I'm Sorry, Dave, I'm Afraid I Can't Do That: Chatbot Perception and Expectations

Jennifer Zamora
Google, Inc.
Mountain View, USA
zamora@google.com

ABSTRACT
Artificial intelligence continues to grow in popularity on mobile platforms, increasing exposure to chatbot apps. Chatbot technology has evolved over time, yet the purpose and added value that chatbots offer has not been clearly defined. In order to design a chatbot that provides a meaningful experience, we must first understand what expectations people have for this technology, and what opportunities are there for chatbots based on user needs. This study includes qualitative data from 54 participants in the US and India, sharing their expectations and experiences with a chatbot. The research objectives include: 1) understand user perception and expectations of chatbots 2) surface preferences for input modality and 3) identify domains where chatbots can add meaningful purpose.

Author Keywords
Chatbots; Artificial Intelligence; Mobile; Input;

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

INTRODUCTION

Artificial intelligence (A.I.) has grown in popularity for simulating conversations between bots and humans, particularly on mobile platforms. The functionality of these chatbots ranges from utilitarian to entertainment, but the value is often not clearly defined. The purpose and need for these chatbots are often not clearly defined. Curiosity and interest may spark an initial interaction with a chatbot, but to add more value to ongoing interactions we should identify a broadly acceptable role that has a defined purpose. What a chatbot is, and how to use one effectively, are new concepts that many struggle to define.

Chatbot experiences can happen through text or voice interactions, and can be more or less valuable depending on the context. Determining the preferred input modality means understanding the entire context of the individual involved, user end goal, and environmental variables.

Rather than defining a purpose based on the chatbot creator's perspective, we take a user-centered approach to understand how chatbots are perceived and experienced by people in their everyday lives. By understanding how chatbot experiences live up to expectations and how chatbot services compare to alternatives, we can begin to evaluate their performance and purpose. In this study we focus on the following objectives: 1) understand user perception and expectations of chatbots 2) surface preferences for input modality and 3) identify domains where chatbots can add meaningful purpose.

This paper starts with background and relevant research about chatbots. Following is the methodology used to evaluate chatbot perception and usage. We then discuss implications of the qualitative insights and end with areas for future work.

BACKGROUND

Now that chatbots are available on mobile devices, we can expect greater accessibility to this technology. According to comScore, a mobile measurement company, 50% of digital media in the US is consumed on mobile, and people spend most of that time in social and messaging apps [8]. The number of chatbot apps on mobile has steadily grown, as has the number of chatbot functions that live in communication platforms like Facebook Messenger, Slack, Telegram, and Skype. Accessing chatbots has become easier with this spike in development, and the first generational wave of chatbots in mobile chat apps can greatly define how people interact with this type of A.I. for future iterations.

RELATED WORK

Defining what a successful chatbot is can be a challenge in itself. For chatbots that strive to be assistive, they must at a minimum provide some degree of help. A useful virtual assistant can be described as one who is responsive, observes carefully, and provides assistance without interfering [10]. Acting as a successful virtual assistant also includes broader user interface principles like accounting for user goals, minimizing the cost of poor chatbot responses, and adding value in some form [11].

HAI '17, October 17–20, 2017, Bielefeld, Germany

ACM ISBN 978-1-4503-5113-3/17/10.
https://doi.org/10.1145/3125739.3125766

Perception of Human-like Behavior

A chatbot that is conversational is able to provide an experience that many menu-based products lack. Human-like behaviors, gestures, and appearances have been found to elicit social mannerisms from users that typically exist between humans [2,3,4,5,6,17]. A chatbot with human-like characteristics may combat the initial distrust that users can have towards computer-based systems [14].

Based on past literature, we see the importance of chatbots displaying human-like behaviors rather than computer-based behaviors for relationship building. There are many variables that may shape the expectations that a person has for chatbots. Age and behavior trends have been found to influence the types of responses and interactions one has with conversational agents [12]. Past studies have found users will even apply stereotypes when the conversational agents express human traits like gender indicators through voice [13, 17]. Acknowledging these findings places an importance on designing and evaluating chatbots with users who have diverse perspective if we are to achieve success with the general audience.

Input Modality

Designing an interaction as voice or text can be determined by variables such as the context of the task, the limitations of the software or hardware, and language considerations. Input preferences may also be influenced by specific language complexities. For example, voice input has been found to be faster and more accurate in English, Mandarin, and Indic languages when compared to keyboard input [9,19].

Multiple Languages

Blending of languages is another concern that arises during conversational interactions, regardless of input modality. This blending, referred to as code-mixing, can completely break down communication with a chatbot due to limitations in current language processing technology [1,16]. In regions with multiple prominent languages such as India, this may become a technical challenge for chatbots. Past research has also found relationships between language and input error rates [9, 19]. Multilingual users may have a preference for text versus speech input based on the accuracy for a particular language.

What Are Chatbots Good For?

A large segment of chatbots seek to fill the role of "electronic butler" by providing assistance to the user [10]. Common tasks, such as information seeking or other administrative needs, are objective and can be fulfilled by a chatbot. Some chatbots are also designed to attempt to build relationships between human and A.I. Developing a sense of trust is critical for building a healthy relationship. Interaction and visual designs can greatly influence the perception of how trustworthy a chatbot is. When users perceive embodied conversational agents as high quality artifacts, they are more likely to trust them [15]. However, in order to define high quality, we must first define what the expectations are for chatbots.

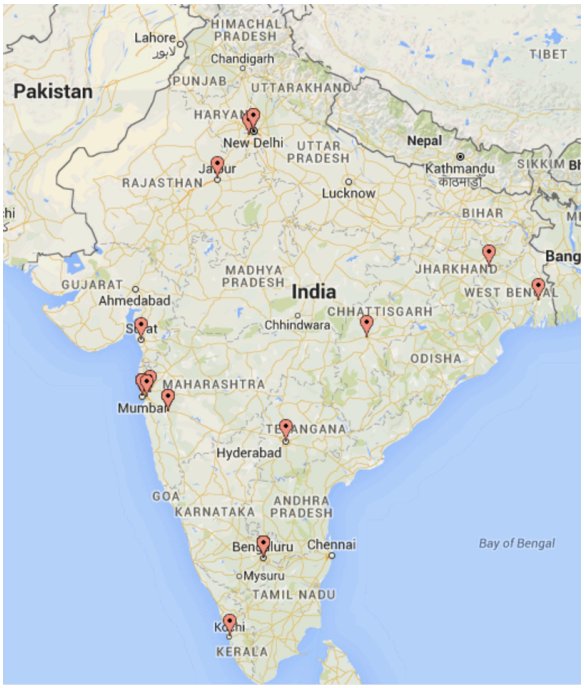

Figure 1. Twenty-eight participants distributed across India.

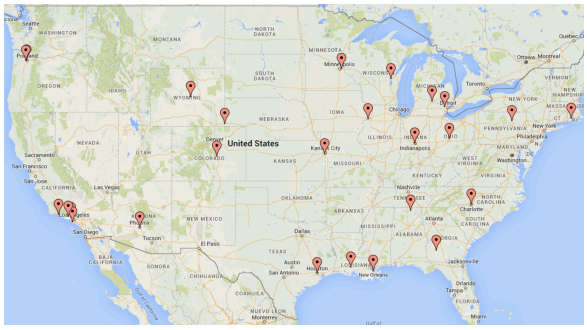

Figure 2. Twenty-six participants distributed across US.

METHOD

This research collected qualitative insights from 54 participants in India and the US over the course of 14 days (see Figures 1 and 2). Participants reported on their initial perspectives on virtual assistants, chatbot experience, text versus voice usage, and general daily routines.

Participant Demographics

Of the 54 participants, 28 were in India while 26 were in the US. The regional mix was intended to capture diverse perspectives and multi-language usage. The combination of India and US offers cultural contrast, and India has high prevalence of multilingualism. English is also a

shared language in both countries, reducing translation challenges.

All of the participants spoke English. The participants in India spoke two or more languages and felt comfortable reading and writing in English. While most participants reported speaking multiple Indic languages, their mother tongue breakdown included: Hindi (12), Kannada (4), Tamil (4), Malayalam (2), Marathi (2), Bengali (2), Gujarati (1), and Kutchi (1).

This study included 20 females and 34 males. Participants were between 24-40 years old and reported middle-class household incomes respective to their regions. Locales were dispersed across each country, but were generally metropolitan or peri-urban neighborhoods. Participants were engaged through a recruiting database and incentivized for their contribution.

Four participants in the US and 10 participants in India had prior experience with chatbots, but none were actively using a chatbot.

Chatbot Apps

Chatbot app regional availability and app store reviews were the most influential factors for selecting which apps to include in each country.

Participants in India were assigned one of the following chatbots: Hike messenger with built-in Natasha bot (n=8), Helpchat (n=10), Niki (n=10). Participants in the US were assigned one of the following chatbots to engage with: Indigo (n=9), Assistant.ai (n=9), and Siri (n=8).

Baseline and chatbot app usage

During this study, participants interacted with two mobile apps. In the first week users interacted with Google Search app that includes the standard search engine and Google Now, an intelligent assistive app that offers text and voice functionality. This experience would act as a baseline and point of reference since it is a common way to seek information to complete tasks. During the second week, users focused on engaging with an assigned chatbot (see Table 1).

Week 1	India & US	Google Search app (n=54)		
Week 2	India	Hike (n=8)	Helpchat (n=10)	Niki (n=10)
	US	Indigo (n=9)	Assistant.ai (n=9)	Siri (n=8)

Table 1. Fifty-four participants in India and US engaged with a baseline search app to reference, and during the second week of the study engaged with an assigned chatbot.

Chatbot Measurements

For both weeks participants completed tasks of their choice in addition to benchmark tasks, including ordering food, booking movie tickets, and looking up the weather.

Measuring perception and expectations

Four core traits were measured at the end of each phase to compare the overall experience of the baseline app and chatbot app. These objectives are based on communication flows between a user and chatbot. Ideally, a chatbot should be easy to engage in conversation, should understand what you are saying, and should provide sufficient, relevant information.

Responses were structured as a 5-point Likert satisfaction scale worded as extremely satisfied, very satisfied, moderately satisfied, slightly satisfied, and not at all satisfied.

Baseline benchmark question: *Thinking of your overall experience with the app, rate the following:*

- *Ease of starting a conversation*
- *Accuracy of understanding your words*
- *Quality of results*
- *Relevancy of content*

Participants expressed their expectations by providing three words describing what they think a good virtual assistant should be, and then followed up with examples supporting their expectations. This perspective was captured during the first day of the study, prior to interacting with the baseline search engine or the chatbot. The intent was to understand current expectations for a chatbot without being influenced by the actual experience. Through open word coding and thematic analysis, we were able to categorize the top traits participants expected from a virtual assistant in the form of a chatbot. To follow up on expectations of the chatbot performance at the end of phase two, users were asked if they plan to continue using the chatbot or revert back to their previous way of getting tasks done.

Measuring input preferences

Perceived input efficiency was used to evaluate speech versus text input. Users in India were asked to rate speech and text input efficiency for English and native Indic language.

Responses were structured as a 5-point Likert efficiency scale worded as extremely efficient, very efficient, moderately efficient, slightly efficient, and not at all efficient.

Perceived input efficiency questions: *How efficient is it to speak to a virtual assistant? How efficient is it to type to a virtual assistant?*

Measuring contextual input and appropriateness

This measurement identifies input preference based on contextual activities, as well as general fit for a virtual assistant. Users were asked to rate input preference for 20 routine tasks.

Response format was multiple-choice which included three options: speaking, typing, or not appropriate for virtual assistants.

Contextual input modality question: *Please select your preferred way of communicating with a virtual assistant when getting the following tasks done.*

Tasks included:

- *Catch up with social networks or latest news*
- *Check finances*
- *Get something done from a service provider*
- *Religious or spiritual needs*
- *Educational learning*
- *Find a job*
- *Pay bills*
- *Research a topic*
- *Order a meal*
- *Browse or research products to purchase*
- *Travel logistics*
- *Find something fun to do nearby*
- *Learn how-to do something*
- *Movie booking*
- *Manage daily commute*
- *Keep up with sports*
- *Communication like text or calls*
- *Take steps to be healthy (i.e. diet, exercise)*
- *Manage schedule*
- *Make a meal*

Data collection and analysis

Participants provided insights through daily surveys for two weeks. Email notifications were sent with a link to an electronic survey. The surveys were a mix of multiple choice and open-ended responses.

Analysis of the responses was a mix of qualitative open coding, scale ratings, and remote task-based evaluation. With this combined effort we were able to identify themes for the core objectives of chatbot expectations, input preferences, and opportunities for chatbots to provide meaningful experiences.

FINDINGS

Perception and Expectations

Participants rated satisfaction for four objectives for both the baseline and assigned chatbot app: ease of starting a conversation, accuracy of understanding your words, quality of results, and relevancy of content (Figure 3). Quality of results for chatbots was rated lower than the baseline due to information inefficiencies.

"It does the basics like sports, weather, time and alarms but general questions are often met with a Google search, so I might as well use the Google app, since it will just work better. The verbal syntax required to get the sports information is more verbose than the Google app and sometimes it still doesn't work." (P3)

In three words, participants expressed their expectations for what makes a good assistant. Through code-analysis, several themes emerged identifying four high-level

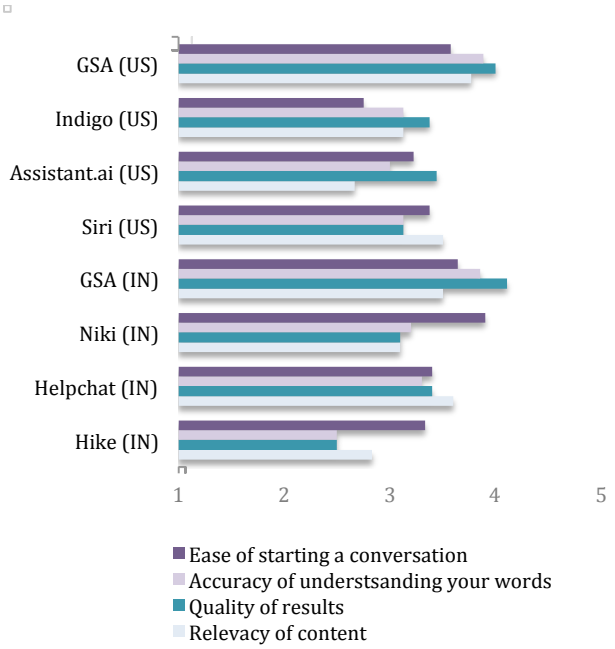

Figure 3. Fifty-four participants rated satisfaction on a five-point scale for four chatbot interaction objectives. Satisfaction is averaged with not at all satisfied coded as 1 and extremely satisfied coded as 5. Google Search App is abbreviated to GSA.

categories of expectations: high performing, smart, seamless, and personable.

High performing had 58 mentions with subcategories about good technical performance like fast, efficient, and reliable.

"Show me the fastest way to do something."(P17)

"It should save time." (P2)

Smart had 40 mentions with subcategories focusing on intelligence like knowledgeable, accurate, and foreseeing.

"It knows answers to all my questions." (P9)

"Anticipate timelines, for example departure times based on appointment location." (P19)

Seamless had 20 mentions with subcategories describing smooth experiences that are easy, and flexible.

"It should be all in one, easy to use." (P5)

"Everyday is different. [The virtual assistant] would have to be reliable and flexible to manage everything for me." (P23)

Personable had 16 mentions with subcategories specifying, "understands me" and likable.

"I like the idea of one who gets to know my personal quirks." (P18)

"It needs more to be truly likable and can anticipate my needs accordingly." (P43)

Meeting Expectations (or Not)

At the end of the study, participants were asked if they plan to continue using the chatbot. Twenty-seven participants reported that they would be unlikely to continue using their assigned chatbot. The most common reason was related to unmet expectations.

"No. The replies are slow and the app keeps bugging me with notifications that I won't care about. Besides, the whole thing doesn't really make my life much easier." (P42)

Another theme for discontinuing use of the chatbot app was due to low performance compared to their pre-chatbot experience. Participants compared the chatbot experience to how they would normally complete tasks. If the experience with the chatbot was not better than the current process, then the perceived value and incentive to engage was low.

"No. It doesn't distinguish itself from what I'm used to." (P4)

"I'll continue the same way things were before, because it's more convenient. My data is already saved in my wallet apps like Paytm and Mobikwik. I can pay bill in 2 clicks. If I get good offers and discounts or Cashback, I can consider it." (P36)

Among the participants, 22 indicated interest in continuing their chatbot app usage. Rationale was often centered on efficiency offered by the chatbot, or convenience.

"Yes, I am likely to use Niki, probably the cab service. It is way way faster than using the individual apps." (P25)

"I plan to continue to use Siri because it's installed on my phone." (P6)

Input Preferences and Multi-language

Language input efficiency was perceived lower for native Indic experiences in comparison to English input modalities (Figure 4). Participants also commonly needed to blend languages for local terms and names. For example when browsing movies, food, and music, it was important that the software interaction understood a blend of English with Indian words. In many cases, there was no translation of the musician, movie title, or local cuisine, so the user had no choice but to use words of another language. This often led to a poor experience where the chatbot was not able to accurately identify the user's request.

Contextual Input Preferences and Appropriateness

Participants rated their input preferences for common tasks and explained why they selected voice, text, or "not appropriate for a chatbot." Themes in input preferences were largely driven by user goal and context.

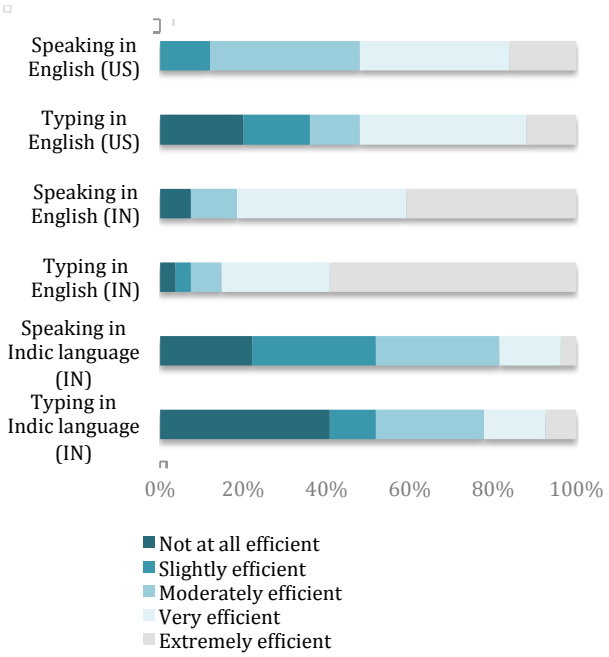

Figure 4. Fifty-two participants rated input modality efficiency for speech and text when communicating with a chatbot. US participants (n=25) rated input efficiency for input in English. India participants (n=27) rated input efficiency in English and their primary Indic language.

Speaking to a chatbot is best when the user is multi-tasking, hands or eyes are occupied, or while they are moving and unable to be stationary. Examples supporting these themes include managing daily commute, making a meal, or observing an instructional tutorial. In each scenario, it is common to have your hands engaged in an activity and your eyes fixed on a stimulus other than the chatbot.

"I find that I mainly use speaking for things like directions, checking the weather, or sports scores when I need to be hands-free (my phone connects to my car via bluetooth)." (P13)

Typing to a chatbot is best when the activity is complex, includes a confirmation step, or requires logic. In each of theses scenarios there is some degree of control that the user must maintain. According to the participants, this control is often manageable through text as opposed to voice input.

"If it is something that required multiple communication between chats, or like for example, looking up information from more than one site." (P14)

"I would say it's best for typing when you need to make sure that the idea you are conveying to the app is perfectly understood. You don't want the app to misunderstand you and do something wrong." (P5)

Things that are not appropriate for a chatbot often revolved around sensitive topics such as financing or social media content.

"Personal data such as bank account details are not appropriate for a virtual assistant." (P53)

"Anything regarding social media is too personal." (P12)

Administrative Tasks and Emotional Needs

Common administrative and menial needs such as scheduling appointments, setting reminders, booking tickets, and sharing traffic or weather updates were highly valued. This is likely related to the fact that most chatbots in the market cater to these needs. A theme for how an assistant chatbot is beneficial is that this technology can take care of the little things while the person focuses on more important, big-picture needs.

"It would be helpful because I am a working mother and all free time is precious to me and spending time with baby." (P29)

"This would be useful because it will reduce unnecessary burden and stress on my mind so that I can work productively." (P7)

Participants expressed an interest in fulfilling emotional needs, such as someone to provide motivation or someone to listen. The idea of a chatbot that a person can talk to is appealing, and one unique characteristic about this experience is the lack of perceived judgment. Several participants noted that a chatbot is ideal for sensitive content that is too embarrassing to ask another human about.

"*Encouragement, motivation, or instructions when needed is the most desirable... Someone who believes in them and motivates." (P48)*

"If a virtual assistant could be someone who I can discuss things with and ask questions to and ask for feedback I would be very happy." (P11)

DISCUSSION

Many respondents pointed out that they were unclear of what the chatbot could do, leaving their expectations undefined. Expectation setting will be imperative as users are likely to shape their satisfaction and intent to engage based on how their expectations were met.

Better, Faster, Stronger

Meeting expectations will not be as simple as checking off each box on a list. Each expectation is relative to the user's current experience meaning that the perceived speed and efficiency will be held to the standards of the alternative services that exist for that individual. In many cases participants reported that the chatbot did not make things faster and often felt like a "middleman." As things currently stand, information seeking or completing complex tasks using a standard search engine may be more efficient and informative than a chatbot. This is partly due to familiarity of interacting with search engines versus a chatbot. In addition, as the technology itself continues to develop and learn from ongoing engagements, the experience has the potential to exceed the current standard of internet queries.

Designing For Input Modalities

Input preferences are highly influenced by the environmental factors in each context. For example, text is most convenient for complex tasks because dialogue can be challenging to formulate. Participants noted that they were not sure what they could say to the chatbot, or how to phrase it so that they could communicate effectively. Providing a secondary input channel such as displaying menu options or including voice input in addition to text allows for better experiences that can reduce error or recovery time [7,18]. Designing the best experience means accounting for the entire context case by case, which can help prioritize which input design makes the most sense.

Multi-language Experience

Nearly all of the participants in India preferred speaking in English with the chatbot, not necessarily because of preferred language, but because of perceived success and error rates. This can have negative consequences if the multilingual individual choses to speak in English for technical accuracy reasons but is not as expressive in English. Having limited vocabulary hinders communication with technology that is already limited in its language flexibility. Across both regions, communication with the chatbot was described as structured because commands need to be specific in order to successfully communicate with the chatbot. Some participants noted that they would actually prefer to use their local dialect if it were possible. Being able to do so could have implications on the relationship building process and trust that can develop between the user and chatbot.

"With English, the results were extremely good. This was not at all the case in Hindi. I'll just stick with English, enough said." (P47)

"I'm so used to having English everywhere in my phone, I wouldn't like to change it. But, having my phone speak Kannada with me will make me feel closer to my phone..." (P50)

Code-mixing

The need to code-mix is important for those who speak multiple languages. This will be critical for regions that use languages such as Hinglish, an evolving blend of Hindi and English. Without this capability, the results and responses from the chatbot were poor. If the chatbot is to be seamless, it will need to be able to easily transition between languages.

A Chatbot As Your Admin and Confidant

Menial tasks that are simple and straightforward appear to be a good fit for chatbots. Users want services that will increase efficiency and save time. Examples were often related to routine daily tasks that require little trust or human logic and result in low consequences if failures occur. However, some menial tasks that have heavier consequences, like social risk or financial responsibilities, are considered to be sensitive and would require more trust development prior to engagement.

Subjective needs, like emotional support, were of interest to participants. Several participants identified this concept as a void to be filled. Providing for emotional needs will be a challenging objective, but also an opportunity not fulfilled by other technology. As seen in past research, there is initial distrust with computer-based systems [14], and doubt surrounding chatbot abilities to "think like a human." Emotional logic is not currently seen to be within chatbot abilities, and building an application intelligent enough to converse about subjective needs like emotional support, reducing anxiety, and responding in a relevant way is still a major challenge for developers.

Judgment and privacy

The ability to have a conversation and gain insights on sensitive topics without being judged is perceived as valuable. The lack of judgment is a unique aspect of this technology that opens opportunities to expand on chatbot development and use cases. However, one note to acknowledge is the theme of privacy concerns. Several participants voiced privacy concerns around content that was viewed as sensitive or high risk. Topics like finance and social media for chatbot interactions made some participants uncomfortable. This is due to concerns around mishandling data, including leaks. While it is possible to build a relationship, developing trust will be required for meaningful interactions, especially if there is high-risk data involved.

"You are giving them access to data that might not even be accessible to someone's closest human companions. If that relationship is respected, and a bond of trust is established and maintained, I would not have any tasks that are too personal." (P51)

"There's freedom to ask a robot things that you would feel embarrassed by a person doing." (P52)

"It would be so easy to get feedback from virtual assistant as it would not be embarrassing for me :)" (P43)

Limitations and Future Work

It was only logistically feasible to recruit participants in two countries for this study. These insights do not represent entire countries, particularly our sample in India because English speakers make up a minority of the general population. This study also focused on initial experiences that were contained in one week's time. For future research, looking beyond English speakers and the first week experience can capture richer insights.

CONCLUSION

We have evaluated the perception of chatbots by defining user expectations for chatbots to be high performing, smart, seamless, and personable. We have identified the influence of contextual cues for input preferences, and perceived challenges of native Indic language input. Lastly, we have recognized opportunities for chatbots to add meaningful experiences through improving efficiency of administrative tasks and the potential for acting as a non-judgmental confidant for sensitive topics.

AKNOWLEDGEMENTS

Thank you to all the participants who shared their experiences, the participant recruiters who staffed the study, the researchers who guided this effort, and the individuals who provided feedback on this paper.

REFERENCES

1. Umair Z Ahmed, Kalika Bali Monojit, Choudhury Sowmya VB. 2011. Challenges in Designing Input Method Editors for Indian Languages: The Role of Word-Origin and Context. In *Proceedings of the Workshop on Advances in Text Input Methods* (WTIM '11).
2. Gene Ball, Dan Ling, David Kurlander, John Miller, David Pugh, Tim Skelly, Andy Stankosky, David Thiel, Maarten van Dantzich, and Trace Wax. 1997. Lifelike computer characters: the persona project at Microsoft. In Bradshaw, J., ed., *Software Agents*. Menlo Park, CA: AAAI/MIT Press.
3. Linda Bell and Joakim Gustafson. 1999. Interacting With An Animated Agent In A Spoken Dialogue System. In *Proceedings of Eurospeech* '99.
4. Justine Cassell and Timothy Bickmore. 1999. Embodiment In Conversational Interfaces: Rea. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '99).
5. Justine Cassell and Katie Thórisson. 1999. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* 13, 519-538.
6. Loredana Cerrato and Susanne Ekeklint. 2002. Different ways of ending human-machine dialogues. In *Proceedings of Embodied Conversational Agents.*
7. Philip R. Cohen 1992. The role of natural language in a multimodal interface. In *Proceedings of the 5th annual ACM Symposium on User Interface Software and Technology, 143-149.*
8. comScore. 2016. Smartphone apps are now 50% of all U.S. social media time spent. Retrieved February 21, 2017 from https://www.comscore.com/Insights/Blog/Smartphone-Apps-Are-Now-50-of-All-US-Digital-Media-Time-Spent
9. Prasenjit Dey, Ramchandrula Sitaram, Rahul Ajmera, and Kalika Bali. 2009. Voice key board: Multimodal indic text input. In *Proceedings of Multimodal Interfaces,* 313-318.
10. Sara Reese Hedburg. 1998. Is AI going mainstream? A look inside Microsoft Research. IEEE *Intelligent Systems*, March/April 21-25.
11. Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '99).
12. Robert Hubal, Diana Fishbein, Monica Sheppard, Mallie Paschall, Diana Eldreth, and Christopher Hyde. 2008. How do varied populations interact with embodied conversational agents? Findings from inner-city adolescents and prisoners. *Computers in Human Behavior, 24, 3, 1104-1138.*
13. Max Louwerse, Arthur Graesser, Shulan Lu, and Heather Mitchell. 2005. Social cues in animated conversational agents. *Applied Cognitive Psychology*, 19, 6, 683-704.
14. Bonnie M. Muir. 1987. Trust between humans and machines, and the design of decision aids. *International Journal of Human-Computer Studies*, 27, 5-6, 527-539.
15. Susanne van Mulken, Elisabeth André, Jochen Müller. 1999. An empirical study on the trustworthiness of life-like interface agents. In *Proceedings of the HCI International* '99 (the 8th International Conference on Human-Computer Interaction).
16. Peter Muysken. 2000. *Bilingual speech: A typology of code-mixing.* Cambridge University Press, Cambridge, UK.
17. Clifford Nass and Scott Brave. 2007. Wired for Speech: How voice activates and advances the Human-Computer Relationship. The MIT Press, Cambridge, US.
18. Sharon Oviatt and Robert vanGent. 1996. Error resolution during multimodal human-computer interaction. In *Proceedings of the International Conference on Spoken Language Processing*.
19. Sherry Ruan, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James Landay (2016) Speech Is 3x faster than typing for English and Mandarin text entry on mobile devices. arXiv preprint arXiv:1608.07323, 2016.