Voice Interfaces for Underserved Communities

Aditya Vashistha and Agha Ali Raza

Abstract

Internet technologies–like social media platforms and discussion forums–have transformed how people communicate with each other. In addition to improving access to information, news, and entertainment, they have impacted governance, politics, civil society movements, crisis response, marketplaces, and healthcare, among other parts of our lives. Although concerns regarding data misuse, privacy breaches, and overuse have grown recently, these technologies are continuing to soar, mostly among literate, urban, and connected communities, all across the world. However, despite their promises (and pitfalls), these technologies currently exclude billions of people worldwide who are too remote to access the Internet, too low-literate to navigate the mostly text-driven Internet, or too poor to afford Internet-enabled devices. This chapter presents the innovation, implementation, and adaptation of voice forums that have evolved over the last two decades, discusses challenges that plagued their growth, and highlights ingenious solutions that were used to address their pain points.

1. Evolution of Voice Forums

Information and Communication Technologies (ICTs) have the potential to enable socio-economic development where information and connectivity are the missing components. ICT-based interventions can lead to better management of available resources, improved monitoring and reporting of corruption and more awareness and connectivity among people. To achieve impact at such a large scale such solutions need to be robust enough to reach the target populations using available means with minimum resource expenditures. The Internet is a phenomenon that has enabled this enormous metamorphosis and services like social media and online discussion forums have transformed how people participate in the information ecology and digital economy.

Unfortunately, most Internet services currently only empower urban, affluent and literate people, and exclude billions of 'othered' people 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. Up to 81% of the people living

Aditya Vashistha
243 Gates Hall, 107 Hoy Road
Ithaca NY, United States - 14850
Assistant Professor, Cornell University
adityav@cornell.edu

Agha Ali Raza
DHA Phase 5, Khayaban e Jinnah,
Lahore, Punjab, Pakistan- 54792
Assistant Professor, Lahore University of Management Sciences (LUMS)
agha.ali.raza@lums.edu.pk

in developing countries are offline (compared to the 13% in the developed world)[1]. This disproportionate access is even worse for marginalized populations within these countries. The Internet access gender gap is as high as 43% in developing countries (compared to 2.3% in developed countries)[2]. Also among these offline communities are low-literates and native speakers of languages that do not have written forms (46% of all languages[3]). Hence, lack of access is strongly associated with poverty, low-literacy, tech-naivety and being marginalized. This makes mainstream Internet services inadequate to meet the needs of billions of people living in underserved settings. Sadly, many of these communities are unable to find alternate means of communication as television and radio are non-interactive, print-media assumes literacy, and computers require stable electricity and Internet connectivity.

Clearly, textual interfaces and Internet access cannot be relied upon to provide information access to the populations identified above. Speech as an alternate communication medium is more suitable especially because speech-based information can even be provided over simple phone calls without any internet connectivity. Given the rapid proliferation of mobile phones in developing countries, speech-over-simple-mobile phones is a viable way to connect underserved populations. In light of these considerations, global development 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 socio-economic barriers by using toll-free (1-800) lines. These services assume no more than the capability to make and receive a voice call from users and let them call a phone number to record and listen to voice messages in their local languages. Over the last two decades such services have included: efficient marketplaces; common-interest groups; message boards, blogs, mailing lists and social networks that facilitate social and political activism; information campaigns on themes of education, training, employment opportunities, health, agriculture, and emergency response; citizen journalism; and, automated surveys and polls to gather information. Because of their accessible and usable design, these services have found applications in diverse domains and have profoundly impacted underserved com-

---

[1] "Measuring Digital Development - Offline population." https://itu.foleon.com/itu/measuring-digital-development/offline-population/ (accessed Aug. 31, 2020).

[2] "Measuring Digital Development - Gender gap." https://itu.foleon.com/itu/measuring-digital-development/gender-gap/ (accessed Aug. 31, 2020).

[3] "How many languages in the world are unwritten? | Ethnologue." https://www.ethnologue.com/enterprise-faq/how-many-languages-world-are-unwritten-0 (accessed Aug. 31, 2020).

munities in low-resource environments. This section follows the evolution of these services over the last decade, and their big challenges and new frontiers.

Voice forums can be designed as: 1) top-down information push services where the content is developed by experts and the interface simply allows individuals to consume the content, 2) peer-to-peer services where individuals can communicate with other individuals via audio messages, and 3) social services where users can record and broadcast content to the community. The first wave of voice-based services focused on improving information access for people in low-resource communities. The target audience of Project HealthLine –one of the first such endeavors–was low-literate community health workers in rural Sindh province, Pakistan (Sherwani et al, 2007). It enabled them to retrieve relevant information by speaking out predefined commands. The goal was to provide telephone-based access to reliable and up-to-date health information, and the speech interface performed well once the health workers were trained to use it via human-guided tutorials. This project also highlighted the challenges in eliciting informative feedback from low-literate users. While services 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.

The impact of the community involvement in voice-forums has been thoroughly studied and has been shown to promote social inclusion among underserved communities. Notable research in this direction includes Avaaj Otalo (an agriculture discussion forum in India), CGNet Swara (a citizen journalism service in India), MobileVaani (a social media service in India), Ila Dhageyso (a civic engagement portal in Somaliland), and IBM's Spoken Web (a user-generated information directory in India) (Patel et al, 2010; Mudliar et al, 2012; Moitra et al, 2016; Gulaid and Vashistha, 2013). For example, Avaaj Otalo was designed to connect farmers in Gujarat, India and offered three features: an open forum where users could post and answer questions, a top down announcement board, and a radio archive that allowed users to listen to previously broadcast radio program episodes. The most popular service turned out to be the open forum, constituting 60% of the total traffic, and users found interesting unintended uses for it like business consulting and advertisement. Mudliar et al. examined participation of rural communities in India on CGNet Swara, an interactive voice forum to record and listen to local news, grievances, and cultural content, and found that it serves as a vehicle for digital inclusion for indigenous communities. Koradia et al, 2012, integrated voice forums with community radio technology to amplify its reach and involve radio listeners in content creation, feedback, and station management. The influence of peer-generated content available on social platforms on the target communities has been compared against expert-generated content by Patel et al.,2012. In a two-week trial, seven agricultural tips were disseminated to 305 farmers in Gujarat, India. Each tip was recorded in the voices of university scientists and farmers. The study showed that farmers preferred to hear agricultural tips

in the voice of their peers; even though in interviews they maintained their more socially acceptable inclination towards scientists. Table 22.1 provides a summary of the comparison of popular voice forums that have been deployed in various developing regions.

An important question in developing voice forums has been that of the input modality: speech vs. key press. Project HealthLine found that speech input performed better than key press in terms of task completion, for both literate and low literate users (Sherwani et al, 2007). However, it provided no clear answer in terms of subjective user preference. Lee et al. reported key press input to be more efficient for linear and simple tasks while speech input to be better for non-linear tasks. Grover et al. found task completion rates for speech and key press input modes to be similar; and tech-literacy being a more important factor than overall literacy for task completion. Patel et al, 2009, found key press input to be more intuitive and reliable than speech. Over the subsequent years, key press input became the de facto standard for IVR services targeting developing regions.

The success of these initial services demonstrated their great potential to enable information access and connectivity among underserved populations in diverse contexts. However, voice forums also presented a wide array of implementation, scalability, and sustainability challenges, including how to: (1) train the users to use speech interfaces, (2) spread and advertise these services to tech-novice and poorly connected masses, and (3) engage users in potentially involved voice-based interactions while keeping them oriented, motivated, and grounded. These traits of training, spread, engagement and retention were found to be necessary for effective transmission of knowledge but were very hard to achieve simultaneously.

2. Scaling Voice-based Services

In 2010, the biggest roadblocks to scaling voice forums among underserved populations were usability, motivation, and spread. Target populations faced difficulties in using even the simplest of voice forums, they did not exhibit interest or trust in using such services, and it was hard to advertise and spread these services to under-connected people. Researchers tried to overcome these barriers by conducting lab-trainings, demo sessions and door-to-door field campaigns, but it was quickly realized that these approaches too are not scalable. The identified challenge was to inform and train the users without the benefit of explicit, in-person sessions. To reach people at a large scale, a need to advertise and promote these voice services was also identified that was challenging using traditional advertising mechanisms as these populations are low-literate and not very well connected with technology. The way forward, unexpectedly was found to be entertainment!

In 2010, Smyth et al. described the remarkable ingenuity exhibited by low-literate users when they are motivated by the desire to be entertained and concluded that such powerful motivation "turns UI (user interface) barriers into mere speed bumps" (Smyth et al, 2010). Inspired by this powerful demonstration, Raza et al. set out to systematically develop practices for entertainment-driven mass familiarization and training of low-literate users in the use of telephone-based services, and created Polly (Raza et al, 2013 a). Polly engaged low-literate and non-tech savvy users in light entertainment to spread useful development-related information to them as they became more comfortable with the voice interface. Polly allowed users to record their voice, modify it using funny voice modifications and send the modified recording to their friends (as shown in Figure 22.2). The entertainment appealed to the target users and allowed Polly to spread virally. It acted as a soft incentive for users to train themselves and overcomes the scalability hurdle of explicit user-training and motivation. It also allowed Polly to organically spread among the population through word-of-mouth as well as scheduled message deliveries from one user to another. As users became more comfortable with the interface, Polly introduced them to development-related services like job search, and health information.

Polly was first pilot-tested in 2011 with 32 low-literate users who were handed out Polly's phone number and were asked to explore its functionality. Within 3 weeks Polly had organically reached over 2,000 users and logged more than 10,000 calls before it was shut down due to insufficient telephone capacity and unsustainable cellular airtime cost. In 2012, Polly was relaunched in Pakistan with an increased telephone capacity. The system was seeded via automated phone calls to 5 users from the pilot launch. Within one week, the 30 phone lines were saturated and usage quotas had to be imposed. A job audio browser had also been introduced in Polly that allowed users to listen to entry-level job ads from local newspapers in their local language. Callers could browse job opportunities and could even forward promising ones to friends. Polly remained online for a year and amassed 165,000 users who participated in 636,000 interactions, including over 200,000 forwarded voice messages and 22,000 forwarded job ads. At its peak, it was spreading to a thousand new users every day. The 728 job ads were listened to 386,000 times by 34,000 users. Polly was used primarily by low-educated young men for entertainment and other creative uses like voicemail, group messaging and telemarketing. Its viral spread crossed gender and age boundaries and also attracted a large number of visually impaired users, but remained primarily in the similar socio-economic strata.

Raza et al. showed that Polly's users improved at using speech interfaces with time. A behavior study of the 165,000 users of Polly revealed that with more experience, users responded faster to menus (using more intentional menu-interruptions aka barge-ins) and made fewer mistakes and abortive attempts (Raza et al, 2013). Users' choice of activity also evolved over time, and with experience

they showed an increasing interest in message sending, became more explorative of the system's capabilities, and better adapted themselves to its constraints. Long-term users engaged in lengthier calls from the start and took a more active interest in voice modification and forwarding features. The forwarding feature that allowed users to send audio messages to friends brought 85% of all new users to the service. An analysis of the geographical spread of Polly showed that there were calls from all over Pakistan (Raza et al, 2013).

Since 2010, Polly has been successfully used in three countries to rapidly spread useful information to underserved populations at a large scale. 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 (Wolfe et al, 2015). The information originated from the US Centers for Disease Control (CDC) and the service was funded by the US Embassy in Conakry. A hurdle to scale information dissemination in the Guinean context was great linguistic diversity and the lack of a widely understood common language. This did not turn out to be a major impediment for voice forums and Polly-Santé was launched in eleven local languages and reached more than 7,000 users within a few months. In India, Polly was used by Babajob.com to advertise a voice directory of available jobs to thousands of low-literate job seekers, and by Jharkhand Mobile Vaani, a popular citizen-radio-over-phone platform, to spread awareness about their platform using a "cross-selling" model of advertisement (Raza et al, 2016). These deployments highlighted the significance of committed local partners, and showed that seeding via promos and advertisements has the potential to induce viral spread, and that the content, mood and tone of the promos plays a vital role in influencing a user's understanding of the service and its capabilities.

Despite their demonstrated impact on user-training, trust, and service-spread, voice forums like Polly face three major challenges that significantly impede their scalability and sustainability: (1) How to moderate, filter and manage the massive user-generated local-language audio content in near real-time, (2) how to retain the users of the service for long-term interactions, as most users of Polly just stop using the platform within days as the novelty of the simple entertainment wears off, and (3) how to manage the high call costs required to subsidize these voice forums for the target users who are mostly not able to afford such expenses on their own. These challenges were further amplified as voice-forums evolved from being peer-to-peer (individual) to broadcast (social) platforms.

3. Managing Local Language Audio Content

Voice forums quickly evolved to be a de facto social media platform for basic phone users where they record, listen to, and vote on audio messages generated in local languages. However, since such voice forums generate a large amount of audio content in low-resource languages and accents that are unsupported by advancements in natural language processing, it is very difficult to automate categorization and moderation of posts and responses, which are needed for respectful use of the service. Lack of categorization makes it very hard for users to browse data and providers to regulate these services. This is a major hurdle for voice modality where users must listen to audio content in a sequential manner and cannot skim it like textual content. Also, voice forums are often deployed in low-resource environments where most people lack technical know-how of social media, making them particularly vulnerable to disinformation and fake news. For example, recent acts of mob violence in India have been attributed to fake WhatsApp[4] messages[5], which can also be easily recorded and shared on voice forums. These reasons make content moderation critical to present users with respectful, accurate, and high-quality recordings.

Large voice forums typically employ a dedicated team of moderators, who screen recordings, offer feedback to contributors, and perform tagging, categorization, and moderation of audio posts. For example, both CGNet Swara and Mobile Vaani currently employ 10–15 full-time moderators who carefully review each submitted post. Although manual moderation is highly accurate, it becomes difficult to scale as these services grow, due to high cost, delayed response, and challenges in hiring moderators who are familiar with local context.

Several news and social networking sites like Reddit and Stack Overflow draw on collaborative filtering and community moderation algorithms to manage user-generated content. They use community votes and recency to determine high quality and contextually relevant user-generated content. Since most of the content on these platforms is textual, a number of natural language processing techniques have been employed to categorize, annotate, and moderate content, and even predict emotions. However, no prior work has focused on using community moderation on a voice forum, which is different in several ways from community moderation on a text-based forum. For example, audio content is more difficult to skim than textual content, meaning that users may lose patience in hearing and ranking lower-ranked posts. An IVR system can track exactly what a user listened to and

[4] S. Biswas, "How WhatsApp helped turn a village into a mob," *BBC News*, Jul. 19, 2018.

[5] V. Goel, S. Raj, and P. Ravichandran, "How WhatsApp Leads Mobs to Murder in India," *The New York Times*, Jul. 18, 2018.

what content they skipped, which is difficult to do on a webpage. Finally, the limited affordances of an IVR interface and the limited digital skills of voice forum users add more constraints to the design of community moderation algorithms.

To address the content moderation challenge, Vashistha et al. harnessed collaborative filtering and community sourcing, 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 (Vashistha et al, 2015a). Recognizing that entertainment drives technology adoption by low-income people, they built a new voice forum called Sangeet Swara, a community-moderated social media service that enabled users to record, listen to, and vote on songs, poems, and other cultural content. Figure 22.3 shows the high-level user interface design of the voice forum. As users listened to messages, Sangeet Swara requested them to annotate the quality and category of the content by pressing phone keys (e.g., press 1 to upvote or 2 to downvote the message) and used collaborative filtering techniques to rank, order, and categorize audio posts based on users' votes. A key aspect of the system was that it *ranked* the posts based on feedback from the community. The ranking aimed to order the posts according to what was most likely to be enjoyed and appreciated by listeners. There was a single, global ranking computed across all posts and all listeners in the system. In addition to the rank order, the system calculated a separate playback order that determined which post a listener heard at a given time. The playback order balanced the interests of listeners (who desired to hear high-quality posts) with the interests of content contributors (who desired to have as large of an audience as possible). Both the rank order and playback order were dynamically updated based on listeners' ratings of the content.

To create awareness about Sangeet Swara in rural and small-town India, Vashistha et al. seeded Sangeet Swara with 15 songs and poems that appeared previously on CGNet Swara and posted a message on CGNet Swara to invite participation from its users. The post was accessible to CGNet Swara callers for two days, during which time it was heard by 393 unique callers. Out of those, 73 people placed a call to Sangeet Swara. In an eight-month deployment in India, Sangeet 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 visual impairments (Vashistha et al, 2015). Many users attributed great value to their interactions on Sangeet Swara. They recorded strong positive sentiments about the service and shared interesting anecdotes about how the service was impacting their lives. They considered it to be a platform where people show their creativity, voice their opinions, and record interesting content. This sentiment was often strongest among visually impaired users. The findings indicate that users took an active role in policing the system, for example, by urging others to record cultural content, follow community guide-

lines, and to avoid posting abusive comments. 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. The ability to vote, comment, and share led to viral spread, deeper engagement, and the emergence of true dialog among participants. Beyond connectivity, Sangeet Swara provided its users with a voice and a social identity as well as a means to share information and get community support. Moreover, it 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.

Although Sangeet Swara demonstrated the feasibility, acceptability, usability, and efficacy of community moderation, several aspects of community moderation remain untested. For example, Sangeet Swara focused on the domain of entertainment, where the content is relatively uncontroversial. Extending to domains such as politics and citizen journalism will require sensitivity to stronger disagreements between callers, which could impact their ratings as well as their flagging of posts for deletion. Similarly, the community moderation algorithm can be improved by making it more sensitive to who is voting and when they are voting. For example, discounting votes of users who acted too soon or those who deviated from community standards (e.g., people who abused other users) could reduce randomness in moderation. Similarly, assigning higher weights to votes of users who call consistently and record high-quality posts could improve the quality of moderation. Future work should identify features to predict casual voting by community members.

4. Increasing User Retention

Unlike Polly users that churn at a high rate as the novelty wears off, Sangeet Swara saw much lower churn rate and high user engagement throughout its deployment. To investigate these effects more systematically, Raza et al. created Baang, a social media voice forum with in-built mechanisms to achieve greater spread and uptake as well as deeper and long-term engagement (Raza et al, 2018). The interface of Baang was inspired by both Polly and Sangeet Swara, and included two novel features. Like Sangeet Swara, Baang allowed users to create and consume audio content and express their feedback via likes and dislikes (see Figure 3). Similar to Polly, users could share audio content with friends and Baang actively called up the recipients to deliver the messages. In addition to these features, Baang also allowed users to record audio comments on posted messages and also allowed them multiple options to control the order of playback of messages (popularity, recency, and trending).

Deployed in Pakistan in 2015, Baang organically reached 10,000 users (69% of them blind) within 8 months who participated in nearly 270,000 calls and contributed more than 44,000 voice messages that were played more than 2.8 million times and received 340,000 votes, 124,000 audio comments and about 95,000 shares. The user-retention of Baang was significantly higher compared to Polly. The differences between the two services became more pronounced after a week when more than 20% of users returned to Baang, while Polly only retained less than 5% of its users beyond the initial week of exposure. Up to 20% of Baang's users (compared to 1% to 3% in case of Polly) kept returning after four weeks. Quantitative analysis of usage patterns revealed that the user retention was highly associated with the act of posting audio comments. Interestingly, posting of comments was found to be a better predictor of continued use compared to any other single action of the platform including posting of messages. The viral spread of Baang was found to be largely through the message sharing feature where 60% of all new users were introduced to the platform through forwarded messages.

The analysis found that Baang created a community of users from diverse socio-economic and linguistic backgrounds including 69% blind people, 10% females and mostly low-educated, unemployed, young men from all over Pakistan. Baang's open community included people from remote areas and linguistic minorities. Social network features like voting, content sharing and voice comments led to viral and enthusiastic uptake of the service, high user engagement and retention, and true dialog among the community. Baang provided a window into the collective values of a community as they raised their voice against disability abuse, female harassment, foul language, hatred, terrorism and united for their rights and in support of the oppressed. Like Sangeet Swara, Baang showed that orality-driven social platforms have the potential to provide under-connected and tech-naive individuals with a voice and social identity.

5. Managing Cost

Although the success of services like Polly, Sangeet Swara, and Baang is very encouraging, there is a growing concern among global development researchers and practitioners about the high operating costs of voice forums especially when they reach a large scale. Providers of these services often pay for the cost of acquiring toll-free lines so that low-income people can access these services for free. However, this cost becomes prohibitive as the usage increases, often putting these services at risk of being shut down. For example, Polly was discontinued several times because of the lack of resources to meet growing call volumes. Many voice forums rely on external funding in the form of grants and awards to subsidize the cost of voice calls, however, the non-deterministic nature of such opportunities makes this approach unsustainable. For example, the founder of CGNet Swara ex-

pressed frustrations about how limited funding to subsidize phone calls may cause them to *"shut down completely"*[6]. Some voice forums, including Sangeet Swara, have conducted experiments to examine users' willingness to bear the cost of voice calls, however, the outcome of such experiments have consistently shown that low-income users are unable to afford the cost of voice calls regardless of the benefits offered by the service (Vashistha et al, 2015).

A few voice-based services such as Kan Khajura Tesan[7] and Mobile Vaani have used advertising revenues to subsidize the cost of voice calls. These services advertise products and services that cater to low-income consumers in rural and peri-urban areas (e.g., small sachets of washing powder, toothpaste, soap). Although these services are existential proof of advertising as a viable approach to financially sustain large-scale voice forums, the initial investment required to gain critical mass for advertising is often beyond the reach of most bottom-up, development-focused voice forums.

Some voice forums such as Ila Dhageyso and 3-2-1 service[8] have partnered with government agencies and mobile network operators to subsidize the cost of voice calls. Although such partnerships greatly reduce the burden of voice call costs, building and maintaining such partnerships is seldom possible due to mismatch in goals, expectations, and values. Given these limitations in existing approaches to financially sustain voice forums, there is an urgent need to find alternatives to reduce the burden of phone calls on voice forum providers.

Recently Vashistha et al. examined an alternative approach to address the financial sustainability challenge. They investigated whether low-income users of these services could complete useful work on their mobile phones to offset their participation costs on services like Sangeet Swara. To do so, they designed and built Respeak, a new crowdsourcing marketplace that works on basic mobile phones and offers tasks that do not require familiarity with English language and advanced skills like typing, unlike mainstream crowdsourcing marketplaces like Mechanical Turk and CrowdFlower. Respeak is a voice-based, crowd-powered speech transcription system that pays users to transcribe audio files vocally (Vashistha et al, 2017; Vashistha et al, 2018; Vashistha et al, 2019). To obtain transcrip-

---

[6] A. S. Writer, "Amid fund crunch, CGNet Swara eyes shift to Bluetooth radio tech," *https://www.livemint.com/*, Sep. 25, 2016. https://www.livemint.com/Politics/UcrYsrB8fIAGTDiIoC452N/Amid-fund-crunch-CGNet-Swara-eyes-shift-to-Bluetooth-radio.html (accessed Sep. 04, 2018).

[7] "Kan Khajura Tesan." http://www.kankhajuratesan.com/ (accessed Feb. 23, 2017).

[8] "3-2-1 -- On-Demand Messaging for Development." http://hni.org/what-we-do/3-2-1-service/ (accessed Sep. 04, 2018).

tion of an audio file, Respeak partitions the file into small audio segments and sends these segments to multiple users. Instead of typing the transcript on a phone's keyboard with constrained physical space, users re-speak (i.e., repeat) audio content into an off-the-shelf speech recognition engine and submit the speech recognition output as a transcript. Once multiple users submit transcripts for a particular segment, Respeak combines the transcripts using sequence alignment algorithms to reduce random speech recognition errors. Based on the overlap between aligned transcript and individual transcripts, Respeak sends rewards to users via mobile airtime or mobile payment, thereby incentivizing them to complete more tasks accurately. Figure 22.4 shows the high-level illustration of Respeak's design.

Before launching the Respeak system widely, Vashistha et al. conducted a range of cognitive experiments, usability studies, and experimental evaluations to assess its feasibility, usability, and acceptability. For example, they investigated how audio segment length and presentation order affects content retention and cognitive load on Respeak users, whether speaking is indeed a more efficient and usable output medium for transcription than typing, and how different phone types, channel types, and modes to review transcripts affect task accuracy and completion time. After iteratively incorporating insights from these evaluations into Respeak's design, they deployed it to 73 low-income students, blind people, and rural residents in India for nearly two months by partnering with Indian Institute of Technology Bombay (IIT Bombay), Enable India, and Nehru Yuwa Sangathan Tisi (NYST), respectively. Collectively, users transcribed 70 hours of audio data by completing 50,000 micro tasks with an average accuracy of 70% and earned INR 31,000 at an hourly rate that exceeds the average hourly wage in India. Respeak then merged transcripts from multiple users to produce the transcript with over 90% accuracy at one-fourth of the market rate, generating sufficient profit to subsidize participation costs of other voice-based services. The analysis indicated that one minute of crowd work on Respeak could subsidize eight minutes of airtime on services like Sangeet Swara. User evaluations also indicated that voice forum users were willing to do tasks on Respeak to get subsidized airtime to use other voice-based services like Sangeet Swara. Also, switching between these two services–Respeak to complete crowd work and Sangeet Swara to use free credits–did not affect the usability and experience of users on both services. Although Vashistha et al. demonstrated the promise of using crowd work to financially sustain voice forums, future work should look at conducting long-term deployments, large-scale field evaluations, and real-time integration with existing voice forums like CGNet Swara and Mobile Vaani to uncover issues that may arise at scale.

6. Replication and Scale

Despite the enthusiasm surrounding voice forums, the unfortunate reality is that it remains quite complex to install and configure them. Many services like CGNet Swara, Avaaj Otalo, and Phone Peti utilize open-source platforms like Asterisk[9] or FreeSwitch[10] for the telephony interface, and require hosting a Web server to connect with moderators (Koradia and Seth, 2012). Although tractable for technology researchers, using these platforms requires Linux expertise that is usually beyond the reach of many grassroots and non-governmental organizations that have no in-house developers.

Another issue in scaling voice forums is the implementation of the underlying architecture. Most voice forums have a centralized architecture that provides a single access point (or calling number) for users, making it difficult to scale the service to new geographic locations. For example, if an organization would like to scale a voice forum operating in region A to another region B, then either users living in B need to make an expensive long-distance phone call to the access point in region A or the organization need to set up a local service in B, thereby disconnecting people in the two locations. Also, most voice forums currently operate in silos and are disconnected from mainstream social media platforms, impairing information exchange between different local communities as well as global audience, which might be desirable in cases like political activism, human rights violation reporting etc.

Current toolkits to build voice forums like Asterisk and FreeSwitch do not offer such features to support distributed, scalable, and connected operations. Although cloud telephony systems–like Twilio, Tropo, Exotel, KooKoo, and engageSPARK–makes it easy for organizations that lack technical expertise to build and maintain voice forums, these services are *very* expensive to use, disconnected from social media platforms, and do not synchronize content across distributed call centers, making them less scalable and disconnected.

To address these bottlenecks in building, replicating, and maintaining voice forums at scale, Vashistha et al. built IVR Junction: a free and open-source toolkit that enables organizations to build and replicate voice forums (Vashistha and Theis, 2012). IVR Junction has three main advantages over existing IVR toolkits like Asterisk, FreedomFone[11], and FreeSwitch.

---

[9] "Asterisk.org," *Asterisk.org*. http://www.asterisk.org/ (accessed Sep. 15, 2016).

[10] FreeSWITCH.org | Communication Consolidation." https://freeswitch.org/ (accessed Sep. 15, 2016)

[11] "Freedom Fone." http://www.freedomfone.org/.

1. Easy to build and set up: IVR Junction makes it easier for organizations with limited technical skills to build, set up, and maintain voice forums. Using IVR Junction, anyone with basic computer literacy can use templates and configure simple options to set up a robust voice forum as an ordinary program on a Window-based commodity machine.
2. Distributed architecture: IVR Junction enables distributed access points, thereby connecting multiple geographically distributed communities via inexpensive local calls as well as enabling robustness to regional power outages or crackdowns by repressive regimes.
3. Global reach: IVR Junction integrates voice forums with free Internet services and social media platforms: recordings contributed over the phone are immediately broadcast on YouTube and Facebook, and posts made on the Internet can also be listened to over the phone. Thus, IVR Junction enables anyone with a basic mobile phone to participate in global social media; low-income populations can record and listen to posts via mobile phone, while the global community can access and contribute recordings via the Internet. This capability enables remote communities to create their own repositories of highly-relevant information, while also sharing them with audiences worldwide.

In the last few years, IVR Junction has been used by many organizations to connect people in low-resource environments and provide them access to information, news, and governance. For example, in Somaliland, IVR Junction was used to build a voice forum that established a direct communication channel between the rural tribal population and government officials to bring transparency and trust in the political processes (Gulaid and Vashishtha, 2013). Somaliland–an autonomous region of Somalia–has fragile political institutions, fragmented and polarized media, and unstable government. Parliamentarians in the capital city were unable to convey their policies and receive feedback from low-literate constituents living in remote, rural, disconnected regions due to misinformation by partisan media. The solution to this intractable problem appeared simple: connect parliamentarians directly with their communities. However, Facebook and Twitter were infeasible solutions in a region with extremely low Internet penetration and adult literacy rate. To overcome these challenges, IVR Junction was used to build Ila Dhageyso, a voice forum that enabled parliamentarians and constituents to call a phone number, and record and listen to asynchronous audio posts in a discussion forum format. Ila Dhageyso also automatically posted these audio messages to a YouTube channel to engage with Somaliland's diaspora. The voice forum was supported by the Office of the Communication of the President and Telesom (largest telecommunication company in Somaliland), and was launched as a toll-free line so that people living in poverty could participate. The deployment received an enthusiastic response both from the constituents and parliamentarians who recorded over 4,300 audio messages in just five months.

In war-torn Mali, the Broadcasting Board of Governors and Voice of America used IVR Junction to provide on-demand, reliable, and up-to-date news in the local language. People in Mali called the service to listen to a three-minute news broadcast by Voice of America, thereby getting access to breaking news and health information as well as sharing their feedback.

In India, IVR Junction was used by women's rights activists in response to a gang rape incident in New Delhi that sparked international outrage. They built a voice petition forum where supporters from all economic backgrounds and varied literacy levels raised their voice for women's safety and empowerment. The recordings, which spanned from support for the victim to plans for sensitizing local communities, were available not only on the voice forum but also on a YouTube channel and Facebook page, thereby dramatically amplifying the local voices.

7. Mechanisms of User Acquisition

Until now we have discussed the unique advantages of voice forums in isolation without establishing a comparison between such forums and other available means in developing regions to reach various development goals. As discussed, technology-based interventions typically require significant resources to achieve scale beyond the pilot stage. Spreading awareness, acquiring users, and retaining them over time are all significant barriers to scale. While we have discussed that voice-based entertainment services can be used as vehicles for spreading development services, it is not clear how well such entertainment-driven proliferation performs in comparison with traditional advertising channels in terms of cost, extent of spread, and quality of user-engagement.

Raza et al. compared various advertisement channels side-by-side as they attempt to scale the same development service–a maternal health hotline, Super Abbu, connecting expectant parents anonymously to trained gynecologists (Naseem et al, 2020). They also show how users acquired through various channels fare in terms of overall activity, engagement, seriousness of purpose and IVR sophistication. An 11-week campaign was conducted where users were acquired for Super Abbu through seven different advertising channels: (1) paper flyers, (2) banners displayed at the back of auto rickshaws, (3) cable TV ads, (4) radio ads, (5) robocalls, (6) sponsored Facebook ads, and (7) an IVR-based entertainment service. Across these channels they reached 21,770 users who engaged in over 32,500 interactions on Super Abbu. To assess the efficacy of the channels, three main user acquisition metrics were considered: conversion rate, cost of user acquisition, and retention rate. Furthermore, to understand whether the IVR users interact with Su-

per Abbu differently from users acquired through other channels, users were compared in terms of their activity, engagement, and IVR use sophistication.

The results show that IVR platforms (robocalls and Polly) performed better than other platforms (Facebook, flyers, radio, cable TV and rickshaw ads) in terms of user acquisition. Users acquired through the IVR entertainment service (Polly) performed better than other channels in all interface-related measures (activity, feature engagement, use of sophisticated interface features, and retention). Only robocalls, Facebook, and the entertainment service were able to acquire users at relatively low cost per user (around $1 or less). In contrast, most users acquired from outside of the entertainment service did not end up becoming long-term users of the development service. Their findings also show the comparatively lower performance of increasingly popular social media advertising platforms (Facebook) to recruit low-income users (91 recruits out of over 100,000 people reached).

This study established that voice forums perform better compared to other means as advertisement channels that recruit users for development intervention. It also revealed the surprising lack of success of mainstream social media platforms for user acquisition in developing regions.

8. Measuring Impact

Generally, impact assessment of information campaigns is carried out via follow up surveys to measure knowledge retention. However, measuring impact presents a unique set of challenges when information is mass disseminated using voice forums. To encourage spread, inclusiveness, and anonymity, such services do not require users to go through any formal recruitment or registration and a significant fraction of users only engage with a handful of calls. As a result, manual or telephonic baseline and end-line surveys are not feasible. In addition, such surveys are hard to scale and prone to delays. Consequently, despite a growing number of telephonic campaigns, there has been little focus on the measurement of retention of the delivered information. Due to a lack of rapid, scalable, and reliable mechanisms to quantify and measure knowledge retention, campaigns mostly resort to measuring only the extent of delivered information (for instance, number of calls and number of users who listened to the information content).

The services described in this section make use of voice-based quizzes to simultaneously spread information and measure its impact on the knowledge of the target users. These services use various incentives to engage the users in answering multiple-choice questions over IVR interactions. As the users provide their responses, they are informed about the correct answers to these questions. Such interactions allow the users to actively participate in learning about topics that are instrumental for their development. Their score as they keep engaging with the

quizzes acts as an indicator of how well the information dissemination influences their knowledge and beliefs.

Raza et al. merge knowledge-gap discovery, information dissemination and knowledge retention measurement into one service, using voice-based quizzes. They show that voice-based quizzes over simple mobile phones, consisting of multiple-choice questions, can be used to simultaneously measure the existing knowledge gaps as well as to disseminate information. Rephrased versions of quiz questions are repeated at regular intervals to measure retention of conveyed information. Long-term user engagement is encouraged by allowing users to contribute their own questions, and with social connectivity, gamification and spirit of competition that make the service engaging and fun for the target audience. The resulting service, Sawaal, allowed its open community of users to post and attempt multiple-choice questions and to vote and comment on them (Raza et al, 2019). Sawaal was designed to spread virally as users challenge friends via shared quizzes and compete for high scores. Administrator-posted questions allowed discovering knowledge gaps, spreading correct information and measuring knowledge-retention via rephrased, repeated questions. Community-contributed quiz content and an ability to play against friends for high scores, led to inclusion and ownership, active collaborative learning and a spirit of competition among the users. Sawaal spread organically among the target audience, received an enthusiastic response, and successfully retained a significant fraction of the users for several weeks. In 14 weeks and with no advertisement, Sawaal reached 3,400 users (120,000 calls) in Pakistan, who contributed 13,000 questions that were attempted over 450,000 times by 2,000 users. Knowledge retention remained significant for up to two weeks. Surveys revealed that 71% of the mostly low-literate, young, male users were blind.

Along a similar theme, Swaminathan et al., created Learn2Earn to spread awareness in rural areas of developing regions about critical issues in health, governance and other instrumental topics (Swaminathan et al, 2019). While Sawaal incentivized platform usage through gamification, competition and social media soft incentives of likes and votes, Learn2Earn, leveraged mobile payments to bolster public awareness campaigns in rural India. Users who interacted with the service listened to a brief tutorial followed by a multiple-choice quiz. Users who passed the quiz received a mobile top-up (approximately $0.14) and got the opportunity to earn additional credits by referring others to the service. The pilot deployment of Learn2Earn reached 15,000 people within seven week via word-of-mouth. Most of the users were young men including a large fraction of students. Learn2Earn was shown as a viable way of building awareness among target users about important topics.

While these techniques of remotely contacting low-income, low-literate populations, engaging them in interactive quizzes and gauging their awareness around

important topics were created originally in the development context, the lessons learned from these services have a broader appeal during crises like COVID-19 where it is difficult to conduct door-to-door surveys. Services like Sawaal and Learn2Earn may be employed to remotely examine knowledge gaps and gather user feedback in such crisis scenarios.

9. Open Challenges

Most of the work focusing on voice forums demonstrate their promise to serve as instruments of inclusion for low-literate people, rural residents, indigenous communities, and people with visual impairments. However, like any other social platform, voice forums 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, evaluation of several voice forums revealed extremely low participation of women despite that these services are designed to be inclusive, accessible, and usable for everyone; CGNet Swara and Sangeet Swara in India have only 12% and 6% female contributors, respectively; Baang and Polly in Pakistan have only 10% and 11% female contributors, respectively; Ila Dhageyso, a voice forum to connect government officials and tribal people in Somaliland, has only 15% female users. Table 22.2 shows the usage of Sangeet Swara and Baang by male and female users.

To examine why the participation of women is almost non-existent on social media voice forums that are designed to be inclusive, accessible, and usable for everyone, Vashistha et al. examined Sangeet Swara and Baang, two widely popular social media voice forums in India and Pakistan, respectively (Vashistha et al., 2019). Using a mixed methods approach spanning quantitative analysis of usage logs, content analysis of 10,361 posts containing 140 hours of audio data, and qualitative analysis of 50 surveys and interviews, they investigated how men and women interacted with each other on these services, what content they posted and voted, and what factors affected their participation.

The analysis found that female users of Sangeet Swara and Baang faced systemic discrimination and harassment in the form of abusive, threatening, and flirty posts directed at them. Most women lacked agency to retaliate due to deep-rooted patriarchal values that discourage them from arguing and questioning others. They were worried of backlash, on the service from predatory men and in real life from family members, if they record threatening responses or responded to flirty posts. On the other hand, many male users perceived women as objects of desire, reinforcing patriarchy in these digital social spaces. Male users who behaved inappropriately ganged up on those who criticized their behavior. Although only a fraction of male users recorded objectionable comments, most other male users

condoned the unruly behavior of other men and disapproved of objectionable messages less strongly than women did.

A large fraction of women in low-income communities only have access to shared mobile devices where usage is directed by male family members. Social media voice forums like Sangeet Swara and Baang have been only marginally successful in reaching to some women in these communities, however, they are still a long way from providing a welcoming, vibrant, safe, and enriching environment to women. The experience of women users of Sangeet Swara and Baang demonstrates that access is just a first step towards actual and meaningful social inclusion, and more is required to address secondary barriers to women's digital inclusion beyond the basic hurdles of literacy, connectivity, and poverty. There is an urgent need to re-think the design of these services and use participatory design approaches to ensure that we provide equitable and inclusive social platforms for women. We also emphasize the importance of using the intersectionality lens while designing technologies aiming for inclusion to ensure that they do not widen existing economic, racial, cultural, and societal inequalities.

Like mainstream social media platforms, voice forums also face grand challenges when tackling misinformation and disinformation, especially since the posts are audio recordings in local languages that are unsupported by the advances in natural language processing. Our ongoing investigations have uncovered the presence of a significant amount of false posts, misinformation, and hoaxes related to the COVID-19 pandemic on Baang. It is important to note that voice forums like Baang differ greatly from mainstream social media platforms like Facebook 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 understanding the composition of misinformation on voice forums, measuring diffusion properties of false posts, examining interaction of low-literate users with suspicious posts, understanding strategies users employ to verify information, and designing new tools and techniques to combat misinformation. The HCI4D community needs to tackle grand challenges like harassment, abuse, and misinformation to make these services truly diverse, inclusive, and impactful.

10. Summary

While Internet services have transformed how people participate in the information ecology and digital economy, these services have discounted the needs and wants of billions of people who experience literacy, language, socioeconomic, and connectivity barriers. To address the information and instrumental needs of these people, several HCI4D practitioners and researchers have designed voice forums

that enable users to interact with others via ordinary phone calls in local languages. Although voice forums have had a demonstrated impact on marginalized communities, most forums operate at a pilot scale because of known challenges in training and retaining users, managing local language content and costs of voice calls, difficulties in building and deploying these services, and measuring impact. This chapter discussed the innovation, implementation, and adaption of voice forums along with several approaches used to scale, sustain, and replicate them.

Discussion Questions

1. Why is speech-over-simple phones a viable strategy to provide information and connectivity to low-literate, non-tech savvy, visually impaired and marginalized segments of the society? How does this strategy compare with other modalities like Television, radio, newspapers, websites, smartphone apps?
2. Using the case studies discussed in the chapter as an inspiration, identify the features of voice forums that can help attain the following objectives.
    a. The service should spread virally among users
    b. The service should promote user engagement
    c. The interface should promote improvements in usage patterns of the users
    d. Users should keep returning to the service for a long time (several weeks, months or years)
    e. Users should retain the information delivered by the service
3. How does viral (person-to-person) spread compare with broadcast spread? Assume that you need to spread a voice forum among people in a rural developing region. You have two options: Schedule automated calls (robocalls) to every phone number in the village or create a viral service and recruit a handful of initial users to seed the spread. Discuss the benefits and shortcomings of each of these approaches.
4. If you have a high accuracy speech recognizer for languages of developing regions, would speech input be preferable over touch-tone (key press) input in voice forums? Discuss the benefits and shortcomings.
5. In addition to speech transcription, which other tasks are suitable for voice forum users to gain airtime and earn money?
6. Would voice forums become obsolete when underserved people gain access to low-cost smartphones and affordable Internet access?
7. How could voice forums be designed to be more inclusive, safe, and welcoming to women?
8. What lessons could be used from this chapter to design a voice-based social media platform for people in developed regions of the world?

References

Agarwal, S. K., Jain, A., Kumar, A., Nanavati, A. A., & Rajput, N. (2010). The spoken web: a web for the underprivileged. *ACM SIGWEB Newsletter*, (Summer), 1-9.

Ahmad, S. S. O., Naseem, M., & Raza, A. A. (2017). Maternal Awareness for Low-Literate Expecting Parents via Voice-Based Telephone Services. HCI.

Gulaid, M., & Vashistha, A. (2013, December). Ila Dhageyso: an interactive voice forum to foster transparent governance in Somaliland. In *Proceedings of the Sixth International Conference on Information and Communications Technologies and Development: Notes-Volume 2* (pp. 41-44).

Koradia, Z., Balachandran, C., Dadheech, K., Shivam, M., & Seth, A. (2012, March). Experiences of deploying and commercializing a community radio automation system in india. In *Proceedings of the 2nd ACM Symposium on Computing for Development* (pp. 1-10).

Koradia, Zahir, and Aaditeshwar Seth. "PhonePeti: exploring the role of an answering machine system in a community radio station in India." *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development*. 2012.

Mudliar, P., Donner, J., & Thies, W. (2012, March). Emergent practices around CGNet Swara, voice forum for citizen journalism in rural India. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development* (pp. 159-168).

Moitra, A., Das, V., Vaani, G., Kumar, A., & Seth, A. (2016, June). Design lessons from creating a mobile-based community media platform in Rural India. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development* (pp. 1-11).

Naseem, M., Saleem, B., St-Onge Ahmad, S., Chen, J., & Raza, A. A. (2020, April). An Empirical Comparison of Technologically Mediated Advertising in Under-connected Populations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Patel, N. (2011, March). Information service or online community? putting'peer-to-peer'in social media for rural india. In *Workshop on Social Media for Development at ACM Conference for Computer Supported Cooperative Work*.

Patel, N., Chittamuru, D., Jain, A., Dave, P., & Parikh, T. S. (2010, April). Avaaj 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 (pp. 733-742).

Patel, N., Shah, K., Savani, K., Klemmer, S. R., Dave, P., & Parikh, T. S. (2012, March). Power to the peers: authority of source effects for a voice-based agricultural information service in rural India. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development* (pp. 169-178).

Patel, N., Agarwal, S., Rajput, N., Nanavati, A., Dave, P., & Parikh, T. S. (2009,

April). 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* (pp. 51-54).

Smyth, T. N., Kumar, S., Medhi, I., & Toyama, K. (2010, April). Where there's a will there's a way: mobile media sharing in urban india. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 753-762).

Raza, A. A., Ul Haq, F., Tariq, Z., Pervaiz, M., Razaq, S., Saif, U., & Rosenfeld, R. (2013, April). Job opportunities through entertainment: Virally spread speech-based services for low-literate users. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2803-2812).

Raza, A. A., Rosenfeld, R., Haq, F. U., Tariq, Z., & Saif, U. (2013, January). Spread and sustainability: The geography and economics of speech-based services. In *Proceedings of the 3rd ACM Symposium on Computing for Development* (pp. 1-2).

Raza, A. A., Kulshreshtha, R., Gella, S., Blagsvedt, S., Chandrasekaran, M., Raj, B., & Rosenfeld, R. (2016, June). Viral spread via entertainment and voice-messaging among telephone users in india. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development* (pp. 1-10).

Raza, A. A., Saleem, B., Randhawa, S., Tariq, Z., Athar, A., Saif, U., & Rosenfeld, R. (2018, April). Baang: A viral speech-based social platform for under-connected populations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Raza, A. A., Tariq, Z., Randhawa, S., Saleem, B., Athar, A., Saif, U., & Rosenfeld, R. (2019, May). Voice-Based Quizzes for Measuring Knowledge Retention in Under-Connected Populations. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Swaminathan, S., Medhi Thies, I., Mehta, D., Cutrell, E., Sharma, A., & Thies, W. (2019). Learn2Earn: Using Mobile Airtime Incentives to Bolster Public Awareness Campaigns. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1-20.

Sherwani, J., Ali, N., Mirza, S., Fatma, A., Memon, Y., Karim, M., ... & Rosenfeld, R. (2007, December). Healthline: Speech-based access to health information by low-literate users. In *2007 International Conference on Information and Communication Technologies and Development* (pp. 1-9). IEEE.

Vashistha, A., & Thies, W. (2012). {IVR} Junction: Building Scalable and Distributed Voice Forums in the Developing World. In *Presented as part of the 6th USENIX/ACM Workshop on Networked Systems for Developing Regions*.

Vashistha, A., Garg, A., & Anderson, R. (2019, May). ReCall: Crowdsourcing on basic phones to financially sustain voice forums. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Vashistha, A., Garg, A., Anderson, R., & Raza, A. A. (2019, May). Threats, abuses, flirting, and blackmail: Gender inequity in social media voice fo-

rums. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Vashistha, A., Sethi, P., & Anderson, R. (2017, May). Respeak: A voice-based, crowd-powered speech transcription system. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 1855-1866).

Vashistha, A., Sethi, P., & Anderson, R. (2018, April). BSpeak: An accessible voice-based crowdsourcing marketplace for low-income blind people. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Vashistha, A., Cutrell, E., Borriello, G., & Thies, W. (2015, April). Sangeet swara: A community-moderated voice forum in rural india. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 417-426).

Vashistha, A., Cutrell, E., Dell, N., & Anderson, R. (2015, October). Social media platforms for low-income blind people in india. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (pp. 259-272).

Wolfe, N., Hong, J., Raza, A. A., Raj, B., & Rosenfeld, R. (2015, September). Rapid development of public health education systems in low-literacy multilingual environments: combating ebola through voice messaging. In *SLaTE* (pp. 131-136).

Wang, H., Raza, A. A., Lin, Y., & Rosenfeld, R. (2013, December). Behavior analysis of low-literate users of a viral speech-based telephone service. In *Proceedings of the 4th Annual Symposium on Computing for Development* (pp. 1-9).

Figure 22.1: High-level user interface design of CGNet Swara.

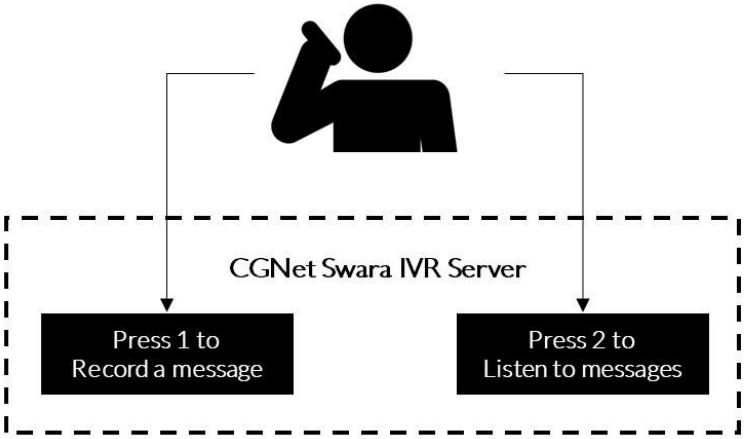

Figure 22.2: High-level simplified user interface design of Polly.

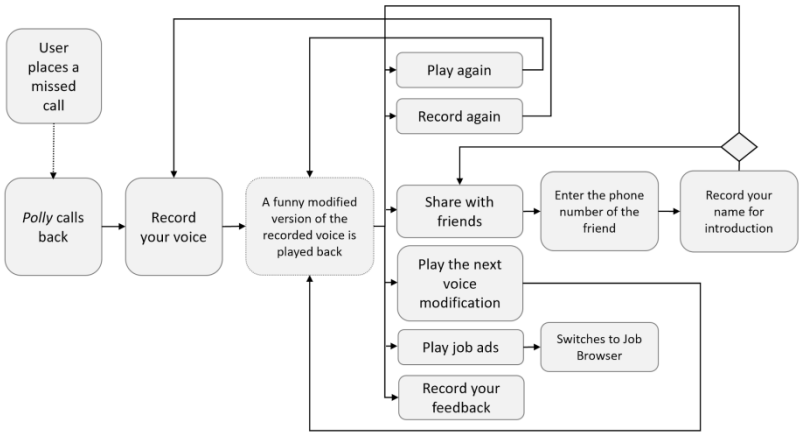

Figure 22.3: High-level simplified user interface design of Swara and Baang. Additional components in Baang's UI design are represented in red color.

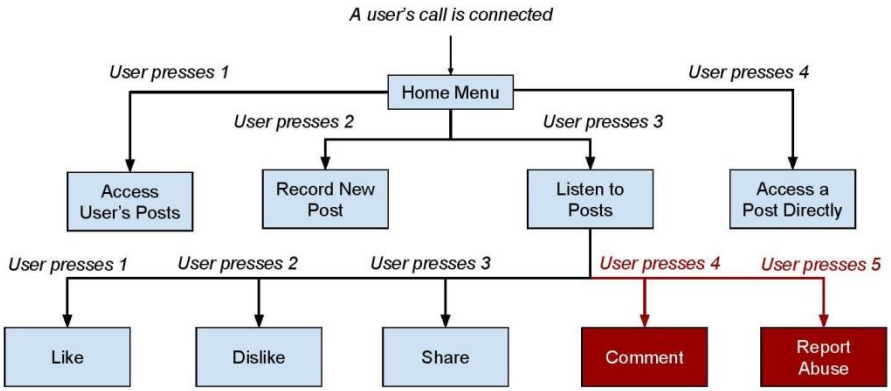

Figure 22.4: A high-level illustration of Respeak's design. Areas inside dotted lines represent the processes of the engine.

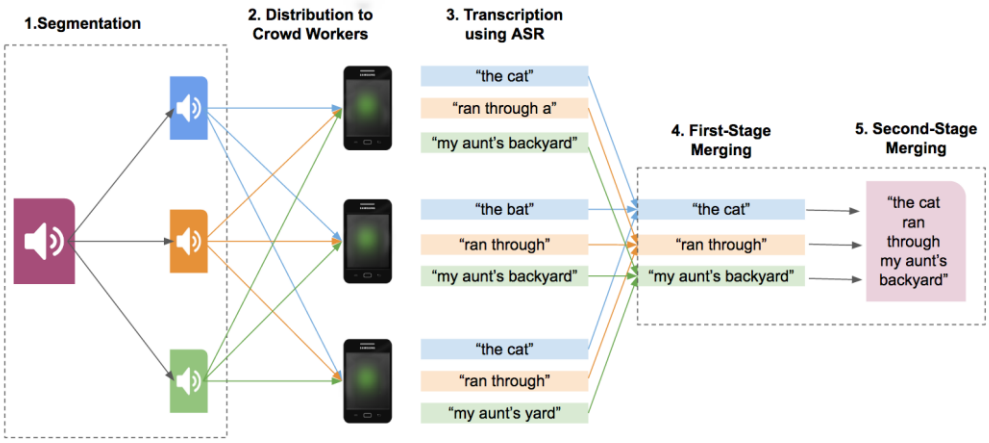

Table 22.1: Summarized comparison of major voice-forums

	Avaaj Otalo	CGNet Swara	Ila Dhag-eyso	Mobile Vaani	Gurga-on Idol	VoiKi-osk	Polly	Sang-eet Swara	Baang
Domain	Agriculture	Citizen Journalism	Civic En-gagement	Grievance Redressal	Social Media	Social Media	Social Media	Social Media	Social Media
Subsidized airtime	Yes	Yes	Yes	Yes	No	N/A	Yes	Yes	Yes
Voting	No	No	No	Yes	Yes	No	No	Yes	Yes
Audio comments	Yes	No	Yes	Yes	No	No	No	No	Yes
Sharing posts	No	No	No	Yes	No	No	Yes	Yes	Yes
Deployment length	7 months	2009 – now	5 months	2012 – now	1 month	4 months	1 year	4 months	8 months
Calls	6,975	137,000	N/A	10,000/day	306	20,499	636,000	25,000	269,468
Users	45	100,000	N/A	1.5 Mil-lion	252	976	165,000	13,500	10,721
Posts	896	13,595	4,300	300/day	31	2,532	387,301	5,000+	44,178
Female us-ers	0%	21%	15%	N/A	Low	N/A	11%	6%	10%
Users with visual im-pairment	N/A	N/A	N/A	N/A	N/A	N/A	< 1%	26%+	69%

Table 22.2: Usage statistics by gender for Sangeet Swara and Baang for random 5,000 posts.

Voice forum	Gender	Total posts	Unique users	Likes	Dislikes	Shares	Comments	Reports
Sangeet Swara	Male	4,764	419	21,630	58,644	189	Not Appli-cable	
	Female	275	31	270	2,636	15		

Baang	Male	4,142	376	8,181	5,541	778	1,942	2,061
	Female	325	31	508	253	2	25	25