



Designing a Smart Speaker for Emergent Users: Human Plus AI Response

Shashank Ahire

Human-Computer Interaction

Leibniz University Hannover

Hannover, Germany

shashank.ahire@hci.uni-hannover.de

ABSTRACT

This paper reports on the development of a smart speaker for the home setting of ‘emergent’ users – those whose technology experience and resource availability are low. Earlier research has shown that AI (Artificial Intelligence) powered smart speakers struggled in recognising many emergent users requests. On the other hand, smart speakers powered by human responses were more accurate but slower. In this study, we began by determining, given a choice, emergent users prefer a smart speaker enabled by a human response or an AI response, and what are their preference criteria. We found that they were not completely inclined towards either of those choices. Rather they preferred a smart speaker based on three factors: first, the language of the request, second, the length and complexity of the request, and third, the urgency of response. We developed an integrated version of the smart speaker and evaluated it with emergent user families. From our analysis, it was evident that, when combined, AI and human responses complement each other and provide an elaborate and richer response for emergent users.

CCS CONCEPTS

• **Human-centered computing** → **Sound-based input / output; Human computer interaction (HCI)**.

KEYWORDS

emergent users, speech interfaces, human-AI interaction, smart speaker

ACM Reference Format:

Shashank Ahire. 2022. Designing a Smart Speaker for Emergent Users: Human Plus AI Response. In *Proceedings of the 13th Indian Conference on Human Computer Interaction, 2022 (India HCI 2022), November 9–11, 2022, Hyderabad, India*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3570211.3570217>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

India HCI 2022, November 9–11, 2022, Hyderabad, India

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9982-1/22/11...\$15.00

<https://doi.org/10.1145/3570211.3570217>

1 INTRODUCTION

Speech offers several advantages over a typical GUI – for instance, speed of input, natural and intuitive interaction, provision of performing a task from a distance. Importantly, speech has no direct correlation with literacy. Thus, it is beneficial not only for mainstream English speaking users but also for ‘emergent’ users of technology, who are less educated, economically disadvantaged, and culturally heterogeneous [5]. Due to less education or education in regional languages, emergent users struggle to use English-dominated technological devices. Previously, many studies have demonstrated speech as a valuable interface for emergent users in various domains, such as healthcare [15], agriculture [4, 13], education [7], and community engagement [6].

Smart speakers are increasingly popular in emerging markets like India. In India, sales of smart speakers have increased from 14% to 39% in 2018 [9]. There are several reasons for an increase in sales, but one primary contributing factor is support for Indian languages. English-dominated AI smart speakers, such as Amazon’s Alexa and Google Home, are now supporting prominent Indian languages like Hindi, Punjabi, and Marathi [8, 17], and are aiming to add support for more Indian languages [16]. Although the reach of these smart speakers is increasing in India, it remains unknown how well these smart speakers perform in recognizing Indian languages and how well they can be embedded in an Indian house-hold setting.

Few studies have investigated voice assistant usage among low-income groups in India. Bhalla et al. [3] explored the usage of voice assistants and voice search with middle income and low-income groups in India. Authors reported that literacy, language, and privacy played a vital role in the Indian context. Likewise, a survey conducted in “Rise of the Chatbots” [18] claimed that a chatbot that is distinctively communicating in a local language made users more responsive towards it.

In the “StreetWise” study, Pearson et al. [12] compared human and AI responses in a public setting (uncontrolled environment) of Dharavi, Mumbai. They found a higher error rate as a significant issue with AI-enabled smart speakers. But, it is unknown which factors led to higher error rates (noisy environment, voice recognition, or request interpretation issues). Further, systems were deployed 1 km away from each other. Thus, users did not have an opportunity to compare, evaluate and understand the limitations of the both systems. However, in our study we evaluated prototypes that were inspired by previous work [12], but in the home setting (a controlled environment) of emergent users. Evaluation in home environments allowed us to identify the causes of errors, user’s preferences, and factors that are influencing their preferences.

In this work, we aim to address three main research questions: (1) Given a choice, do emergent users prefer a human-powered smart speaker or an AI-powered smart speaker response? (2) What are the factors that influence the selection of a human-powered smart speaker or an AI-powered smart speaker response? (3) Does providing a combined human-plus-AI response prove beneficial to emergent users?

In this paper, we started by performing a simultaneous deployment of human and AI-powered smart speakers in the home setting of emergent user families. In this deployment over 14 days, we intended to learn smart speaker usage by emergent users and their preferences for a system. From post-deployment interviews, we found that each smart speaker has its limitation and benefits. Hence, the users did not have an overall inclination towards a single system. Instead, the preference was based on type of request to be directed. Additionally, from the data collected in this deployment, we determined factors that played a crucial role in deciding the preference of emergent users towards AI and human-powered systems. Subsequently, we built a smart speaker that provided a human and an AI response for each question. After the evaluation of the smart speaker for ten days. The evaluation showed that human and AI responses were comprehensive and satisfactory for emergent users.

This paper has three major contributions; first, we performed a comparative study to determine the emergent user’s preferences for an AI-powered and human-powered smart speaker. Second, we identified factors that influenced emergent user’s preferences while directing a request. Third, we performed an investigation to identify if the human and AI responses were relevant and satisfactory for emergent users.

2 PILOT DEPLOYMENT

2.1 Prototypes

For comparison, we selected two open-source prototypes, Google Voice Kit¹ (GVK) and Human Power Delayed² (HPD). GVK was powered by AI for answering the questions, whereas HPD had a human responding the questions. In the following sections we explain the working of both the prototypes

2.2 GVK Interaction

GVK is an open-source voice assistant kit by Google (shown in figure 1). GVK consist of a simple button-based interaction. First, the user has to push the translucent button, then direct their request. Next, GVK captures the request and instantly responds to the user.

2.3 HPD Interaction

The HPD prototype (shown in figure 2) is based on the Wizard-of-Oz technique. To direct a question to HPD, users have to push a blue button and then ask their question. Following this, the prototype will provide a unique reference number for each question. A human moderator will receive the question and respond with an appropriate answer in the span of 10 minutes. After 10 minutes the user has to input a reference number to retrieve an answer that is provided by the moderator.

¹<https://aiyprojects.withgoogle.com/voice/>

²<https://github.com/reshaping-the-future/streetwise>



Figure 1: GVK prototype



Figure 2: HPD prototype

2.4 Participant Families

We recruited three families located in Phule Nagar – a slum-based setting near Powai, Mumbai. The families represented typical emergent user families located in a slum: multilingual, less educated, and with a limited income source. They owned a 15×15 feet house. In total the families consisted of 9 members (5M, 4F); three members were children in the age-group of 5-13 years, 5 members were in the age-group of 35-50, and one member was above 60 years old. All members were fluent in the Marathi language and reasonably competent in speaking the Hindi language. Children had a decent command of the English language. To ensure the confidentiality of participants, we assigned a participant identifier [P#] to each participant.

2.5 Method

To determine emergent user preference towards AI and human-powered smart speakers, we deployed the GVK and HPD prototypes neighbouring each other for a period of 14 days. We trained the participants on how to interact with both of the prototypes. We demonstrated each prototype on different types of questions (Basic fact, General informative, Contextual questions and Domain specific) stated in Robinson et al. [14]. Later, participants were given a chance to try their own questions. Further, they were informed about what data was collected by the prototypes. Each family was compensated with an honorarium of \$12/week.

2.6 Data Collection

We audio-recorded each request and response with their respective timestamps. We transcribed the audio recordings in text form for the analysis. We also audio-recorded the interviews with families. Later, the recording was transcribed into text for thematic analysis.

2.7 Results

2.7.1 Descriptive Analysis. Overall, there were 607 requests in total. Among these, 528 requests were directed to the GVK, and 79 requests were posed towards the HPD prototype. Table 1 shows the comparison of GVK and HPD prototypes on various factors. On comparing unanswered responses, the GVK failed miserably in answering many questions. It delivered a sorry response to more than half of the requests (57%) – out of 528 questions, 300 requests were unanswered. The number of requests directed towards the GVK prototype were 6.5 times the number of requests directed to the HPD prototype.

The HPD prototype had 76% of requests in the Hindi language, whereas GVK had only 51% requests in the Hindi language. Similarly, GVK received 27% English requests while HPD had only 4%. The request’s word count on both systems had a significant disparity. The mean word count for GVK requests was 2.9 words, whereas, mean word count for HPD requests was 6.2 words. The maximum length of requests in the HPD prototype was 18 words; comparatively, GVK had a maximum length of only ten words.

Table 1: Comparison of GVK and HPD on various factors

Factors	GVK	HPD
Total request	528	79
Request in Hindi	51%	76%
Request in English	27%	4%
Unanswered requests	57%	0%
Average word count of requests	2.9	6.2
Average word count of responses	18.18	13.01

We categorised questions using the Robinson et al. [14] categorization criteria. According to the categorization criteria, the questions related to the information about a particular person or a place, belonged to the category of ‘Basic fact questions’. For example, ‘Who is the father of our nation?’ or ‘How many districts does India have?’. The questions specific to the place, time, or an artifact were categorized as ‘Contextual questions’. For instance ‘Tomorrow’s weather’ or ‘Teachings of Mahabharata’. ‘Domain specific’ questions were linked to requests such as ‘Find a place’, ‘Play me a song’, or asking for an update about the specific movie or a movie character. For instance, ‘Motu Patlu jokes’ or ‘Play Bahubali Cartoon’. Next, the ‘philosophical questions’ category had questions like ‘Who is the father of Google Maharaj’ or ‘What is your name (system)’. Also, there were many instances in which questions were half captured or were meaningless. In this case they were labelled as ‘Not-a-Question’.

In addition to the categories mentioned in Robinson et al. [14], there were questions which did not belong to either of these categories. For example, some questions were not specific and had only

Table 2: Comparison of GVK and HPD on different categories of questions

Categories	GVK	HPD
Basic Fact	31%	45%
General Informative	22%	22%
Contextual Questions	2%	2%
Domain-Specific	31%	17%
Philosophical Questions	2%	1%
Conversational	2%	0%
Not-a-Question	10%	1%

Table 3: Comparison of answered and unanswered questions directed towards GVK

	Unanswered	Answered
Basic Fact	29.90%	32.89%
General Informative	23.59%	20.18%
Contextual Questions	2.99%	3.95%
Domain-Specific	29.24%	27.19%
Philosophical Questions	1.66%	2.19%
Conversational	0.66%	5.70%

one or two words. These questions were either the name of the place, person or state, for example ‘Rangoli’, ‘Madhuri Dixit’. For such questions, a new category was created called ‘General Informative’. Also, there were requests in the form of a conversation such as ‘Will talk to you (system) tomorrow’, ‘Good morning’ and ‘Hello’. Such instances were grouped in the category of ‘Conversational Instances’. Table 2 illustrates the proportion of different categories of questions in GVK and HPD

In category-wise comparison (as shown in table 2), GVK had an equal number of domain-specific and basic fact questions (31%). HPD mainly consisted of basic fact questions with a coverage of 45%. General information had the same number of questions on both the prototypes. On the other hand, GVK received 10% of ‘Not-a-Question’ requests.

GVK responded to 57% of the questions with a ‘Sorry’ response. Similarly, in case of ‘Streetwise’ [12] it was 62% of the questions that had an irrelevant response. So it would be intriguing to investigate answered and unanswered responses as per their category. Table 3 shows the comparison of answered and unanswered questions for GVK. From the values in each category, it is evident that answered and unanswered responses in each category are close to each other. Furthermore, one way ANOVA showed no significant difference in the values of either category ($F(6,7) = 65.15, p = 8.69$).

2.7.2 Qualitative Analysis. In the case of GVK prototype, users criticized its consistent inability to recognize their requests. One of the family members stated:

“It does not understand my questions; it understands 1 out of 100 requests. It is not accurate and sometimes gave random answers, which discouraged me to use the prototype.” [P3]

Moreover, a second family member complained: “On directing a same question, it didn’t reply me, but it replied to my brother.” [P4]

In the HPD prototype, users did not like the concept of remembering the reference number and inputting it later to retrieve their response. One member commented *“The prototype was boring, waiting time is high for some questions.”* [P6]

Another member added *“It didn’t consider urgency for responding questions, e.g.: What is the current time? Such requests should be responded instantly.”* [P8] A female adult member was also unhappy with the size of HPD, *“HPD is bigger and consumes more space, in comparison to GVK. Due to the small housing area, the space occupied by a particular object is important for us.”* [P9]

Despite facing issues in GVK, participants were impressed with the clear voice and its size. Since, they could place it anywhere in their house. On the other hand, in the case of HPD, although they found it bulky in size and were annoyed to remember request numbers, the users were impressed with the informative and richer response.

2.8 Discussion

2.8.1 Accent and dialect. The GVK system failed to detect the accent and dialect of family members. When two family members posed a same question, GVK replied to one of the member but responded ‘sorry’ to other family members. GVK’s inability to understand their accent discouraged some family members from using it. The consistent failure of the speech recognition engine to detect different accents was also noted in earlier studies [10–12]. On the other hand, HPD answered all questions irrespective of accent and dialect.

2.8.2 Preference criteria. HPD had 76% of requests in the Hindi language, whereas GVK had only 51% Hindi requests. Moreover, considering the average length of the requests – GVK (2.9 words) and HPD (6.2 words). It is evident that, HPD had long, complicated and Hindi language requests. Whereas GVK was only preferred for short, uniform and English language requests. Lastly, they also preferred a smart speaker based on the urgency of a response.

2.8.3 No overall inclination. GVK prototype was primarily used for requests, which required an instant response. However, users were unhappy with its failure rate and quality of response. On the other hand, the HPD prototype better understood requests and delivered a relevant response which users found helpful. However, noting down the request number and inputting it caused them inconvenience. Also, they were disappointed with the delay in receiving a response, which caused them inconvenience. Overall, from the analysis, it is evident that users preferred both prototypes. Due to no complete inclination, towards any prototype we decided to build a smart speaker consisting of both human and AI responses.

3 DESIGN OF THE SMART SPEAKER

From the data analysis, we found that AI-powered and a human-powered smart speaker had their limitations and benefits. We were motivated to design a smart speaker which could deliver responses instantly and also had a low error rate. Smart speakers should be able to respond to every request of the user. Hence, we decided to combine the human and AI responses.

Further in qualitative interviews, the emergent users highlighted the issue of space occupied by the prototypes, particularly the HPD

device. Due to the small housing area, the space occupied by a particular object is always given a priority while buying it. From these interviews, it was clear that the users desired a smart speaker which is appropriate to their house setting. Considering this, we decided to design a smart speaker for emergent users.

Our smart speaker (Figure 3) consists of two buttons: a request button (translucent) and response button (blue). Due to diverse age groups in participant families, we used two separate buttons to keep the interaction simple and intuitive. The request button was used to pose a question, and the response button was meant to retrieve a human response. Similar to the pilot prototype, an AI response was delivered instantly after posing a question. To build a GVK and HPD integrated smart speaker, we synchronized their process, such that they function simultaneously. We created individual processes for each system, known as GVK-P (Google Voice Kit - Process) and HPD-P (Human Powered Delayed - Process). Both processes were initiated on a request button push.

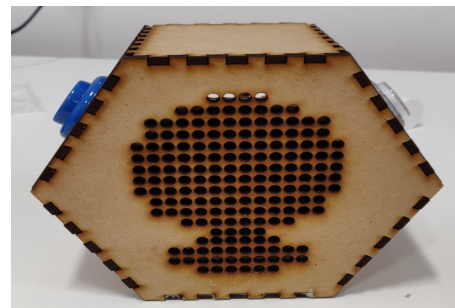


Figure 3: Design of the ‘Human plus AI’ smart speaker

3.1 Smart Speaker Interaction

The operation of the prototype is divided into three stages: posing a request, downloading a response, and delivering a response.

3.1.1 Posing a Request. On a request button push, the button LED lights up as a cue to the user to direct their request. Simultaneously GVK-P and HPD-P will initiate capturing the user’s request. Subsequently, the smart speaker will acknowledge the request by playing *“Thank you for your request”*.

3.1.2 Response Delivery. Like the pilot prototype, GVK-P provides an answer instantly, whereas HPD-P forwards the question to a moderator. After approximately ten minutes, when a participant presses the blue button, the HPD process plays the latest downloaded ten responses in descending order of arrival. After hearing a sought response, the user could again press the blue button to break the responses loop.

3.1.3 Response Download. At an interval of every two minutes, the smart speaker pings the HPD server to check for a new response. If a new response is uploaded, it will download the response. To notify the user of the downloaded response, the smart speaker will lighten up the button LED and play *“New answer has been downloaded”*

4 FINAL DEPLOYMENT

We performed a longitudinal evaluation with the same families for a period for ten days. We trained the family members on our smart speakers and informed them about the LED and audio cues. Further, we developed a video tutorial.



Figure 4: ‘Human plus AI’ smart speaker pictured in-situ during deployment

4.1 Findings

4.1.1 Primitive analysis. Overall there were 129 interactions with the smart speaker. Out of 129 interactions, 53 (41%) requests, GVK-P replied with the ‘sorry’ response. However, in the case of HPD-P, only 4 (0.3%) requests had no response. The mean response length for HPD-P and GVK-P was 6.9 and 8.6 words, respectively. Therefore, GVK-P had a 20% lead in the response length comparison. The total mean word length for each response was 15.5 words per request.

4.1.2 Response Discrepancy. Users encountered response discrepancy at multiple instances, GVK-P answered questions in English but failed to answer the same question in the Hindi language. For instance, “Which snake builds a nest?”, it responded accurately when the question was posed in English but replied with a “Sorry” response when asked in Hindi.

4.1.3 Multilingualism. HPD-P responded to the questions which were in dual languages. Particularly, “Who built Taj Mahal?”. In this question, the word “who” was in the Hindi language, whereas the word “built” was in the Marathi language. The amalgamation of two languages is frequent in a multilingual country like India and specifically in metropolitan cities like Mumbai. HPD-P was efficient in responding to multilingual, grammatically incorrect questions and with diverse dialects and accents.

4.1.4 Power consumption concern. Power consumption is an essential factor for emergent users. The adult member of the family was concerned about power consumption by the device. The member expressed his concern by stating “It works well, but I do not know how much power it consumes. Like smartphone, it should work on the battery and should give clear indication of battery consumption.” [P8].

4.1.5 Performance balance and enhancement. Users admired the idea of receiving diverse responses from a single source. They were impressed by the variation and detailing they received from two

responses. GVK-P responded with multiple responses for a single question. For the question “Which snake builds a nest?” the GVK-P gave following three responses: (1) King Cobra (2) Here’s a summary from the website livescience.com: (3) The king cobra, will build a nest for its young. Out of 76 delivered responses by GVK-P, it delivered multiple responses for 26 requests. One user commented, “Human answers were short and precise whereas the Google answers were long and elaborate.” [P2]

To this other user added “Smart speaker helped me to get detailed information on the particular topic. This variation in response helped me to understand a complicated topic.” [P1]

On several instances, participants observed that integration of both techniques balanced the performance of smart speaker. When GVK-P failed to deliver a response, HPD-P gave a delayed but relevant answer. Similarly when user required an urgent response GVK-P delivered it instantly.

5 DISCUSSION

In pilot deployment, overall emergent users did not incline to a human-powered or AI-powered system. Rather, they preferred a particular system based on three factors: first, the language of a question to be addressed, second, the length of a question and third urgency of the response. Request in the Hindi language were 25% (shown in Table 1) higher in the HPD system than in the GVK system. Therefore, it is evident that users preferred HPD for non-English requests. Also, GVK consisted of 27% requests in the English language, but HPD had only 4% requests in English. Hence, they preferred a system based on the language of a request, i.e. GVK for English, whereas HPD for Hindi requests. Further, the mean word count of HPD requests was 6.2 words, whereas, for GVK, it was 2.9 words. A difference of 3.3 words is a clear indication of preference for HPD for long and complex requests and for GVK for short requests. Moreover, the user’s preference was also based on the urgency of the request. If they required a response instantly, they would prefer the GVK system.

From the analysis, it is evident that a symbiosis of human and AI responses added value to the overall strength of the smart speaker. In total, users received an average response of 15.5 words per question, which illustrates the comprehensiveness and richness of smart speaker responses. We found that two factors magnified the efficiency of the system: first, multiple responses delivered by Google, due to its large corpus of data. Second the knowledge of multiple languages and contextual settings by a human in HPD-P responses. In a multilingual country like India, it is frequent to mix two languages in a sentence. It is difficult for an AI-powered smart speaker to interpret a sentence containing phrases in two languages.

Although the speech interface has been proved beneficial for emergent users, the current mainstream smart speaker’s are substantially incapable for addressing requests of emergent users. With more than 50% error rate, AI-powered smart speaker has failed to achieve the expectation of emergent users. With significantly higher error rates it is likely that AI-powered smart speakers could be abandoned by emergent users. Thus, it is essential to provide an alternative mechanism for non-English, multilingual speakers. The integration of human and AI responses exhibited as an efficient

technique to reduce error rates and, provide richer and relevant response to emergent users.

6 FUTURE WORK

Improving the privacy of the ‘Human plus AI’ smart speaker is a potential area of future work. This could be achieved by implementing distributed computing architecture consisting of ‘Human plus AI’ smart speaker and emergent user’s smartphone. Emergent user can use their personal device for directing and receiving personal response from the smart speaker [2]. It is also important to identify and minimize the cognitive load during an interaction with ‘Human plus AI’ smart speaker [1]. Lastly, another area of future work is the integration of human and AI answers, i.e. providing human answers only for those questions which failed to be answered by AI.

7 CONCLUSION

Our study helped in understanding the usage of smart speakers in the domestic setting of emergent users. We started by performing a study to recognize the limitations and advantages of human-powered and AI-powered prototypes. Based on the pilot study findings and inputs, we built a ‘Human plus AI’ integrated smart speaker. The smart speaker was then tested with emergent users in an field study. From the results of the final deployment, we conclude that integrating the two concepts proved beneficial to emergent users. ‘Human plus AI’ smart speaker was able to provide instant, elaborate, and precise responses to emergent users.

ACKNOWLEDGMENTS

I would like to thank Simon Robinson, Matt Jones, Thomas Reitmaier and Jennifer Pearson of Swansea University for their supervision and guidance during this project. Also, I would like to thank Michael Rohs, Leibniz University Hannover for his valuable inputs.

REFERENCES

- [1] Shashank Ahire, Aaron Priegnitz, Oguz Önbas, Michael Rohs, and Wolfgang Nejdl. 2021. How Compatible is Alexa with Dual Tasking? – Towards Intelligent Personal Assistants for Dual-Task Situations. In *Proceedings of the 9th International Conference on Human-Agent Interaction* (Virtual Event, Japan) (HAI ’21). Association for Computing Machinery, New York, NY, USA, 103–111. <https://doi.org/10.1145/3472307.3484165>
- [2] Shashank Ahire, Michael Rohs, and Simon Benjamin. 2022. Ubiquitous Work Assistant: Synchronizing a Stationary and a Wearable Conversational Agent to Assist Knowledge Work. In *2022 Symposium on Human-Computer Interaction for Work* (Durham, NH, USA) (CHIWORK 2022). Association for Computing Machinery, New York, NY, USA, Article 3, 9 pages. <https://doi.org/10.1145/3533406.3533420>
- [3] Apoorva Bhalla. 2018. An Exploratory Study Understanding the Appropriated Use of Voice-based Search and Assistants. In *Proceedings of the 9th Indian Conference on Human Computer Interaction* (Bangalore, India) (IndiaHCI’18). ACM, New York, NY, USA, 90–94. <https://doi.org/10.1145/3297121.3297136>
- [4] Sebastien Cuendet, Indrani Medhi, Kalika Bali, and Edward Cutrell. 2013. VideoKheti: Making Video Content Accessible to Low-Literate and Novice Users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (CHI ’13). Association for Computing Machinery, New York, NY, USA, 2833–2842. <https://doi.org/10.1145/2470654.2481392>
- [5] Devanuj and Anirudha Joshi. 2013. Technology Adoption by ‘Emergent’ Users: The User-usage Model. In *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction* (Bangalore, India) (APCHI ’13). ACM, New York, NY, USA, 28–38. <https://doi.org/10.1145/2525194.2525209>
- [6] Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q. Vera Liao, Khai Truong, and Shwetak Patel. 2018. FarmChat: A Conversational Agent to Answer Farmer Queries. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 170 (Dec. 2018), 22 pages. <https://doi.org/10.1145/3287048>
- [7] Anuj Kumar, Pooja Reddy, Anuj Tewari, Rajat Agrawal, and Matthew Kam. 2012. Improving Literacy in Developing Countries Using Speech Recognition-Supported Games on Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (CHI ’12). Association for Computing Machinery, New York, NY, USA, 1149–1158. <https://doi.org/10.1145/2207676.2208564>
- [8] Thuy Ong. 2017. Google now recognizes 119 languages for voice-to-text dictation. <https://www.theverge.com/2017/8/14/16142786/google-recognises-119-languages-dictation-voice-typing>
- [9] Accenture Organization. 2018. *Accenture Digital Consumer Survey 2018*. https://www.accenture.com/t20180302T094127Z_w_/us-en/_acmedia/PDF-69/Accenture-2018-Digital-Consumer-Survey-Findings.pdf#zoom=50
- [10] Neil Patel, Sheetal Agarwal, Nitendra Rajput, Amit Nanavati, Paresh Dave, and Tapan S. Parikh. 2009. A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) (CHI ’09). Association for Computing Machinery, New York, NY, USA, 51–54. <https://doi.org/10.1145/1518701.1518709>
- [11] Neil Patel, Deepti Chittamuru, Anupam Jain, Paresh Dave, and Tapan S. Parikh. 2010. Avaj Otalo: A Field Study of an Interactive Voice Forum for Small Farmers in Rural India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) (CHI ’10). Association for Computing Machinery, New York, NY, USA, 733–742. <https://doi.org/10.1145/1753326.1753434>
- [12] Jennifer Pearson, Simon Robinson, Thomas Reitmaier, Matt Jones, Shashank Ahire, Anirudha Joshi, Deepak Sahoo, Nimish Maravi, and Bhakti Bhikne. 2019. StreetWise: Smart Speakers vs Human Help in Public Slum Settings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI ’19). ACM, New York, NY, USA, Article 96, 13 pages. <https://doi.org/10.1145/3290605.3300326>
- [13] Madelaine Plauche, Udhayakumar Nallasamy, Joyojeet Pal, Chuck Wooters, and Divya Ramachandran. 2006. Speech Recognition for Illiterate Access to Information and Technology. *2006 International Conference on Information and Communication Technologies and Development* (2006). <https://doi.org/10.1109/ictd.2006.301842>
- [14] Simon Robinson, Jennifer Pearson, Shashank Ahire, Rini Ahirwar, Bhakti Bhikne, Nimish Maravi, and Matt Jones. 2018. Revisiting “Hole in the Wall” Computing: Private Smart Speakers and Public Slum Settings. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI ’18). ACM, New York, NY, USA, Article 498, 11 pages. <https://doi.org/10.1145/3173574.3174072>
- [15] Jahanzeb Sherwani, Alan Black, and Raj Reddy. 2009. Speech Interfaces for Information Access by Low Literate Users. (01 2009).
- [16] Vikas SN. 2018. Amazon Alexa & Google Assistant’s next frontier: Understand and speak Indian languages. <https://tech.economicstimes.indiatimes.com/news/internet/amazon-alexa-google-assistants-next-frontier-understand-and-speak-indian-languages/65599967>
- [17] Sen Sunny. 2017. The backstory of Alexa’s Indian makeover: desi, agnostic, politically independent and... work in progress. <https://factordaily.com/amazon-alexa-india-makeover-review/>
- [18] Jennifer Zamora. 2017. Rise of the Chatbots: Finding A Place for Artificial Intelligence in India and US. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces Companion* (Limassol, Cyprus) (IUI ’17 Companion). ACM, New York, NY, USA, 109–112. <https://doi.org/10.1145/3030024.3040201>