A Fully Annotated Corpus for Studying the Effect of Cognitive Ageing on Users' Interactions with Spoken Dialogue Systems

Kallirroi Georgila(1), Maria Wolters(1), Vasilis Karaiskos(1), Melissa Kronenthal(1), Robert Logie(2), Neil Mayo(1), Johanna D. Moore(1), Matt Watson(3)

(1) Human Communication Research Centre, University of Edinburgh
(2) Human Cognitive Neuroscience-Psychology, University of Edinburgh
(3) Psychology, University of Sunderland

Thanks to: The Scottish Funding Council

Research Questions

- How does cognitive aging affect users' interaction with spoken dialogue systems?
- How can we design dialogue systems to adapt to
  - range of cognitive abilities
  - older people's interaction styles
- In particular, which dialogue strategies lead to more successful/efficient/satisfying interactions?

MATCH Corpus

- Designed to enable in-depth analyses of older users' interaction with SDSs
- Includes
  - 447 dialogues with older and younger users interacting with a WoZ appointment scheduling system
  - comprehensive range of cognitive measures for each participant
  - extensive user satisfaction assessments for each dialogue
  - user responses to recall question

Cognitive Assessments

- Participants assessed for
  - Fluid intelligence (linked to abstract reasoning, Raven's Progressive Matrices)
  - Crystallised intelligence (linked to acquired knowledge, Mill Hill vocabulary test)
  - Speed of information processing (Digit/Symbol Substitution)
  - Working memory capacity

Dialogue Strategies

- Manipulated two aspects of dialogue policy:
  - the number of options that users were presented with (one option, two options, four options)
  - the confirmation strategy employed (explicit confirmation, implicit confirmation, no confirmation)
- The combination of these 3 X 3 design choices yielded 9 different dialogue policies (or systems)
- Participants interact with Wizard-of-Oz (WoZ) system that implements the 9 policies

Appointment Scheduling Task

- Users asked to book health-care appointments by providing:
  - health care professional (e.g., diabetes nurse)
  - half-day (e.g., Monday morning)
  - time slot (e.g., 10:30 a.m.)
- After each scheduling dialogue, user
  - filled out a 39-item questionnaire
    - perceived task completion
    - overall impression
    - user satisfaction
  - responded to question asking them to recall 4 items
    - health professional, day, time, location

WoZ System Design

- The human wizard took over the function of:
  - speech recognition
  - language understanding
  - dialogue management
- Natural language generation using templates
- Text-to-speech using Cerevoice synthesiser (Aylett et al., 2006), which has been shown to be intelligible to older users (Wolters et al., 2007)

Presentation of Options

- 1 Option (Yes/No)
  - *Would you like to see the occupational therapist?*
- 2 Options
  - *Would you like to see the occupational therapist or the community nurse?*
- 4 Options
  - *Would you like to see the occupational therapist, the community nurse, the physiotherapist or the diabetes nurse?*

Confirmation Strategies

- Explicit:
  - User: I would like to see the occupational therapist please.
  - Sys: *You would like to see the occupational therapist. Is this correct?*
- Implicit:
  - User: I would like to see the occupational therapist please.
  - Sys: *When would you like to see the occupational therapist, on Monday afternoon or on Friday morning?*
- No confirmation:
  - User: I would like to see the occupational therapist please.
  - Sys: *When would you like to come, on Monday afternoon or on Friday morning?*

Example Dialogue (2 options, explicit confirmation)

System: *Good morning, this is your appointment scheduling system. Would you like to see the community nurse or the physiotherapist?*

User: The physiotherapist, please, and good morning to you.

System: *You would like to see the physiotherapist. Is that correct?*

User: That is correct, the physiotherapist, thank you.

System: *On which day would you like to come, Tuesday afternoon or Thursday afternoon?*

User: Tuesday afternoon is fine and Thursday afternoon is fine, you can just give me a time.

System: *You would like to make an appointment on Tuesday afternoon. Is that correct?*

User: That is correct. Anytime between two and four o'clock would be fine.

System: *At what time would you like to make an appointment, 2:00 p.m or 3:00 p.m?*

User: Three P.M.

MATCH Corpus

- Dialogues from
  - 26 older users (50-85)
  - 24 younger users (20-30)
- Total of 447 dialogues
  - 232 with older users
  - 215 with younger users
- Includes each user's
  - Results on cognitive assessments
  - Responses to user satisfaction surveys
  - Responses to recall question

Participant Statistics

	Demographics			Cognition			
	#Users	Age	%Female	MillHill	DSST	Ravens	WMS
Younger	24	22 ± 3	71%	42	75	54	37
Older	26	66 ± 9	61.5%	52	51	49	28
Δ Sig.	n/a	**	n.s	**	**	**	.

Participant statistics .: p<0.05, **: p<0.001 or better

Corpus Annotation

- All dialogues were:
  - recorded digitally with a sampling frequency of 48 kHz
  - transcribed orthographically by a human annotator with the tool Transcriber following the guidelines of the AMI meeting corpus (Carletta, 2007)
  - semi-automatically annotated with dialogue acts and dialogue context (Georgila et al., 2005; Georgila et al., 2008)
  - hand-corrected by human annotators
  - all transcriptions and annotations are stored in NXT format (Carletta et al., 2003)

Dialogue Act Annotations

Includes speech act and task, e.g.,

Sys: *Would you like to come on Monday afternoon?*
User: Monday afternoon please but not at two, better at four.

Dialogue Act:

[accept_halfday, social_polite, block_slot, provide_slot]

Speech Act:

[accept_info, social, block_info, provide_info]

Task:

[halfday, polite, slot, slot]

3 annotators annotated the same 36 dialogues (18 older, 18 younger, 4 for each dialogue system)

- Kappa = 0.82

Example User Speech Acts

- **Accepting/Rejecting System Suggestions**
  - accept_info (accept option suggested by the system)
  - confirm_pos (user confirms an option when asked for confirmation)
- **Correcting System / Indicating Misunderstandings**
  - correct_info (user corrects previously provided information)
  - request_info (user requests help, clarification, repetition)
- **Taking Initiative**
  - provide_info (user provides information about possible options)
  - reprovide_info (user provides information again in the same utterance or turn)
- **Social Interaction with the System**
  - acknowledgement (user shows that s/he understands system)
  - social (e.g. 'goodbye', 'thank you')

Information State Annotations

Each dialogue annotated as a sequence of Information States (Larsson and Traum, 2000).

Each IS includes the following fields:

- **Dialogue Level:** Speaker, TurnNumber, UtteranceNumber, DialogueAct, SpeechAct, UserInput, SystemOutput
- **Task Level:** Task, FilledSlot, FilledSlotValue, BlockedSlot, BlockedSlotValue, ConfirmedSlot, GroundedSlot
- **Low Level:** Segmentation
- **History Level:** FilledSlotsStatus, FilledSlotsValuesStatus, BlockedSlotsStatus, BlockedSlotsValuesStatus, ConfirmedSlotsStatus, GroundedSlotsStatus, DialogueActsHist, SpeechActsHist, TasksHist, FilledSlotsHist, FilledSlotsValuesHist, BlockedSlotsHist, BlockedSlotsValuesHist, ConfirmedSlotsHist, GroundedSlotsHist

Example Information State (not all fields are displayed)

DIALOGUE LEVEL
TurnNumber: 4
Speaker: user
DialogueAct: [accept_halfday,social_polite,block_slot,provide_slot]
SpeechAct: [accept_info,social,block_info,provide_info]
UserInput: Monday afternoon please but not at two better at four.

TASK LEVEL
Task: [halfday,polite,slot,slot]
FilledSlot: [halfday,slot]
FilledSlotValue: [monday pm,four pm]
BlockedSlot: [slot]
BlockedSlotValue: [two pm]
GroundedSlot: [hp]

LOW LEVEL
Segmentation: [monday afternoon],[please],[but not at two],[better at four]

HISTORY LEVEL
FilledSlotsStatus: [hp],[halfday],[slot]
FilledSlotsValuesStatus: [physiotherapist],[monday pm],[four pm]
BlockedSlotsStatus: [slot]
BlockedSlotsValuesStatus: [two pm]
GroundedSlotsStatus: [hp]
DialogueActsHist: greeting,suggest_hp_2,[accept_hp,social_polite],
suggest_halfday_2_implicit,[accept_halfday,social_polite,block_slot,provide_slot]

Differences between User Groups

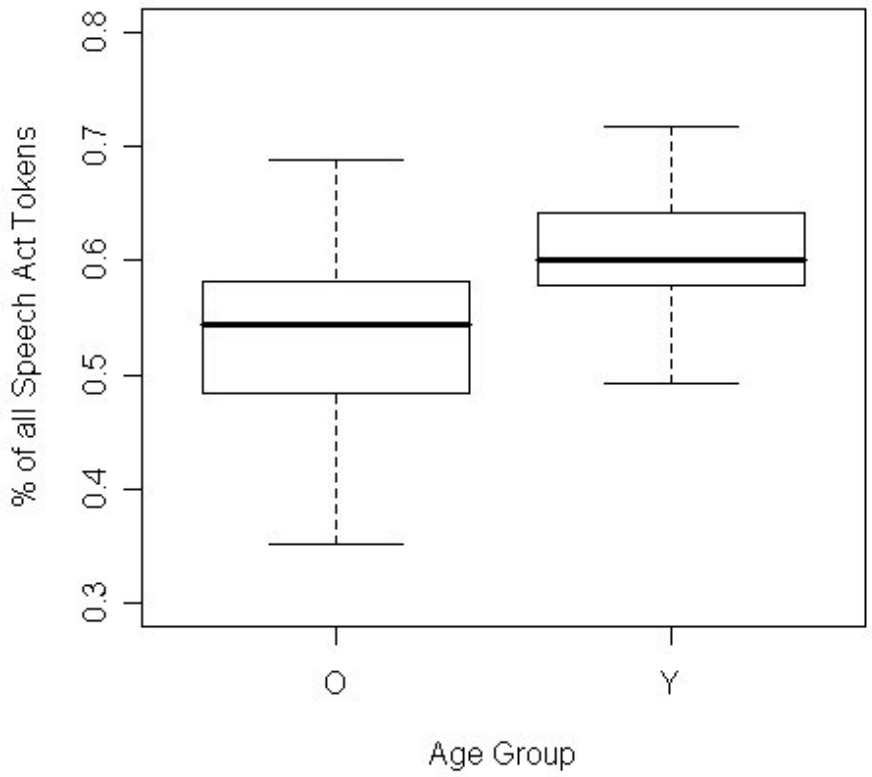
- No effect of age on task success
  - older users perform as well as younger users, no matter what the dialogue strategy (Wolters et al., submitted)
- No effect of cognitive ability levels on task success (Wolters et al., submitted)
- Strong effect of age on efficiency
  - older users less efficient than younger users (Wolters et al., submitted)
- Strong effect of age on interaction style (this presentation)

Overall Dialogue Level Differences

	Older	Younger	Sig.
# Turns	79	59	***
# Word Types	81	30	***
# Word Tokens	312	102	***
# Speech Act Types	14	9	***
# Speech Act Tokens	126	73	***

Counts summed over all dialogues and divided by the number of users (***: $p<0.0001$ or better)

Relative Frequency of 3 Most Frequent Speech Acts



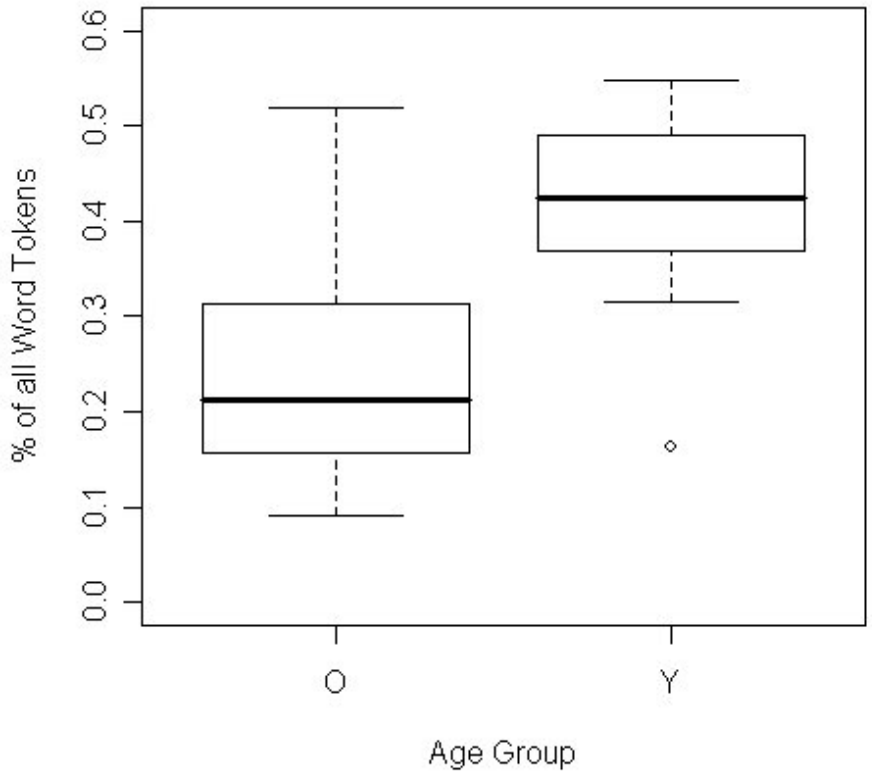
Older participants use wider range of speech acts

Differences in Relative Frequencies of Speech Acts

Speech Act	Older	Younger	Sig.
Accept_*	22.1	**32.1**	***
Confirm_*	28.3	**41.5**	***
Social	**17.9**	5.3	***
Acknowledge	**0.8**	0.0	***
Provide_*	**7.8**	3.4	*
Reprovide_*	1.8	0.2	**
Block	0.5	0.0	*
Garbage	**3.2**	0.5	***

*: $p<0.01$; **: $p<0.001$; ***: $p<0.0001$ or better

Relative Frequency of 3 Most Frequent Words



Older participants use a richer vocabulary

Differences in Relative Frequencies of Lexical Categories

Lexical Category	Older	Younger	Sig.
YesNo	13.0	33.8	***
PosNeg	4.1	1.8	*
SocWords	7.7	3.6	*
Thanks	2.3	0.3	***
Bye	1.2	0.2	***
Please	4.0	2.9	n.s.
Sorry	0.2	0.1	n.s.

*:p<0.01, **:p<0.001, ***:p<0.0001 or better

Summary

- MATCH corpus is a collection of fully annotated interactions of older and younger users with 9 WoZ dialogue systems in the appointment-scheduling domain
- The corpus is useful for:
  - investigating how older users interact with dialogue systems (Wolters et al., 2008)
  - assessing the impact of cognitive ageing on human-machine interaction
  - learning dialogue management strategies
  - creating realistic user simulations (Georgila et al., 2008)
  - adapting speech recognisers to older voices
- Corpus available at end of project (Dec 2009)