COMMUNICATION SCIENCES & DISORDERS

Original Article

| ISSN 2288-0917 (Online) | Commun Sci Disord 2026;31(2):401-413

https://doi.org/10.12963/csd.260185

Comparing Discourse Production and Conversational Engagement in Older Adults across Robot- and Tablet-Assisted Conversational Agents

Yae Rin Yoo[a,b], Hyunseo Ryu[b], ChangHwan Kim[b], Jee Eun Sung[a,c,d], Yoonseob Lim[b]

[a]*Department of Communication Disorders, Ewha Womans University, Seoul, Korea*
[b]*Intelligence and Interaction Research Center, Korea Institute of Science and Technology, Seoul, Korea*
[c]*Department of Brain and Cognitive Sciences, Ewha Womans University, Seoul, Korea*
[d]*Ewha Brain Institute, Ewha Womans University, Seoul, Korea*

Correspondence: Jee Eun Sung, Ph.D.
Department of Communication Disorders, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea
Tel: +82-2-3277-2208
Fax: +82-2-3277-2122
E-mail: jeesung@ewha.ac.kr

Co-Correspondence: Yoonseob Lim, Ph.D.
Korea Institute of Science and Technology (KIST), 5 Hwarang-ro 14-gil, Seongbuk-gu, Seoul 02792, Korea
Tel: +82-2-958-6641
Fax: +82-2-958-6641
E-mail: yslim@kist.re.kr

Received: December 30, 2025
Revised: March 5, 2026
Accepted: March 5, 2026

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT (RS-2022-NR070151, RS-2024-00461617), by the Ewha Global Excellence Program (EGEP), and by the Korea Institute of Science and Technology (KIST) Institutional Program under Grant 2E3384.

Objectives: Discourse assessments are widely used in clinical settings to identify cognitive and linguistic impairments associated with normal aging and neurogenic communication disorders. Because richer spontaneous speech facilitates more reliable discourse analysis, effective methods for eliciting naturalistic language are needed. As the social roles of robots expand, their potential as tools for language and cognitive assessment has gained increasing attention. This study examined whether a conversational robot elicits greater discourse production and conversational engagement in older adults compared to a tablet and explored the clinical usability of robots as tools for discourse-based assessment. **Methods:** Fourteen healthy older adults participated in unstructured conversations on three daily topics with both a robot and a tablet. The robot incorporated non-verbal backchannel behaviors, such as nodding and facial expressions, and interacted with older adults using the ChatGPT model, while the tablet offered only verbal communication. Five linguistic measures were analyzed, including turn-taking frequency, total and mean utterances, and total and mean speech duration. Following each condition, participants completed a self-report survey assessing usability and interaction experience. **Results:** Results showed that older adults produced significantly more total and mean utterances when interacting with the robot. Additionally, participants reported a greater sense of encouragement to engage with the robot. **Conclusion:** These findings suggest that the robot's physical embodiment and interactive behaviors enhance conversational engagement, making it a clinically promising tool for eliciting spontaneous speech and collecting discourse data for language and cognitive assessment in aging populations.

Keywords: Human-robot interaction, Linguistic measures, Older adults, Open conversation, Robot, Tablet

The growing aging population has led to increased interest in the linguistic and cognitive declines that can be caused by normal aging or neurodegenerative diseases such as Alzheimer's disease (AD) or mild cognitive impairment (MCI) as a transitional stage to AD. The Mini-Mental State Examination (MMSE) (Folstein et al., 1975) and Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) are commonly utilized paper-based assessments to detect cognitive decline. In addition to the traditional paper-based evaluations, many researchers have collected patients' spontaneous speech and analyzed linguistic and/or acoustic features to detect early symptoms of cognitive decline (Chang et al., 2022; Huang et al., 2024; Khodabakhsh et al., 2015; Pastoriza-Domínguez et al., 2022; Skirrow et al., 2022; Tóth et al., 2018; Yamada et al., 2020; Yoshii et al., 2023). For example, Khodabakhsh et al.

(2015) examined 20 prosodic and 18 linguistic features from spontaneous speech and found significant differences in variables, such as the response time or duration of silent periods between normal older adults and people with AD. Tóth et al. (2018) also reported significant differences in the duration of speech periods, silent periods, and fillers in spontaneous speech between healthy adults and people with MCI. Similarly, Huang et al. (2024) conducted an automatic speech analysis in three groups of community-dwelling older adults in Shanghai (healthy controls, MCI, and AD). During a *Cookie Theft* picture description task, both acoustic features (e.g., pitch, jitter, shimmer, formants) and linguistic features (e.g., part-of-speech distributions, type–token ratio, counts of information words and units) were extracted from participants' spontaneous speech. They compared the screening performance of acoustic, linguistic, and combined feature sets using machine learning classifiers. Their results showed that linguistic features derived from discourse production contributed significantly to the automated discrimination of cognitive status, supporting the utility of speech-based automated approaches for detecting cognitive decline in older adults.

To identify sensitive linguistic markers of cognitive decline, it is essential to elicit a sufficient amount of spontaneous speech, as longer speech samples provide more reliable and informative linguistic and acoustic cues than shorter ones (Petti et al., 2023; Segkouli et al., 2025). Petti et al. (2023) demonstrated that sample length in spontaneous speech significantly affects the extraction and longitudinal tracking of language features, with longer speech yielding more stable measures of cognitive-related change. In addition, recent reviews indicate that continuous parameters of spontaneous speech—such as silence duration and speech segment frequency—are sensitive to early cognitive change, further supporting the need for sufficiently rich speech samples in analysis (Segkouli et al., 2025).

To obtain sufficiently rich discourse data, the selection of the task is an essential methodological consideration. Previous studies have commonly collected spontaneous speech using discourse tasks such as story retelling (Taler et al., 2021), picture description (Ambadi et al., 2021), and conversational interaction (Dijkstra et al., 2004). Although each task type offers distinct advantages, conversational speech is generally regarded as the most naturalistic form of discourse, as it encourages extended, self-initiated language production that closely resembles everyday communication. Consequently, conversational discourse has been shown to provide particularly valuable insights into the relationship between cognitive functioning and both linguistic and interactional features of spontaneous speech (Dodge et al., 2015).

Although conversational discourse tasks are well-suited for eliciting naturalistic and sufficiently rich speech and would be highly valuable for repeated monitoring of cognitive–linguistic changes, they are difficult to implement frequently in real-world clinical settings. Collecting discourse data typically requires in-person administration by trained examiners, scheduled appointments, and travel to clinical facilities. These requirements can impose substantial burdens on older adults, including time constraints, financial costs, psychological stress, and physical mobility limitations, which may ultimately reduce participation and limit the feasibility of repeated assessments (Ferrar et al., 2021; Skirrow et al., 2022; Woods et al., 2015). However, more frequent and regular assessments are essential for the early detection of cognitive decline, particularly before noticeable memory problems begin to interfere with daily functioning (Lyons et al., 2015).

To address these limitations and increase accessibility, researchers have begun exploring automated approaches to discourse-based assessment. Recently, researchers have adopted an artificial intelligence-based dialogue system as the agent of conversation instead of human examiners, conducting conversations with a virtual agent on a computer (Hall et al., 2019) or a tablet (Tanaka et al., 2017). While these screen-based or voice-only systems increase accessibility, they lack physical embodiment and dynamic nonverbal feedback, which are important components of natural face-to-face interaction. In contrast, robots that more closely resemble human conversational dynamics are being employed as conversational partners for older adults and people with neurodegenerative diseases (Chang et al., 2022; Khoo et al., 2023; Miller & McDaniel, 2022; Schüssler et al., 2021; Seaborn et al., 2022; Yoshii et al., 2021). Several unique features of robots are known to have a positive impact on the overall interaction experience with a human. The human-like appearance and physical embodiment of robots contribute to regarding them as real communication partners and feeling a sense of familiarity (Fischer et al., 2012). Addi-

https://doi.org/10.12963/csd.260185

tionally, the facial expressions and movements of the robot help subjects focus more on the conversation (Rosenthal-von der Pütten et al., 2018). The backchannel (BC) behaviors, such as nodding or blinking in response to a speaker, also may lead users to perceive that the robot understands or attends to the conversation (Seok et al., 2023). Despite the potential of robots, several studies have demonstrated the effectiveness of tablets and computers for language assessment (Dougherty et al., 2010; Trustram & de Jager, 2014). However, there are no direct comparisons between robots and tablets in linguistic analysis other than for posing healthcare-related questions (Mann et al., 2015).

The purpose of this study was to examine whether an embodied conversational robot with dynamic BC behaviors elicits differences in objective linguistic measures of conversation compared to a tablet-based agent. In addition, the study aimed to determine whether the robot can elicit significantly more speech from older adults compared to a tablet, thereby assessing its potential as an alternative to human examiners in language assessments. Accordingly, five linguistic measures were used to quantify the engagement of interactions and the discourse productivity: turn-taking frequency, two utterance-related measures (total number of utterances and utterances per turn), and two response duration measures (total duration of the user's speech and mean speech duration per turn). Furthermore, a self-report survey on usability and interaction was conducted to compare older adults' subjective perceptions of their conversational experiences with the robot versus the tablet.

METHODS

Conversational Robot with Backchannel Behavior

This study used MyBom Mini, developed by Roaigen[1], for social interaction at home. The robot has one speaker, neck motors (2 Degrees of Freedom), a motor control board, and a PC with an i5-8259U processor and 8GB of memory.

The robot system is implemented in Python using ROS (Robot Operating System) on Linux 16.04. It incorporates OpenAI APIs for speech recognition and generation, including Whisper, GPT-3.5-Turbo, and TTS (Text-to-speech). As shown in Figure 1, the system captures user speech via a microphone, converts it to text using the STT (Speech-to-text) module, and generates responses with the GPT-3.5-Turbo model from ChatGPT. We prompted the model to engage in small talk to establish rapport and foster a sense of intimacy, using a friendly, soft, and polite tone throughout the open and continuous conversations with older adults. The response text is then synthesized into speech via the TTS module. After the robot speaks, the system automatically resumes recording the user's speech.

The behavior manager module determines the robot's gesture and its corresponding timing. Each gesture is the combination of

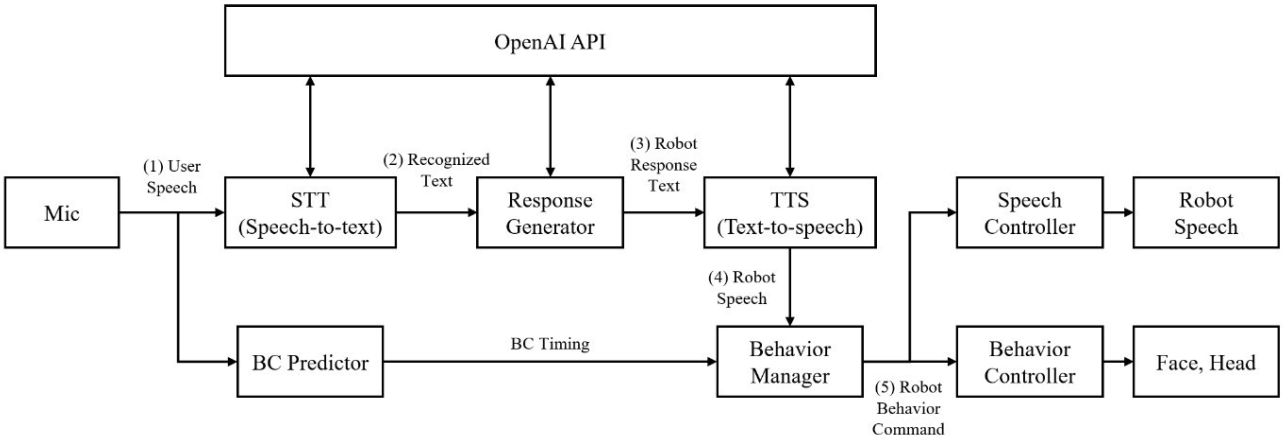

Figure 1. Architecture of the robot system.
BC = backchannel.

[1] http://roaigen.com

an eye-blinking face and a head movement, which is randomly selected by the behavior manager. 10 different head movements, including actions such as moving the robot's head to look around and nodding, were pre-designed using 3ds Max. Throughout the robot's speech, the behavior manager module continuously selects a new gesture and its corresponding timing, with random intervals ranging from 4 to 7 seconds. The gestures with nodding movements are used as BC behaviors during the user's speech. We adopted the optimal BC timing prediction model from Seok et al. (2023), which was developed using human-human interaction data. Based on this model, the BC predictor module analyzes acoustic features in real-time, allowing the behavior manager to instruct the robot to nod appropriately during user speech.

Participants

Fourteen healthy older adults (60–73 years old) participated in this study (Table 1). All participants satisfied the criteria including 1) native Korean speakers with no report of neurologic or psychiatric disorders based on the Health Screening Questionnaire (Christensen et al., 1991), 2) the normal range of the Korean Mini-Mental State Examination (>16%ile) (Kang, 2006), and 3) the normal range (>16%ile) of the Seoul Verbal Learning Test (SVLT) from the Seoul Neuropsychological Screening Battery, 2nd Edition (SNSB-II) (Kang et al., 2012). The study was approved by the Institutional Review Board (IRB) of the Korea Institute of Science and Technology (Approval No. KIST-202209-HR-001).

Experiment

In this study, participants had free conversations with two agents, a tablet and a robot (Figure 2). A tablet (Apple iPad) and a Bluetooth speaker (BOSE) were adopted for the tablet condition. The tablet showed only a still image of the robot's face. The Bluetooth speaker delivered the robot's speech during the tablet condition. A wireless microphone was used to record the subject's speech to improve the quality of the speech recognition.

The study was conducted using a within-subjects design. The order of each condition was counterbalanced, and the participants conducted the experiment twice with at least a 14 day interval and a different agent form each day.

The experiment was divided into two parts: a practice session and a main session. During the practice session, participants engaged in conversations on two topics with the agent to become accustomed to turn-taking interactions. The topics of the practice session were today's weather and food.

The main session covered three different topics of conversation. Each topic of conversation started with a question initiated by the agent. The three questions were: *1) If you had a week of free time, where would you go and what would you do? 2) Would you tell me about your most unforgettable travel experience? 3) Could you tell me the most memorable gift that you have received and the reason for it?* The order of questions was randomized for each participant. Except for the agent's first speech, all the other agents' speeches were generated by ChatGPT. Participants were allowed to have multiple turns within a single conversation topic if the total length of the conversation did not exceed the token limit of ChatGPT. The end of the participant's speech at each turn was manually determined by the experimenter and the robot was triggered to generate the next speech. Unlike younger adults, older adults exhibit distinct speech patterns, such as extended pauses between utterances and the occasional addition of remarks after seemingly completing a statement. To account for these speech

Table 1. Demographic information of participants

Group	Age (SD) (year)	Education (SD) (year)	Gender (Male : Female)
Old (N = 14)	66.07 (4.37)	14.28 (3.0)	4 : 10

Values are presented as mean (SD).
SD = Standard deviation.

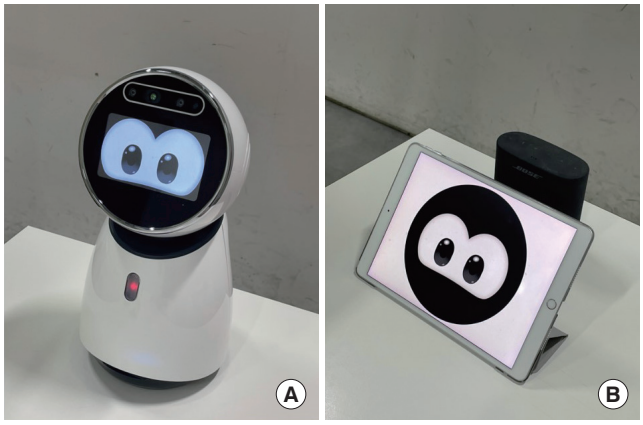

Figure 2. Example of agents setting in the experiment environment. A robot 'MyBom Mini' (A) and a tablet with a Bluetooth speaker (B).

https://doi.org/10.12963/csd.260185

characteristics and avoid premature interruptions, we implemented manual intervention in the conversation. A licensed speech-language pathologist with clinical experience in discourse analysis identified the end of each speaking turn by carefully assessing the context of the utterances and allowing a 5-second pause after the final spoken word. Importantly, participants were unaware of any researcher involvement during the interactions, as the entire experiment was conducted under a Wizard of Oz setup. Both participants and the conversational agents were placed separately from the researchers throughout the experiment. We also instructed participants to say, *"Let's move on to the next question/topic."* to end each topic of conversation if there was nothing more to say about the topic. Upon receiving the request, the experimenter controlled the robot to move on to the next conversation topic.

Measurements

Recordings and transcriptions of the entire experiment were automatically collected. All the conversational linguistic measures were analyzed by a speech-language pathologist with Praat (version 6.4.05). The study employed five conversational linguistic measures to analyze interactional engagement and the productivity of spontaneous speech: the frequency of turn-taking (FTT), total utterances (TU), mean utterances (MU), total speech duration (TSD), and mean speech duration (MSD) (Table 2).

FTT reflects the engagement and dynamism of conversational exchange and represents the total number of speaking turns taken by both the user and the agent. A *'turn'* is defined as a sequence of conversational exchanges. For example, if person A speaks, person B responds, and person A speaks again, the FTT is three.

Two utterance-related measures (TU and MU) capture the extent of the user's discourse elaboration and speech output. TU represents the total number of utterances produced by the participant, whereas MU indicates the average number of utterances per speaking turn. Utterance segmentation followed the criteria of Kim et al. (1998), dividing utterances based on four markers: sentence-closing endings, conjunctions, pauses exceeding two seconds, and sudden changes in intonation.

In addition, two response duration measures (TSD and MSD) provide temporal indices of speech sustainability and response expansion of the user. TSD represents the total speaking time (in seconds) produced by the user across the conversation, and MSD indicates the average duration of speech per speaking turn. All linguistic measures were averaged across the three conversation topics. TSD and MSD measure the total speaking time (seconds) of the user across the conversation and the average duration of the user's speech per speaking turn, respectively. All the linguistic measures were the average values across the three conversation topics.

After the experiment, a self-report survey was conducted in each agent condition, with the questions modified from previous studies (Khodabakhsh et al., 2015; Seaborn et al., 2022) (Table 3). The survey consisted of a total of seven questions, including prior experience with AI devices, usability, and the quality of the interaction (naturalness, enjoyment, encouragement, friendliness, and meaningfulness). Except for the question about prior experience with AI devices, all questions were on a 5-point Likert scale. To

Table 3. List of questions of self-report survey

Section	Question
Prior experience with AI device	Have you ever had a conversation with an AI device?
Usability	Do you want to have a conversation with this device again?
Interaction	
Naturalness	Do you think the conversation was natural?
Enjoyment	Do you feel fun while talking to this device?
Encouragement	Do you think the device encouraged you to engage in the whole conversation?
Friendliness	Do you consider the device as your friend?
Meaningfulness	Do you think the conversation was meaningful?

Table 2. Conversational linguistic measures and its description

Conversational linguistic measures	Description
Frequency of turn-taking	Sum of participants' speaking turns and agent's.
Total number of utterances	Sum of utterance numbers appeared in participants speaking turns.
Number of utterances by each turn	Total number of utterances / the frequency of participant's turn-taking
Total duration of user's speech (seconds)	Sum of length of time from the start to the end of a participant's speech during a turn.
Mean speech duration for each turn (seconds)	Total response duration / the frequency of participant's turn-taking

evaluate subjective opinions on each device, the participants were asked the following question at the end of the experiment: *"If you could bring only one device to your home so that you could have a conversation with it at any time, which one would you choose and why?"*

Statistical Analysis

The Wilcoxon signed-rank test was conducted to confirm the differences in conversational linguistic measures and survey score results between the two agent conditions using IBM SPSS 29.0.

RESULTS

Differences in Conversational Linguistic Measures between the Robot and the Tablet

The overall paired Wilcoxon signed-ranks test results are shown in Figure 3. There was no significant difference in the FTT between the robot (M=13.12, SD=3.80) and tablet (M=11.98, SD=4.71) conditions (Z=-.973, p=.331). However, participants demonstrated a significantly higher TU in the robot condition (M=24.62, SD=14.44) than in the tablet condition (M=17.55, SD=10.05) (Z=-2.229, p=.026). Participants also generated significantly greater MU in the robot condition (M=3.80, SD=2.11) than in the tablet condition (M=3.04, SD=1.53) (Z=-2.480, p=.013). The TSD was marginally different for each condition (Z=-1.915, p=.056). For the MSD, there was no significant difference between the robot and tablet conditions (Z=-1.412, p=.158). An example of the actual conversation between a participant and each type of agent (tablet and robot) is provided in a supplementary file and by video links for each interaction (https://url.kr/e6mifp).

Differences in Usability and Interaction Survey Results between the Robot and the Tablet

The usability survey results indicated no significant difference

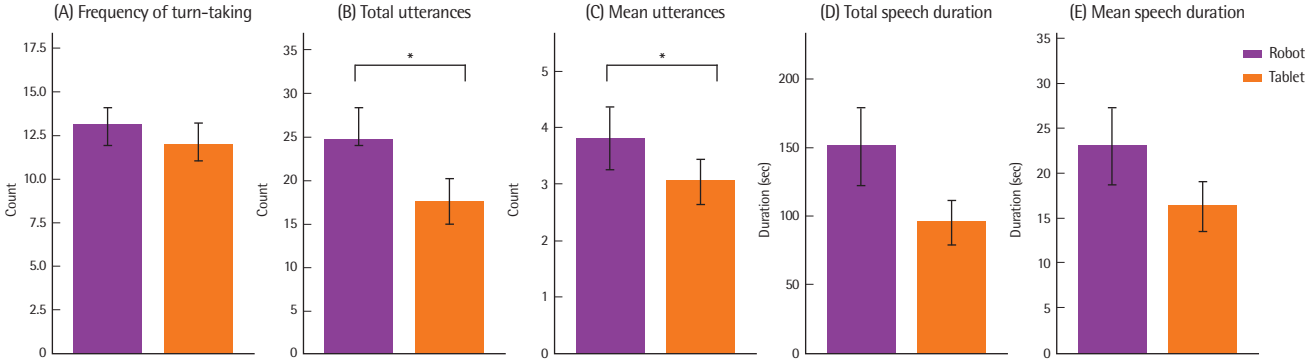

Figure 3. Differences in conversational linguistic measures between robot and tablet agent conditions.
Mean values and standard errors are presented.
*p<.05.

Table 4. Survey results of conversational experience with the robot and the tablet

Section	Question	The robot condition	The tablet condition
Prior experience with AI device	Have you ever had a conversation with an AI device?	Yes : No=10 : 4	
Usability	Do you want to have a conversation with this device again?	3.92 (0.829)	3.78 (1.122)
Interaction			
Naturalness	Do you think the conversation was natural?	3.78 (0.892)	3.43 (1.016)
Enjoyment	Do you enjoy talking to this device?	3.78 (0.975)	3.64 (1.081)
Encouragement*	Do you think the device encouraged you to engage in the whole conversation?	4.07 (0.730)	3.78 (0.975)
Friendliness	Do you consider the device as your friend?	3.71 (1.069)	3.5 (1.092)
Meaningfulness	Do you think the conversation was meaningful?	3.64 (1.216)	3.36 (1.151)

Values are presented as mean (SD).
SD=Standard deviation.
*p<.05.

https://doi.org/10.12963/csd.260185

between the two agents (Z = -.632, p = .527), suggesting that participants found the tablet and the robot to be equally usable during the conversation. Among the five interaction-related survey items (naturalness, enjoyment, encouragement, friendliness, and meaningfulness), a significant difference between the two conditions was found only for perceived encouragement, whereas the remaining items did not differ significantly (Table 4). Regarding the naturalness (Z = -1.311, p = .190) and enjoyment (Z = -.707, p = .480) of the interaction, there was no significant difference between the two agents. However, in the questionnaire assessing whether the agent motivated participants to engage in the conversation, the result showed that the robot significantly encouraged more interaction than the tablet (Z = -2.000, p = .046). The participants' responses for the last two questions on the friendliness of the agent (Z = -1.134, p = .257) and the meaningfulness of the conversation (Z = -1.633, p = .102) also confirmed there were no significant differences between the two agent conditions. For the question of choosing one device to bring home and having a conversation with it at any time, thirteen out of fourteen participants selected the robot. Most participants explained that the reasons for their choice were that the robot's appearance was more familiar than the tablet, and they also preferred the robot's nodding gesture as they perceived it as attentive listening, creating a sense of conversing with a close friend.

DISCUSSION

This study examined whether a robot could elicit more conversation and engagement from healthy older adults compared to tablets. Results showed that participants produced significantly more TU and more MU in the robot condition, with marginally longer TSD. As collecting sufficient spontaneous speech data is a critical prerequisite for identifying sensitive linguistic markers, these results suggest that the robot may serve as a more effective tool than a tablet for administering discourse-based language assessments—potentially replacing the role of a human examiner.

These findings support and extend prior research demonstrating the utility of automated cognitive assessments, including the use of socially interactive robots as platforms for administering language and cognitive tasks (Chang et al., 2022; Skirrow et al., 2022). For example, although not implemented via a robot, Skirrow et al. (2022) demonstrated the feasibility of daily, self-administered automated story recall tasks using participants' own smart devices over a period of 7–8 days. The protocol included 18 short and 18 long stories, and participants' retellings were automatically transcribed and scored using text-similarity metrics that quantified overlap between the source texts and recall responses to derive a generalized match score. Their findings showed that automated story recall measures were sensitive to subtle episodic memory impairments, successfully differentiating individuals with MCI and AD from cognitively healthy older adults. Furthermore, Chang et al. (2022) conducted social robot *RoBoHoN*-administered cognitive tasks to assess verbal fluency, episodic memory, prospective memory, and other aspects of executive functions (e.g., working memory, inhibition, and shifting). The robot autonomously delivered standardized verbal instructions, presented stimuli, and recorded participants' spoken responses across multiple brief cognitive tasks, including animal category fluency, backward digit span, word-list learning with delayed recall and recognition, an event-based prospective memory task, and a modified color–word interference task. They found that the social robot successfully distinguished between healthy older adults and those with MCI, demonstrating greater sensitivity than the MMSE in detecting subtle cognitive differences, thereby highlighting the potential of robot-administered assessments as an alternative to traditional paper-based tools.

Prior research has demonstrated that socially interactive robots can effectively administer structured cognitive tasks and detect subtle cognitive differences in older adults. However, discourse-based linguistic and cognitive assessments require sufficiently rich spontaneous speech to enable reliable analysis of language features. In this regard, the present study extends existing work by examining the role of an embodied and dynamically responsive robot in an open-ended conversational context. The finding that older adults produced more utterances in the robot condition may provide empirical evidence supporting the potential of robots as effective tools for discourse-based language and cognitive assessment.

Beyond linguistic measures, survey responses indicated older adults' clear preference for the robot over the tablet. Participants reported significantly higher encouragement scores in the robot

condition, and 92.8% selected the robot for future interactions. Notably, the encouragement item specifically assessed the extent to which the device facilitated participation in the interaction, rather than the perceived quality of the conversational content itself. This preference was attributed to the robot's familiar, human-like appearance and its BC behaviors, such as nodding and blinking, which may enhance perceived social presence and engagement during interaction. These findings are consistent with previous studies showing that socially expressive robots facilitate more positive and meaningful conversational experiences for older adults (Miller & McDaniel, 2022; Seaborn et al., 2022). For example, Miller and McDaniel (2022) reported that older adults enjoyed, engaged in, and found unstructured conversations with the social robot *Misty* meaningful. Similarly, Seaborn et al. (2022) found that older adults rated the usability and overall experience of a physically embodied robot more positively than a voice-only AI speaker in group conversations. Notably, we also observed that participants openly expressed emotions and discussed personal challenges with the robot, including concerns about retirement, family issues, and the loss of a loved one—topics often difficult to broach.

While the findings offer promising insights, this study also has several limitations. The small sample size and relatively demanding protocol (two sessions over two weeks, each lasting over 50 minutes) may reduce the generalizability of the results, as older adults may face practical barriers to participation. In addition, some participants perceived the agent's response time to be longer than expected, largely due to response generation processing delays from ChatGPT and network instability. Occasional errors from GPT's well-known hallucination issues may have disrupted the conversational flow. Future work should explore more accessible study designs for older adults, incorporate technical improvements to reduce response delays by leveraging OpenAI's streaming service, and consider several strategies to mitigate hallucinations in large language models (Chen et al., 2023; Sun et al., 2023).

Another consideration relates to the characteristics of the participant sample. The participants in the present study had a relatively high level of education (mean years of education = 14.28), which may have influenced their familiarity and comfort with digital devices such as tablets and robots. Higher educational attainment has been associated with greater exposure to and adaptability toward novel technologies, potentially facilitating smoother interaction and conversational engagement with digital agents. Therefore, future research should recruit participants with more diverse educational levels and from varied regional or socioeconomic backgrounds to further examine the generalizability of the results.

In addition, although participant inclusion criteria involved screening with the Health Screening Questionnaire (Christensen et al., 1991), and all participants reported no history of neurologic or psychiatric disorders, such as psychiatric hospitalization or medication use, the present study did not directly employ a standardized measure to quantify depressive symptoms. Given the well-documented association between depression and cognitive function in older adults, subtle depressive symptoms may have influenced cognitive or conversational performance. Future studies should incorporate validated depression scales, such as the Short Form of the Geriatric Depression Scale (SGDS; Kang et al., 2012), to more rigorously control for the potential impact of depressive symptoms on discourse production and cognitive assessment outcomes.

Moreover, the present study focused primarily on quantitative conversational measures to ensure objective comparison across agents. As a result, qualitative aspects of discourse—such as thematic coherence, narrative structure, lexical richness, and pragmatic appropriateness—were not examined. Although quantitative indices provide important information about engagement and productivity, qualitative analyses may offer deeper insights into how conversational agents influence the content and organization of spontaneous speech. Future studies should incorporate qualitative and content-based analyses to further investigate differences in discourse quality across interaction agents.

In summary, our findings suggest that older adults engage more actively with the robot and show a clear preference for it over the tablet, highlighting the potential of robots as effective tools for discourse assessment and as complementary alternatives to human examiners. To expand the application of robots in automated assessment systems, future research should examine their validity in remote settings compared to human examiners, and extend investigations to clinical populations such as individuals with MCI or

 https://doi.org/10.12963/csd.260185

AD. In addition, a qualitative analysis of sensitive linguistic markers across various discourse tasks is needed to support their integration into reliable automated scoring systems.

REFERENCES

Ambadi, P. S., Basche, K., Koscik, R. L., Berisha, V., Liss, J. M., & Mueller, K. D. (2021). Spatio-semantic graphs from picture description: applications to detection of cognitive impairment. *Frontiers in Neurology, 12*, 795374.

Chang, Y. L., Luo, D. H., Huang, T. R., Goh, J. O., Yeh, S. L., & Fu, L. C. (2022). Identifying mild cognitive impairment by using human-robot interactions. *Journal of Alzheimer's Disease, 85*(3), 1129-1142.

Chen, Y., Fu, Q., Yuan, Y., Wen, Z., Fan, G., Liu, D., ..., & Xiao, Y. (2023). Hallucination detection: robustly discerning reliable answers in large language models. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), Association for Computing Machinery,* 245-255.

Christensen, K. J., Multhaup, K. S., Nordstrom, S., & Voss, K. (1991). A cognitive battery for dementia: development and measurement characteristics. *Psychological Assessment, 3*(2), 168-174.

Dijkstra, K., Bourgeois, M. S., Allen, R. S., & Burgio, L. D. (2004). Conversational coherence: discourse analysis of older adults with and without dementia. *Journal of Neurolinguistics, 17*(4), 263-283.

Dodge, H., Mattek, N., Gregor, M., Bowman, M., Seelye, A., Ybarra, O., ..., & Kaye, J. A. (2015). Social markers of mild cognitive impairment: proportion of word counts in free conversational speech. *Current Alzheimer Research, 12*(6), 513-519.

Dougherty, J. H., Cannon, R. L., Nicholas, C. R., Hall, L., Hare, F., Carr, E., ..., & Arunthamakun, J. (2010) The Computerized Self Test (CST): an interactive, internet accessible cognitive screening test for dementia. *Journal of Alzheimer's Disease, 20*(1), 185-195.

Ferrar, J., Griffith, G. J., Skirrow, C., Cashdollar, N., Taptiklis, N., Dobson, J., ..., & Munafò, M. R. (2021). Developing digital tools for remote clinical research: how to evaluate the validity and practicality of active assessments in field settings. *Journal of Medical Internet Research, 23*(6), e26004.

Fischer, K., Lohan, K. S., & Foth, K. A. (2012). Levels of embodiment: linguistic analyses of factors influencing HRI. *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, 463-470.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Hall, A. O., Shinkawa, K., Kosugi, A., Takase, T., Kobayashi, M., Nishimura, M., ..., & Yamada, Y. (2019). Using tablet-based assessment to characterize speech for individuals with dementia and mild cognitive impairment: preliminary results. *AMIA Joint Summits on Translational Science Proceedings, 2019*, 34-43.

Huang, L., Yang, H., Che, Y., & Yang, J. (2024). Automatic speech analysis for detecting cognitive decline of older adults. *Frontiers in Public Health, 12*, 1417966.

Kang, Y. (2006). A normative study of the Korean mini-mental state examination (K-MMSE) in the elderly. *Korean Journal of Psychology, 25*(2), 1-12.

Kang, Y., Jang, S., & Na, D. L. (2012). *Seoul neuropsychological screening battery* (2nd ed., SNSB-II). Seoul: Human Brain Research & Consulting Co.

Khodabakhsh, A., Yesil, F., Guner, E., & Demiroglu, C. (2015). Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. *EURASIP Journal on Audio, Speech, & Music Processing, 2015*(1), 9.

Khoo, W., Hsu, L. J., Amon, K. J., Chakilam, P. V., Chen, W. C., Kaufman, Z., ... & Sabanović, S. (2023). Spill the tea: when robot conversation agents support well-being for older adults. *Proceedings of the Companion of the 2023 ACM/IEEE International Conference on Human-robot Interaction,* 178-182.

Kim, H., Kwon, M., Na, D. L., Choi, S. S., & Chung, C. S. (1998) Decision making in fluency measures of aphasic spontaneous speech. *Korean Journal of Communication & Disorders, 3*(1), 5-19.

Lyons, B. E., Austin, D., Seelye, A., Petersen, J., Yeargers, J., Riley, T., ..., & Kaye, J. A. (2015). Pervasive computing technologies to continuously assess Alzheimer's disease progression and intervention efficacy. *Frontiers in Aging Neuroscience, 7*, 102.

Mann, J. A., MacDonald, B. A., Kuo, I. H., Li, X., & Broadbent, E. (2015). People respond better to robots than computer tablets delivering healthcare instructions. *Computers in Human Behavior, 43*, 112-117.

Miller, J., & McDaniel, T. (2022). I enjoyed the chance to meet you and I will always remember you: Healthy Older Adults' Conversations with Misty the Robot. Proceedings of the *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 914-918.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ..., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Jour-*

nal of the American Geriatrics Society, 53(4), 695-699.

Pastoriza-Domínguez, P., Torre, I. G., Diéguez-Vide, F., Gómez-Ruiz, I., Geladó, S., Bello-López, J., ..., & Hernández-Fernández, A. (2022). Speech pause distribution as an early marker for Alzheimer's disease. *Speech Communication, 136*, 107-117.

Petti, U., Baker, S., Korhonen, A., & Robin, J. (2023). How much speech data is needed for tracking language change in Alzheimer's disease? A comparison of random length, 5-min, and 1-min spontaneous speech samples. *Digital Biomarkers, 7*(1), 157-166.

Rosenthal-von der Pütten, A. M., Krämer, N. C., & Herrmann, J. (2018). The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. *International Journal of Social Robotics, 10*(5), 569-582.

Schüssler, S., Zuschnegg, J., Paletta, L., Lodron, G., Steiner, J., Pansy-Resch, S., ..., & Holter, M. (2021). Effects of coach robot pepper versus tablet training on psychosocial and physical outcomes of persons with dementia: a mixed-methods study. *Alzheimer's & Dementia, 17*, e053453.

Seaborn, K., Sekiguchi, T., Tokunaga, S., Miyake, N. P., & Otake-Matsuura, M. (2022). Voice over body? Older adults' reactions to robot and voice assistant facilitators of group conversation. *International Journal of Social Robotics, 15*(2), 143-163.

Segkouli, S., Gkioka, M., Kokkas, S., Votis, K., Valero, S., Miguel, A., ..., & Manias, G. (2025). Linguistic markers in spontaneous speech: insights into subjective cognitive decline. *Healthcare, 13*(22), 2888.

Seok, S., Jeon, T. H., Chae, Y. J., Kim, C., & Lim, Y. (2023). Explorative study on the non-verbal backchannel prediction model for human-robot interaction. *Proceedings of the 15th International Conference on Social Robotics (ICSR)*, 264-275.

Skirrow, C., Meszaros, M., Meepegama, U., Lenain, R., Papp, K. V., Weston, J., & Fristed, E. (2022). Validation of a remote and fully automated story recall task to assess for early cognitive impairment in older adults: longitudinal case-control observational Study. *JMIR Aging, 5*(3), e37090.

Sun, W., Shi, Z., Gao, S., Ren, P., de Rijke, M., & Ren, Z. (2023). Contrastive learning reduces hallucination in conversations. *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, 13618-13626.

Taler, V., Davidson, P. S., Sheppard, C., & Gardiner, J. (2021). A discourse-theoretic approach to story recall in aging and mild cognitive impairment. Aging, *Neuropsychology, & Cognition, 28*(5), 762-780.

Tanaka, H., Adachi, H., Ukita, N., Ikeda, M., Kazui, H., Kudo, T., & Nakamura, S. (2017). Detecting dementia through interactive computer avatars. *IEEE Journal of Translational Engineering in Health & Medicine, 5*, 1-11.

Tóth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatlóczki, G., Bánréti, Z., ..., & Kálmán, J. (2018). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research, 15*(2), 130-138.

Trustram Eve, C., & de Jager, C. A. (2014). Piloting and validation of a novel self-administered online cognitive screening tool in normal older persons: the cognitive function test. *International Journal of Geriatric Psychiatry, 29*(2), 198-206.

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the internet: a tutorial review. *PeerJ, 3*, e1058.

Yamada, Y., Shinkawa, K., & Shimmei, K. (2020). Atypical repetition in daily conversation on different days for detecting Alzheimer's disease: evaluation of phone-call data from regular monitoring service. *JMIR Mental Health, 7*(1), e16790.

Yoshii, K., Nishimura, M., Kimura, D., Kosugi, A., Shinkawa, K., Takase, T., ..., & Higashi, S. (2021). A study for detecting mild cognitive impairment by analyzing conversations with humanoid robots. *Proceedings of the 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, 347-350.

Yoshii, K., Kimura, D., Kosugi, A., Shinkawa, K., Takase, T., Kobayashi, M., ..., & Nishimura, M. (2023). Screening of mild cognitive impairment through conversations with humanoid robots: exploratory pilot study. *JMIR Formative Research, 7*(1), e42792.

 https://doi.org/10.12963/csd.260185

Appendix 1. An example of the actual conversation between a participant and the robot

Speaker	Transcription
Robot	Prioritizing safety is great! Could you share the safety precautions you took while trekking and especially, how you handled risky situations? As mentioned earlier, it could be beneficial for children!
Participant	Risky situations, well, when crossing a stream with strong currents, never cross alone. Even if I do, I wait and use a rope to cross together when the next hiker arrives. When walking through deep snow, I wait until some of it has melted before proceeding. It's considered safe to wait until conditions improve.
Robot	Yes, being cautious in risky situations is crucial. Collaborating with other hikers to cross or waiting for safer conditions is a wise choice. Your emphasis on safety is truly commendable! Always prioritize safety on your future travels. Where would you like to travel next?
Participant	I achieved the dream of many long-distance hikers worldwide, the Triple Crown in the U.S. But back then, it was a rushed experience. Now, I want to have some time and take it easy, and I want to try again.
Robot	Wow, the Triple Crown is an incredible achievement! Going back to the trail and taking it slow sounds like a great plan. Any specific trail you want to walk slowly again, or maybe explore a different place? What is the most important thing you consider while traveling?
Participant	Regardless, safety comes first. Even in the Himalayas, they say the ultimate summit is coming back alive. I always prioritize the utmost safety.
Robot	Emphasizing safety is really good. Whether trekking in the Himalayas or any other travel, always prioritizing safety is key. Your safety-conscious mindset can positively influence other travelers. When you embark on future treks, keep prioritizing safety, and I hope your travels are always enjoyable. Do you have any other travel-related questions? I might have helpful information.
Participant	By the way, I cycled across the U.S. round trip. Now, I want to cycle from the southern to the northern tip of Japan. I'm curious about the expenses involved and how long it would take.
Robot	Oh, cycling round trip across the U.S. is impressive! Now, planning to cycle through Japan from south to north sounds fascinating. The expenses and duration of travel stay can vary based on factors like budget, travel route, and resting places. I recommend considering these factors while planning. Generally, it might take a few weeks to a month to complete the journey. You can check travel information websites or the official Japan tourism website for more detailed information. If you face any difficulties in planning, feel free to ask me. I'll be happy to help!

Appendix 2. An example of the actual conversation between a participant and the tablet

Speaker	Transcription
Tablet	Wow, those are really amazing travel destinations! Can you share one of the most special and enjoyable memories among them?
Participant	It was interesting... Because most of them were unfamiliar to me, each one was interesting, and they remain in my memories.
Tablet	I see! Traveling to unfamiliar places for the first time is always special and enjoyable. So, what kind of travel plans do you have next? Where would you like to go?
Participant	Right now, at the top of my bucket list, it's sailing around the world on a yacht.
Tablet	Wow, that's a really fantastic dream! While sailing around the world, are there any countries or cities you definitely want to visit? Where would you like to go?
Participant	I don't have any specific ones.
Tablet	That's understandable, especially if it's your first time. Traveling is about discovering and experiencing new places continuously. So, what type of travel would you like to try next? For example, cultural trips or adventures in nature.
Participant	Adventures in nature.
Tablet	Ah, adventures in nature are a fascinating choice! Where in nature would you like to have an adventure? What activities do you want to do? For example, hiking, trekking, or sailing.

https://doi.org/10.12963/csd.260185

국문초록

대화 매체에 따른 노년층의 담화 산출 및 대화 참여도 비교 연구: 로봇과 태블릿을 중심으로

유예린[1,2] · **유현서**[2] · **김창환**[2] · **성지은**[1,3,4] · **임윤섭**[2]

[1]이화여자대학교 언어병리학과, [2]한국과학기술연구원 지능·인터랙션연구센터, [3]이화여자대학교 뇌인지과학부, [4]이화여자대학교 뇌융합과학연구원

배경 및 목적: 담화의 수집 및 평가는 정상 노화뿐 아니라 신경언어장애군의 인지 및 언어 결함을 파악하기 위해 임상 현장에서 빈번하게 이루어지며, 대상자의 자발화가 많이 나타날수록 담화 분석이 용이하게 이루어진다. 본 연구는 대화 매체에 따라 노년층의 담화 산출과 대화 참여도에 차이가 나타나는지를 검증하고, 로봇 매체의 임상 평가 및 자료 수집 도구로서의 활용 가능성을 태블릿과 비교하여 살펴보고자 하였다. 방법: 노년층 14명이 로봇 및 태블릿 조건에서 일상 주제에 대한 자유 대화를 실시하였다. 로봇은 백채널 행동을 통해 청자 반응을 구현할 수 있으며, 목 움직임이나 표정 변화가 포함된 상호작용을 제공하였다. 태블릿 조건에서는 음성 기반 상호작용이 주로 이루어졌다. 담화는 대화 차례 빈도, 발화 수, 발화 시간을 포함한 5개의 언어적 지표로 분석되었고, 대화 종료 후 상호작용 경험에 대한 주관적 사용성 설문을 실시하였다. 결과: 태블릿에 비해 로봇 조건에서 총 발화 수와 대화 차례 당 평균 발화 수가 유의하게 많았으며, 대상자들은 태블릿보다 로봇 매체가 대화 참여를 유도하는데 효과적이었다고 평가하였다. 논의 및 결론: 로봇의 신체화와 비언어적 상호작용 행동은 노년층의 담화 산출과 대화 참여를 촉진하는 요인으로 작용하였으며, 로봇 매체는 담화 기반 언어·인지 평가를 위한 임상적 도구로 활용될 가능성이 있음을 시사한다.

핵심어: 로봇, 태블릿, 노년층, 담화

본 연구는 한국과학기술연구원 기관고유사업(No. 2E33841), 과학기술정보통신부의 정부 재원으로 수행된 한국연구재단(NRF)의 연구과제(RS-2022-NR070151, RS-2024-00461617) 및 Ewha Global Excellence Program (EGEP) 연구과제의 지원을 받아 수행되었습니다.

참고문헌

강연욱 (2006). K-MMSE (Korean-mini mental state examination)의 노인 규준 연구. **한국심리학회지: 일반, 25**(2), 1-12.

강연욱, 장승민, 나덕렬 (2012). **서울신경심리검사 2판.** 서울: 휴브알엔씨.

김향희, 권미선, 나덕렬, 최상숙, 이광호, 정진상 (1998). 실어증환자 자발화의 유창성 연구. **언어청각장애연구, 3**(1), 5-19.

ORCID

유예린(제1저자, 대학원생, 학생연구원 https://orcid.org/0009-0005-6501-1864); **유현서**(공동저자, 인턴 https://orcid.org/0009-0006-8748-5865);
김창환(공동저자, 책임연구원 https://orcid.org/0000-0002-2443-0174); **성지은**(교신저자, 교수 https://orcid.org/0000-0002-1734-0058);
임윤섭(교신저자, 책임연구원 https://orcid.org/0000-0003-4754-6038)