DOI:10.1145/3343452

BY ADITYA VASHISTHA, UMAR SAIF, AND AGHA ALI RAZA

The Internet of the Orals

INTERNET SERVICES LIKE social media, online discussion forums, and crowdsourcing marketplaces have transformed how people participate in the information ecology and digital economy. These services empower mostly urban, affluent, and literate people, and improve their reach to information and instrumental needs. However, these services currently exclude billions of people worldwide who are too poor to afford Internet-enabled devices, too remote to access the Internet, or too low literate to navigate the mostly text-driven Internet.

In India and Pakistan alone, there are nearly 1.1 billion people offline. Although 70% of their populations have access to mobile phones, most people still use basic or feature phones, making it difficult to extend existing Internet services on these devices running custom operating systems. Even when people can afford smartphones and the Internet, literacy barriers prevent 26% of adults in India and 42% of adults in Pakistan from using text-based interfaces. Most South Asian languages and dialects are still unsupported by the advancements in natural language processing ruling out the use of voice interfaces like Siri and Alexa.

In light of these constraints, Human-Computer Interaction for Development (HCI4D) researchers and practitioners have used interactive voice response (IVR) technology to create voice-based services that overcome connectivity barriers by using ordinary phone calls, literacy barriers by using local language speaking and listening skills, and socioeconomic barriers by using toll-free (1-800) lines. These services let users call a phone number to record and listen to voice messages in their local languages. Because of their accessible and usable design, these services have found applications in diverse domains and have profoundly impacted marginalized communities in low-resource environments. This article follows the evolution of these services over the last two decades (see the accompanying figure), and their big challenges and new frontiers.

First Wave: Access and Inclusion

The first wave of voice-based services focused on improving information access for people in low-resource communities. For example, HealthLine enabled low-literate frontline health workers in Pakistan to retrieve relevant information by speaking out predefined commands.[6] While initial efforts like HealthLine allowed users to only consume information, subsequent services took the form of voice forums and enabled marginalized communities to also produce and share information. This included Avaaj Otalo (an agriculture discussion forum in India),[3] CGNet Swara (a citizen journalism service in India),[2] MobileVaani (a social media service in India), Ila Dhageyso (a civic engagement portal in Somaliland),[1] and IBM's Spoken Web (a user-generated

A blind user of Sangeet Swara recording a voice message.

information directory in India). The success of these initial services demonstrated their great potential to enable information access and connectivity among underserved populations in diverse HCI4D contexts. However, the vast majority of these services ran into the hurdles of user training and technology adoption.

Second Wave: Training and Spread

Nearly a decade ago, the biggest roadblocks to designing voice forums were usability, motivation, and spread; target populations faced difficulties in using even the simplest of speech-based telephone interfaces, they did not exhibit interest or trust in using such services, and it was difficult to advertise and spread such services to underconnected people. Researchers tried to overcome these barriers by conducting lab-trainings as well as door-to-door field campaigns, but it was quickly realized that these approaches were not scalable. Raza et al. used a ludic design approach to train users and promote usability and spread. They built Polly, a voice-based entertainment service that lets users make a short audio recording, apply funny voice modifications to it, and share it with their friends via automated voice calls.[5] They deployed Polly to five low-income people in Pakistan in early 2012. Within a year, Polly spread virally to over 165,000 users via 636,000 calls without any outreach efforts. Polly's ludic interfact design trained users to navigate IVR interfaces, and also led to its viral adoption. Raza et al. then used Polly to share instrumental information with users to aid their socioeconomic development. In an initial test, 34,000 Polly users listened to 728 job advertisements nearly 386,000 times within a year.

Over the last seven years, Polly has been successfully used in multiple countries to rapidly spread useful information to underserved populations. In 2014, at the peak of the Ebola crisis in West Africa, Polly-Santé (Polly-Health) was deployed as an emergency disaster-response service in Guinea to spread reliable information about prevention, symptoms, and cure of Ebola.[12] The information originated from the Centers for Disease Control and the service was funded by the U.S. Embassy in Conakry. One of the hurdles to information dissemination in the Guinean context is great linguistic diversity and the lack of a widely understood

Because of their accessible and usable design, voice-based services have found applications in diverse domains and have profoundly impacted marginalized communities in low-resource environments.

common language. Fortunately, this is not a major impediment for voice forums. Polly-Santé was launched in 11 local languages and reached more than 7,000 local mobile phone users within a few months. In 2014, Polly was also used in India by Babajob.com to advertise a voice directory of available jobs to thousands of low-literate job seekers.

Since 2016, Polly has been active in Pakistan as a gateway to maternal health information for underconnected expectant parents. Polly advertises a hotline called Super Abbu (Super Dad) that allows expectant parents to record health questions that are answered by volunteer doctors. Such private and anonymous access to trained gynecologists allows parents to ask questions about pregnancy and childbirth that are often considered sensitive and even taboo topics in the local context. The service specifically targets fathers to promote paternal participation and allow them to share their experiences with their peers. In its initial deployment, Super Abbu reached 21,000 users (96% of them men) in just two months, uncovering a pent-up demand for maternal health information and giving the target population an agency to anonymously access culturally sensitive yet lifesaving reliable information.

Despite their demonstrated impact, large-scale voice forums like Polly face two challenges that significantly impede their scalability and sustainability: how to manage user-generated content in local languages, and how to manage the cost of voice calls from users to access these services.

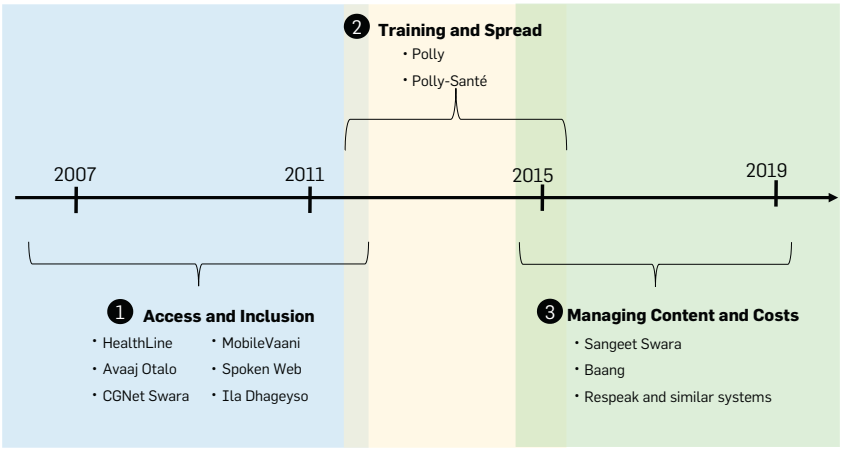
Third Wave: Managing Content and Costs at Scale

Voice forums deployed in low-resource environments often receive large volumes of user-recorded content in local languages and accents that have no speech corpora and recognition models. Consequently, it is very difficult to moderate, search, and index such content at large scale. Various voice forums often hire a dedicated team of moderators who listen to messages, categorize them, and review the quality. However, manual moderation is difficult to scale if these services grow by orders of magnitude. To address this challenge, Vashistha et al. harnessed crowdsourcing and showed that the users of voice forums, although socioeconomically marginalized and technologically inexperienced, can themselves be entrusted with the tasks of audio content moderation and categorization. In 2014, they built Sangeet Swara, a community-moderated social media voice forum that lets users record, listen to, and vote on songs, poems, and other cultural content.[10] As users listen to messages, Swara requests them to annotate the quality and category by pressing phone keys (for example, press 1 to upvote or 2 to downvote the message) and uses collaborative filtering techniques to rank, order, and categorize audio messages based on users' votes.

In an eight-month deployment in India, Swara received 53,000 phone calls from 13,000 users who submitted 6,000 voice messages in 11 languages as well as 150,000 votes. Nearly 80% of users had never used any social media platform before, 50% lived in low-income environments in rural India, and 25% were people with vision impairments (as shown in the opening image). Community moderation was 98% accurate in content categorization, made meaningful distinctions between high- and low-quality posts, and performed judgments that were in 90% agreement with expert moderators.

Deriving inspiration from Swara, Raza et al. used community moderation to manage content on Baang, a voice-based social media platform that encouraged users to record and share audio messages of diverse genres.[4] Baang allowed users to also record threaded audio comments on voice messages and added a Polly-like sharing mechanism. Deployed in Pakistan in 2015, Baang organically reached 10,000 users within eight months who contributed more than 44,000 voice messages that were played more than 2.8 million times, and received nearly 340,000 votes and 124,000 audio comments. The ability to vote, comment, and share led to viral spread, deeper engagement, and the emergence of true dialog among participants. Beyond connectivity, Swara and Baang provided its users with a voice and a social identity as well as a means to share informa-

Three waves of voice forums in low-resource environments.



tion and get community support. Moreover, they demonstrated that a community of low-income, low-literate people can moderate themselves without any outside support, thereby addressing the content management challenge of these voice forums.

The second key challenge in scaling voice forums is the airtime cost. Often, these services use expensive toll-free lines to remain accessible to low-income users. The resultant cost poses a huge burden to sustainability, often putting these services at risk of being shut down as the usage grows. While a few services sustain themselves through advertisements, grants, and partnerships with telecoms or governments, these options are often beyond the reach of most voice forum providers. To make these services financially sustainable, Vashistha et al. examined whether low-income users of voice forums could complete useful work on their mobile phones to offset their participation costs. In 2016, they created Respeak, the first voice-based crowdsourcing marketplace that pays users to transcribe audio files vocally.[7–9] Respeak sends short audio segments to multiple voice forum users and pays them via mobile airtime for each submitted transcript. Instead of typing the transcript, users respeak audio content into an off-the-shelf speech recognition engine and submit the autogenerated transcript. Respeak combines the transcripts for each segment from multiple users using sequence-alignment algorithms to reduce random speech recognition errors. It then pays users in mobile airtime based on the accuracy of transcripts submitted in them. In the last three years, Respeak has been used by low-income students, blind people, and rural residents in India to produce speech transcriptions with over 90% accuracy at one-fourth of the market rate, generating sufficient profit to subsidize their participation costs. One minute of crowd work on Respeak enable users to earn eight minutes of airtime.[8]

Grand Challenges: Harassment, Misinformation, and Disinformation

Voice forums, like any other social platform, come with their own pitfalls. They end up reflecting the existing sociocultural norms and values of the society, including its shortcomings and biases. For example, while Swara and Baang served as instruments of inclusion for low literate, rural, indigenous, and visually impaired communities, they failed to create a welcoming environment for female users.[11] Women faced systemic discrimination and harassment in the form of messages that contained abuses, threats, and flirtatious behavior.

Both mainstream social media platforms and voice forums face grand challenges when tackling misinformation, disinformation, harassment, and abuse. These platforms and forums differ greatly in terms of scale, features, interfaces, supported languages, and target users. Consequently, solutions to tackle these challenges on a platform like Facebook might be ineffective for voice forums, and vice versa. This presents interesting research challenges of identifying indecorous content in local language audio, filtering out spreaders of disinformation, and addressing situations where the collective ignorance of community members eclipse their collective intelligence. The HCI4D community must tackle these grand challenges to make the Internet of the orals more diverse, inclusive, and impactful.

References

1. Gulaid, M. and Vashistha, A. Ila Dhageyso: An interactive voice forum to foster transparent governance in Somaliland. In *Proceedings of the 6th Intern. Conf. Information and Communications Technologies and Development: Notes, Vol. 2* (Cape Town, South Africa, 2013), 41–44.
2. Mudliar, P. et al. Emergent practices around CGNet Swara, voice forum for citizen journalism in rural India. In *Proceedings of the 5th Intern. Conf. Information and Communication Technologies and Development* (Atlanta, GA, USA, 2012), 159–168.
3. Patel, N. et al. Avaaj Otalo: A field study of an interactive voice forum for small farmers in rural India. In *Proceedings of the SIGCHI Conf. Human Factors in Computing Systems* (Atlanta, GA, USA, 2010), 733–742.
4. Raza, A.A. et al. Baang: A viral speech-based social platform for under-connected populations. In Proceedings of the 2018 CHI Conf. Human Factors in Computing Systems (Montreal, QC, Canada, 2018), 643:1–643:12.
5. Raza, A.A. et al. Job opportunities through entertainment: Virally spread speech-based services for low-literate users. In *Proceedings of the SIGCHI Conf. Human Factors in Computing Systems* (Paris, France, 2013), 2803–2812.
6. Sherwani, J. et al. Healthline: Speech-based access to health information by low-literate users. *Inter. Conf. Information and Communication Technologies and Development* (Bangalore, India, 2007), 1–9.
7. Vashistha, A. et al. BSpeak: An accessible voice-based crowdsourcing marketplace for low-income blind people. In *Proceedings of the 2018 CHI Conf. Human Factors in Computing Systems* (Montreal, QC, Canada, 2018), 57:1–57:13.
8. Vashistha, A. et al. ReCall: Crowdsourcing on basic phones to financially sustain voice forums. In *Proceedings of the 2019 CHI Conf. Human Factors in Computing Systems* (Glasgow, Scotland, U.K., 2019).
9. Vashistha, A. et al. Respeak: A voice-based, crowd-powered speech transcription system. In *Proceedings of the 2017 CHI Conf. Human Factors in Computing Systems* (Denver, CO, USA, 2017), 1855–1866.
10. Vashistha, A. et al. Sangeet Swara: A community-moderated voice forum in rural India. In *Proceedings of the 33rd Annual ACM Conf. Human Factors in Computing Systems* (Seoul, South Korea, 2015), 417–426.
11. Vashistha, A. et al. Threats, abuses, flirting, and blackmail: Gender inequity in social media voice forums. In *Proceedings of the 2019 CHI Conf. Human Factors in Computing Systems* (Glasgow, Scotland, U.K., 2019).
12. Wolfe, N. et al. Rapid development of public health education systems in low-literacy multilingual environments: Combating Ebola through voice messaging. In *Proceedings of the ISCA Special Interest Group on Speech and Language Technology in Education* (Leipzig, Germany, 2015).

Aditya Vashistha is an assistant professor at Cornell University, Ithaca, NY, USA.

Umar Saif is UNESCO Chair, ICTD, Lahore, Pakistan.

Agha Ali Raza is an assistant professor at Information Technology University, Lahore, Pakistan.