Viral Spread via Entertainment and Voice-Messaging Among Telephone Users in India

Agha Ali Raza[1], Rajat Kulshreshtha[2], Spandana Gella[3], Sean Blagsvedt[4], Maya Chandrasekaran[4], Bhiksha Raj[1], Roni Rosenfeld[1]

[1] Carnegie Mellon University, Pittsburgh, PA
{araza, bhiksha, roni}@cs.cmu.edu

[2] IIT Guwahati,
rajat.kul@iitg.ac.in

[3] Microsoft Research India, Bangalore,
spandana@spandanagella.com

[4] Babajob.com,
{sean, maya}@babajob.com

ABSTRACT

We explore how development-related, voice-based, information services could organically spread among low-literate masses in the developing world. We report lessons learned from a *remote deployment* of "Polly" in India (from the US) to spread job-related information. Polly is an entertainment driven, voice-based service, available over simple phones that is aimed at familiarizing people with speech interfaces and mass-dissemination of development related information to low-literate users. In 2012, Polly had become viral in Pakistan and successfully spread recorded newspaper job ads to thousands of mobile phone users. Remotely deployed in India, Polly did not take off immediately as it did in Pakistan. Instead, it initially entered a six-month long phase of fluctuating, intermittent activity. We experimented with various forms of seeding and it eventually transitioned into a *viral phase*, with sustained transmission that continued for five months but without (exponential) growth. Finally, interface adjustments in response to user feedback enabling *plain-voice* asynchronous voice-messaging resulted in an abrupt *exponential* and *viral* growth amassing 10,349 phone calls by 1,613 users over a span of seven days. Of these, 299 users also transitioned to the job service. User feedback and surveys suggest possible reasons for each phase. We study the challenges of remote deployment and the interplay of user interface; language of the system; seeding mechanisms and active response to user feedback towards the uptake of the service. We also report a detailed comparison of viral spread in the two countries.

CCS Concepts

• Human-centered computing ~ Natural language interfaces • Human-centered computing ~ Sound-based input / output • Human-centered computing ~ User interface design • *Human-centered computing ~ User studies* • *Human-centered computing ~ Accessibility systems and tools.*

Keywords

HCI4D; ICT4D; Speech Interfaces; illiteracy; low-literate; cellular phones; viral; exponential spread; job search; mobile phones; telephone; entertainment; information services; communication services; low-skill jobs; remote deployment challenges.

ICTD '16, June 03-06, 2016, Ann Arbor, MI, USA
 ISBN 978-1-4503-4306-0/16/06…$15.00
DOI: http://dx.doi.org/10.1145/2909609.2909669

1. INTRODUCTION

Most ICTD projects design interfaces suitable for users who are low-literate and inexperienced with technology. Such projects typically require explicit user training (e.g. Health Line [20, 22], Avaaj Otalo [12]) and as a result are restricted to a moderate number of users. Based on the idea of using entertainment as a tool to help users overcome interface barriers (Smyth et al. [23]), Raza et al. [16, 17] made use of speech-based, viral, entertainment services for low-SES telephone users as a vehicle for disseminating core development-related services. This enables developing practices for entertainment-driven mass familiarization and training of low-literate users in the use of telephone-based systems, and using the viral platform to spread development services.

The ultimate goal of this line of research is to disseminate speech-based, development-related information and communication services to low-literate telephone users throughout the developing world. Such services may be used not only to provide information access to users but also to gather information from them in real-time and to allow them to share useful data amongst themselves.

Some uses of such services may include but are certainly not limited to: facilitating social and political activism via speech-based message boards and blogs; speech-based mailing lists; user surveys and polls to gather up-to-date information about health conditions, public sentiment, demands, needed facilities (health, social, infrastructure), grievances, available workforce and skilled labor (unemployed or looking for employment) etc.; citizen journalism; speech-based efficient marketplace; speech-based access to health, employment, skill-training, agricultural and other useful information. Very few such services are currently available to the low-literate and low-SES communities.

In [16], Raza et. al reported the pilot deployment of *Polly*, a viral entertainment service, in Pakistan. Polly is a simple telephone-based service that allows a user to record a short message, modify it using a variety of funny voice-manipulations, and forward it to friends. It was introduced to 32 low-literate office workers (the phone number was handed out with minimal to no explanation) and within 3 weeks attracted 2,000 users and resulted in more than 10,000 interactions (calls and voice-message deliveries).

The second (large-scale) deployment of Polly ([17]) remained online in Pakistan for a full year. It featured increased 30-line calling capacity and included Polly's first development-related service: an *audio job browser*. It was introduced through automated phone calls to 5 people. Within a year Polly amassed 165,000 users and resulted in over 636,000 interactions, including

200,199 forwarded voice messages and 22,104 forwarded job ads. At its peak it was spreading to 1,000 new users every day. The 728 job ads were listened 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, attracted a lot of blind users but remained primarily in the same socio-economic strata.

Wang et. al [27] found that, with experience, Polly's users respond faster to menus; make fewer mistakes and abortive attempts; show more interest in message sending; become more explorative of the system's capabilities, and better adapt themselves to its constraints. Some new users come familiar with the interface, presumably through offline introductions and demos by friends.

This paper reports Polly's first launch in India in collaboration with a commercial job portal, *babajob.com*, that maintains an active listing of thousands of informal and entry level jobs. Our team was not present on-ground and Polly was hosted in the US from where it was remotely launched in India. Based on the lessons learned during Polly's year-long deployment in India, and in comparison with its deployment in Pakistan, we attempt to understand factors impacting *virality* (defined as long, sustained chains of transmission to new users) and *exponential spread* of telephone-based speech services among low-SES people.

Unlike Pakistan [17], our initial attempts at seeding Polly in India did not lead to viral spread, and activity dwindled within a few weeks to a mere handful of daily calls (a *"sputtering"* phase) that persisted for nearly six months. This was in spite of careful bug fixes, various forms of seeding, multiple focus groups, and several changes to the interface and language of Polly. We eventually achieved virality following one of our seeding attempts, though the system still did not take off exponentially. This *viral-non-exponential* phase continued for five months, during which time we actively responded to feedback collected through Polly and conducted telephonic surveys of existing users. Based on feedback of active users, we changed Polly's interface to highlight its plain-voice asynchronous voice messaging capabilities. The immediate result was the hoped-for abrupt exponential spread comparable to the growth and spread of Polly in Pakistan.

The primary contribution of this paper is an understanding of some of the challenges involved in remote deployment of voice-based, telephone-based, information services in developing countries with very limited on-ground support. Remote deployment is a powerful mechanism that allows for launching such services in any country within a matter of days with minimal local support. This could be useful, for instance, to disseminate vital information in response to disasters/emergencies; may be using a modified form of spread suitable for such a context e.g. voice-based message boards [28]. Among the secondary contributions of this paper is the first reproduction of a service that had become viral in Pakistan, in a new geographical, linguistic and cultural setting (India); a side-by-side quantitative comparison of its spread patterns in the two countries and impacts of user-interface and system language on service uptake.

1.1 Research Questions

Our original research question was of reproducibility: *can Polly become viral in a different country/culture*? We hypothesized that the design that was found extremely successful in Pakistan will also be successful in India, a country similar in many ways. We were also interested in studying the challenges involved in deploying Polly remotely (with our team not being present on-ground and Polly making international calls from another country). In addition, we were interested in measuring the impact of the development related back-end service. However, as Polly did not immediately take off in India as it did in Pakistan, we became interested in the following questions:

- What are the challenges of remote deployment of IVR services?
- Which factors impact virality and exponential spread?
- During the non-exponential, viral phase: How was a daily stream of new users sustained yet without achieving exponential growth?
- Once virality is achieved: How does the spread in India compare with the spread in Pakistan?

The next section summarizes related work on the use of spoken dialog systems for development and on viral services in the developing world. Section 3 describes the design and user interface of Polly. Section 4 and 5 provide a detailed analysis of the one year long deployment, including usage patterns over time, demographics, user feedback, user behavior in response to interface changes and seeding attempts, and the eventual virality and exponential spread. Section 6 compares the virality and exponential spread in India and Pakistan. We conclude with a summary of our findings, lessons learned and discussion of future plans.

2. RELATED WORK

We find several attempts of user-interface design for the low-literate and tech-shy in the literature. Plauché et al [13] deployed information kiosks, supporting multimodal input (speech and touch screen) and output (speech and display) in Tamil Nadu, India, to disseminate agricultural information to farmers. The 50 low-literate participants, who had received some initial training (including short training sessions and group sessions), exhibited mixed preference towards speech vs. touch screen input. Speech data gathered from spoken interactions was used to further improve the Automatic Speech Recognition used in the kiosks [14]. *Warana Unwired* [26] replaced computer-based kiosks with SMS to disseminate agricultural information to sugarcane farmers. In a study conducted in three slums of Bangalore, Medhi et al [9] compared textual and non-textual interfaces for digital maps and job search systems for low-literate users. Their work highlighted the importance of consistent help options in the interface and confirmed user preference towards abstracted non-textual and voice based systems over textual ones.

Most efforts to provide speech-based information and communication services to the low-literate strongly rely on explicit user training. In *Project HealthLine* [20, 22] low-literate community health workers in rural Sindh (Pakistan) were trained (using human-guided tutorials) to use a telephone-based speech service to access reliable healthcare information. The speech interface performed well once the health workers were trained. This project highlighted the challenges involved in eliciting useful feedback from low-literate users. *Avaaj Otalo* [12] is another successful example of a speech interface for low-literate farmers. After an initial tutorial, the service was pilot-launched with 51 users in Gujarat, India. It offered three services: an open question/answer forum, an announcement board and a radio archive that allowed users to play broadcast radio programs. The open forum turned out to be the most popular service. Constituting 60% of the total traffic, the forum motivated users to find

interesting unintended uses like business consulting and advertisement.

Voice-based media has been shown to promote social inclusion among underserved communities. Mudliar et al. [10] examined participation of rural communities in India via citizen journalism using *CGNet Swara*, an interactive voice forum that became popular among its target audience. Koradia et al. [6] reported involving listeners of a community radio in voice content creation, feedback and station management. Vashista et al [24] explore community moderation in voice forums (*Sangeet Swara*) for entertainment-related content. They also explore the use of social media among their blind users and compare it with the use of voice-telephone-based forums [25]. Heimerl et al. [5] explored the utility of voice messaging in ten villages of rural Uganda and found it to be uniformly preferable over SMS and a good substitute to live calls in areas of poor coverage and intermittent connectivity. They also found voice messaging to be easier than SMS for visually impaired users.

Explicit training is not feasible when a service is oriented towards a large user base. An alternative is to rely on peer-training and on viral spread. Baker [1] lists some conditions for viral spread (albeit in the context of literate users and web-based services). *SMS-all* [2], a group text-messaging service in Pakistan, is an example of a virally spreading text based mobile service with two million users and four hundred thousand groups [2]. However, the use of text presumes a certain level of literacy.

Input modality: speech vs. push-button (DTMF) is another important question in developing telephone based interfaces. Project HealthLine [19, 20]====22] reports that speech performs better in terms of task completion for both literate and low-literate users. However, in terms of subjective user preference it provided no clear answer. Sharma's [18] user studies in Botswana with HIV health information systems for the semi and low-literate populations suggest user preference towards push-button over speech input while both modes perform comparably in terms of task completion. On the other hand [12] and [11] (conducted in a controlled environment) report that push-button performs better than speech in terms of both task completion and performance improvement. Patel et al [11] report the problem of transitioning between push-button and speaking as a major challenge and suggest that numerical input is more intuitive and reliable than speech. From these reports it seems that push-button is a better choice if user perception is vital for system adoption, especially where training and tutorials cannot be relied on.

A major hurdle to effective speech-based input is the lack of local linguistic resources and expertise for training a speech recognizer with the languages of the developing world. This is especially true in regions of great linguistic diversity like Pakistan and India, where even neighboring villages may speak different languages or dialects. The *Salaam* method [15] can be used for services requiring a small input vocabulary, as it provides high recognition accuracy in any language for up to several dozen words.

Affordable smart phones are rapidly gaining popularity in the developing world. Several researchers are exploring the use of text-free graphical interfaces [8] and multimodal (spoken and graphical) interfaces [4] for the low-literate, however, user's literacy and experience using smart phones plays an important role in the usability of these interfaces. Chaudry et al. [3] report that chronically ill patients of varying literacies are able to use text-free graphical interfaces and prefer the ones with more prominent buttons. A comparison of textual and text-free interfaces by Medhi et al. [7] shows that textual interfaces are problematic for novice low-literacy users; a live-operator is ten times more accurate than textual interfaces; task completion is the highest with graphical interfaces while spoken dialog improves user's efficiency, speed and comfort when system's language and dialect is understandable. In *Video Kheti*, Cuendet et al. [4] explored graphical interfaces used in conjunction with speech and touch-tone to allow low-literate farmers in rural India to find and watch agricultural videos in their own language and dialect. Their field study based on 20 farmers shows that although *Video Kheti* is usable and farmers are enthusiastic about it yet task success largely depends upon user's education level.

3. WHAT IS POLLY?

As described in [16] and [17]: *Polly is a telephone-based, voice-based application which allows users to make a short recording of their voice, modify it, and send the modified version to friends.*

3.1 Design

As described in [16] and [17]: Polly was initially conceived and designed via focus groups and surveys among low-literate office workers in a university in Lahore. Our first proposed application, *Songline*, failed to attract enough interest as users expressed privacy concerns about its broadcast nature and controversial cultural views towards a service aimed solely at singing and music. This led us to explore simple, non-controversial forms of entertainment and a voice message system based on funny voice modifications emerged as a promising candidate in which our subjects also exhibited a lot of interest. Initial surveys and focus groups guided our interface choices leading us to design with shallow call trees; simple, informal language; fewer menu options; local-dialing format for phone number entry etc. The voice mods were ranked and selected based on user preferences.

Following the initial success of Polly [16], our design process in all subsequent deployments involves launching Polly with bare-minimum options, gathering explicit and implicit user feedback [17, 27] and modifying its design accordingly.

3.2 User Interface

Polly's user interface is described in detail in [17]. The interaction starts when a user places a *missed call* to Polly's phone number. Polly calls back and after a short greeting (and before requiring any touch tone input from the user) casually prompts them to say something after the *beep*. As soon as the recording is finished (10 seconds, or shorter if the user presses # or remains silent for 4 seconds), user's voice is modified in a funny way and played back. User is then allowed to listen to the recording again, to forward it to friends (by entering their phone numbers), to hear a different voice modification, to give us feedback or to get transferred to the jobs service. Currently the following voice modifications are offered in the given order that users can cycle through:

1. A *Male to female* voice mod, achieved via raising the pitch and increasing the pace.

2. A *Female to male* voice mod, achieved via lowering the pitch and decreasing the pace.

3. A *drunk chipmunk* mod, achieved with pitch and pace modification,

4. An *I-have-to-run-to-the-bathroom* mod, achieved by a gradual pitch increase,

5. The original, *unmodified* voice of the user

6. *Whisper*, achieved by replacing the excitation source of user's voice with white noise

7. And *background music* added to the recording.

When users choose to forward their recording to a friend they are required to enter a phone number, record the name of their friend and also their own name to be played as a part of the introductory greetings when the message delivery call is placed. Message recipients can also choose to hear the phone number of the sender to prevent prank message deliveries. Message recipients are allowed to respond with a message of their own; to forward the recording to others; to create their own recordings or to get transferred to the jobs service.

Polly's contact information is shared with message recipients via SMS after their first two interactions as an additional mechanism for viral spread. It is also played during the phone call itself. User Feedback is elicited in the form of an unconstrained, unstructured recording (up to 60 seconds, with a silence timeout) from repeat users.

3.2.1 Interface Modifications in India

Polly-India initially supported voice prompts in Kannada since we aimed it at Bangalore. We later switched the prompts to Hindi as we found that Kannada was suitable for only a subset of our target population in Karnataka, and Hindi allowed it to spread to larger parts of India.

In India, the job service option allowed users to transfer to an IVR service run by *babajob.com* and choose a job category of interest, browse openings available in that category, register or leave messages for employers who advertised the opening. In the currently described launch only jobs from Bangalore were available.

Based on user feedback, we also added an *audio speed-dial* feature; an explicit menu option to forward messages using unmodified voice and longer recording intervals. Section 4 presents details of these features.

3.3 Remote Deployment in India

Polly was launched in India using a remote setup. In this setup, Polly receives call-back requests from users (via missed calls) on a local phone number set up in the target country and then calls them back directly from the US. The database, audio storage, call scheduler, automatic monitoring, reporting and error recovery mechanisms are all hosted in the US. Hence, the only on-ground hardware/software support that this setup needs is a phone capable of rejecting incoming calls and forwarding the caller's phone number to Polly's servers in the US over the internet. Remote deployment allows us to test Polly before going for a large-scale local deployment (as described in [17]) or even to run Polly long-term in case a local deployment is not feasible.

4. RESULTS AND ANALYSIS

4.1 Seeding Strategies and Interface Changes to Induce Viral Spread

We initially attempted to make Polly viral in India with minimal on-ground support but a few months into the project it became clear to us that this had little chance of success. We needed local partners to help with seeding and demos; culturally appropriate translation of prompts in local languages; prompt recording with careful attention to tone and style; system maintenance and testing; reporting and fixing of bugs; conducting of focus groups and surveys and translation of the automatically collected user feedback. Six months into the project we teamed up with *Microsoft Research, India (MSRI)*, although Polly remained hosted in the US throughout the project.

As shown in Figure 1, Polly was launched in Bangalore, India on July 03, 2013. The initial seeding was done via automated calls to 100 frequent users of babajob.com. As it did not lead to sustained call traffic, a few days later we tried manual seeding where a few office workers at babajob.com gave a demo of Polly to their family members. It is clear from the figure that our initial seeding did not lead to sustained activity. We focused our efforts to ascertain that the system was technically sound and bug free. Existing users were informed via automated calls following all bug fixes and interface changes.

As frequent "manual" (human) seeding was not feasible with minimal on-ground presence, we tried seeding Polly via "cold calling" random numbers in Hindi speaking areas (avoiding *do-not-disturb* subscribers). These *cold-seeding* attempts were made between Oct 23 and Nov 26 (blue lines in Figure 1). Another such attempt was made on Jan 03, 2014. We tracked the activity of all seeded users and found that none of our cold-seeding attempts generated any noticeable activity.

To induce exponential spread, we experimented with tweaking several interface features: language and gender of the system prompts; cold seeding on various times-of-day and days-of-week; cold-seeding with and without initial introduction of the service; an audio speed-dial feature; adding the job service option from the very first interaction of a user vs. only adding it *after* a user gains some experience with Polly; changing the order of voice modifications; servicing the call-back requests more quickly. Some of these attempts resulted in minor fluctuations of call and user traffic but none led to sustained activity or spread.

On Jan 09, 2014, with the help of MSRI we seeded Polly with undergraduate students who also advertised it through their personal connections (friends, Facebook pages, blogs etc.). This led to the beginning of the viral-non-exponential phase. With the help of our new partners we also started getting the accumulated recorded user feedback translated and began responding to it. We also started conducting telephonic interviews of existing active users. Following are the major changes that we made in response to the gathered feedback:

1. Audio Speed Dial: In order to facilitate user interaction we added an optional audio speed-dial feature as surveys revealed that users find it difficult to enter long phone numbers repeatedly. The speed-dial remembers the most recently entered nine phone numbers and allows users to assign names to them. Users can later choose a number by pressing a key (e.g. for John, press 1, for David, press 2,…). Users are also allowed to forward their voice to multiple recipients with the same or different modifications applied.

2. Messaging using Unmodified Voice:

a. On May 19, 2014 we added a new option in Polly's menu that explicitly offered forwarding a message without voice modification. *This option was always implicitly available to the users* via cycling to the fifth voice modification, but Polly's menu or instructions did not explicitly announce it. It is also of note that we did not advertise this new menu option upfront and it took most of the users several days to discover it and start using it. Call traffic starting growing soon afterwards.

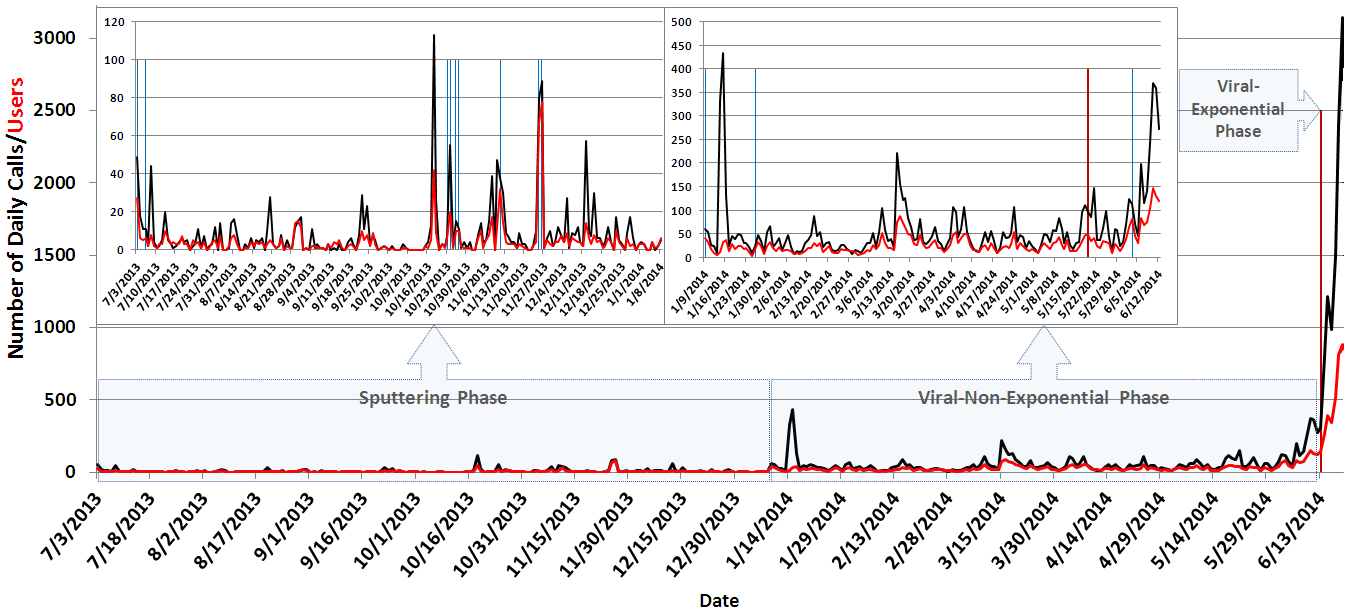


Figure 1: Polly's Activity Levels in India (Blue vertical lines: Seeding attempts; Dark red line: Major interface changes.)

b. On Jun 13, 2014, based on popular demand, we increased the recording duration from 10 seconds to 25 seconds. Polly's call and user volumes abruptly started growing exponentially that very day. 299 users also transitioned to the job service from Polly during this period. On June 19 we had to suspend the service after exhausting our telecommunication budget.

5. USER SURVEYS AND FEEDBACK ANALYSIS TO UNDERSTAND SPREAD

5.1 Telephonic Surveys

During the *viral* phase, we did two rounds of surveys, manually calling up several users of Polly (the calls were mostly placed from the US) to gather feedback and to find out the reasons for lack of exponential spread.

We conducted the first survey three months after the Jan 09 seeding (which was done through undergraduate students over Facebook, blogs etc.). We called 20 of our most active users, reaching 16 of them. To our surprise, all of them were young men (6), associated with skilled labor (10), low-SES (10), mostly low-literate (7) and living in Delhi (7), West Bengal (3), Mumbai (1) and Jaipur (1). None of the students who seeded the system were from any of these cities. The interviewees' professions were: carpenter (1), marble polisher (1), tinsmith (1), goldsmith (1), tailor (1), embroider (3), diamond-assorter (1), self-employed (1) and one unemployed person. Most of them had heard about Polly from friends (from other cities in two cases). They used Polly for free voice messaging (5) and fun (7) (without using voice modifications).

The interviewed users were cost conscious as some kept confirming if Polly is indeed free. Interestingly, 9 out of 20 interviewees initially denied ever using the service, claiming that their friend/brother/someone else might have used it. Following this "disclaimer", they nevertheless showed intimate familiarity with Polly and its various uses. The formal denial suggested to us a fear of being asked to pay for the service (of course there could have been other reason that we are not aware of).

Most common suggestions included: Not to morph the voice as the modified voice is unclear, difficult to recognize and it confuses users into thinking that the service is not working properly; to increase the message recording interval and to make Polly available in local languages, especially Bangla.

We conducted another set of 250 survey calls in mid-May 2014. Only 71 calls were answered as we tried calling several user types (super-spreaders, spreaders, non-spreaders etc.) including several who had not shown any interest in using Polly. We found that the best time to get a response from users in on weekend evenings and nights.

One major reason for the lack of interest cited by non-spreaders was the language barrier. Polly supported Hindi prompts while the majority of our users came from West Bengal and Calcutta and understood little to no Hindi. More motivated users got instructions from Hindi speaking friends and operated Polly accordingly.

Category	# Responses	Outcome
Literacy	28	12+ Years: 16 Under 10 Years: 10 None: 2
Age	35	Under 20: 5 20-30: 26 30+: 3 80: 1
Locations	12	Delhi: 7, Calcutta: 2, West Bengal: 1, Mumbai: 1, Jaipur: 1
What do you use it for?	31	Messages: 23 Messages. Fun. Jobs: 6 Messages and Jobs: 3

Table 1: User Demographics of Survey Calls

Users complained about call audio quality (22%); requested the ability to send unmodified messages (55%); increased recording interval and better message sending abilities (32%); improved job ads service (32%) and availability of Polly in Bangla (19%). Almost all users were using Polly only to send and receive voice messages. Very few reported using Polly to access jobs, and

around 40% had no idea that the job service even exists. We were not able to ascertain if anyone actually got a job through Polly.

Table 1 summarizes the demographics of surveyed users. Among the contacted users we also encountered several school/college students and three blind men. Most of the blind men belonged to a blind institute in Calcutta and were all prolific users. They claimed that a lot of their blind friends are also using Polly. Some very long term users reported having used Polly for several months.

5.2 Feedback Collected through Polly

Table 2 summarizes automated user feedback collected through Polly. Users who provided the feedback were mostly Bangla-speaking men who could also speak some Hindi. Even the recordings marked as Hindi often contain several Bangla words.

Of the 1,029 recorded feedback files: - 773 were empty, noise, messages by confused users.
Feedback and Suggestions
There were 256 files with real feedback. Of those: - 174 simply contained praise for Polly and its features - 82 contained suggestions and requests. Of those (the following are not mutually exclusive): • Audio quality is bad: 21% • Key-presses are not recognized correctly sometimes: 10% • Allow sending unmodified voice messages: 22% • Increase message recording interval: 18% • Make Polly available in Bangla: 16% • Demanded/Suggested new voice modifications: 9% • Job service related feedback: 6% • Demanded other features, mostly making Polly more suitable for voice messaging: 21%
Gender
Of 419 recordings successfully annotated for gender: • Female: 13% • Male: 87%
Language
Of 376 recordings annotated for Language: • Bengali: 38% • Hindi: 47% • English: 7% • Mixture of Hindi, Bengali, English: 8%

Table 2: User Feedback

Most users seemed to like Polly a lot. Several low-SES blind users said that it was a much needed service for them as they could not use regular SMS. They used it to remain connected with their friends. Bad audio quality was the most common complaint. Later testing revealed that indeed the audio quality degraded in calls to Indian phone numbers (as opposed to when tested from the US) which critically reduced the entertainment appeal of the voice mods. The reasons are not clear to us, but we suspect it to be a mix of sporadic low quality of international IP calls and bad signal reception in some regions. Some users also reported problems with key press (DTMF) recognition, a problem also mentioned anecdotally by other researchers working with IVR in India.

The improvements suggested by users all aimed to tweak Polly into a better voice messaging service (several users even referred to it as "voice sms"). Among the most popular suggestions were to increase the recording interval; allow unmodified messages; make unmodified voice the default option; provide a way for users to directly talk to their friends through Polly (like a conference call). Bangla speakers complained that they do not understand Hindi; the service does not speak like "one of them" and most people around them cannot use Polly because of the language barrier. Some users complained that the job ads were not suitable for their location; were not available in Bangla and were not updated frequently enough.

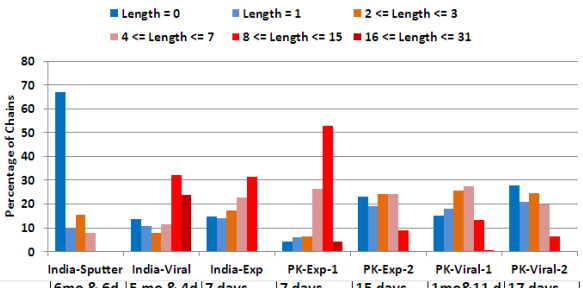
6. POLLY IN PAKISTAN AND INDIA – A COMPARISON

This section focuses on a very different question: Once obtained, how does the spread in India compare with the spread in Pakistan in terms of virality, exponential growth, user retention and choice of options during the interaction?

In what follows we compare the three phases of Polly in India with the following phases of Polly's Pakistan deployment (see [17] for figures and details): *PK-Exp* is the first exponential growth phase in Pakistan that started from the initial seeding on May 09, 2012 and continued till May 15, 2012. *PK-Exp2* is the second growth phase in Pakistan that started on Jun 27, 2012 with the fixing of a telecom bug (that reduced capacity) and continued till July 11, 2012. The viral-non-exponential phase *PK-Viral* is defined as May 16, 2012 to Jun 26, 2012 [17]. We also define a *steady-state* phase from Pakistan that represents a six month period when we stopped subsidizing message deliveries and imposed a quota of one subsidized call per day for each user (publication pending).

6.1 Virality and Exponential Spread

Polly spreads mostly through forwarded voice messages that accounted for 72% of its new (*Polly-Introduced*) users in India. Another 6% were introduced through our seeding attempts. The remaining 22% were introduced by *word of mouth* i.e. informed about Polly by friends or family face-to-face or through a regular phone call or text message. Sometimes the introduction was also accompanied by a demo. We were able to assign 8% of these users to their putative introducers by tracing their message passing activity. If a *Word-of-Mouth introduced* (WoM) user *Y* sends or receives a Polly message from a user *X* who started using Polly before *Y*; we assign *X* as the word-of-mouth introducer of *Y*.



Period	6mo & 6d	5 mo & 4d	7 days	7 days	15 days	1mo&11 d	17 days
R0	0.2	0.7	1.2	1.3	0.8	0.7	0.7
Avg Chain	0.9	9.3	5	8.1	3	3.7	2.5
Max Chain	7	26	15	21	19	23	19

Figure 2: R_0 and Chain Lengths

Viral spread is characterized by long, sustained chains of transmission to new users. The **Basic Reproductive Rate** of spread, R_0, is defined as the expected number of new users introduced by a current user over its lifetime, in a fully susceptible population. Exponential spread occurs when $R_0 > 1$ while activity dies out quickly if $R_0 < 1$. We estimate R_0 for a given cohort of users by summing up introductions over a user's first 7 days of activity. This establishes a lower bound on R_0 because: (1) a small subset of Polly's users continues to introduce new users for several weeks and beyond; and (2) there remains a large group of unassigned, and therefore unaccounted for, Word-of-Mouth users.

Figure 2 shows R_0 estimates and chain lengths of new user introductions during the various phases. Only *PK-Exp* and *India-Exp* qualify as truly exponential as $R_0 > 1$. *PK-Exp-2* appears exponential based on call and user traffic but is more close to the viral-non-exponential phases because it was induced by an increase in traffic due to increased capacity and not because of true growth. All but the *India-Sputtering* phase are characterized by long chains of transmission and are clearly viral.

6.2 User Retention and Fecundity

Figure 3 compares user retention across phases, based on users who initiate call-back requests to Polly. During *India-Exp* and *PK-Exp* phases a large number of users kept returning to Polly. Nearly half of the users called back on their second day, more than 30% on their third day and this trend continued for several days. All viral-non-exponential phases are also similar and are characterized by lesser retention as compared to exponential phases. Sputtering phase exhibits very little retention. All but the sputtering phase have a non-negligible fraction of long-term users (not shown).

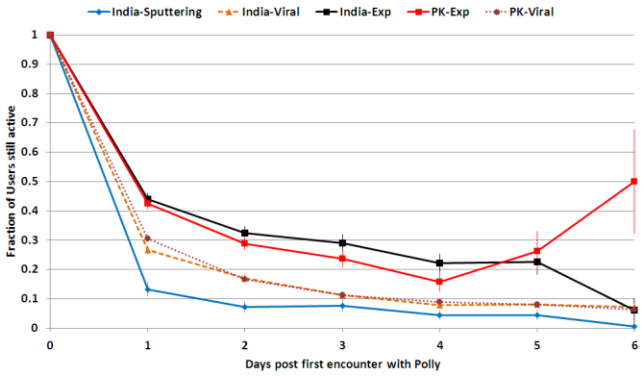


Figure 3: User Retention

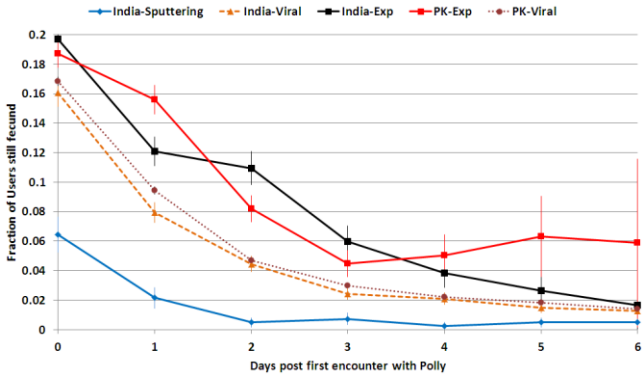


Figure 4: User Fecundity

Figure 4 shows *user fecundity* - the tendency of a user to introduce new users, as a function of days post first encounter with Polly. The cohort of potentially fecund users is composed of all users who take part in at least one successful call with Polly. We see that even on their first day, only 18%-20% of users spread Polly to new people during the exponential phases, and around 16% do so during the viral-non-exponential phases. These fractions drop quickly as the users 'age', but a small fraction of users continues bringing in new users several days after their first use of Polly. During the sputtering phase, only around 6% of users introduce new users on their first day, but even this phase has a small fraction of long-term fecund users.

This shows that virality and exponential spread are due to several users returning to the service and contributing towards its spread and not just a handful of *super-spreaders*.

6.3 Choice of Voice Modifications

Figure 5 shows the percentage of voice modifications chosen by users in their forwarded messages in the various phases. Here we have further divided the viral phase into two sub-phases: *Viral before unmodified voice menu option* and *Viral after unmodified voice menu option*. The first option (*male-to-female)* is very dominant across all systems and phases, presumably because (1) it is the first option; (2) it is funny; (3) some users do not know that other modifications exist; (4) it is clear enough to be used for serious message passing; or (5) frequent users prefer not to cycle through even for serious messages. There is significant increase in the use of the unmodified voice for message delivery after the explicit option to do so was introduced. In the exponential phase the percentage of such messages increases threefold, indicating a clear user preference.

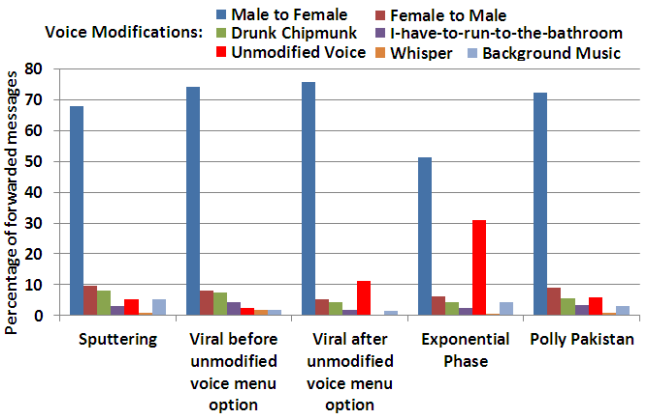


Figure 5: Choice of Voice Modifications

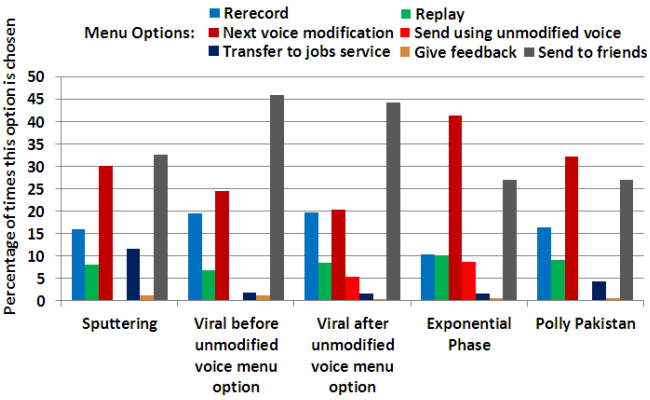


Figure 6: Choice of Menu Options

6.4 Choice of Menu Options

Figure 6 shows the prevalence of main menu selections during the various phases. The *transfer to job service* option is used a lot during the sputtering phase but not afterwards. One reason is that our jobs service only has a listing of jobs from Bangalore and after the sputtering phase Polly spread out to other areas of India. As a result, users did not remain as interested in this option. Another reason we discovered is that our initial seeding in Bangalore brought in some *job-brokers* who used to call in just to browse job ads, and presumably sell this information to others.

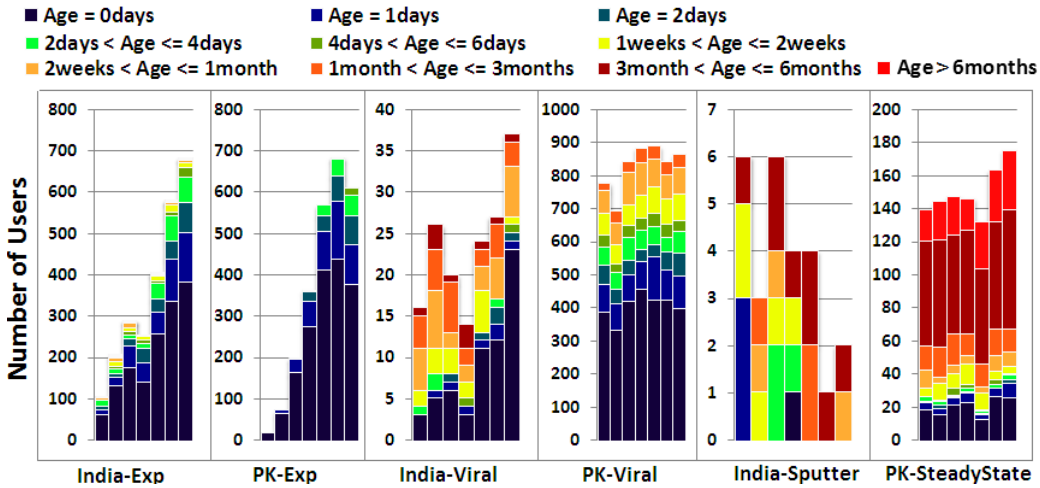


Figure 7: Distribution of Users by Polly Age (Days post first encounter with Polly). The horizontal axis shows consecutive days of a prototypical week from each phase.

In the viral phases users showed much interest in forwarding messages and much less interest in exploring the voice modifications. With the introduction of explicit option to send unmodified messages users not only started using this option but also started cycling through the mods to hunt for the unmodified voice (as is evident from Figure 5). It is as if they got a hint from this menu option that it is possible to send unmodified messages using Polly. The use of rerecord key is reduced to half during the exponential phase as the recording interval was increased from 10 to 25 seconds and users no longer needed several attempts to fit their message in the recording interval

6.5 Distribution of Users by Experience

Figure 7 compares distribution of daily users by their *Polly-Age* in prototypical week-long periods of activity from each phase. Here *Polly-Age* is defined as *days post first encounter of a user with Polly*. The height of each bar shows the number of users who used Polly on that date while the colors represent their Polly-ages.

During the exponential phases, user-traffic is almost exclusively composed of users who are new to Polly or have very recently started using it. During the viral-non-exponential phase in India we find nearly half of the daily users to be more than a week old with a significant number of 2weeks-, 1month-, and even more than 3month old users. The viral-non-exponential phases in Pakistan are not very different, though there we find around 30% of the users to be more than a week old.

The most interesting phases are India-sputter and PK-Steady State, where user traffic is dominated by long-term users. In India-sputtering phase, the few users who called back every day are mostly very old users. In PK-Steady State more than 70% of the daily traffic comprised of users who were introduced more than a week before, nearly 50% were more than 3 months old, and around 10% of daily users were more than 6 months old! Thus, during these phases of low-activity, the service is kept alive mostly by very long-term, loyal users.

7. SUMMARY AND DISCUSSION

Our main question regarding reproducing the viral spread of Polly in India using a remote-deployment model was eventually answered in the affirmative. But we discovered several reasons, some of them directly linked to remote deployment and lack of on-ground presence which delayed the onset of exponential spread:

Ineffective Seeding, either not being able to reach appropriate 'seeds' (people to initiate viral transmission), or not introducing the service "appropriately".

In Pakistan Polly was only seeded once, with low-literate office boys who were handed pieces of paper with Polly's phone number written on them. They were instructed to explore what this service is about, after explaining that it is a fun service that is free to use. The "seeders" also remained available for a few days to help the seeds with their questions and technical difficulties. In addition, the seeds may have felt obliged to spread the service to their friends out of their respect and relationship with the seeders, whom they personally knew. Regretfully, we were not able to use these tactics in India, and had to rely on several less intensive, less personal introductions with little or no success:

Our initial seeds of July 03 were active users of the jobs service. They were informed about Polly via automated calls. As the calls originated from the phone number of the job service, and also mentioned Polly as being a service launched by them: the seeds ended up using Polly only to access the job-service, some probably not even realizing that Polly has other options to offer. Also they had to pay for the service if accessed directly while they could so for free through Polly. As a result it is not a surprise that many of them called periodically and only listened to job ads. We also suspect some of them to be job brokers, who therefore hardly ever attempted to explore the voice modifications or to spread Polly to friends.

The second round of seeding with the friends and family of the workers at our partners' office, also suffered from the same basic flaw: the seeds continued to see Polly as another way of accessing the job-service.

In Pakistan, the second (large-scale) deployment of Polly was seeded only once via cold-calls to 5 old users. These calls were placed from a new phone number, a whole year after they had used the first version of Polly. Of course there was the benefit that these users already knew about Polly and could figure out that it's the same service if they did not hang up immediately.

Based on this we attempted cold calls in India but without much success. Unlike Pakistan, Indian cellphone users are faced with a large number of daily spam calls and text messages. Frauds perpetrated through mobile services have also deteriorated users'

trust. This explains the lack of interest and mistrust exhibited by the recipients of Polly's *cold-calls* – automated calls from an unknown phone number.

Inappropriate voice prompts: Polly's voice prompts in India were not suitable for the majority of its potential users.

In India, Polly was initially seeded in Bangalore, Karnataka where Kannada is the official state language. However, Bangalore is a metropolitan city with people from diverse linguistic backgrounds and Kannada is only appropriate for a modest subset of them. Additionally, we later found through user feedback that our Kannada system prompts were too formal-sounding and literary, and were therefore unsuitable for low-literate Kannada speakers. Our choice of language also prevented Polly's spread beyond the state of Karnataka until we changed it to Hindi. The main lessons are:

- The prompts should be translated using very simple and casual language that the majority of low-literate people can understand (as opposed to formal and educated language). All prompts should be friendly and easy to deliver.
- The voice artist should visualize as if informally (yet politely) talking to a friend; urging them to use the service, so that they are no longer shy or afraid after listening to these prompts.

Another interesting observation is that during the non-exponential phases, call traffic never died out completely. This was surprising, because standard epidemiological theory predicts either an exponential growth or an exponential decay. Based on the analysis and survey results presented earlier, we believe the following to be the major factors responsible:

During the *sputtering* phase, Polly was occasionally called by users of the job service (presumably professional job brokers) who wanted free access to the job ads. A handful of interested users were responsible for generating the rest of the fluctuating call traffic.

During the *viral-non-exponential* phase, activity was mostly due to long-term users (as shown in Figure 7) who had discovered some utility in the service, most likely free voice messaging. Interestingly, a fraction of such utility-oriented users kept introducing new users for many months post their first introduction to Polly.

Contrary to our expectation, Polly users in India were not as interested in the entertainment aspect of Polly (voice modifications) as were their Pakistani counterparts. This could have been the result of poor call audio quality in India. Quality of the call is directly linked to how funny the voice mods would end up sounding and hence the overall entertainment-appeal of the service. Or it might have been due to an intrinsic difference between the people reached by Polly in the two countries. Clearly, Polly's users in India wanted a good voice messaging platform. User feedback kept reflecting that demand and as soon as we tweaked Polly accordingly (by introducing a way to send unmodified messages and increasing the recording interval), Polly immediately achieved exponential spread.

7.1 Lessons Learned

Our experience in India shows that remote deployment and management of speech-based telephone services without committed local partners has little chances of success. Local partners play a pivotal role in both the initial deployment as well as long-term system maintenance. However, remote hosting of voice services could work well as long as enough local support is available to do appropriate translation and recording of prompts, understanding the local culture and to attempt various forms of seeding and frequent testing of the system

Another important lesson is that people with very similar demographics, linguistic and cultural background could still have very different needs. On the positive side, our eventual success in India shows that services like Polly that allow rapid adaptation of design based on user feedback do have a chance of becoming popular in new regions with different user needs as long as user feedback is vigilantly monitored and promptly responded to.

We found that cold calling does not play any significant role towards service uptake. Initial seeding must be personal, face-to-face and accompanied by demos. The seeders should ideally remain available for a few days to resolve issues faced by seeds. Voice prompt translation must be done with careful attention to appropriate wording, tone, style and quality.

Virality is easier to obtain as compared to exponential spread and both result from significant fraction of users who keep returning to the service and keep spreading it for several days, as opposed to a handful of *super-spreaders*.

Specific to Polly we found that once virality is obtained its organic spread follows a unique, characteristic pattern in terms of basic reproductive number; chain lengths; socio-economic background, gender and age of users; fraction of spreaders and distribution of new vs. old users. Our original "ring-of-fire" hypothesis, that all of Polly's users lose interest after a few days of interaction, fails to explain why activity does not die out in steady state, non-exponential phases (when $R_0 < 1$). We found the explanation by observing user distribution by Polly age that there are a significant number of very long term users who keep the service alive by returning to it even after several months.

8. ACKNOWLEDGMENTS

Partial support for the project was provided by the *U.S. Agency for International Development* under the Pakistan-U.S. Science and Technology Cooperation Program, the *Fulbright Program* and *Higher Education Commission* of Pakistan. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity. We are grateful to *William Thies*, *Indrani Medhi Theis* at *Microsoft Research, India* and *Rita Singh* at *Carnegie Mellon University, Pittsburgh* for their help and guidance throughout this project.

9. REFERENCES

[1] Interview with edward baker about the viral factor | entrepreneurial minded. http://michaelrucker.com/thought-leader-interviews/edward-baker-the-viral-factor/. Accessed: Sunday, Apr 03, 2016.

[2] Sms-all cheapest group sms service. http://www.smsall.pk/. Accessed: Sunday, Apr 03, 2016.

[3] Chaudry, B. M., Connelly, K. H., Siek, K. A., and Welch, J. L. Mobile interface design for low-literacy populations. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium* (2012), ACM, pp. 91–100.

[4] Cuendet, S., Medhi, I., Bali, K., and Cutrell, E. Videokheti: making video content accessible to low-literate and novice

users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), ACM, pp. 2833–2842.

[5] Heimerl, K., Honicky, R., Brewer, E., and Parikh, T. Message phone: A user study and analysis of asynchronous messaging in rural uganda. In *SOSP Workshop on Networked Systems for Developing Regions (NSDR)* (2009), pp. 15–18.

[6] Koradia, Z., Balachandran, C., Dadheech, K., Shivam, M., and Seth, A. Experiences of deploying and commercializing a community radio automation system in india. In *Proceedings of the 2nd ACM Symposium on Computing for Development* (2012), ACM, p. 8.

[7] Medhi, I., Patnaik, S., Brunskill, E., Gautama, S., Thies, W., and Toyama, K. Designing mobile interfaces for novice and low-literacy users. *ACM Transactions on Computer-Human Interaction (TOCHI) 18*, 1 (2011), 2.

[8] Medhi, I., Sagar, A., and Toyama, K. Text-free user interfaces for illiterate and semi-literate users. In *Information and Communication Technologies and Development, 2006. ICTD'06. International Conference on* (2006), IEEE, pp. 72–82.

[9] Medhi, I., Sagar, A., and Toyama, K. Text-free user interfaces for illiterate and semiliterate users. *Information Technologies and International Development 4*, 1 (2007), 37–50.

[10] Mudliar, P., Donner, J., and Thies, W. Emergent practices around cgnet swara, a voice forum for citizen journalism in rural india. In *International Conference on Information and Communication Technologies and Development (ICTD)* (2012), vol. 2.

[11] Patel, N., Agarwal, S., Rajput, N., Nanavati, A., Dave, P., and Parikh, T. A comparative study of speech and dialed input voice interfaces in rural india. In *Proceedings of the 27th international conference on Human factors in computing systems* (2009), ACM, pp. 51–54.

[12] Patel, N., Chittamuru, D., Jain, A., Dave, P., and Parikh, T. Avaaj otalo: a field study of an interactive voice forum for small farmers in rural india. In *Proceedings of the 28th international conference on Human factors in computing systems* (2010), ACM, pp. 733–742.

[13] Plauché, M., and Nallasamy, U. Speech interfaces for equitable access to information technology. *Information Technologies and International Development 4*, 1 (2007), 69–86.

[14] Plauché, M., Nallasamy, U., Pal, J., Wooters, C., and Ramachandran, D. Speech recognition for illiterate access to information and technology. In *Proc. of 2006 International Conference on Information and Communication Technologies and Development*.

[15] Qiao, F., Sherwani, J., and Rosenfeld, R. Small-vocabulary speech recognition for resource-scarce languages. In *Proceedings of the First ACM Symposium on Computing for Development* (2010), ACM, p. 3.

[16] Raza, A., Milo, C., Alster, G., Sherwani, J., Pervaiz, M., Razaq, S., Saif, U., and Rosenfeld, R. Viral entertainment as a vehicle for disseminating speech-based services to low-literate users. In *International Conference on Information and Communication Technologies and Development (ICTD)* (2012), vol. 2.

[17] Raza, A. A., Ul Haq, F., Tariq, Z., Pervaiz, M., Razaq, S., Saif, U., and Rosenfeld, R. 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* (2013), ACM, pp. 2803–2812.

[18] Sharma Grover, A., Plauché, M., Barnard, E., and Kuun, C. Hiv health information access using spoken dialogue systems: touchtone vs speech.

[19] Sherwani, J. Speech interfaces for information access by low-literate users in the developing world. *PhD Thesis* (May 2009).

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

[21] Sherwani, J., Palijo, S., Mirza, S., Ahmed, T., Ali, N., and Rosenfeld, R. Speech vs. touch-tone: Telephony interfaces for information access by low literate users. In *International Conference on Information and Communication Technologies and Development, ICTD* (2009), IEEE, pp. 447–457.

[22] Sherwani, J., Tongia, R., Rosenfeld, R., Ali, N., Memon, Y., Karim, M., and Pappas, G. Health-line: Towards speech-based access to health information by semi-literate users. *Proc. Speech in Mobile and Pervasive Environments, Singapore* (2007).

[23] Smyth, T., Kumar, S., Medhi, I., and Toyama, K. Where there's a will there's a way: mobile media sharing in urban india. In *Proceedings of the 28th international conference on Human factors in computing systems* (2010), ACM, pp. 753–762.

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

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

[26] Veeraraghavan, R., Yasodhar, N., and Toyama, K. Warana unwired: Replacing pcs with mobile phones in a rural sugarcane cooperative. *Proceedings of ICTD* (2007).

[27] Wang, H., Raza, A. A., Lin, Y., and Rosenfeld, R. Behavior analysis of low-literate users of a viral speech-based telephone service. In *Proceedings of the 4th Annual Symposium on Computing for Development* (2013), ACM, p. 12.

[28] Wolfe, N., Hong, J., Raza, A. A., Raj, B., and Rosenfeld, R. Rapid development of public health education systems in low-literacy multilingual environments: Combating ebola through voice messaging. In *ISCA Special Interest Group on Speech and Language Technology in Education (SLaTE)* (2015), INTERSPEECH.