

# Evaluating Older Users' Experiences with Commercial Dialogue Systems: Implications for Future Design and Development

Libby Ferland<sup>1</sup> and Thomas Huffstutler<sup>1</sup> and Jacob Rice<sup>2</sup> and Joan Zheng<sup>1</sup>  
and Shi Ni<sup>1</sup> and Maria Gini<sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455

<sup>2</sup> Wayzata High School, 4955 Peony Ln N, Plymouth, MN 55446

## Abstract

Understanding the needs of a variety of distinct user groups is vital in designing effective, desirable dialogue systems that will be adopted by the largest possible segment of the population. Despite the increasing popularity of dialogue systems in both mobile and home formats, user studies remain relatively infrequent and often sample a segment of the user population that is not representative of the needs of the potential user population as a whole. This is especially the case for users who may be more reluctant adopters, such as older adults.

In this paper we discuss the results of a recent user study performed over a large population of age 50 and over adults in the Midwestern United States that have experience using a variety of commercial dialogue systems. We show the common preferences, use cases, and feature gaps identified by older adult users in interacting with these systems. Based on these results, we propose a new, robust user modeling framework that addresses common issues facing older adult users, which can then be generalized to the wider user population.

## Introduction

In the last decade, use of spoken dialogue systems (SDSs) has been rapidly increasing in popularity. Voice-controlled smart home devices, such as Amazon Echo devices and Google Home are currently being used in millions of households. Dialogue agents are also seeing increasingly heavy use as part of smart phone operating systems, with agents like Apple's Siri and Genie on Android devices leading the charge. Despite their popularity, most commercial SDSs are primarily designed with healthy younger users in mind, overlooking older adult users and other large segments of the user population. This shortfall in design will become more readily apparent when considering future population trends. By 2050, over half of the US population is predicted to be over the age of fifty (Ortman et al. 2014), and many of these adults which will likely interact with smart home devices and other dialogue systems on a regular basis. Furthermore, with this expected large increase in the size of the aged population, greater demands for care services will be made on already taxed healthcare systems. Dialogue systems, especially smart home technologies, have been proposed as excellent resources to help alleviate strain on care providers.

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

However, despite this apparent need and demand, there remains a lack of current user studies that focus on older adult users' interactions with smart devices and dialogue systems and what they imply for the needs of this population.

In this study, we surveyed adults over age fifty on their experiences with commercially available dialogue systems. We asked users to describe their interactions with the systems, including current use cases, failures and shortcomings, and any feature gaps they perceive and would like addressed. The results of this study provide substantive, quantitative evidence that support many of the current focuses and research directions in the design and development of dialogue systems. The results also reveal some very specific features, namely integrated, unimodal help features, that are most desirable to older users and may help serve as guideposts for directing the focus of future research.

In this paper, we first present some of the motivating and related work in the design and development of dialogue systems for a variety of users. Next, we present key results from our study revealing important user behaviors and perceived and actual feature gaps. We further discuss the implications of some of these results in terms of key research questions for the field, and propose that the need for context-sensitive, voice-activated help functions for these systems reveals a need to further focus research on developing robust user models and better predictive capabilities for user behavior. Last, we detail some of the current and future work we are doing in applying deep learning techniques to larger bodies of user interaction data to develop a stronger user model.

## Related Work

Motivation for this work is drawn from general problems in the field of dialogue agents in addition to problems specific to older users. The main motivation for undergoing this study stems from the desire to identify areas of acute interest in the development of dialogue systems as assistive technology, especially as support systems for elderly users at home.

## Identifying Areas of Improvement

Much literature has been devoted to identifying broad areas of interest in the development of conversational agents and assistive technology, independent of demographic-specific considerations. In general, conversational agents have to face three different types of problems, including completing

tasks, answering questions, and interacting socially (Gao, Galley, and Li 2018). These are extremely broad problems, but more and more approaches to creating effective dialogue systems acknowledge that they cannot be treated as wholly independent, and a complete end-to-end system must be expected to address all the different subtasks that make up the whole (Gao, Galley, and Li 2018) (Rao, Ture, and Lin 2018).

It is specially important for assistive technologies to adopt this holistic view of task completion for dialogue systems, which requires a deeper examination of some potential agent architectures and computational problems involved. The design of intelligent personal assistants in particular provides some insight into potential areas of focus. Milhorat et al. identify several key focus areas in the development of conversational agents as personal assistants (Milhorat et al. 2014). These include: 1. Tracking longer dialogue histories; 2. Improving context awareness (ideally solely through dialogue, without involving additional sensor data); 3. Dynamic real-time adaptation to user behavior; and 4. Improved theoretical frameworks for problems such as task handling and task hierarchies. Determining which of these areas is of acute interest to users is a different problem altogether, and one which we hope to address with the results of this study.

### **Older Users Interacting with SDS**

A number of small-scale studies of older users interacting with smart technology help to narrow the list of areas of general improvement to those problems that are of particular importance to an older user demographic (Schlögl, Garschall, and Tscheligi 2014). Previous studies show that in general, older users do interact with dialogue systems in a manner that is quite distinct from other groups. In particular, older users tend to interact with dialogue agents in a manner much more consistent with interactions with living social agents, as opposed to younger users who are often much more task-driven (Georgila et al. 2008) (Möller, Götter, and Wolters 2008). The most important aspects of interaction style in attempting to isolate key areas of improvement are that older users in general tend to participate in much more turn-taking in conversation, and use a much richer, more varied vocabulary with a greater number of social cue words.

There are also mechanical features of speech for older people that may further dictate important areas of focus in the improvement of dialogue agents. Older speakers often exhibit a greater number of speech disfluencies that can cause issues with successfully completing tasks using SDSs (Bortfeld et al. 2001). First, older speakers use conversational repair strategies such as backtracking and error correction more frequently than do younger speakers, which dialogue systems may not handle gracefully. A greater mechanical source of frustration, however, may be that older speakers tend to exhibit more frequent and longer pauses in speech. These may cause issues with timeouts or reprompts, depending on the setup of a given dialogue system.

### **Design and Acceptance with Older Users**

There are a number of different elements of user acceptance for older adults using assistive technology, especially

technology with a social or intelligent component. These include, but are not limited to trust, perceived adaptability, perceived usefulness, and perceived ease of use, among others (Heerink et al. 2010). These dimensions are not often fully assessed for older users when designing new assistive technology. A further, perhaps greater barrier to acceptance and adoption among older users of new technology is the failure to include representative users in design and testing phases. This has been shown in a variety of studies, but is especially evident in the development of assistive technology (Heerink et al. 2010) (Dodd, Athauda, and Adam 2017).

## **Experimental Design**

### **Setting**

This study was conducted in the D2D (Driven to Discover) facility at the Minnesota State Fair. The facility is designed for researchers to access a larger and more diverse pool of participants from fairgoers that is a representative sample of the states population. The majority of participants recruited in this setting were native English speakers and residents of Minnesota and Wisconsin.

### **Participants**

The participants that were actively recruited were adults of both sexes aged 50 or older, as well as those with severe vision impairment, as defined by commonly accepted occupational and legal standards (Scheiman, Scheiman, and Whittaker 2007). Participants of any age with severe sensory impairment were included both as a point of comparison with an even wider age range, and as a representative sample to help pinpoint any issues that may specifically pertain to users of dialogue systems with the types of sensory impairments commonly seen in an older population.

### **Procedure**

Potential participants were given an introduction to the background and objectives of the study. Those who wished to participate were asked to sign a consent form were shown to a table where they were given either an iPad or a paper copy to complete the survey based upon their preference.

The most significant portion of the data collected was with regards to the participants' experiences with smart devices (i.e., Alexa, Siri, Google Home, etc.) as well as their thoughts on current shortcomings of smart devices and potential improvements that could be made. Participants were first asked about their experience with smart devices in terms of frequency. Based on this response, if they identified themselves as users of voice activated technology either currently or in the past, they were directed to section one. Section one asked participants about which specific smart devices they used the most, the application of the smart devices they used, and most thoroughly, their thoughts on the voice interaction experience of using the smart devices. The interest in the users experience with voice interactions of smart devices was with regards to the difficulty of the interaction, their preferences for voice interaction, and their thoughts on the current constraints of the technology. Section two of the survey focused on participants who initially responded

that they were non-users of smart technology. Questions in that section asked about experiences with smart devices they have used, reasons for their limited use of smart technologies, impressions on the voice capabilities of smart devices, and tasks for which they would like to use smart devices.

Questions from both sections of the survey were designed to follow the Almere model for assessing acceptance and acceptability of assistive technology in older adults, as proposed by Heerink et al. (Heerink et al. 2010) The Almere model identifies ‘intention to use’ as a heavy contributing factor in whether or not older users actually adopt assistive technology. Intention to use in turn is a combination of user perceptions, including perceptions of enjoyment, usefulness, ease of use, and trust and trustworthiness. The questions in this study are designed specifically to develop a more detailed picture of how well commonly available dialogue systems address these factors, and to further identify major acceptability gaps. These questions were also specifically chosen in the hopes of developing a more detailed picture of the connections between any identified acceptability gaps and the current state-of-the-art and relevant research questions in the field.

## Results

The survey instrument in this study contained two sections, one pertaining to experiences with ‘smart technology’ in general, and the second pertaining to experiences with voice interaction specifically. Questions in both sections were fairly similar as a form of quality control and verification of consistency in participant responses. For the sake of brevity and maintaining focus on the most relevant results for this work, this paper will only discuss the responses to the second, voice specific, section.

All free response questions had two independent raters. All questions were optional, so response rate may also vary per question. The population size for each question is given. Most free response questions allowed multiple answers per participant.

### Demographics

The total number of respondents was  $N = 174$ . Of these, 145 participants indicated at least some experience with dialogue systems. 36 participants were primarily users of smart home systems, 102 were users of mobile devices, and 7 were not otherwise categorized. Table 1 contains a summary of the primary device type and frequency of use for respondents.

### User Preferences Over Interaction Style

Participants were asked to identify their preferred interaction style with dialogue agents. Responses for this question were hard-coded and covered a variety of social, personally adaptable, or impersonal interaction styles. Responses are once again grouped by primary device type.

Among smart home device users, 69% preferred an agent with some ability to adapt to user interaction styles and history. 42% preferred a social or relational agent. 25% of users had a preference for strictly impersonal interactions. Among smart phone or other mobile device users, 68% preferred an

Device Type	Frequency of Use			
	Daily	Weekly	Monthly	Rarely
Smart Home	0.43	0.27	0.17	0.13
Smart Phone	0.35	0.29	0.23	0.13
Other	0.09	0.09	0.00	0.82
Total	0.33	0.25	0.18	0.23

Table 1: Device type and usage frequency among participants ( $N = 126$ .  $N = 30$  smart home device users, 94 smart phone users)

Device Type	Interaction Preference					
	A, $\neg S$	A, S	S	I	All A	All S
Smart Home	0.33	0.36	0.06	0.25	0.69	0.42
Smart Phone	0.45	0.23	0.13	0.20	0.68	0.35
Other	0.33	0.17	0.33	0.17	0.50	0.50

Table 2: Dialogue system interaction style preferences by primary device type. Options abbreviated for length. A = adapted to user interaction style and history, S = social or conversational agent, I = impersonal agent, no preference over adaptability to user.  $N = 144$  respondents.  $N = 36$  smart home users, 102 mobile device users.

adaptable agent, 35% preferred a social or relational agent, and 20% of users preferred strictly impersonal interactions. There is no significant difference in any of the preferences between the different device types.

### Frequent Use Cases

Participants were asked to identify their most common or favorite uses of voice interaction with their primary device.

Device Type	Interaction Type						
	Ph.	Dict.	Mus.	Gen.	Struct.	O	NS
Smart Home	0.03	0.09	0.11	0.51	0.20	0.06	0.00
Smart Phone	0.16	0.24	0.03	0.24	0.19	0.11	0.03
Other	0.00	0.00	0.00	0.50	0.50	0.00	0.00

Table 3: Most frequent uses for older users of dialogue systems. Ph. = phone, Dict. = text dictation, Mus. = music, Gen. = general information (unstructured queries), Struct. = structured queries (weather, time, etc.), O = other, NS = not sure.  $N = 134$ .  $N = 35$  smart home users, 95 mobile device users

We differentiate between structured and unstructured queries. Structured queries are highly transactional interactions that have shallow dialogue trees (i.e., asking

for weather r directions, querying for time). Unstructured queries are interactions to find more general information with a high likelihood of follow-up questions (i.e., web searches). Surprisingly, despite frequent frustrations identified elsewhere in the survey, users most often interact with dialogue systems to obtain general information.

### Frequent Difficulties

We next asked participants to identify their most common sources of difficulty when interacting with dialogue systems.

Device Type	Source of Difficulty						
	GQ	Dict.	UL	SR	U(NS)	O	NS
Smart Home	0.29	0.03	0.03	0.03	0.23	0.26	0.13
Smart Phone	0.18	0.10	0.07	0.07	0.18	0.30	0.09
Other	0.14	0.14	0.00	0.00	0.29	0.14	0.29

Table 4: *Most frequent difficulties experienced by users in voice interactions with dialogue systems. GQ = general/complex questions, Dict. = text dictation, UL = unusual language or names, SR = speech recognition, U(NS) = understanding (non-specific), O = other, NS = not sure. N = 134. N = 31 smart home users, 96 mobile device users.*

Most difficulties identified in this question relate to issues with ‘understanding.’ In this work, we are interested in examining differences between experiences with language understanding (i.e., intent recognition, NLU, NLG) and speech recognition. For this and all other free response questions, the coding scheme presented here is strict, and responses on ‘understanding’ that do not differentiate between sound and language are categorized as understanding (non-specific).

As might be expected from frequent use cases, the most commonly identified source of difficulty with SDSs is in unstructured queries and (non-specific) understanding. The indicated difficulties do not vary significantly between smart home and mobile technologies when proportions are taken into account, except for the case of text dictation.

### Desired Features and Improvements

Participants were asked to identify any perceived feature gaps in the commercial SDS they use most often. This question had the lowest rate of response out of the entire study, and many responses were highly specific and thus categorized as ‘other.’ Most of the commonly identified desired features are existing features in the SDSs most commonly seen in this study. Interestingly, respondents in this question did not specifically indicate speech recognition, unlike in every other question. Instead participants gravitated towards issues of NLU and NLG.

### Attractive Qualities of Commercially Available SDS

The last two questions of the study asked participants to discuss features, issues, or impressions that most attract or repel

Device Type	Desired Features						
	CS	UH	MD	LU	UNS	O	NS
Smart Home	0.14	0.11	0.14	0.04	0.07	0.25	0.25
Smart Phone	0.04	0.08	0.20	0.04	0.04	0.23	0.37
Other	0.00	0.50	0.25	0.00	0.00	0.00	0.25

Table 5: *Commonly desired features and improvements for commercial dialogue systems. CS = context switching, UH = user help functions, MD = multiple device integration, LU = language understanding, UNS = understanding (non-specific), O = other, NS = not sure. N = 108. N = 28 smart home users, 76 mobile device users*

them when using SDS and interacting with intelligent agents via voice.

Device Type	Attractive Aspects					
	Speed	EoU	Conven.	HF	O	NS
Smart Home	0.40	0.33	0.06	0.15	0.06	0.03
Smart Phone	0.34	0.25	0.06	0.26	0.17	0.01
Other	0.60	0.60	0.00	0.00	0.20	0.00

Table 6: *Most attractive elements of SDS and voice activated technology among older users. EoU = ease of use, Conven. = convenience, HF = hands-free, O = other, NS = not sure. N = 133. N = 33 smart home users, 95 mobile device users. This question was multiple response.*

Most responses to this question were fairly generic and did not vary widely. It is worth noting, however, that despite clearly indicated frustrations and issues with fundamental elements of SDS technology, like speech recognition and language understanding, many users still view the technology as easy to use and convenient.

### Undesirable Qualities of Commercially Available SDS

The last question in the survey echoed earlier, more general questions about smart technology and asked participants to identify their biggest source of frustration or least desirable feature in using SDSs.

Specificity in responses to this question varied wildly (“Not understanding me” vs. “Siri does not understand who the Packers are”), but follow the same general themes as other questions. Poor understanding of all types was by far the most commonly identified undesirable quality of SDS.

### Discussion

This study focused on older users, who as a user group are generally less targeted for study and perhaps more hesitant to adopt new technology. However, they are also the fastest

Device Type	Desired Features						
	SR	LU	UNS	SP	LC	O	NS
Smart Home	0.07	0.17	0.20	0.03	0.13	0.37	0.07
Smart Phone	0.21	0.12	0.23	0.04	0.04	0.28	0.09
Other	0.00	0.17	0.00	0.33	0.33	0.00	0.33

Table 7: Commonly identified sources of frustration and undesirable aspects of using SDS in home and mobile systems. SR = speech recognition, LU = language understanding, UNS = understanding (non-specific), SP = security and privacy, LC = learning curve/knowledge gap, O = other; NS = not sure. N = 126. N = 30 smart home users, 90 mobile device users. Multiple responses possible.

growing segment of the population worldwide, and a number of studies and applications indicate that an aging population may derive some of the greatest benefit from artificial assistants and other technological support in a variety of ways, including memory support and maintaining independence. This user group, therefore, cannot be overlooked and should be viewed as a great source for information about what a general user population truly wants out of dialogue systems.

While some of the findings of this study may echo previous work, the strength of this study is two-fold. First, the population size examined here is many times larger than most similar studies, which often boast population sizes of 30 or less. Second, the fact that this study focuses on a less well-examined portion of the user population provides insight into what features of dialogue systems are truly generalizable, and how developers and researchers can focus efforts to create systems tailored to the needs of relatively untapped potential user groups.

The absolute conclusion of this study is that user modeling and user intent recognition are perhaps the two most important problems underlying many of the issues and desires of the wider user population. Time and again, responses to questions indicate a desire for features and abilities that are underpinned by the need for more robust capabilities in these two areas. Interaction style preferences, perception of feature gaps, and common user frustrations all point to the fact that an agent that is better capable of predicting user actions, and tailoring its behavior to match the behavior of individual users, should be the most immediate goal in the development of further dialogue systems.

### Implications of User Preferences Over Interaction Style

A natural conversational relational agent is of course the end goal in developing dialogue systems and indeed almost any artificial agent. However, in this study users indicate fairly strongly that it is not the social ability of dialogue agents that interest them so much as the ability of agents to adapt to them. In fact, a significant number of users have a distinct preference for sterile agents that are not at all social. This indicates that more immediate efforts in research and

development should be directed towards problems like user intent recognition, context switching, and the development of a robust, adaptable user model that captures unique patterns in user behavior revealed by spoken interactions and transaction histories.

### Implications of Frequent Use Cases

One of the more surprising results of this study was that despite any perceived or actual capability gaps in dialogue systems, users have already embraced them as participants in less-structured vocal interactions. By far the most commonly indicated use for dialogue agents on both smart home and mobile platforms is in answering general or free-form queries such as might appear in a web search. The inclusion of general queries as one of the biggest sources of user frustration further supports that this is the use case users really want these systems to be sufficient for. The obvious conclusion from this result is that natural language understanding should remain one of the most important areas of research in developing dialogue systems. However, this also supports the conclusion that developing agents with better predictive capabilities based on a deeper understanding of user behavior and use patterns should be a major focus in assistive agent research.

### Implications of Common Frustrations and Perceived Feature Gaps

The root of many user frustrations is naturally in difficulties with both speech recognition and natural language understanding. However, a number of user frustrations and perceived gaps in agent capability actually stem from a lack of awareness of what these agents can actually do. A number of responses to this survey also indicated that the accessibility of educational materials is a barrier in both learning to use and being willing to learn to use these systems.

Part of the onus in addressing these issues certainly lies with the manufacturers and developers of commercial systems. However, the results of this study obliquely suggest that one of the best ways to address issues with the availability of educational materials is to keep user education as unimodal as possible. This is especially helpful in the case of elderly or otherwise impaired users who may have sensory deficits that act as barriers to accessing more traditional media avenues for user materials like tutorials and help menus. A number of studies already exist that indicate that dialogue systems are ideal for training exercises in a variety of applications and for all age groups, and may be especially helpful for older users in terms of cognitive support and skill maintenance (Dodd, Athauda, and Adam 2017) (Greenaway, Duncan, and Smith 2013). Literature in supporting adults with memory issues further paints intelligent assistants as ideal helpmates in situations where questions may be repetitive and users may need frequent reminders and guidance in procedural tasks (Hawkey et al. 2005) (König et al. 2016), such as might be the case with naïve users learning to use a new piece of technology.

## Conclusions and Future Work

In this work we present key results of a very large user study with older users of dialogue systems. The responses from study participants shows that while many wants, needs, and current use cases for older users mirror those of other user groups, older users do have some unique needs and desires, namely in learning to use new and continuously evolving systems. These results indicate that older users and other naïve users of SDS technology may benefit from more closely integrated system help and guidance, which means that problems like robust user modeling and intent prediction are vital in addressing the needs of a wider user population.

Based on the results of this study, we will focus on the development of a robust user model based on pattern detection and extraction from extended user interaction histories. To do this, we will be collecting user interaction data from naïve users of dialogue systems interacting with an intelligent personal assistant. After developing this dataset, we will use deep learning techniques to find features of voice interactions, including user intents and context tracking, that are unique to individual users and can be used as the basis for a generalized user model that can be learned and improved by dialogue agents with each successive user interaction.

## Acknowledgements

The authors would like to thank Charlotte Marien for assistance in data collection. The authors also gratefully acknowledge the support of an NSF NRT training grant, “Graduate Training Program in Sensory Science: Optimizing the Information Available for Mind and Brain” (Grant DGE-1734815) and the ROSE-HUB I/UCRC in undertaking this study.

## References

Bortfeld, H.; Leon, S. D.; Bloom, J. E.; Schober, M. F.; and Brennan, S. E. 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and speech* 44(2):123–147.

Dodd, C.; Athauda, R.; and Adam, M. T. P. 2017. Designing User Interfaces for the Elderly: A Systematic Literature Review. *Australasian Conference on Information Systems* 1–11.

Gao, J.; Galley, M.; and Li, L. 2018. Neural approaches to conversational ai. *arXiv preprint arXiv:1809.08267*.

Georgila, K.; Wolters, M.; Karaiskos, V.; Kronenthal, M.; Logie, R.; Mayo, N.; Moore, J. D.; and Watson, M. 2008. A fully annotated corpus for studying the effect of cognitive ageing on users’ interactions with spoken dialogue systems. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)* (1):938–944.

Greenaway, M. C.; Duncan, N. L.; and Smith, G. 2013. The memory support system for mild cognitive impairment: Randomized trial of a cognitive rehabilitation intervention. *International Journal of Geriatric Psychiatry* 28(4):402–409.

Hawkey, K.; Inkpen, K. M.; Rockwood, K.; McAllister, M.; and Slonim, J. 2005. Requirements gathering with

alzheimer’s patients and caregivers. In *Proc. 7th Int’l ACM SIGACCESS Conference on Computers and Accessibility*, 142–149.

Heerink, M.; Kröse, B.; Evers, V.; and Wielinga, B. 2010. Assessing acceptance of assistive social agent technology by older adults: the almere model. 2:361–275.

König, A.; Malhotra, A.; Hoey, J.; and Francis, L. E. 2016. Designing personalized prompts for a virtual assistant to support elderly care home residents. In *Affective Interaction with Virtual Assistants within the Healthcare Context*, 278–282.

Milhorat, P.; Schlogl, S.; Chollet, G.; Boudy, J.; Esposito, A.; and Pelosi, G. 2014. Building the next generation of personal digital assistants. In *Advanced Technologies for Signal and Image Processing (ATSIP), 2014 1st International Conference on*, 458–463.

Möller, S.; Gödde, F.; and Wolters, M. K. 2008. A corpus analysis of spoken smart-home interactions with older users. *LREC* 735–740.

Ortman, J. M.; Velkoff, V. A.; Hogan, H.; et al. 2014. *An aging nation: the older population in the United States*. US Census Bureau, Economics and Statistics Administration, US Department of Commerce.

Rao, J.; Ture, F.; and Lin, J. 2018. Multi-task learning with neural networks for voice query understanding on an entertainment platform. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 636–645. ACM.

Scheiman, M.; Scheiman, M.; and Whittaker, S. 2007. *Low vision rehabilitation: A practical guide for occupational therapists*.

Schlögl, S.; Garschall, M.; and Tscheligi, M. 2014. Designing Natural Language User Interfaces with Elderly Users. *Workshop on Designing Speech and Language Interactions at the ACM SIGCHI Conference on Human Factors in Computing Systems* 1–4.