

Original Paper

# Effect of Speech Recognition on Problem Solving and Recall in Consumer Digital Health Tasks: Controlled Laboratory Experiment

Jessica Chen, BEng, MBIomedEng; David Lyell, PhD; Liliana Laranjo, MD, MPH, PhD; Farah Magrabi, BEng, PhD  
Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, North Ryde, Australia

**Corresponding Author:**

Jessica Chen, BEng, MBIomedEng  
Centre for Health Informatics  
Australian Institute of Health Innovation  
Macquarie University  
Level 6  
75 Talavera Rd  
North Ryde, NSW 2113  
Australia  
Phone: 61 2 9850 2400  
Email: [jessica.a.chen@mq.edu.au](mailto:jessica.a.chen@mq.edu.au)

## Abstract

**Background:** Recent advances in natural language processing and artificial intelligence have led to widespread adoption of speech recognition technologies. In consumer health applications, speech recognition is usually applied to support interactions with conversational agents for data collection, decision support, and patient monitoring. However, little is known about the use of speech recognition in consumer health applications and few studies have evaluated the efficacy of conversational agents in the hands of consumers. In other consumer-facing tools, cognitive load has been observed to be an important factor affecting the use of speech recognition technologies in tasks involving problem solving and recall. Users find it more difficult to think and speak at the same time when compared to typing, pointing, and clicking. However, the effects of speech recognition on cognitive load when performing health tasks has not yet been explored.

**Objective:** The aim of this study was to evaluate the use of speech recognition for documentation in consumer digital health tasks involving problem solving and recall.

**Methods:** Fifty university staff and students were recruited to undertake four documentation tasks with a simulated conversational agent in a computer laboratory. The tasks varied in complexity determined by the amount of problem solving and recall required (simple and complex) and the input modality (speech recognition vs keyboard and mouse). Cognitive load, task completion time, error rate, and usability were measured.

**Results:** Compared to using a keyboard and mouse, speech recognition significantly increased the cognitive load for complex tasks ( $Z=-4.08$ ,  $P<.001$ ) and simple tasks ( $Z=-2.24$ ,  $P=.03$ ). Complex tasks took significantly longer to complete ( $Z=-2.52$ ,  $P=.01$ ) and speech recognition was found to be overall less usable than a keyboard and mouse ( $Z=-3.30$ ,  $P=.001$ ). However, there was no effect on errors.

**Conclusions:** Use of a keyboard and mouse was preferable to speech recognition for complex tasks involving problem solving and recall. Further studies using a broader variety of consumer digital health tasks of varying complexity are needed to investigate the contexts in which use of speech recognition is most appropriate. The effects of cognitive load on task performance and its significance also need to be investigated.

(*J Med Internet Res* 2020;22(6):e14827) doi: [10.2196/14827](https://doi.org/10.2196/14827)

**KEYWORDS**

speech recognition software; consumer health informatics; ergonomics

## Introduction

Recent advances in natural language processing and artificial intelligence have led to improvements in and widespread adoption of speech recognition technologies [1]. Speech recognition is an input modality that translates human speech into computerized text [2]. In consumer applications, speech recognition is usually applied as a way to interact with conversational agents, which are systems that mimic human conversation using text or spoken language [3,4]. Consumer conversational agents (such as Amazon Alexa and Google Assistant) can improve patient workflow by allowing patients to call nurses [5]. In health care, conversational agents have been utilized for a variety of purposes, including data collection, decision support, and patient monitoring [5-8].

Problems with the use of digital health technology represent a well-documented safety concern in the literature [9,10]. However, little is known about the problems associated with conversational agents that pose actual or potential risks of harm to consumers [9]. One study in which the participants were asked to interact with conversational agents identified significant safety concerns arising from the quality of information provided in response to health-related questions [11]. For example, incorrect information provided by a conversational agent in response to a question about the amount of alcohol that could be consumed while taking oxycodone could lead to severe harm, including death. However, few studies have evaluated conversational agents in the hands of consumers [3,4].

In other consumer-facing tools, cognitive load has been observed to be an important factor affecting the use of speech recognition technologies. Users reported finding it more difficult to think and speak at the same time when compared to typing, pointing, and clicking [12,13]. However, the effects of speech recognition use on cognitive load when performing health tasks has not yet been explored. Cognitive load is the amount of workload imposed on the brain's working or short-term memory, which has limited capacity [14,15] and a short duration [16], particularly when performing tasks requiring problem solving and recall [14,17]. An example of a problem-solving task is using basic arithmetic to calculate nutritional information, whereas recall involves memorizing and reporting exercise information. Due to the cognitive load, certain tasks may be more difficult to perform because speaking shares the same cognitive resources in the brain as those required for problem solving and recall (ie, working memory) [12]. Therefore, a possible challenge with the use of speech recognition is that it can increase the cognitive load in tasks requiring more problem solving and recall.

Although studies in other domains have investigated the effects of speech recognition use on cognitive load, to our knowledge, no study has measured its effects in digital health tasks [12,18,19]. Thus, the aim of the present study was to evaluate the use of speech recognition for documentation in consumer digital health tasks such as recording diet and exercise information in comparison to using a conventional keyboard and mouse. The following hypotheses were tested: (1) cognitive load is higher for speech recognition compared to keyboard and mouse use in complex tasks requiring more problem solving and recall; (2) the percentage of errors is higher for speech recognition compared with keyboard and mouse use in complex tasks requiring more problem solving and recall; (3) task completion time is lower for speech recognition compared with keyboard and mouse use in simple tasks requiring less problem solving and recall; and (4) speech recognition is less usable than a keyboard and mouse for both simple and complex tasks requiring more problem solving and recall.

These findings will shed light on the characteristics of consumer digital health tasks that make them most suitable for using speech recognition as an input modality.

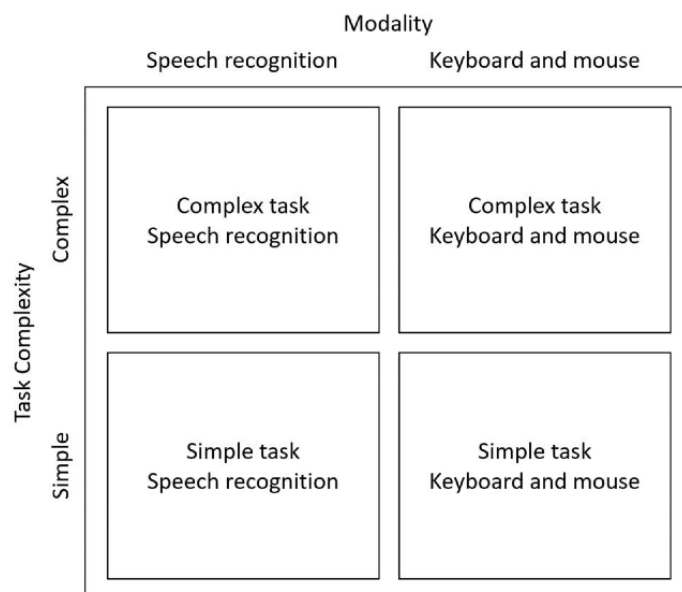
## Methods

### Participants

Fifty-two university students and staff participated in this study (see Results for a summary of the participant demographics). The participants were either students or staff who met the minimum English language proficiency for admission to a university program or workplace. These participants also had working knowledge of computer technology and systems as required for their degree or profession. Hence, there were no requirements regarding level of English language, health literacy, or technology proficiency for inclusion. The participants responded to advertisements sent by email or published in a university newsletter. Consenting adults aged 18 years or older were eligible to participate. Ethical approval was granted by Macquarie University's Human Research Ethics Committee to recruit people within the Faculty of Medicine and Health Sciences. Participants were not offered any remuneration or gifts to incentivize participation.

### Experimental Design and Tasks

The study included two within-subject factors: human-computer interaction modality (speech recognition vs keyboard and mouse) and task complexity (simple vs complex) providing four experimental conditions (Figure 1).

**Figure 1.** Experimental design.

Each participant was asked to complete four consumer digital health tasks to document nutritional and exercise information: two simple and two complex tasks, using speech recognition and a keyboard and mouse to interact with a simulated conversational agent (Figure 1). For each documentation task, participants were asked to adopt a persona within a hypothetical scenario focusing on physical activity and diet, and to answer the conversational agent's questions (see Multimedia Appendix 1). The tasks were designed to assess problem solving and recall using the information provided within the scenario.

Complexity was measured by the number of information items that participants needed to manipulate in working memory. The relationship between the number of information items requiring manipulation in working memory and cognitive load has been well established in the literature [20]; that is, human performance is affected when the cognitive load is too high or exceeds the limits of working memory [21,22]. For simple tasks, the conversational agent displayed the hypothetical scenario on the same screen as the questions. Complex tasks were designed to impose a higher cognitive load by increasing the number of pieces of information needed to perform arithmetic and commit to memory (6 total items for simple tasks vs 17 total items for complex tasks; see Multimedia Appendix 1). Simple tasks required participants to problem solve using 5 items of information, including 5 days, 20 minutes, line dancing, 10:00-10.50 am, whereas complex tasks required participants to recall the same 5 items and were provided with 12 items for problem solving. A higher cognitive load was also imposed by displaying the scenario in a pop-up modal window, requiring

participants to commit key information in the scenario to working memory when transferring between the windows [23]. Participants could not copy and paste their answers.

The tasks were developed in consultation with a health informatics researcher (DL) and a primary care physician (LL). Pilot testing was performed by asking 7 individuals to complete the tasks using a prototype of the system and provide feedback. Any issues were iteratively fixed before the next pilot test. This pilot testing ensured that the prescribed tasks and the system were understandable and functional. Individuals who assisted with pilot testing were excluded from participating in the experiment. The correct answers to tasks were predetermined and validated by the health informatics researcher.

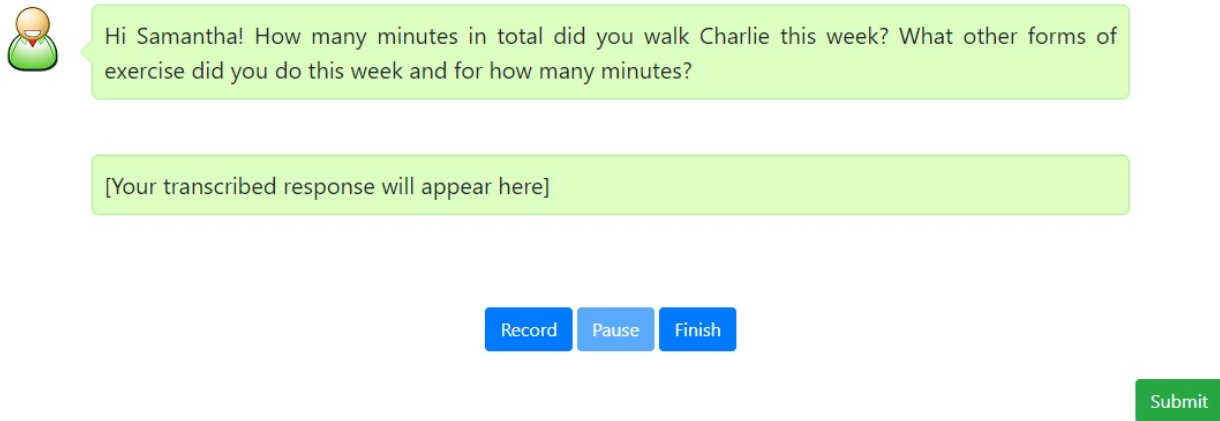
### Simulated Conversational Agent

Participants attended a computer laboratory at the university where a workstation was set up with the tasks running on a web app connected to a keyboard, mouse, and microphone. Presented as a text message conversation, user responses could either be typed using the keyboard and mouse or transcribed using speech recognition depending on the experimental condition (Figure 2). Participants were able to view their responses before submitting. Web Speech API, a JavaScript-based general purpose speech recognition application programming interface, was used to implement a live audio transcription function for this system [24]. To activate speech-to-text using speech recognition, the participants were instructed to press the "Record" button. The system was launched using Google Chrome on a local machine.

**Figure 2.** Graphical user interface of the simulated conversational agent on Google Chrome.

### Persona 1

Samantha is a 68-year-old woman who is 150 cm tall and weighs 48 kgs. She owns a pet dog called Charlie and is retired. She has walked her dog for five days this week for 20 minutes in the morning and 20 minutes in the afternoon. She is also enrolled in a line dancing club that holds classes every Tuesday at 10:00-10:50 am.

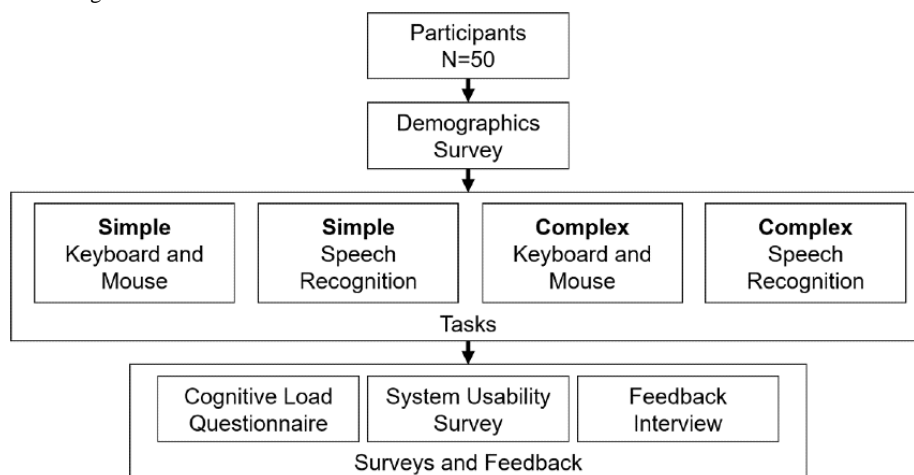


### Procedure

After obtaining informed consent, the participants completed a short survey about use of speech recognition technology and provided demographic information. They were then briefed that the tasks consisted of problem solving and recall and given instructions about how to use the speech recognition interface before commencing the experiment. Training continued until the participant clearly understood all aspects required to perform the experimental tasks; no practice tasks were undertaken.

Participants completed two simple and two complex tasks, half of which were randomly assigned to using speech recognition (Figure 3). At the end of each task, cognitive load was assessed. To avoid order effects, the assignment of simple and complex tasks for the two modalities and task sequence were randomized. All voice recordings for the speech recognition tasks were captured independently to determine errors. At the end of all four tasks, the System Usability Scale (SUS) questionnaire [25] was completed for each modality followed by a feedback interview.

**Figure 3.** Experimental flow diagram.



### Outcome Measures and Analysis

Participants’ responses to the lifestyle management tasks with speech recognition and keyboard and mouse were compared using data extracted from the computer log, audio recordings of participant responses, and paper surveys.

Cognitive load was measured using a cognitive load inventory that was adapted from a validated instrument to reflect the nature of the tasks in the present study [26] (Multimedia Appendix 2).

Self-ratings of using this inventory have shown it to be reliable, unobtrusive, and sensitive to small differences [23]. This inventory has been widely used [23,26,27], including in controlled studies of clinical decision making [28]. The inventory was administered on paper at the end of each condition.

Task completion time was measured in seconds and calculated from computer logs as the difference between the task start and end time.

The error rate for a given task was defined as the number of failed recall and problem-solving responses, calculated as a percentage of the maximum score for the task. Participants were asked to answer each of the conversational agent's questions correctly by recalling information or solving for numerical answers. Each task had predefined correct answers along with numbers associated with correct problem-solving and recall responses (Multimedia Appendix 1). For example, if the correct answers to a task were "line dancing" and "50 minutes," then the maximum score was 2. If a participant incorrectly recalled "line dancing" but correctly answered "50 minutes," then the error rate would be calculated as 50%. To ensure that transcription errors were not mistakenly attributed to participants, the voice recording on each response was used to validate the answers for tasks that were completed with speech recognition.

Usability was assessed using the SUS, a validated 10-item questionnaire [25], resulting in overall usability scores and scores of subelements (usability and learnability).

### Statistical Analyses

The effects of speech recognition on cognitive load, task completion time, error rate, and SUS scores were tested using repeated-measures analysis to control for individual differences. The Wilcoxon signed-rank test was used because the results were not normally distributed. We compared speech recognition and keyboard and mouse across both levels of complexity. Cohen criteria were used to calculate and interpret effect sizes ( $r$ ), where 0.1 indicates a small effect, 0.3 indicates a medium

effect, and 0.5 indicates a large effect [29]. We estimated that a sample size of 42 was required to detect a difference of 25% and a standard deviation of 20% for each outcome measure for all tasks with 90% power and  $P < .05$  [30]. No baseline measures could be derived from the literature. Descriptive statistics were used to summarize demographic information. All statistical analyses were undertaken using SPSS v24.0.0.0 software (IBM Corp, Armonk, NY, USA).

## Results

### Participants

Fifty-two university staff and students participated in the experiment. One participant did not complete the experiment and data from another was lost due to a technical error, leaving a total of 50 participants for inclusion in the analysis. The participants were aged 18-64 years, 30% (15/50) were 25-34 years and 54% (27/50) were women. More than half of the participants (56%, 28/50) reported never using speech recognition in their daily lives, 28% (14/50) reported using it once or twice a week, 14% (7/50) reported using it more than multiple times a week, and 4% (2/50) used speech recognition multiple times a day.

### Effects of Speech Recognition on Cognitive Load

Participants experienced a significantly higher cognitive load when using speech recognition to perform the prescribed tasks (Table 1). These findings were consistent across both levels of task complexity, although the effect size increased from medium for simple tasks to high for complex tasks.

**Table 1.** Comparison of cognitive load between keyboard and mouse and speech recognition use by task complexity.

Task complexity	Keyboard and mouse, median (IQR)	Speech recognition, median (IQR)	Z value <sup>a</sup>	P value	Effect size ( $r$ )
Complex	3.2 (2.1-4.2)	5.2 (3.6-6.3)	-4.22	<.001	0.60
Simple	2.2 (2.2-4.4)	3.3 (2.2-5.6)	-2.24	.02	0.32

<sup>a</sup>Wilcoxon signed-rank test.

### Task Completion Time and Error Rate

Participants took significantly longer to complete tasks using speech recognition than the keyboard and mouse for complex tasks; however, there was no difference observed for simple tasks (Table 2). For complex tasks, there was a statistically significant increase in task completion time with a medium

effect size ( $r=0.36$ ); however, there was no difference for simple tasks.

There was no difference in error rates for both simple and complex tasks (Table 2). For complex tasks, we examined error rates by their type and found no difference for both problem solving ( $Z=-1.96$ ,  $P=.05$ ) and recall error rates ( $Z=-1.55$ ,  $P=.12$ ).

**Table 2.** Comparison of task completion time and error rates between keyboard and mouse and speech recognition use by task complexity.

Task complexity	Keyboard and mouse	Speech recognition	Z value <sup>a</sup>	P value
<b>Complex, median (IQR)</b>				
Completion time (seconds)	162 (124-192)	173 (136-223)	2.52	.01
Error rate (%)	11 (0-25)	11 (0-36)	1.80	.07
<b>Simple, median (IQR)</b>				
Completion time (seconds)	90 (74-124)	83 (68-11)	7.90	.43
Error rate (%)	0 (0-33)	0 (0-33)	-0.33	.74

<sup>a</sup>Wilcoxon signed-rank test.



## Usability and Participant Perceptions About Speech Recognition

Participants found speech recognition to be significantly less

usable than the keyboard and mouse. This was consistent with the SUS factor analysis, which revealed that speech recognition was perceived to be significantly less usable and harder to learn with medium and large effect sizes, respectively (Table 3).

**Table 3.** Usability (SUSa scores) of speech recognition compared to keyboard and mouse.

SUS category	Keyboard and mouse, median (IQR)	Speech recognition, median (IQR)	Z value <sup>b</sup>	P value	Effect size (r)
Overall	85 (72-90)	75 (62-85)	-3.30	.001	0.47
Usability	100 (97-100)	100 (75-100)	-2.98	.003	0.42
Learnability	81 (68-88)	72 (59-84)	-3.54	<.001	0.50

<sup>a</sup>SUS: System Usability Scale.

<sup>b</sup>Wilcoxon signed-rank test.

Overall, participants commented that the simulated conversational agent was “very straightforward” to use. Some reported that not having to think about spelling and grammar was an advantage of speech recognition. Although the participants found speech recognition to be a possibly more convenient way to enter free-text information, many observed that the major pitfalls of speech recognition were any transcription errors generated by the software, an inability to retract and edit answers once a sentence was uttered, and unwanted filler utterances such as “um” and “er” appearing in their responses. Participants also commented about the extra time taken to check and correct the output of speech recognition for transcribing errors when such errors could be prevented by using a keyboard and mouse in the first place. Others raised privacy issues as they were self-conscious about strangers eavesdropping on their responses.

When commenting on difficulties, some participants stated that they had struggled to formulate answers for scenarios and construct sentences to dictate to the conversational agent via speech recognition at the same time. They also commented that this action consumed more time to complete the task. By contrast, typing was perceived to be easier because it was “what we do every day.” This also allowed for using the textbox input to record answers as opposed to “keeping more in the brain” when storing information in cognitive, working memory. Speech was also perceived as a novelty because many were familiar with using a keyboard and mouse and saw it as a standard mode for human-computer interaction.

## Discussion

### Main Findings

When using speech recognition, participants reported a higher cognitive load for both simple and complex tasks compared to using a keyboard and mouse. Some participants reported an inability to think while speaking, which is consistent with previous observations [12,13]. A possible reason is due to an extra step introduced in processing the information and then having to form complete sentences before speaking. This extra processing step, which occurs once participants begin to formulate responses to the simulated conversational agent’s questions, may have involved working memory and contributed

to participants experiencing a higher cognitive load while using speech recognition.

The higher cognitive load also provides a potential explanation for the more time required to complete complex tasks when using speech recognition and for it to be perceived less usable. Participants also noted a time delay required for speech recognition to process responses. It is thus possible that the extra time was introduced by the system and the time participants spent in checking the live transcription provided by speech recognition. If this was true, a consistent increase in time across all conditions would be expected. However, there was no significant time difference found for simple tasks. This could mean that the difference more likely arose from the task characteristics themselves. Further experiments are required to examine the interaction effect of modality and task complexity on cognitive load.

Despite previous reports about higher error rates with speech recognition for clinical documentation tasks [31], we found no difference in error rates between the two modalities for both simple and complex tasks. One possible reason that the second hypothesis was not supported is that most participants were observed to formulate their responses with caution by double-checking their answers when using speech recognition. For simple tasks, this meant looking at the scenario section that was on the same screen. For complex tasks, participants repeatedly clicked the “review scenario” button until they were satisfied with their answer.

Speech recognition was perceived to be less usable than a keyboard and mouse because it was harder to learn. A major factor contributing to this effect is that many participants reported being more comfortable with typing because it was an everyday human-computer interaction, especially in their profession. In addition, 56% of the participants did not use speech recognition in their daily lives, which possibly meant that they are less proficient in using speech recognition. Another possible factor is that the conversational agent was purposely designed to limit participants by not allowing users to correct transcription errors. This may be a common source of frustration that affects perceptions of usability for conversational agents. However, the error rate was unaffected because we used independent voice recordings to determine errors and the impact on usability was captured in the overall score.

Some participants reported lack of privacy as another major factor for speech recognition being less usable than a keyboard and mouse. A major advantage with typing was that a third party could not eavesdrop on interactions with the conversational agent. Thus, less confidence was placed on using the conversational agent with speech recognition compared with a keyboard and mouse. Although there was no difference in error rates, the increased cognitive load and task completion time may also have affected user experience.

### Implications

Our findings suggest that speech recognition may not be uniformly suitable for the different contexts of health care. An important implication is for system designers to consider task characteristics and the resulting impact on cognitive load when selecting modalities for humans to interact with computers. In general, lower cognitive load, fewer errors, less time, and better usability are desirable. Speech recognition may be more suitable for frequent tasks such as documentation of notes in electronic health records (EHRs), which may generally not involve problem solving and recall (eg, when a doctor has already established a diagnosis during the patient consultation and uses the EHR to record their notes). However, it may not be suitable for occasional tasks such as incident reporting, which involves problem solving and recall, requiring clinicians to recollect the sequence of events and identify problems that led to an incident [32]. For such complex documentation tasks, the higher cognitive load and greater time required to use speech recognition, along with the lower usability, suggest that a keyboard and mouse may be a better input modality. By contrast, speech recognition may be necessary for contexts that require use of the hands and eyes.

One practical strategy for designers to assess the suitability of speech recognition as an input modality is to test user interfaces in the prototype stage with the cognitive load inventory, which is readily applicable to different modalities and systems. Importantly, designers need to consider privacy requirements when using speech recognition for busy health care settings, especially when sensitive health information is being handled. Our findings also suggest that privacy requirements may present a barrier for the use of speech recognition in some contexts such as health apps that require users to document their personal health information. These considerations will need to be evaluated on a case-by-case basis.

### Comparison With the Literature

To the best of our knowledge, no previous studies have measured the effects of speech recognition in consumer digital health tasks. In clinical applications, the use of speech recognition for clinical documentation was found to increase error rates and task completion time in a controlled laboratory setting [31,33,34]. One possible reason for the disparity with our results may be the difference in the source of complexity. In the previous studies, complexity was distributed between the clinical scenario and the user interactions with an EHR, which required users to navigate to different sections of the record. By contrast, in our study, the type of user interaction remained the same for the different task types. The complexity instead

arose from the scenario itself, which required users to problem solve and recall information from memory.

### Limitations

There are several limitations to the design of the current study. We focused on the use of speech recognition as an input modality within a computer laboratory, which may not be representative of a real-world setting where environmental factors such as background noise and interruptions affect consumer interactions with digital health technologies. A general-purpose speech recognition engine that was not specifically optimized for the consumer health domain was tested on a desktop computer. This may have affected participant perceptions about usability and time required to use speech recognition. Participants were university students and staff, almost half of whom reported not using speech recognition technology in their daily lives. Therefore, our sample may not be representative of the general population of health consumers who might use conversational agents. However, because the participants were from a cohort that regularly used a keyboard and mouse, we were able to undertake a realistic assessment of the effort to learn and use speech recognition in the context of the conversational agent. For regular speech recognition users, expectations about the robustness and accuracy could have affected SUS scores. Although it is possible that individual differences such as health literacy, native language, pronunciation, fluency, and experience with speech recognition and a keyboard and mouse may have impacted the outcome variables, we attempted to control for these differences by using a within-subjects design. Further studies are needed to explore the influence of health and nutritional literacy. It is also possible that the quality of the speech recognition could have affected task completion time, but the effect would be consistent across experimental conditions. The error rate was unaffected because we used independent voice recordings to determine correct answers; therefore, the risk of the speech recognition mistranscribing speech by recording responses was controlled.

Despite these limitations, this study has contributed evidence related to the use of speech recognition as an input modality in human-computer interaction, particularly in a consumer digital health context. These results provide baseline measures of cognitive load in using speech recognition. Further studies using a more representative population of conversational agent users are needed to investigate the effects of cognitive load on task performance when speech recognition is integrated with consumer digital health technologies in real-world settings, including mobile devices such as smartphones and tablet computers.

### Conclusions

This study found that using a keyboard and mouse was preferable to speech recognition for complex tasks involving problem solving and recall. This may be due to the higher cognitive load reported when using speech recognition and that the participants were more comfortable using a keyboard and mouse. Our results suggest that task characteristics need to be considered by designers when selecting the most appropriate input modality for human-computer interaction. Further studies using a broader variety of consumer digital health tasks of

varying complexity are needed to investigate the contexts in which use of speech recognition is appropriate. The effects of cognitive load on task performance and its significance also need to be investigated.

## Acknowledgments

We thank the participants who gave up their time to participate in this study. We thank Vitaliy Kim and Owen Sun for sharing their expertise to build the simulated conversational agent.

## Authors' Contributions

FM conceptualized the study design. JC conducted the research and primary analysis, and drafted the paper. LL helped to formulate the scenarios and DL helped to adapt the cognitive load inventory. JC is responsible for the integrity of the work as the guarantor of the study. All authors participated in writing and revising the paper. All aspects of the study (including design; collection, analysis, and interpretation of data; writing of the report; and decision to publish) were led by the authors.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Tasks.

[\[DOCX File , 22 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Cognitive load inventory.

[\[DOCX File , 16 KB-Multimedia Appendix 2\]](#)

## References

1. Hinton G, Deng L, Yu D, Dahl G, Mohamed A, Jaitly N, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process Mag* 2012 Nov;29(6):82-97. [doi: [10.1109/MSP.2012.2205597](#)]
2. Hodgson T, Magrabi F, Coiera E. Evaluating the usability of speech recognition to create clinical documentation using a commercial electronic health record. *Int J Med Inform* 2018 May;113:38-42. [doi: [10.1016/j.ijmedinf.2018.02.011](#)] [Medline: [29602431](#)]
3. Kumah-Crystal Y, Pirtle C, Whyte H, Goode E, Anders S, Lehmann C. Electronic Health Record Interactions through Voice: A Review. *Appl Clin Inform* 2018 Jul 18;9(3):541-552 [FREE Full text] [doi: [10.1055/s-0038-1666844](#)] [Medline: [30040113](#)]
4. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc* 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: [10.1093/jamia/ocy072](#)] [Medline: [30010941](#)]
5. Black L, McTear M, Black N, Harper R, Lemon M. Appraisal of a conversational artefact and its utility in remote patient monitoring. 2005 Presented at: 18th IEEE Symposium on Computer-Based Medical Systems; June 23-24, 2005; Dublin. [doi: [10.1109/cbms.2005.33](#)]
6. Levin E, Levin A. Evaluation of spoken dialogue technology for real-time health data collection. *J Med Internet Res* 2006 Dec 11;8(4):e30 [FREE Full text] [doi: [10.2196/jmir.8.4.e30](#)] [Medline: [17213048](#)]
7. Beveridge M, Fox J. Automatic generation of spoken dialogue from medical plans and ontologies. *J Biomed Inform* 2006 Oct;39(5):482-499 [FREE Full text] [doi: [10.1016/j.jbi.2005.12.008](#)] [Medline: [16495159](#)]
8. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. *JAMA Intern Med* 2016 May 01;176(5):619-625 [FREE Full text] [doi: [10.1001/jamainternmed.2016.0400](#)] [Medline: [26974260](#)]
9. Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. *J Am Med Inform Assoc* 2017 Mar 01;24(2):246-250. [doi: [10.1093/jamia/ocw154](#)] [Medline: [28011595](#)]
10. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. *J Am Med Inform Assoc* 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: [10.1093/jamia/ocz175](#)] [Medline: [31599936](#)]
11. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. *J Med Internet Res* 2018 Sep 04;20(9):e11510 [FREE Full text] [doi: [10.2196/11510](#)] [Medline: [30181110](#)]



12. Shneiderman B. The limits of speech recognition. *Commun ACM* 2000;43(9):63-65. [doi: [10.1145/348941.348990](https://doi.org/10.1145/348941.348990)]
13. Ruan S, Wobbrock JO, Liou K, Ng A, Landay JA. Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2018 Jan 08;1(4):1-23 [FREE Full text] [doi: [10.1145/3161187](https://doi.org/10.1145/3161187)]
14. Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 2001 Feb;24(1):87-114. [doi: [10.1017/s0140525x01003922](https://doi.org/10.1017/s0140525x01003922)] [Medline: [11515286](https://pubmed.ncbi.nlm.nih.gov/11515286/)]
15. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 1956;63(2):81-97. [doi: [10.1037/h0043158](https://doi.org/10.1037/h0043158)]
16. Peterson L, Peterson MJ. Short-term retention of individual verbal items. *J Exp Psychol* 1959;58(3):193-198. [doi: [10.1037/h0049234](https://doi.org/10.1037/h0049234)] [Medline: [14432252](https://pubmed.ncbi.nlm.nih.gov/14432252/)]
17. Sweller J. Cognitive load during problem solving: effects on learning. *Cognitive Sci* 1988;12(2):257-285. [doi: [10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)]
18. Limerick H, Moore J, Coyle D, editors. Empirical evidence for a diminished sense of agency in speech interfaces. 2015 Presented at: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; : ACM; 2015; Seoul p. 3967-3970. [doi: [10.1145/2702123.2702379](https://doi.org/10.1145/2702123.2702379)]
19. Leitner G, Ahlström D, Hitz M. Usability of Mobile Computing in Emergency Response Systems – Lessons Learned and Future Directions. 2007 Nov Presented at: Symposium of the Austrian HCI and Usability Engineering Group; November 22, 2007; Graz. [doi: [10.1007/978-3-540-76805-0\\_20](https://doi.org/10.1007/978-3-540-76805-0_20)]
20. Sweller J. Cognitive load theory, learning difficulty, and instructional design. *Learn Instruct* 1994 Jan;4(4):295-312. [doi: [10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)]
21. Ayres P, Sweller J. Locus of Difficulty in Multistage Mathematics Problems. *Am J Psychol* 1990;103(2):167-193. [doi: [10.2307/1423141](https://doi.org/10.2307/1423141)]
22. Ayres P. Systematic Mathematical Errors and Cognitive Load. *Contemp Educ Psychol* 2001 Apr;26(2):227-248. [doi: [10.1006/ceps.2000.1051](https://doi.org/10.1006/ceps.2000.1051)] [Medline: [11273658](https://pubmed.ncbi.nlm.nih.gov/11273658/)]
23. Sweller J, Ayres P, Kalyuga S. Intrinsic and Extraneous Cognitive Load. In: *Cognitive Load Theory. Explorations in the Learning Sciences, Instructional Systems and Performance Technologies Vol. 1*. New York: Springer; 2011:57-69.
24. Beverloo P, Bringert B, Capiel G, Carter J, Kharidi N, Mazzoni D. W3C Community. 2012. Web Speech API Specification URL: <https://www.w3.org/community/speech-api/> [accessed 2018-03-03]
25. Brooke J. SUS: a 'quick and dirty' usability scale. In: *Usability Evaluation In Industry*. London: Taylor & Francis; 1996:189-194.
26. Paas FG. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J Education Psychol* 1992;84(4):429-434. [doi: [10.1037/0022-0663.84.4.429](https://doi.org/10.1037/0022-0663.84.4.429)]
27. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Education Psychol* 2003 Mar;38(1):63-71. [doi: [10.1207/S15326985EP3801\\_8](https://doi.org/10.1207/S15326985EP3801_8)]
28. Lyell D, Magrabi F, Coiera E. The Effect of Cognitive Load and Task Complexity on Automation Bias in Electronic Prescribing. *Hum Factors* 2018 Nov 25;60(7):1008-1021. [doi: [10.1177/0018720818781224](https://doi.org/10.1177/0018720818781224)] [Medline: [29939764](https://pubmed.ncbi.nlm.nih.gov/29939764/)]
29. Pallant J. Non-parametric statistics. In: *SPSS Survival Manual, 3rd ed*. Adelaide: McGraw-Hill Education; 2010:210-231.
30. Faul F, Erdfelder E, Lang A, Buchner A. G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007 May;39(2):175-191. [doi: [10.3758/bf03193146](https://doi.org/10.3758/bf03193146)] [Medline: [17695343](https://pubmed.ncbi.nlm.nih.gov/17695343/)]
31. Hodgson T, Magrabi F, Coiera E. Evaluating the Efficiency and Safety of Speech Recognition within a Commercial Electronic Health Record System: A Replication Study. *Appl Clin Inform* 2018 Apr 16;9(2):326-335 [FREE Full text] [doi: [10.1055/s-0038-1649509](https://doi.org/10.1055/s-0038-1649509)] [Medline: [29768633](https://pubmed.ncbi.nlm.nih.gov/29768633/)]
32. Sun O, Chen J, Magrabi F. Using Voice-Activated Conversational Interfaces for Reporting Patient Safety Incidents: A Technical Feasibility and Pilot Usability Study. *Stud Health Technol Inform* 2018;252:139-144. [Medline: [30040696](https://pubmed.ncbi.nlm.nih.gov/30040696/)]
33. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. *J Am Med Inform Assoc* 2016 Apr 17;23(e1):e169-e179 [FREE Full text] [doi: [10.1093/jamia/ocv152](https://doi.org/10.1093/jamia/ocv152)] [Medline: [26578226](https://pubmed.ncbi.nlm.nih.gov/26578226/)]
34. Hodgson T, Magrabi F, Coiera E. Efficiency and safety of speech recognition for documentation in the electronic health record. *J Am Med Inform Assoc* 2017 Nov 01;24(6):1127-1133. [doi: [10.1093/jamia/ocx073](https://doi.org/10.1093/jamia/ocx073)] [Medline: [29016971](https://pubmed.ncbi.nlm.nih.gov/29016971/)]

---

## Abbreviations

**EHR:** electronic health records

**SUS:** System Usability Scale

---

*Edited by G Eysenbach; submitted 28.05.19; peer-reviewed by M Lee, G Zuccon, L Kelly; comments to author 03.10.19; revised version received 27.11.19; accepted 29.03.20; published 01.06.20*

*Please cite as:*

*Chen J, Lyell D, Laranjo L, Magrabi F*

*Effect of Speech Recognition on Problem Solving and Recall in Consumer Digital Health Tasks: Controlled Laboratory Experiment*  
*J Med Internet Res 2020;22(6):e14827*

*URL: <https://www.jmir.org/2020/6/e14827>*

*doi: [10.2196/14827](https://doi.org/10.2196/14827)*

*PMID: [32442129](https://pubmed.ncbi.nlm.nih.gov/32442129/)*

©Jessica Chen, David Lyell, Liliana Laranjo, Farah Magrabi. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 01.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.