.e., eliciting yes/no responses).
    4. Simplify sentences using **reduced syntactic complexity**.
    5. Give one question or **one direction at a time**.
    6. **Use pronouns minimally**.

How can we mimic this in a robot?
How will people with AD respond?

Data

- 10 individuals (6 female) with AD recruited at Toronto Rehab.
  - Age: 77.8 years ($\sigma = 9.8$)
  - Education: 13.8 years ($\sigma = 2.7$)
  - MMSE: 20.8/30 ($\sigma = 5.5$)
- Three phases:
  - **Familiar human-human** dyad (during informed consent),
  - **Human-robot** dyad (during *tea-making*), and
  - **Unfamiliar human-human** dyad (during post-study interview).

Speech interface

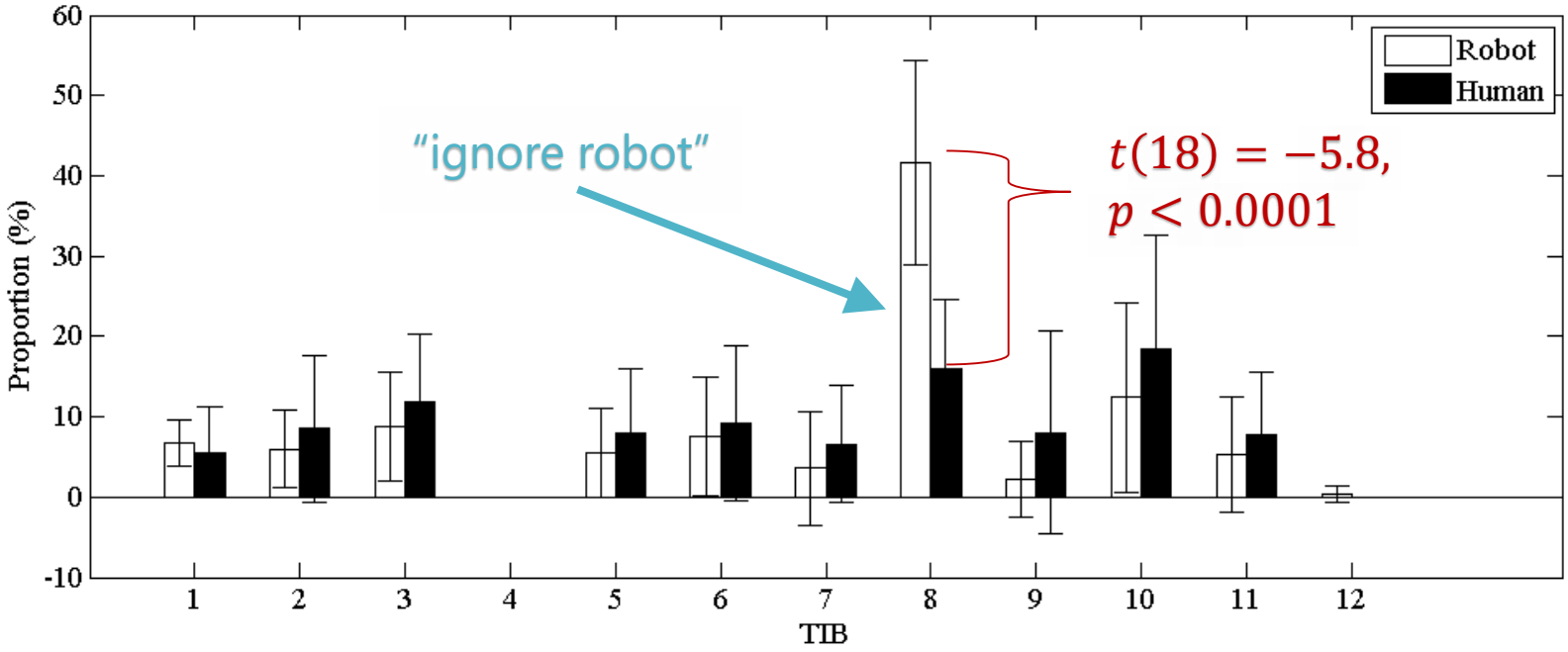
- **Synthetic speech**: 'David' voice from Cepstral.
  - Qualitative feedback was **positive**; 
  - Despite being '**robotic**', the voice was '***clear***' and '***confident***'.

- We **split** the tea-making task into **phases**.
  - **(1)** *go to kitchen*, ..., **(6)** *put teabag in cup*, ...
  - We recorded audio (+video) prompts for **each phase**, at several **levels of detail**.

- A human navigator followed a **flowchart** of **scripts**.
  - **Respond to questions** with **pre-recorded prompts**;
  - When possible, engage in **novel social conversation**.

Language use and interaction

- A speech-language pathologist (SLP) **transcribed** all of the data and **annotated** TIBs.
  - For sanity, a second SLP annotated 20%; Fleiss' $\kappa = 0.84$.



Understanding each other

- People with AD were **much *more* likely**(*) to have ***no* TIB** when interacting with a **robot** (18.1%) than with a non-familiar **human** (6.7%).

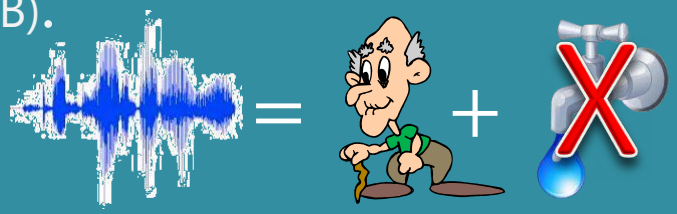
- But it's not *really* interacting with a robot, is it?
  - A human is recognizing the speech.
  - A human is recovering from errors.
  - A human is choosing what to say next (albeit with a script).

(*) $t(18) = -4.78$, $p < 0.0001$

Speech recognition and automation

- We developed methods that **automatically identify** TIBs in speech with **>80%** accuracy (Rudzicz et al., 2014).
  - Indicative features are mostly things like **skewness of the derivatives** of particular Mel-frequency cepstral coefficients, but some have more clinical value e.g., **phonation rate**.

- **ASR** is a standard HMM with mixtures of Gaussians.
  - Data are very noisy (SNR $[-3.42..8.14]$ dB).
    - *LSAE spectral noise subtraction*

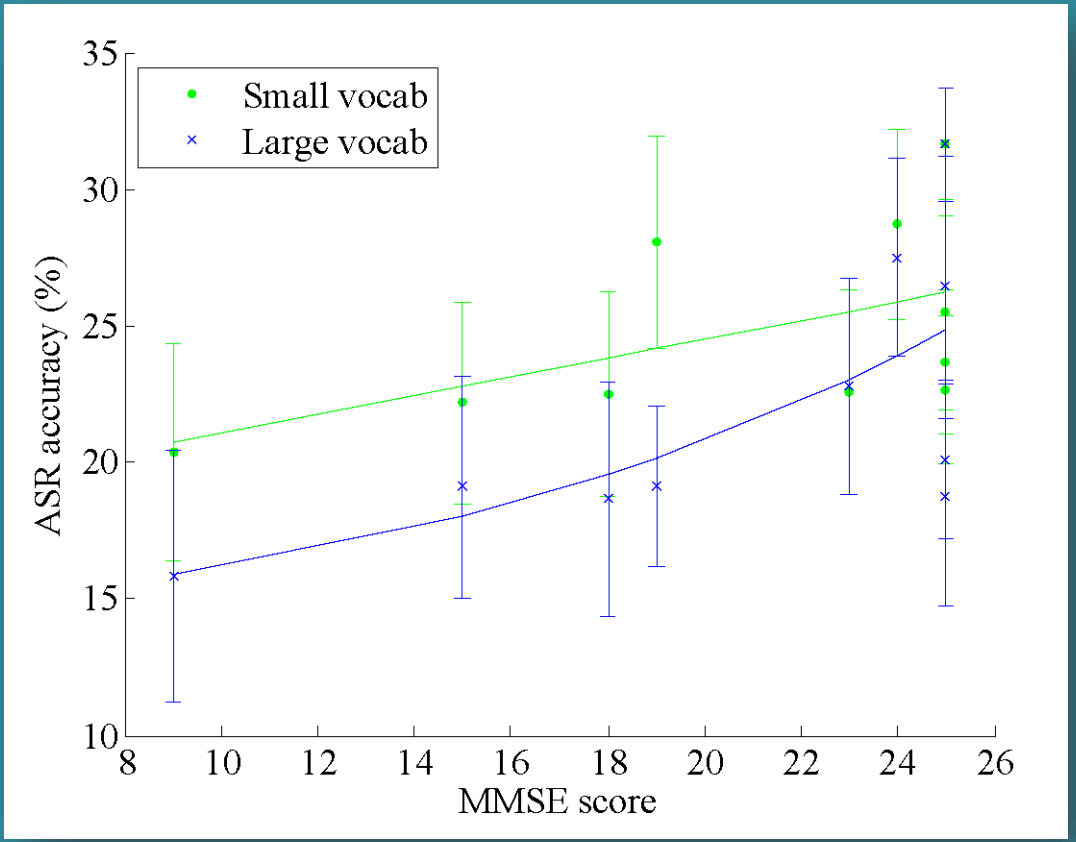


  - Two LMs derived from English Gigaword corpus:
    - *Large: 64,000 words* *Small: 5000 words*

Speech recognition and cognition

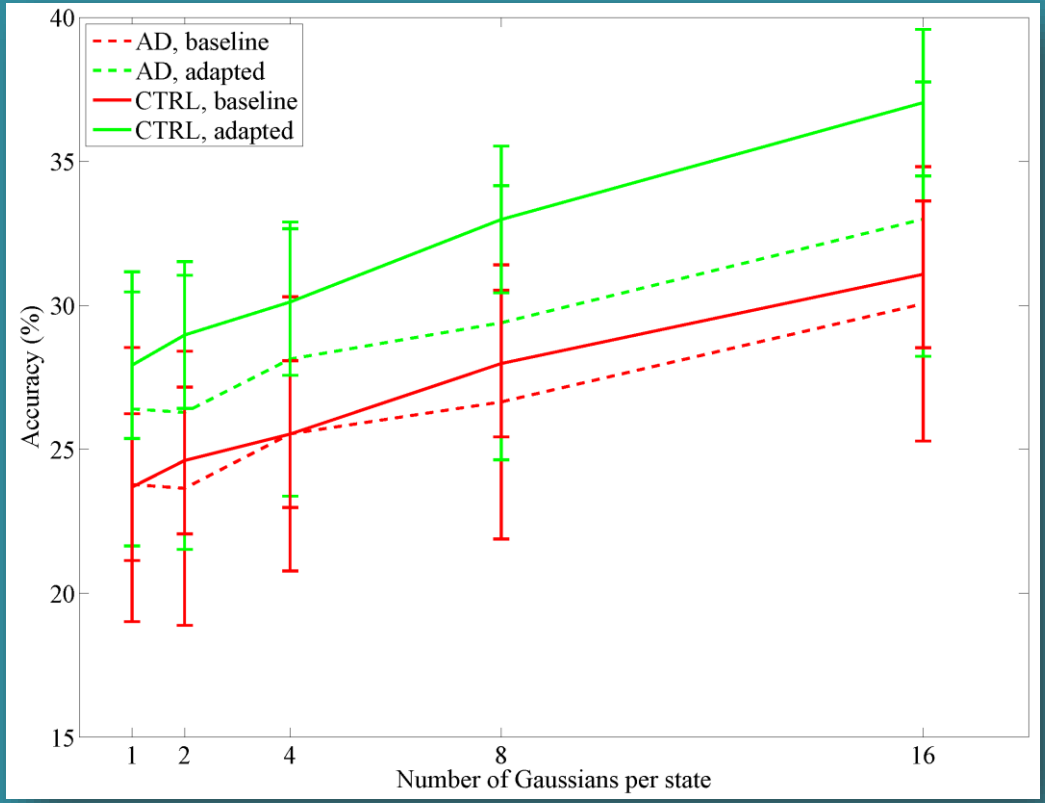
- Clear increases in accuracy with MMSE.

ANOVA: $F_1 = 47.07, p = 0.164$.



Adapting ASR to older voices

- We adapted ASR using data from DementiaBank and Carolina Conversations, and varied model complexity.

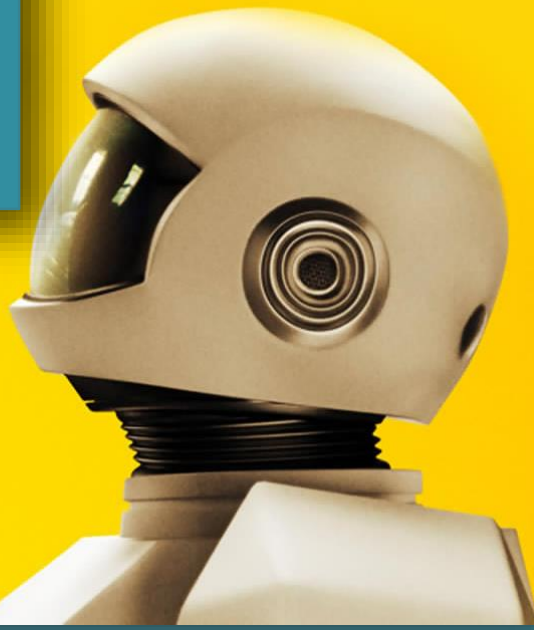


Automating choice of response

Silicon friends for golden years

Speech is increasingly important for interaction.

Our robot friends will need to be sensitive to differences in language as we age.



Special thanks: Raibul Huq & Colin Harry (robot builders),
Jen Boger & Goldie Nejat (study design).

NSERC CRSNG AGEWELL NUANCE Société Alzheimer Society CANADA

SLPAT

- Joint **Special Interest Group** of
  the Association for Computational Linguistics (**ACL**) &
  the International Speech Communications Association (**ISCA**)
- **S**peech and **L**anguage **P**rocessing for **A**ssistive **T**echnologies.
  - *Yearly workshops (next: w/ Interspeech in SanFran).*
  - *Recent special issue of TACCESS.*
  - *Possible Jelinek JHU workshop.*

www.slpat.org