Integrating a Voice User Interface into a Virtual Therapy Platform

Yun Liu
University of Washington, Seattle, WA 98195, USA
liuyun6@uw.edu

Lu Wang
University of Washington, Seattle, WA 98195, USA
wangl33@uw.edu

William R. Kearns
University of Washington, Seattle, WA 98195, USA
kearnsw@uw.edu

Linda E Wagner
University of Washington, Seattle, WA 98195, USA
lw1mhcid@uw.edu

John Raiti
University of Washington, Seattle, WA 98195, USA
jraiti@uw.edu

Yuntao Wang
Tsinghua University, Beijing 100084, China & University of Washington, WA 98195, USA
yuntaowang@tsinghua.edu.cn

Weichao Yuwen
University of Washington, Tacoma, WA 98402, USA
wyuwen@uw.edu

ABSTRACT

More than 1 in 5 adults in the U.S. serving as family caregivers are the backbones of the healthcare system. Caregiving activities significantly affect their physical and mental health, sleep, work, and family relationships over extended periods. Many caregivers tend to downplay their own health needs and have difficulty accessing support. Failure to maintain their own health leads to diminished ability in providing high-quality care to their loved ones. Voice user interfaces (VUIs) hold promise in providing tailored support family caregivers need in maintaining their own health, such as flexible access and handsfree interactions. This work is the integration of VUIs into a virtual therapy platform to promote user engagement in self-care practices. We conducted user research with family caregivers and subject matter experts, and designed multiple prototypes with user evaluations. Advantages, limitations, and design considerations for integrating VUIs into virtual therapy are discussed.

CCS CONCEPTS

• **Human-centered computing** → Human computer interaction (HCI); Interaction paradigms; Natural language interfaces.

KEYWORDS

Voice User Interface, Mental Health, Dialog Systems, Conversational Agent

CHI '21 Extended Abstracts, May 08–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8095-9/21/05…$15.00
https://doi.org/10.1145/3411763.3451595

ACM Reference Format:
Yun Liu, Lu Wang, William R. Kearns, Linda E Wagner, John Raiti, Yuntao Wang, and Weichao Yuwen. 2021. Integrating a Voice User Interface into a Virtual Therapy Platform. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts), May 08–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3411763.3451595

1 INTRODUCTION

Chronic diseases, such as diabetes and cancer, are the leading causes of death and disability in the United States (U.S.). Treatment and management of chronic diseases cost 2.6 trillion dollars annually [1]. The management burden falls largely upon patients and their families - over 50 million family caregivers providing an estimated $470 billion unpaid care services in the U.S. every year [2–4]. Almost half of them considered their caregiving situation highly stressful, and one-quarter found it either difficult to maintain their own health or that their health has worsened [2]. When caregivers are healthy, they can provide high-quality preventative care to those with chronic conditions, reducing the risk of complications and, consequently, healthcare costs [5].

Research has revealed that many caregivers tended to downplay or ignore their own needs – or worse yet, many reported not thinking of themselves as "caregivers" at all [6–8]. In Washington State, only 5% of caregivers make use of current support and services provided by public programs, citing that they are difficult to access, overwhelming in scope, lack personalization and engagement, or require too great an investment of time and money to be utilized [9]. Conventional face-to-face support and services are not reaching this critical population.

This work is part of an on-going study to develop Caring for Caregivers Online (COCO), a platform that delivers an AI-augmented application providing on-demand, empathetic, and tailored caregiving support to promote the health and wellbeing of families with chronically ill members [10]. COCO combines a conversational agent for automated therapy with a telenursing service to

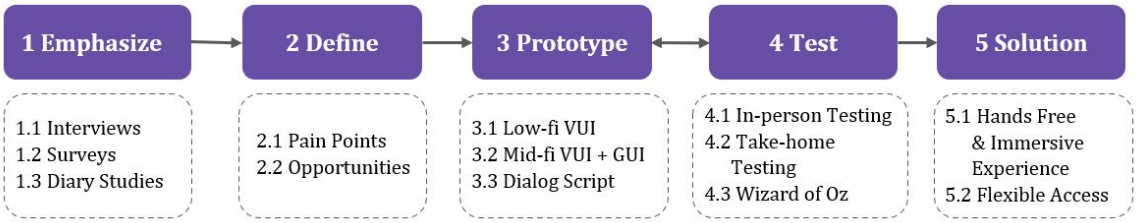

Figure 1: Design Process

maximize access and quality. These integrated services build on top of a state-of-the-art AI layer that detects and tracks mental states to predict empathic responses to events. Formative research with caregivers suggests tools need to be on-demand and flexible for multitasking activities.

Voice-based technology is a component of pervasive computing that has been integrated into everyday lives. Approximately 46% of U.S. adults reported using voice assistants in their smartphones or other smart devices [10, 11]. Voice Assistants continue to evolve and broaden its application in health with strengths including broad reach, ability to personalize and tailor support, deliver content via multimedia formats, and provide real-time feedback [12]. This is an opportune time to explore how voice user interfaces (VUIs) could be valuable in supporting these critically underserved family caregivers.

In this paper, we present the COCO voice assistant, an exploration of integrating voice interfaces into the COCO platform. The voice assistant built upon the same virtual therapy framework of COCO with a focus on providing suggestions and guiding users to practice self-care techniques through voice conversations. To explore how VUIs might be a valuable addition for the COCO and how to design the VUIs for virtual therapy, we developed prototypes based on the design thinking methodology (Figure 1) proposed by d.school [13]. Findings are summarized to direct future exploration. Main contributions include: 1. We compared VUIs, GUIs, and multimodal interfaces for a virtual therapy chatbot to identify scenarios where VUIs would be preferred; 2. We identified unique advantages VUIs might have in improving caregivers' engagement for self-care and design considerations for integrating VUIs into chatbots.

2 BACKGROUND USER STUDIES

2.1 Methods

Three research methods were selected to assess the pain points and issues directly affecting how caregivers care for themselves from different perspectives. The participants consisted of a convenience sample of family caregivers and subject matter experts.

2.1.1 Interviews. We interviewed six subject matter experts in family caregiving: (1) one nursing assistant, (2) one paid caregiver, (3) two social workers, (4) one long-time family caregiver who is also a caregiver coach and the author of a book about family caregiving, (5) one assistant professor whose research focuses on children with asthma. These individual virtual interviews helped us understand caregivers' challenges from close stakeholder views.

2.1.2 Surveys. Through online crowdsourcing tools, we used convenience sampling and recruited 135 unpaid family caregivers of children (age 3-12) to learn the perceived challenges, self-care needs, and priorities as a caregiver. The majority (60.7%) of participants ranged in age from 30 to 40, 22.2% were below 30, and 17% were above 40. Questions covered the medical condition of the children, weekly caring time, self-assessment on self-care conditions, self-care practices, etc.

2.1.3 Diary Studies. We recruited 5 participants from online family caregiver support groups for an observational diary study to understand their daily struggles and coping methods. The participants (four females and one male) were family caregivers of children (age 3-12) with asthma and ASD. During the 10-day study, participants recorded their emotions, the most stressful moment, and how they dealt with stress. They also recorded caregiving tasks performed and technology tools used for caregiving during the study period.

2.2 Results

Through triangulation, we identified two main reasons that contribute to poor self-care. Firstly, family caregivers are often required to pay attention continually, visit hospitals frequently, and perform complicated caregiving tasks (injection, medication, etc.). Over half of the caregivers in our survey thought that they do not have time for self-care, and 18% were unsure how to care for themselves. This lack of time and knowledge makes self-care difficult. Another factor that stops caregivers from caring for themselves is insufficient motivation. Many caregivers were not aware of the importance of self-care and even felt guilty about spending time for themselves when thinking their loved ones are in pain.

Making self-care practice convenient and straightforward is essential to motivating caregivers to care for themselves and maintaining self-care plans. VUIs may offer a more flexible interaction channel for caregivers to get self-care information and guide them through self-care practices even when they have to multitask.

Our diary study showed that "talking to other people (friends, families, etc.)" is the most common method caregivers use to reduce stress, which indicates talking to a virtual therapy chatbot might be preferred than other self-care activities. Moreover, most participants (89%) in our survey had used voice-based products such as Alexa, Siri, or Google Assistant and had a positive experience with voice interactions, which shows potential for this group to adopt voice-based products.

Table 1: Prototype Evaluation Methods

Round	Form	Device	Research Questions	Design Decisions Made
1	Low-fidelity VUI mockup	Alexa-enabled smart speaker	If the voice interaction user flow legible & easy to use for people.	Add mobile graphic user interface
2	Mid-fidelity VUI+GUI	Alexa-enabled smart speaker & Android phone	If the voice and text dialog structures, content and wording make sense to users? What kind of user interface (GUI, VUI, & multimodal user interface) is preferred?	Refine dialog structure and wording; Use VUI only for short & simple conversations
3	Wizard of Oz	Alexa Skills developer text to speech tool	How can we make the voice conversations more intuitive and effective for users?	Finalize conversation design & simplify user flow

Table 2: Major Attributes and Associated Sub-attributes of Findings

Major Attributes	Associated Sub-attributes
Accessibility	Flexibility, navigability, controllability
Efficiency	Minimal memory load, user guidance, error handling, prevention
Satisfaction	User's preference, appropriate output sentence, output quality

3 PROTOTYPE EVALUATION AND ITERATION

Building upon the existing COCO platform where the users learn to identify their caregiving-related symptoms and make a self-care plan, we determined our design goal to encourage user engagement by using VUIs as a more natural, flexible, and efficient interaction channel.

3.1 Methods

We conducted three rounds of prototype iteration with user testing (Table 1). Because our testing mainly focused on the user flow and interaction rather than the actual therapeutic content, and we do not want to add additional work for already over-burden family caregivers during the pandemic, we tested the prototypes with proxy users and general people remotely via Zoom.

The first round of user testing aimed to test the overall user flow of the voice interaction. We recruited five people from the general public who are fluent in English. Participants were asked to follow a pre-designed script to talk with the voice assistant on a smart speaker. In the second round, we selected six proxy users who have similar characteristics as our target users: young to middle-age adults who were busy in life with stress. Participants practiced scheduled meditation exercises using the mobile app two times and then practiced using a smart speaker for two times during the 3-day testing period. Post-interviews were conducted to gather participants' feedback on the dialog structures, content, and interfaces. For the third round, we recruited six new proxy users. Participants were provided with two scenarios to talk with the voice assistant in their own words. Team members acted as the Wizard behind the scene to select among the predefined responses.

3.2 User Testing Results

As VUIs have substantial differences from traditional graphical user interfaces, we adapted general mobile software usability & UX measurement framework [14] based on VUI evaluation guidelines [15] to evaluate and refine our prototypes. We identified three major attributes and eleven associated sub-attributes to summarize our findings (Table 2).

3.2.1 Accessibility. We started with developing the VUIs for several new features that can be added to COCO. We found that voice interfaces alone are more challenging for users to navigate and control. Although participants were able to complete the tasks during the first round of user testing, they did not feel they could get a full picture of the prototype. Most participants were more comfortable learning a new product with graphical user interfaces (GUIs) and have VUIs as an option. Considering the flexibility and portability of mobile devices, we decided to add the same functions for the mobile text-based chatbot to provide users with better control of how, when, and where to use the prototype (Figure 2).

3.2.2 Efficiency. As the voice is a transient type of information [15], it is easy for users to miss information or get confused by complicated questions. When a user gave an unexpected answer, the conversation might end. Detailed questions with full explanations would help reduce confusion and minimize memory load, but users tended to get impatient with repetitive information and long sentences. Moreover, Alexa has a default waiting time before it stops listening. Complicated questions might lead to a longer thinking time and discontinue the conversation.

To facilitate the conversation, we also designed graphical interfaces for smart speakers with a screen (e.g., echo show). The screen shows text prompts like "Try 'Alexa, tell COCO I'm anxious."'. Users tended to follow the prompts on the screen when there is one, which would reduce unexpected user responses. However, multimodal interfaces might lead to more confusion when the perceived information from different interfaces is inconsistent, which results in the users not knowing which prompt to follow.

To prevent and better handle errors during conversations, we simplified the conversation structure and refined wording to guide

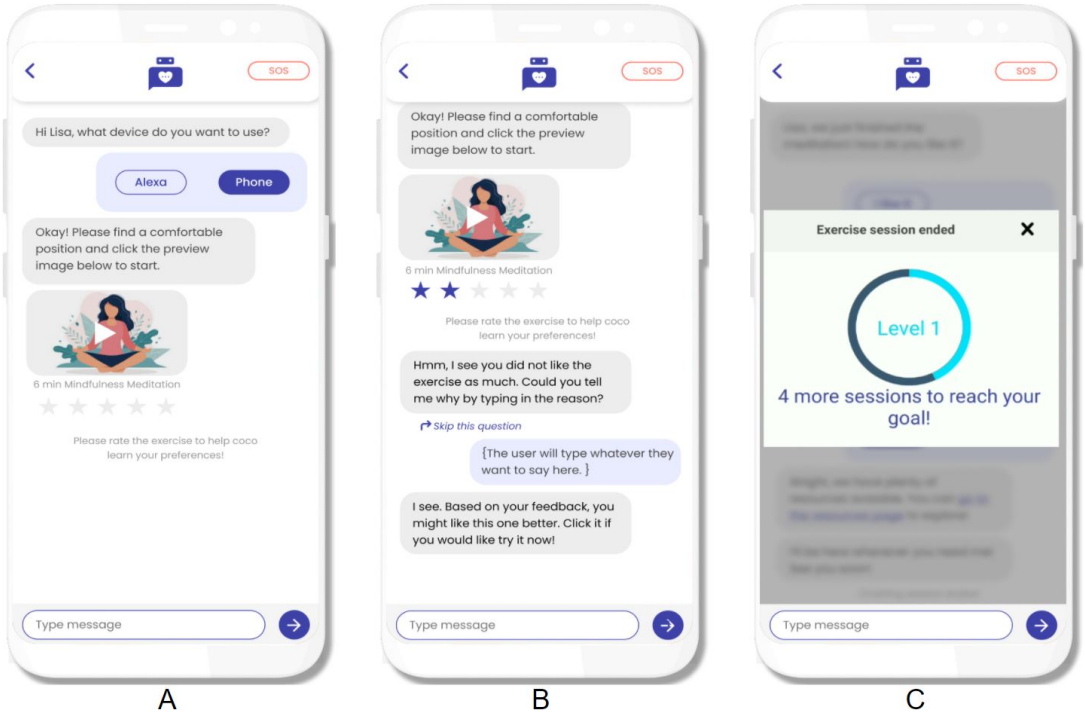

Figure 2: Mobile text-based Chatbot Screenshots. A: Ask users to choose a device and start a recommended self-care activity. B: Ask for user feedback and provide empathetic responses. C: Provide encouragement for future practice

users to say the proper keyword by repeating the question or providing more clarity on expected keywords that the Natural Language Processing (NLP) model can recognize. For example, when asking the user to rate a scale of one to five, and the user's answer is "zero," the voice assistant would say, "sorry, please give a number between one to five".

3.2.3 Satisfaction. Users preferred VUIs through a smart speaker as it helps them concentrate and relax by getting rid of unwanted distractions from phones. VUIs also allow users to start a self-care session when they need to multitask. For users who are sensitive to sound quality, smart speakers provide better sound quality and create a more immersive environment, enhancing the user experience. Technical problems (e.g., unexpected quitting, inappropriate output responses) and wording confusions can significantly undermine the user experience. From the Wizard of Oz testing, we found that understanding the user's input and providing relevant responses are more important in enhancing user satisfaction than quick responses.

4 SOLUTION

Our solution is a voice assistant developed on Alexa that expands the text-based chatbot in the COCO platform. COCO voice assistant facilitates the practice of scheduled self-care activities and provides quick responses to alleviate users' immediate symptoms (e.g., stress, anxiety, worrisome, etc.) (Figure 3). COCO voice assistant is developed as an Alexa skill using Amazon Alexa Conversations. By syncing data from the same databases connected to the text chatbot on the mobile app, the voice assistant provides a flexible access channel for users to practice guided self-care activities efficiently. VUIs also allow users to multitask in some cases and be used as shortcuts to specific functions.

Users can initiate a voice conversation by asking the voice assistant to start a scheduled session ("Alexa, tell COCO to start today's session") or tell the voice assistant their current feeling ("Alexa, tell COCO I'm feeling stressed."). Then they can simply follow the voice prompts to choose a recommended activity and start practicing. After each practice session, the voice assistant asks for the user's feedback and provides the corresponding next-step suggestions. At the end of the conversation, the voice assistant would say motivating words to encourage future practice. All the feedback and practice history is recorded for improving future recommendations (Figure 4).

5 DISCUSSION

We designed a voice assistant to provide self-care suggestions and lead self-care activities as a starting point to explore the integration of VUIs into a virtual therapy platform. We tested the voice assistant in two defined scenarios and found that VUIs worked better for short and straightforward conversations. Significant benefits from integrating VUIs are: (1) allowing multitask for busy caregivers by freeing their hands, (2) creating a more immersive environment so that users can focus on themselves at the moment, and (3) enabling shortcuts to specific functions to improve efficiency.

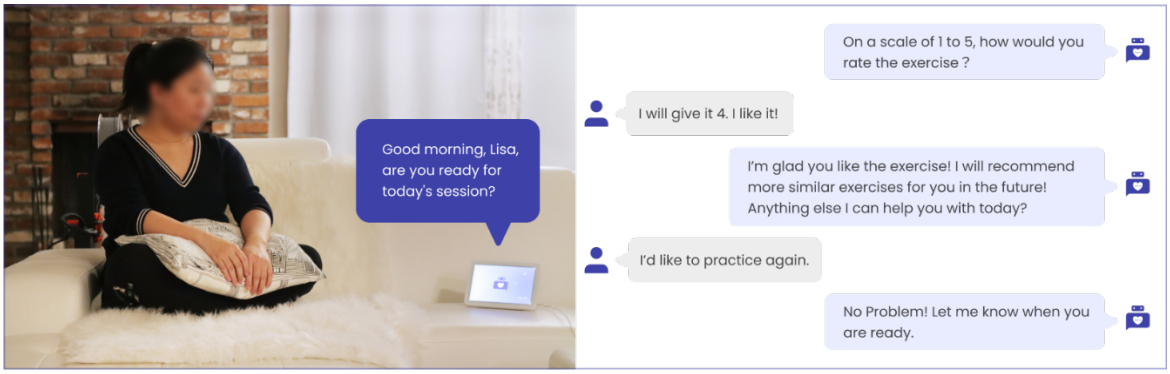

Figure 3: Voice-controlled Interaction. This figure contains a picture showing the scene of a person talking to COCO Alexa Skill using an echo device, and the script of a short sample dialogue.

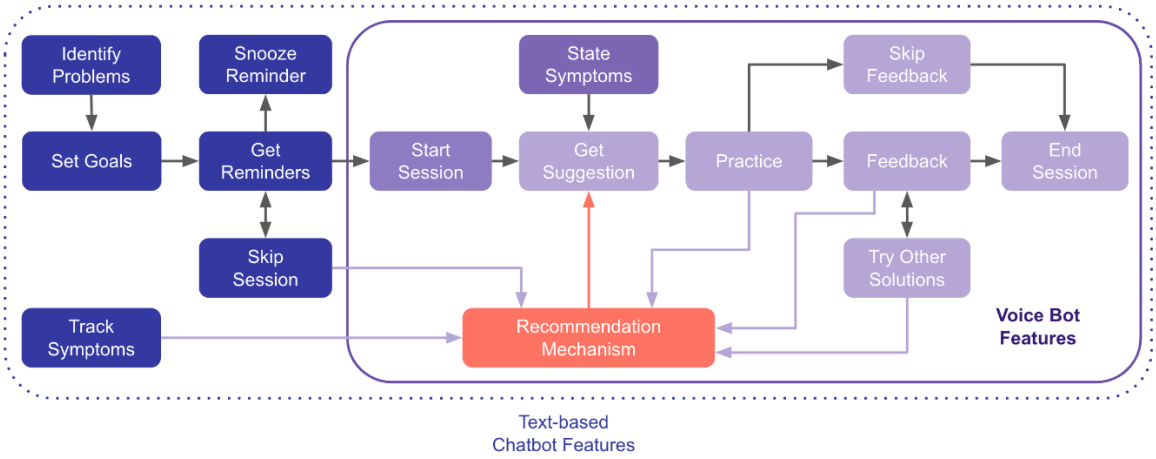

Figure 4: COCO Voice Assistant User Flow. Step by step voice interaction process from starting a session to ending a session.

We also identified limitations of VUIs due to the transient characteristic of voice information and limited accessibility of smart speakers. Long and complicated conversations are difficult to be carried out efficiently through VUIs. Multimodal interfaces would help users say the right keywords but might also distract users with increased cognitive load. Therefore, we think not all the features on COCO are suitable for voice integration. For example, the problem-solving therapy framework education conversations would not be ideal for a VUI due to the process's complexity and length. Moreover, current NLP technology cannot support a free-flowing conversation, limiting the voice assistant's ability to talk like a real human. All conversations were structured with clear expected answers. Users were not able to redirect the conversation to any direction they wanted.

Additional work is warranted to design additional features and test the efficacy of VUI integration in promoting user engagement. As we only developed the voice conversation model on Alexa, we did not test VUIs for reminders considering smart speakers' location constraints. Moreover, although appropriate tone and speed are essential factors in VUIs, Alexa only supports a few varieties with no speed control, limiting our ability to test out how tone and speed would affect the user experience. Finally, as users get familiar with the product, their preference for voice conversations might change (e.g., they might want to skip repetitive explanations). A user testing with a longer duration would be critical to deciding how the VUIs should be personalized over time and whether the VUI integration would increase long-term user engagement.

6 CONCLUSION

This work expands on the growing body of research that supports the use of conversational agents to improve health and wellness [16–19] by illuminating the cases in which voice or multimodal inputs may better serve users than text-based interactions. Further, it identifies the limitations of VUIs, namely the high cognitive load required for complex conversations and the appropriateness of voice-based interactions on sensitive topics in specific environments. Together, these findings support future work to address caregivers' specific needs when their hands are not free and support immersive self-care experiences that lower user screen time.

ACKNOWLEDGMENTS

We would like to thank our peer Huichen Li who contributed to the project. We would also like to thank all our study participants for their time, effort, and patience. An exceptional thanks to Liying Wang, who helped us create tailored exercise scripts, Myra F. Divina for recording source audios, and Stanley Wang for contributing to the mobile app software development. The study was partially funded by the University of Washington Population Health Initiative Pilot Fund (W.Y.) and National Institute of Nursing Research, National Institute of Health, P30NR016585 (W.R.K. and W.Y., Center PI: Ward, TM & Heitkemper, M.), and Launch Project Fund from the University of Washington Global Innovation Exchange (Y.L. and L.W.).

REFERENCES

[1] Centers for Disease Control and Prevention. 2019. About Chronic Diseases. Retrieved December 29, 2020 from https://www.cdc.gov/chronicdisease/about/index.htm

[2] The National Alliance for Caregiving (NAC) and AARP. 2020. Caregiving in the U.S. 2020. Retrieved December 22, 2020 from https://www.aarp.org/content/dam/aarp/ppi/2020/05/full-report-caregiving-in-the-united-states.doi.10.26419-2Fppi.00103.001.pdf

[3] Family Caregiver Alliance. 2019. Caregiver Statistics: Demographics. Retrieved December 29, 2020 from https://www.caregiver.org/caregiver-statistics-demographics

[4] AARP Public Policy Institute. 2019. Valuing the Invaluable: 2019 Update. Retrieved December 29, 2020 from https://www.aarp.org/content/dam/aarp/ppi/2019/11/valuing-the-invaluable-2019-update-charting-a-path-forward.doi.10.26419-2Fppi.00082.001.pdf

[5] Washington State Department of Social and Health Services (DSHS). 2019. Family caregiver support program fact sheet. Retrieved December 29, 2020 from https://fortress.wa.gov/dshs/adsaapps/about/factsheets/

[6] Sunny Chieh Cheng, Uba Backonja, Benjamin Buck, Maria Monroe-Devita, and Elaine Walsh. 2020. Facilitating pathways to care: A qualitative study of the self-reported needs and coping skills of caregivers of young adults diagnosed with early psychosis. Journal of Psychiatric and Mental Health Nursing, 7, 4 (2020), 368–379. DOI: http://dx.doi.org/10.1111/jpm.12591

[7] Weichao Yuwen, Frances M. Lewis, Amy J. Walker, and Teresa M. Ward. 2017. Struggling in the Dark to Help My Child: Parents' Experience in Caring for a Young Child with Juvenile Idiopathic Arthritis. Journal of Pediatric Nursing 37 (2017). DOI: http://dx.doi.org/10.1016/j.pedn.2017.07.007

[8] Wenzhe Hua, Liying Wang, Chenxing Li, Jane M. Simoni, Weichao Yuwen, and Liping Jiang. 2020. Understanding preparation for preterm infant discharge from parents' and healthcare providers' perspectives: Challenges and opportunities. Journal of Advanced Nursing (2020). DOI: https://doi.org/10.1111/jan.14676

[9] Washington State Department of Social and Health Services (DSHS). 2019. Family caregiver support program fact sheet. Retrieved from https://fortress.wa.gov/dshs/adsaapps/about/factsheets/

[10] William R. Kearns, Neha Kaura, Myra Divina, Cuong Vo, Dong Si, Teresa Ward, and Weichao Yuwen. 2020. A Wizard-of-Oz Interface and Persona-based Methodology for Collecting Health Counseling Dialog. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–9. DOI: https://doi.org/10.1145/3334480.3382902

[11] Pew Research Center. 2017. Nearly half of Americans use digital voice-based conversational agents, mostly on their smartphones. Retrieved December 14, 2020 from http://pewrsr.ch/2kquZ8H

[12] Emre Sezgin, Lisa K. Militello, Yungui Huang, and Simon Lin. 2020. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. Translational Behavioral Medicine 10, 3 (2020), 606–628. DOI: http://dx.doi.org/10.1093/tbm/ibz141

[13] Hasso Plattner Institute of Designat Stanford University. 2010. An Introduction to Design Thinking: PROCESS GUIDE. Retrieved January 4, 2021 from http://web.stanford.edu/~mshanks/MichaelShanks/files/509554.pdf

[14] Jia Tan, Kari Rönkkö, and Cigdem Gencel. 2013. A Framework for Software Usability and User Experience Measurement in Mobile Industry. 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement (October 2013), 156–164. DOI: http://dx.doi.org/10.1109/iwsm-mensura.2013.31

[15] Valeria Farinazzo, Martins Salvador, Andre Luiz S. Kawamoto, and Joao Soares De Oliveira Neto. 2010. An Empirical Approach for the Evaluation of Voice User Interfaces. User Interfaces (2010). DOI: http://dx.doi.org/10.5772/9490

[16] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Mental Health 4, 2 (2017). DOI: http://dx.doi.org/10.2196/mental.7785

[17] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. 2018. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR mHealth and uHealth 6, 11 (2018). DOI: http://dx.doi.org/10.2196/12106

[18] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y S Lau, Enrico Coiera. 2018. Conversational agents in healthcare: a systematic review. Journal of the American Medical Informatics Association 25, 07 (2018), 1248 - 1258. DOI: https://doi.org/10.1093/jamia/ocy072

[19] William R. Kearns, Nai-Ching Chi, Yong K. Choi, Shih-Yin Lin, Hilaire Thompson, and George Demiris. 2019. A Systematic Review of Health Dialog Systems. Methods of Information in Medicine 58, 06 (2019), 179–193. DOI: http://dx.doi.org/10.1055/s-0040-1708807