Job Opportunities through Entertainment: Virally Spread Speech-Based Services for Low-Literate Users

Agha Ali Raza[1], Farhan Ul Haq[2], Zain Tariq[2], Mansoor Pervaiz[3], Samia Razaq[2], Umar Saif[2], Roni Rosenfeld[1]

[1]Language Technologies Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA
{araza, roni}@cs.cmu.edu

[2]Department of Computer Science
School of Science & Engineering
Lahore Uni. of Mgmt. Sciences
Lahore, Pakistan
{farhan.haq, zain.tariq, umar}@lums.edu.pk

[3]Bouve Col. of Health Sciences &
Col. of Computer & Info. Science
Northeastern University
Boston, Massachusetts, USA
mansoor@ccs.neu.edu

ABSTRACT

We explore how telephone-based services might be mass adopted by low-literate users in the developing world. We focus on speech and push-button dialog systems requiring neither literacy nor training. Building on the success of *Polly*, a simple telephone-based voice manipulation and forwarding system that was first tested in 2011, we report on its first large-scale sustained deployment. In 24/7 operation in Pakistan since May 9, 2012, as of mid-September Polly has spread to 85,000 users, engaging them in 495,000 interactions, and is continuing to spread to 1,000 new people daily. It has also attracted 27,000 people to a job search service, who in turn listened 279,000 times to job ads and forwarded them 22,000 times to their friends. We report users' activity over time and across demographics, analyze user behavior within several randomized controlled trials, and describe lessons learned regarding spread, scalability and sustainability of telephone-based speech-based services.

Author Keywords

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

ACM Classification Keywords

H.1.2 [**User/Machine Systems**]: *Human factors and Human information processing*; H.5.2 [**User Interfaces**]: *Natural language*.

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 ([1], [2]), Avaaj Otalo [3]) and as a result are restricted to a moderate number of users. Recently, Smyth et al. [4] described the remarkable ingenuity exhibited by such users when they are motivated by the desire to be entertained. Inspired by this powerful demonstration, we set out to systematically develop practices for entertainment-driven mass familiarization and training of low-literate users in the use of telephone-based services.

Our ultimate goal is to disseminate speech-based, development-related information and communication services to low-literate telephone users throughout the developing world. Such services may include: facilitating an efficient marketplace (speech-based Craig's List); facilitating social and political activism (speech-based message boards and blogs); sending/receiving group messages (speech-based mailing lists); citizen journalism. All of these services are already available, in textual form, to affluent people via the web, and some of them are also available to non-affluent but literate people via SMS. Very few such services are currently available to the low-literate.

We aim to introduce and popularize speech interfaces among low-literate users to serve as a delivery vehicle for core development services. We envision speech-based viral entertainment as an ongoing component of a telephone-based offering, drawing people into the service, where they can periodically be introduced to the more core-development oriented services listed above.

In [5] we described a simple telephone-based, voice-based entertainment service, called Polly, that allowed any caller to record a short message, choose from several entertaining voice manipulations, and forward the manipulated recording to their friends. Introduced in 2011 among low-skilled office workers in Lahore, Pakistan, in 3 weeks Polly spread to 2,000 users and logged 10,000 calls before we shut it down due to insufficient telephone capacity and unsustainable cellular airtime cost. In analyzing the traffic, we found that Polly was used not only for entertainment but also as voicemail and for group messaging, and that Polly's viral spread crossed gender and age boundaries but not socio-economic ones.

CHI 2013, April 27–May 2, 2013, Paris, France.

Research Questions and Experimental Goals

We posed the following research questions: (1) Is it possible to virally spread awareness of, and train people in, speech-based services, in a largely low-literate population, using entertainment as a motivation? (2) Is it possible to leverage the power of entertainment to reach a large number of people with other speech-based services?

These led to the following experimental goals:

1. Determine whether a system like Polly can be scaled up to engage and support a much larger user base, for an extended period of time, while at the same time becoming cost efficient.

2. Use Polly as an experimental test bed to answer the following questions:

a. How do usage patterns change with respect to gender, age, socio-economic status (SES), experience with the system, and cost to user?

b. Spread vs. Cost: how to spread Polly as widely as possible at the smallest possible cost?

3. In line with our long-term goals described above, attempt for the first time to add a development-focused service ('payload') to Polly's offering.

In what follows, the next section summarizes related work on the use of spoken dialog systems for development and on viral services in the developing world. We then describe the new setup of Polly: user interface, functionality, software, hardware and telephony setup. The following section provides detailed analysis of the first 4 months of this still ongoing experiment, including usage patterns over time and across demographics, user behavior within several randomized controlled trials, and the successful introduction of a job search service. We conclude with a summary of our findings and discussion of future plans.

RELATED WORK

Several attempts to design user interfaces for low-literate users have been reported in the literature. Plauché et al [6] deployed information kiosks in community centers across six rural sites in Tamil Nadu, India to disseminate agricultural information to farmers. The kiosks allowed multimodal input (speech and touch screen) and output (speech and display). The reported study involved around 50 participants. Various forms of user training were employed, including short training sessions and group sessions. Low-literate users exhibited a mixed preference towards speech vs. touch screen input. The speech data gathered during spoken interactions was used to semi-automatically train acoustic models for each village for the ASR used in these kiosks [7]. In *Warana Unwired* [8], PC based kiosks used for distributing agricultural information to sugarcane farmers were replaced by mobile phones. The information was transferred to the farmers using SMS. Medhi et al [9] compare textual and non-textual interfaces for applications like digital maps and job search systems for low-literate users. The study was conducted in three slums of Bangalore, and highlighted the importance of consistent help options in the interface. It also confirmed that abstracted non-textual and voice based systems are preferred by low-literate users over textual one.

Most of the work done to date in providing speech-based communication services to low-literate users relied on explicit user training. In *Project HealthLine* ([1], [2]) the target audience was low-literate community health workers in rural Sindh province, Pakistan. The goal was to provide telephone-based access to reliable, spoken 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.

Avaaj Otalo [3] is another successful example of a speech interface serving low literate users, in this case farmers. The 51 users of the system were shown how to use Avaaj Otalo before its launch. This telephone based system was pilot-launched in Gujarat, India and offered three services: 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.

Patel et al. [10] recently identified three major factors enabling peer-to-peer services in the context of developing countries: access cost, subject matter or type of exchange and the influence of the administering institution. While subject matter builds the main perception about the service among users, moderation and encouragement can play a vital role in improving and refining the details of peer-to-peer interactions. In a follow up study, Patel et al. [11] compared the influence of peer-generated vs. institutional authority-generated content on farmers. 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. Based on the number of follow up calls to listen to the remainder of the tip, it was concluded. 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.

Voice-based media has also been shown to promote social inclusion among underserved communities. Mudliar et al. [12] examined participation via citizen journalism by rural communities in India using *CGNet Swara*, an interactive voice forum. It enabled users to record and listen to messages of local interest and became popular among the target audience. Koradia et al. [13] involved listeners in voice content creation, feedback and station management via community radio, and showed that it can also be used to

provide ICT solutions in the developing world. They highlighted the need to prioritize hardware stability over cost, diagnostic tools and remote technical assistance to solve most of the equipment related problems. Wyche et al, in their empirical study of professionals living in Nairobi, Kenya [14], highlight factors to guide ICT work in infrastructure poor settings with an emphasis on collective consideration: limited bandwidth; high access cost; varying perceptions of responsiveness and threats to physical and virtual security.

When dealing with a large user base, explicit training is not feasible. One alternative is to rely on learning from peers and on viral spread. Baker [15] lists conditions for viral spread (albeit in the context of literate users and web-based services). A successful example of cellphone based (though not speech-based) viral spread is *SMS-all* [16], a group text-messaging service in Pakistan. Users can also create new access-controlled groups and join already existing ones. As of last report [16] the service has over 2 million users and four hundred thousand groups, and more than 3 billion messages have been sent out. People use this service to share information and discuss hobbies and other interests. However, the use of text assumes a level of literacy which is not common in our target population.

An important question in developing speech based telephone interfaces is the preferred input mode: speech vs. DTMF (push button). Project HealthLine ([17], [1] and [2]) found that speech input performed better than DTMF in terms of task completion, for both literate and low literate users. However, it provided no clear answer in terms of subjective user preference. In fact, [1] found that low literate users preferred DTMF input over speech input, although they performed better on average with the latter. User studies conducted in Botswana by Sharma et al [18] with HIV health information systems for the semi and low literate populations also suggest user preference towards touchtone over speech while both systems perform comparably. In contrast, [3] and [19] (which were conducted in a controlled environment) both report that DTMF and numerical input perform better than speech in terms of task completion and performance improvement. Patel et al [19] also report the problem of transitioning between DTMF and speaking as a major challenge. But overall, the study suggests that numerical input is more intuitive and reliable than speech. It seems from both of these reports that DTMF may be a better choice if user perception is vital for system adoption, especially in a situation where training and tutorials cannot be relied on.

Speech based input presents another major hurdle when dealing with the languages of the developing world: lack of local linguistic resources and expertise for training a speech recognizer. This is especially true in regions of great linguistic diversity as is the case in Pakistan, where even neighboring villages may speak different languages or dialects. However, for applications or services requiring only a small input vocabulary, the *Salaam* method [20] can be used, as it provides high recognition accuracy in any language for up to several dozen words.

Figure 1: Sample low skilled job ad from paperpk.com scanned from Daily Jang

SYSTEM DESCRIPTION

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. In Urdu, Polly is called "Miyan Mithu", which has a meaning similar to "Polly the Parrot". The theme of light entertainment using funny modifications of a voice recording is non-controversial and easy to understand as was discussed in [5].

The system we used in the current study represents a substantial extension of the one we deployed in 2011. The most important changes are:

- Telephony capacity was increased from 1 to 30 channels (up to 30 concurrent phone calls, incoming and outgoing).
- Three different telephone numbers were assigned to a 'hunt group' consisting of these 30 channels, to support flexible, dynamic allocation for multiple application types.
- All the software resides on a single server hosted on location in Pakistan by a local telecom (in 2011, software was distributed and split across two continents). Consequently, outgoing call airtime costs were reduced from $0.126/minute to $0.023/minute.
- Most menus, prompts and recordings can now be skipped by the user by pressing any button on their phone.
- The number of voice manipulations offered was increased to six by adding male-to-female and female-to-male options.
- Extensive logging and real-time monitoring were added.

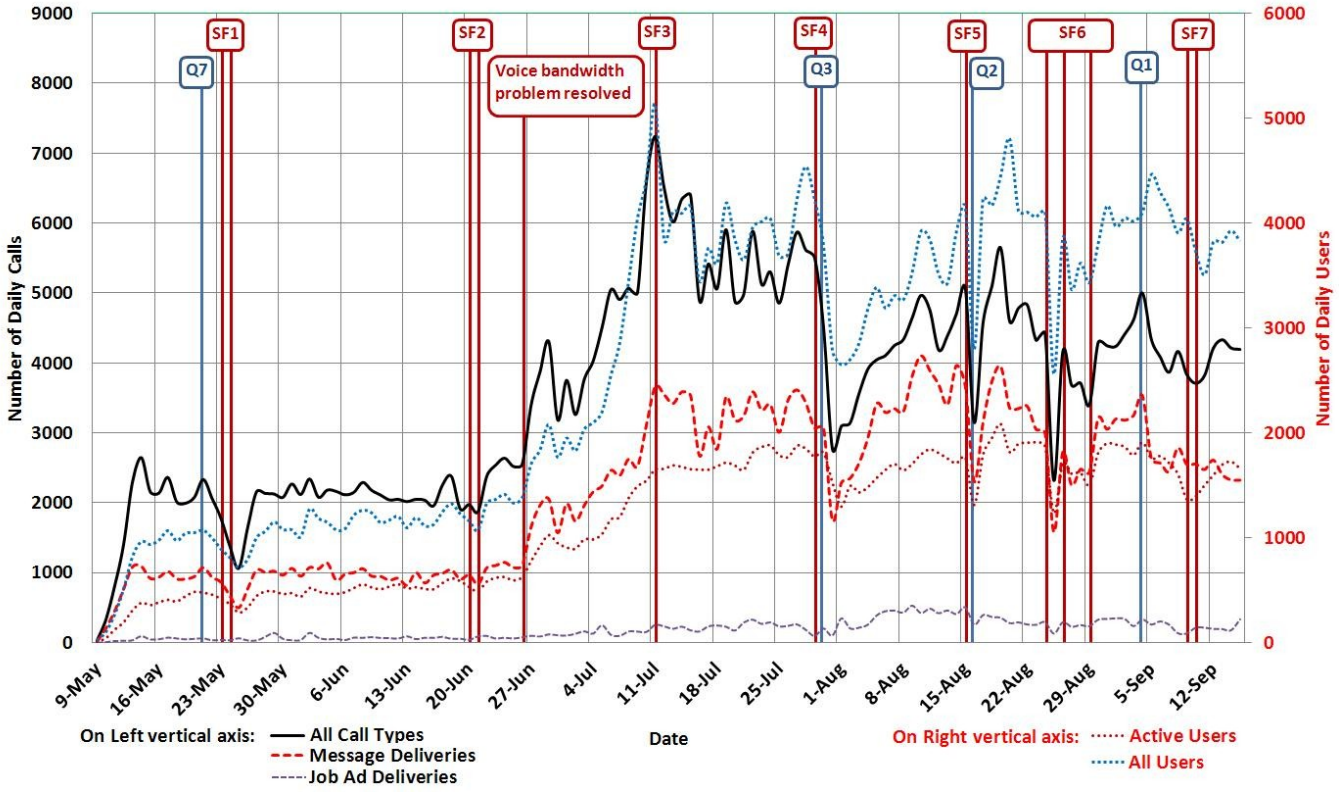

Figure 2: Polly's Activity Levels. "SF": system failures (Table S1). "Q": onset of quota experiments. For term definitions, see "System Activity Level".

• The most important addition in the current system is the introduction of a development-related application (what we call 'the payload') as part of the dialog menu: an audio-browsing of advertisements, collated from Pakistani newspapers, for jobs that are appropriate for low-skilled, low-literate workers (see "Polly's first payload: speech-based job ad browsing" below).

User Interface

Polly's current user interface is an extension of the one described in [5]. For convenience and completion, we provide a full description here. For video and audio demonstrations, see [21].

May 9 -Jun 27	Telcom bug reduced Polly's effective capacity to 10 channels and degraded voice quality during peak hours.
May 25-26	SF1: Disk space crunch reduced system responsiveness / availability
June 21	SF2: Application server remained down as new phone lines were being installed
Jul 12	SF3: A bug reduced the effective channel capacity to 0
Jul 31	SF4: Upgraded software platform severely reduced channel capacity
Aug 16	SF5: OS error crashed the server
Aug 25	SF6: Server ran out of disk space
Aug 29	SF6: Licensing problems reduced channel capacity to 2
Aug 28-30	SF6: SMS provider's glitch stopped all outgoing sms
Sep 11-12	SF7: SMS provider's glitch stopped all outgoing sms

Table S1: Major System Failures

Informed Consent: Before the start of interaction every caller is informed that the call is being recorded and may be analyzed, and is given the opportunity to hang up. All users must listen to this prompt as it cannot be skipped by any key press.

At the start of the call, the user is prompted to make a short recording of their voice (15 seconds, or shorter if the user presses # or remains silent for 4 seconds). A funny voice transformation of the recording is immediately played back to them. The user is then given an option to hear the recording again, rerecord, try another voice manipulation effect, forward their modified voice to friends, give feedback to Polly or listen to the latest job ads. We offer the following voice modifications effects, in the following order, all achieved with a standard audio processing utility:

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

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

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

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

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

6. Converting the voice to a *whisper*, achieved by replacing the excitation source of user's voice with white noise

7. Adding *background music*.

If the user chooses to forward their recording to a friend, they are prompted for the phone number, the name of their friend, and their own name for introduction. Only the phone number is confirmed for correction. Recordings of names are terminated by silence detection and a 4 second hard time-out. The user is allowed to forward their voice to multiple recipients with the same or different modifications applied. The message forwarding request is then added to the system queue, and will be executed as soon as a channel becomes available. When Polly calls the intended recipient to deliver the recorded message, the sender's name (in their own unmodified voice) is immediately played to the listener to prevent confusion regarding the identity of the caller, and the recipient can also choose to hear the phone number of the sender. After hearing the message, the recipient can then choose to replay the message, record a reply, forward the recording to others, create their own recordings, or listen to job ads.

As an additional mechanism for viral spread, text (sms) messages containing Polly's contact information are sent to all of Polly's recipients on their first two interactions with the system. Polly's phone number is also played during the phone call itself.

We also elicit **User Feedback,** in the form of an unconstrained recording (up to 60 seconds, with a silence timeout) from repeat users during their interactions with Polly. Feedback is requested only when a user actively initiates a call. Feedback is requested in two manners:

System Prompted Feedback – Every user is prompted for feedback on their fifth interaction with Polly, and on every 20th interaction thereafter.

User Initiated Feedback – Following the user's fifth interaction, the menu is augmented with an explicit option to give feedback.

Polly's first payload: speech-based job ad browsing

We daily scan Paperpk.com for advertisements that appeared in Pakistani newspapers for jobs that are appropriate for low-skilled, low-literate workers. An example of such an ad is in Figure 1. Although we focus on jobs not requiring any literacy, we also select some that require up to 10 years of education. We then record these ads in Urdu, and make them available for audio-browsing as part of Polly's menu. Specifically, we record each ad in three separate parts: date and newspaper source; details and requirements; and contact information. These are then made available to the user via an interface option in the main menu of Polly. Ads are played latest first. Any ad can be skipped, repeated (as a whole or just the contact information), browsed and forwarded to friends.

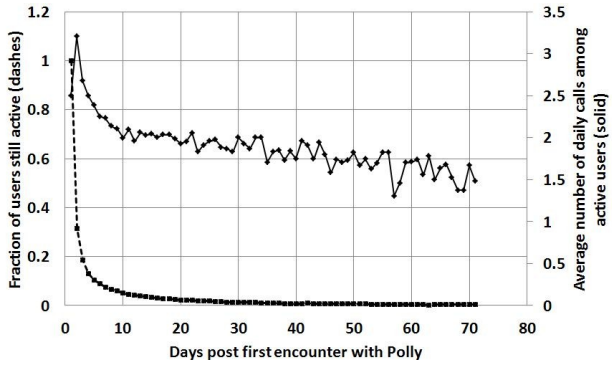

Figure 3: User Retention

A user who receives a call from Polly with a forwarded ad is first informed of the sender, then given Polly's phone number (also by SMS) for future use. After hearing the forwarded ad they can also choose to browse the job ad list.

Software Setup

Polly's IVR code consists of PHP scripts that run on Voxeo's PRISM platform [22] using Voxeo's Tropo [23] as the interpreting engine. Speech prompts and audio files are hosted on an Apache server. The application server also maintains detailed textual and audio logs of all phone calls. The database is managed using MYSQL. Praat scripts are used for manipulating audio files.

Telephony setup

All the software is installed on a single 2U server that is hosted in the data center of Wateen [24], a local internet and telephone service provider. The first of our three phone numbers is used for toll-free calls to Polly by means of a "missed calls" mechanism: a call to that number is interpreted as a "call-me-back" request. It is rejected and a call-back request is added to the queue. As soon as a channel becomes available, Polly calls the user back. The second phone number is used as a caller-paid service which is answered immediately (see below). The third phone number is currently reserved for deployment of future services.

Real-Time Monitoring

To get the pulse of the system as well as an overall picture of activity, we built a real-time monitoring system that provides cumulative, daily, hourly and per minute statistics [21]. Our system calculates and reports statistics on overall traffic volume, answered calls, deliveries of messages and job ads, ads in the system, ads listened to by users, user feedback and other categories. Hourly and per minute statistics on queue lengths of the different request types, channel utilization, inbound and outbound calls and free hard disk space allow us to detect problems in real-time and to schedule maintenance. The system also sends sms and email alerts and recovery reports about server crashes, automatic server restarts, hard disk space crunch, etc.

Annotations of Recordings and Feedback

One graduate student and one undergraduate student listened to a uniformly selected sample of 5388 recordings and created detailed annotations based on their subjective assessments. Each selected recording was annotated by a single annotator. As in [5], each recording was annotated as to the speaker's gender, language(s) used, estimated age (child, young, middle aged, old), estimated Socio-Economic Status (SES) (low, middle, high), and whether the message appears to be recorded for entertainment or utility (not mutually exclusive). The inter-annotator agreement rates were later found to be adequate for gender (97.8%) and language (81.5%) but poor for SES (40.4%), age (62.6%) and primary use (51.6%). We therefore decided not to use the last three in our analysis. Instead, we conducted follow up surveys as described in the next section to get information on these variables. In addition, starting on August 4, 2012 feedback recordings were listened to daily, briefly paraphrased, and categorized as complaints, requests or suggestions. Finally, the annotators were also encouraged to note in unstructured comments any interesting subjective observations, such as unusual functions for which Polly appears to be used.

Follow up Surveys

We called 207 randomly chosen users of Polly to collect more reliable demographic information. Engaging the users in a friendly chat we asked them about their experience using Polly, literacy level, age, rough location (city, small city, village etc.), primary uses of Polly and their experience using Polly's job search (See "Demographics and Primary Use from Follow Up Survey").

RESULTS & ANALYSIS

Seeding

Polly was seeded on May 09, 2012 at 8:13pm Pakistan Time by placing automated calls to 5 of the most frequent callers from Polly's 2011 study. These calls briefly announced that Polly is back online and launched into Polly's main interaction. No attempt to further promote the system was ever made. Polly has been up continuously since then (as of this writing, January 2013), with minimal interruptions (see Table S1).

System Activity Level

In Figure 2, "All call types" shows the overall volume of successful calls; out of these: "Message Deliveries" and "Job Ad Deliveries" depict the number the successful calls made by Polly to deliver messages and ads forwarded by users to their friends. "All Users" shows the total number of unique users of Polly (including those who just received messages without ever calling Polly). "Active Users" are the users who called Polly or used it to schedule deliveries.

Polly's activity level (Figure 2) rose exponentially and saturated our system's capacity within 7 days. On June 27, a telecom system bug was discovered that had been keeping our effective capacity at 10 channels and degrading voice quality during peak hours. Once the telecom company fixed this bug, activity level again climbed exponentially and stabilized within 10 days. Activity levels appear to be limited only by channel capacity. System failures (see Table S1) usually resulted in a commensurate drop in call volume, but the latter always recovered quickly once the problem was fixed. Another source of volume drop was our quota experiments, to be discussed later.

As of mid-September, 2012 Polly has had more than 495,000 telephone interactions with more than 85,000 users. There have been 103,250 failed calls (user busy, out of signal reach, phone switched off, incorrectly entered numbers etc.). The message forwarding feature (where a user forwards a received message) was used in 31,740 calls with the longest chain consisting of 40 forwards and an average chain length of 2.33. Among the most forwarded messages were funny voices (baby laughing, duck sound etc.), quiz-like questions ("what is x called in English?"), a guy whistling a tune.

Delivery requests were placed in 182,652 calls (1.6 delivery requests per call on average), including 55,543 delivery requests for multiple recipients. On average there were 1.7 delivery requests placed for every personal message and 61 delivery requests for every job ad.

Choice of Voice Modifications

On average, users listened to 1.83 voice modifications per interaction. The modifications chosen for delivery of the 291,504 messages were (in menu order): male-to female conversion 72%; female-to-male 9%; drunk-chipmunk 6%; I-have-to-run-to-the-bathroom 3.5%; unmodified 5.3%; whisper 0.7%; and background music 3.5%.

User Activity Profile

Figure 3 depicts user-initiated interactions as a function of the number of days since they were first introduced to Polly. Most users interact with Polly for only a few days. Only 31% of users returned to Polly on the second day, 19% on the third, 13% on the fourth, and 10% on the fifth. Participation continues to drop logarithmically, e.g. to 1% after 36 days and to 0.5% after 55 days. Among users who do continue to use Polly, average daily activity peaks at 3.2 calls on day 2, then drops gradually to around 1.5 calls.

Average call duration was 160 seconds. It takes no more than 40 seconds to start experiencing the first voice modification. Of the 495,000 Polly interactions: 87% lasted 40 seconds or more.

There were a total of 1,023,824 menu options selections (by key presses) during the 495,000 interactions, out of which 4.5% were invalid choices. Note that Polly's IVR tree was designed to require no key presses until after the first voice modification is played back. No keys were pressed in 36% of the interactions. Of the remaining interactions, 91% completed without any invalid key presses.

	Toll Free	Caller Paid
	Count	Count
Annotated recordings	5388	399
By gender		
Male	4713(87.3%)	359 (90.0%)
Female	590(10.9%)	26 (6.5%)
Unclear	93(1.7%)	14 (3.5%)
By language		
Urdu	1135 (21.1%)	64 (16.0%)
Punjabi	3480 (64.6%)	194 (48.6%)
English	23 (0.4%)	4 (1.0%)
Pushto	703 (13.1%)	100 (25.1%)
other/mixed	45 (0.8%)	37 (9.3%)

Table 1: User Demographics from annotated recordings

User Demographics from Recording Annotation

To understand our users' demographics and their use of Polly's message delivery capability, we selected a sample of user recordings uniformly across time and annotated them as described in "Annotations of Recordings and Feedback". Results are in Table 1.

Demographics and Primary Use from Follow Up Survey

Table 2 summarizes our survey's results. Out of the 207 survey calls a 106 resulted in useful information of some type. Low to low-mid SES people having up to 10 years of education comprised 77% of the interviewed users: mostly shopkeepers, fruit sellers, farmers, laborers, carpenters, and craftsmen. Another 14% had 11-12 years of education and belonged to medium SES while the remaining 8% had more than 16 years of education. Majority of the users belonged to villages or small cities. For heuristic mapping of education level to SES, see [25].

Although around 57% of the respondents had browsed Polly's Job ads, only a handful reported applying for those jobs. This was mostly because the ads were either not of interest to them or they did not trust them. Two users claimed that their friends got jobs through Polly, but we were unable to verify this.

The respondents who described their primary use of Polly as "fun" gave examples like making prank calls to friends, hello-hi/random messages, poetry and even browsing job ads as a pastime. More serious users defined Polly as a voice messaging system that they use to send occasion (holiday, birthday) greetings, to request a call-back, to know a friend's whereabouts or to browse and apply for jobs. Four blind users defined Polly as an "alternative to text messaging" and praised it profusely. Females were mostly reluctant/shy to talk.

Cost vs. Spread: RCTs and the Quota Experiments

With an outgoing call airtime cost of $0.023/min, and with each interaction lasting an average of 3 minutes, at its peak Polly was costing us some $400/day in airtime alone. A variety of mechanisms can eventually be used to offset this

Total number of survey calls attempted				207
Number of calls resulting in any information				106
Number of calls that successfully gathered:				
a. Gender Information				106
b. Information about primary use				63
c. Information about use of Polly's Job Browser				65
d. Age Information				60
e. Literacy/SES information				70
Primary Use				
		Fun	**Utility**	**both**
Among the 63 users		34	17	12
Gender				
Male	98	33	17	12
Female	8	1	Did not answer	
Age				
Less than 25 years	30	20	6	3
25-35 years	22	9	6	5
More than 35 years	8	2	2	1
Highest Education Level Attained, SES				
None, low SES	17	30	14	10
Primary (5 years), low SES	15			
Matric. (10 years), low-mid SES	23			
Intermediate (12 years) , mid SES	9			
University (16 years), high SES	6	3	3	0

Table 2: Results of Survey Calls

cost including banner ads, carrier revenue sharing agreements, and content-providing sponsors. However, in the current experiment, one of our goals was to test how far we can reduce our airtime costs directly while maintaining the system's viral spread. We view our ongoing airtime expense as simply the cost of reaching new users, and are interested in strategies that maximize the cost-effectiveness of that investment. We are also interested in understanding how the airtime cost structure affects our target users' behavior.

We did not want to eliminate the toll-free option because that would have biased the user base away from low socioeconomic users, who are our prime target. We also avoid a reliance on SMS messages, so as not to deter low-literate users. Instead, we experimented with imposing daily quotas on the number of toll-free calls for each user (based on their caller id).

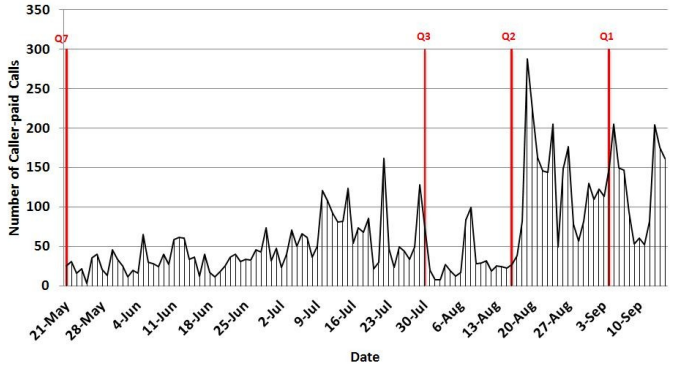

Figure 4: Daily caller-paid calls

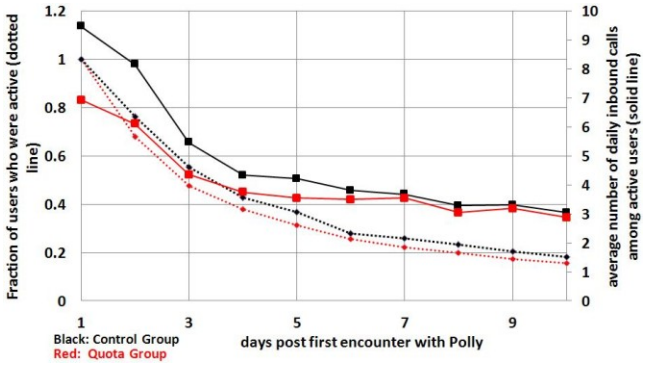

Figure 5: Experiment with a quota of 7 toll-free calls/day

One of the benefits of a large-volume, dynamically controlled system such as ours is the ability to run randomized controlled trials (RCTs). Our first goal was to avoid subsidizing high-volume users, and to nudge them towards a caller-paid model. The main research questions asked here were:

1. How much can we reduce our airtime charges (our main operating expense) while maintaining system activity and spread, and how? Sub-question: What partial subsidy scheme may induce people to contribute their own funds to the airtime cost?

2. Is Polly compelling enough for people to spend their own money on it, at least sometimes?

In our first RCT ("Q7") we targeted users who called Polly more than 7 times a day. Once a user attempted to call Polly for the 8th time on the same day, they were alternately assigned to the quota-restricted group or to a control group (and retained that assignment indefinitely). A user in the quota-restricted group, on their 8th daily call to Polly, was told that they exhausted their subsidy for that day, and invited to call Polly on the caller-paid line, where their call would be picked up immediately and their scheduled deliveries would also receive absolute priority. This was a substantial "perk" because users often complained of delays in receiving call-backs and in delivery of messages, due to long queues, especially at the peak evening hours. Subsequent calls by this user on that day to the toll-free number were not answered. Users in the control group were

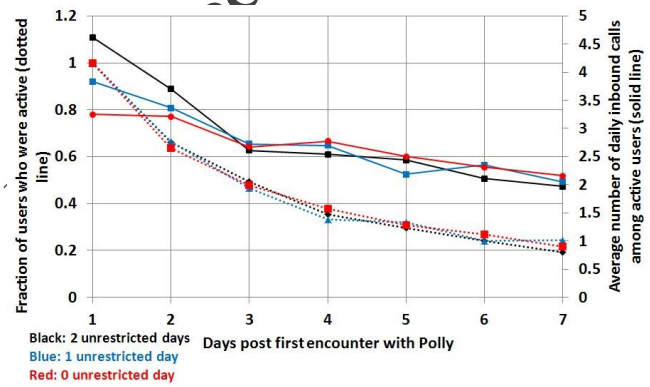

Figure 6: Experimenting with a quota of 3 toll-free calls/day

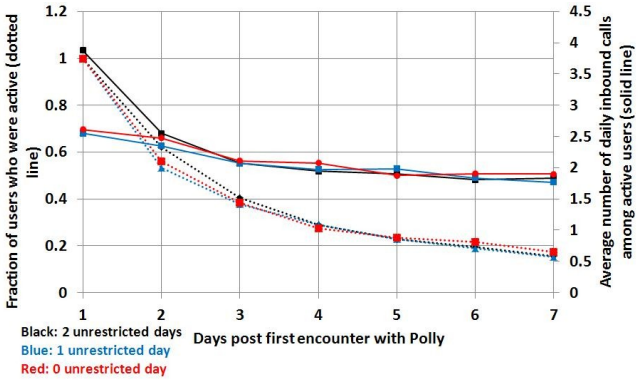

Figure 7: Experimenting with quota of 2 toll-free calls/day

not restricted. We ran this experiment from May 21 through July 30. Results (shown only for the 1,115 high-volume users who were new to Polly during this period) are in Figure 5. Activity on the caller-paid (namely, unsubsidized airtime) line is shown in Figure 4. We can see that the quota indeed reduced toll-free usage by the restricted group, and caused sporadic activity on the caller-paid line. However, the differences from the control group vanished after a week: it appears that most high-volume users reduced their activity substantially within a few days even without the quota.

Our next experiment was to more severely restrict toll-free usage, to a maximum of 3 calls/day, for everyone. We hypothesized that on the first few days of using Polly, a user is likely to send their messages mostly to new people, but that on subsequent days they are more likely to request deliveries to the same recipients. We therefore randomized users into three arms: those on whom the 3/day quota was imposed immediately ("Q3D0"), those on whom it was imposed starting on their second day ("Q3D1", having one day of unrestricted toll-free use of Polly), and those on whom the quota was imposed starting on their third day ("Q3D2", 2 unrestricted days). This experiment was run from July 31 through August 16. Results (shown only for the 486 users who were new to Polly during this period and who attempted to make a 4th daily call) are in Figure 6. The quota reduced activity during the first few days, as expected. Also of note, there was no significant difference

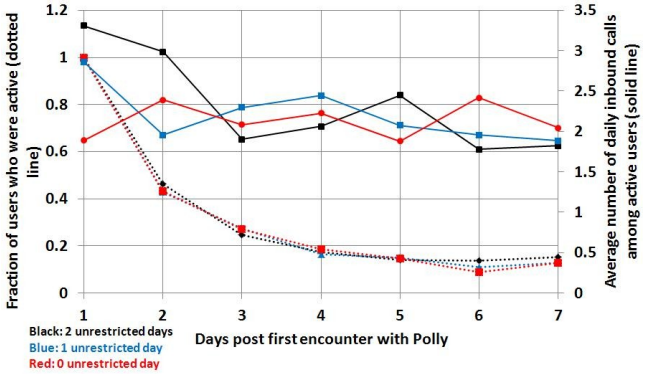

Figure S2: Experimenting with quota of 1 toll-free calls/day

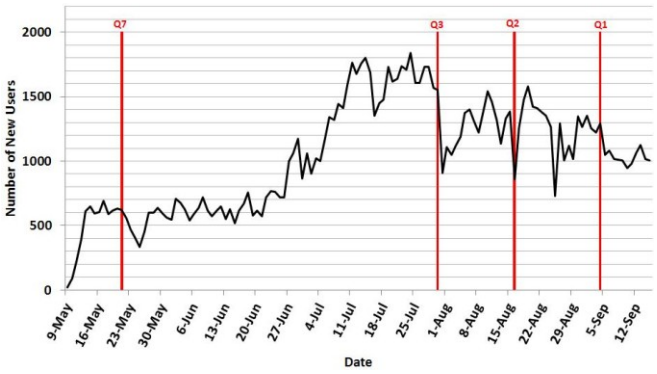

Figure 8: New users added daily

among the behaviors of users from various quota arms once the grace period (in which the users are allowed unlimited daily toll-free calls) expired. (Note: "the number of calls per day" appear to exceed the quota because it also includes the one call during which the quota announcement was played).

Regretfully, the introduction of Q3 required a software platform upgrade, resulting in frequent system crashes (and a sharp drop in all activity) from which we recovered gradually over the following 10 days ("SF4" in Table S1). We therefore cannot measure the short-term impact of the Q3 quota on overall activity level, but we observe that after 10 days activity fully recovered, thanks to a constant supply of new users, and supporting our hypothesis that activity is only limited by our channel capacity. In hindsight, the quick rebound is not surprising, because the few unrestricted days were enough to recruit new users. Note from Figure 3 that 40% of calls to Polly took place during a user's first day of interaction, and fully 56% during their first two days.

Our next experiment, run from August 17 to September 4, was similar but tightened the quota to 2/day, with the same 3-arm randomized assignment (Q2D0, Q2D1, Q2D2). Results (shown only for the 1,029 who were new to Polly during this period and who attempted to make a 3rd daily call) are in Figure 7. Again, users seem to have quickly adjusted to the new quota regardless of whether it was introduced immediately or with delay. Caller-paid calls spiked (Figure 4). Thankfully, the system failure ("SF5" in Table S1) that coincided with this introduction lasted only a day, and overall activity remained high (Figure 1).

Finally, from September 5 on we have been operating under a 3-arm Q1 experiment (Table S2). This time, overall activity level did go down somewhat, although the number of users and number of new users (Figure 8) did not, achieving the same spread with lower costs.

First Payload: Job Ads Service

Perhaps the most meaningful development in our current setup is the addition of a development-related service to our menu – the job ad browser. Since audio-browsing job ads was added as an option at the end of the Polly menu, and since this service was not advertised in any other way, the extent of its use is a direct test of our strategy to reach users via entertainment.

Period of feedback recordings listened to:	Aug 4–Sep 9
Total number of feedback recordings in that period:	7770
Total number of feedback recordings listened to:	2567
Of those, containing thanks or praise to Polly	1797 (70%)
containing other feedback	272 (10%)
Of those containing other feedback:	
System-initiated	183 (67%)
User-initiated	89 (33%)
Complaints about:	189 (70%)
Delay in call-back time	140
Poor sound quality	28
Failure of system to send/receive message	14
Being disturbed by Polly	5
Requests/suggestions:	83 (30%)
Higher quota or more recording time	47
New services: News, Weather, Medical, Voice Chat, Job Ads, Songs.	22
More voice manipulations	8
No voice manipulation	6

Table 3: User Feedback Summary

During the 130 days of Polly operations reported here (May 9 – September 15), we identified and recorded a total of 530 suitable job ads, an average of some 28 ads/week. These ads were listened to, all by user initiative, a total of some 279,000 times, This averages to over 525 playbacks per ad – possibly more than the number of people who read that ad in the newspaper. Some ads were listened to much more than others – the most popular ad was listened to more than 8,400 times, and 73 ads were listened to more than 1,000 times each.

A further indication of the usefulness of this service was the use of job ad deliveries – requesting that a particular job ad be delivered by Polly to a friend. During the period of this reporting, a total of 23,288 such requests were made, and job ads were delivered to 9,475 different users. Even more interestingly, out of all the calls during which job delivery was requested, more than half requested *only* job deliveries (i.e., no regular Polly message deliveries), most likely indicating that the user called Polly specifically to interact with the job ad service.

User Feedback

Table 3 lists the main findings from our feedback annotation process. Several users gave suggestions to improve the user interface. Among these were frequent requests to increase message recording time, remove the voice modifications OR to bring the unchanged voice to the beginning, to display sender's name/number on screen, to keep messages for later listening and to be able to post ads on the job ad system. One guy suggested that Polly should send a text message to the recipient who should call Polly to listen to the message at his convenience (we added this feature).

Anecdotally, among the positive feedbacks, one person said (loosely translated) "after all that is going wrong with the country … well, at least we have Polly … God bless Polly and may the service continue forever".

SUMMARY AND DISCUSSION

Our first goal was to determine whether a system like Polly can be scaled up to engage and support a much larger user base, for an extended period of time, while at the same time becoming cost efficient. With regard to scale and persistence, we believe that the numbers speak for themselves. The long queues for call-me-back and delivery requests, and the quick rebounding of traffic to a fixed level after each disruption, suggest that activity level is resource-bound and that the potential demand for these services is much higher than our current 30-channel capacity.

On the question of cost efficiency, we believe that the jury is still out. We have treaded lightly on limiting the toll-free service because we did not want to scare off poor users, and because we wanted to use the large volume to answer many other questions, some of which we are just now beginning to analyze. However, we have since clamped down more strongly, and have presented the impact in [26]. As we mention above, cost efficiency can be achieved not only by getting the users to pay for airtime, but also by the use of ads, carrier revenue-sharing, and/or content sponsors (e.g. governments or NGOs). We are planning to explore all these options.

Our second goal was to use Polly as an experimental test bed to answer questions about demographics and about spread vs. cost. Regarding demographics, we find that Polly is used predominantly but not entirely by men, who are predominantly young or middle-aged. This was also observed in Polly's 2011 test deployment, and is what led us to select job advertisements as our first development-focused service. Unlike in 2011, we find that Polly has spread significantly into the mid-SES and even high-SES populations.

Usage over time is marked by rapidly declining interest among most users. This was expected given the unchanging nature of the entertainment, although interestingly a still significant number of people continue to use Polly for many weeks and months. Usage grew exponentially because every user spread the system to more than one new user on average. Since the target population is measured in the tens or hundreds of millions, volume will grow exponentially for quite a long time, limited only by the system's carrying capacity. Nonetheless, without significant long-term use eventually activity will indeed decline. In the short term, we are working to increase repeat usage by varying the entertainment content. In the long term, we believe the utility components will draw the users back. We see the entertainment component mostly as a method to spread awareness of the system and train the users in speech-based services, not as a steady-state standalone service in its own right. Additionally, when a new service is added, Polly can call some of its past users and introduce them to it, re-starting a viral spread.

The large volume of users allowed us to use randomized controlled trials to answer some questions regarding users' cost-sensitivity. A high daily quota on user-paid calls did not reduce expenses much. When faced with a lower quota, most users chose not to use their own money to make calls that they would otherwise have made. This can be demonstrated by comparing the average call volume (user-paid and toll-free) of people in the quota-restricted arm to that of people in the control arm. Nonetheless, the total volume of user-paid calls is evidence that at least some people were willing to pay some of the time. During the last week of our reporting period, toll-free calls averaged 2,200/day, whereas caller-paid calls averaged a mere 160/day. It is possible that, once a service is introduced as toll-free, people would always be reluctant to pay for it. We are planning to test this hypothesis by deploying a pure caller-paid system in a new geographic location. We also found that restricting high-volume users does not stymie spread, as measured by the number of new users added (Figure 8).

Anecdotally, quite a few of the user-paid recordings contain strongly worded complaints and even curses regarding the quota. Apparently, people are willing to pay for the service in order to vent their anger at the need to pay for the service.

Our third goal was to add our first development-focused service ('payload') to Polly's offering. We found that users took to the new offering in large numbers, and that many of them started calling Polly specifically for the job information – exactly the result we had hoped for. Our survey calls revealed that around 57% of the interviewed users had used job search but only a handful of them applied. This can be attributed to lack of trust or interest. The former may be by co-branding with familiar government organizations or newspapers. More job ad types can be explored (e.g. jobs for the handicapped) to serve all interest groups.

Limitations of the study: Randomized controlled trials are potentially confounded by offline communication among friends.

Additional benefits: We collected detailed interaction data, which we believe have great potential value as a test bed for analyzing social network dynamics.

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 Voxeo Inc. and Johnny Diggz of *Geeks Without Borders* for generously donating PRISM/Tropo licenses to our project.

REFERENCES

1. J. Sherwani, N. Ali, S. Mirza, A. Fatma, Y. Memon, M. Karim, R. Tongia, and R. Rosenfeld, "Healthline: Speech-based access to health information by low-literate users," in *Information and Communication Technologies and Development, 2007. ICTD 2007. International Conference on*, pp. 1–9, IEEE, 2007.
2. J. Sherwani, R. Tongia, R. Rosenfeld, N. Ali, Y. Memon, M. Karim, and G. Pappas, "Health-line: Towards speech-based access to health information by semi-literate users," *Proc. Speech in Mobile and Pervasive Environments, Singapore*, 2007.
3. N. Patel, D. Chittamuru, A. Jain, P. Dave, and T. Parikh, "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*, pp. 733–742, 2010.
4. T. Smyth, S. Kumar, I. Medhi, and K. Toyama, "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*, pp. 753–762, 2010.
5. A. Raza, C. Milo, G. Alster, J. Sherwani, M. Pervaiz, S. Razaq, U. Saif, and R. Rosenfeld, "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)*, vol. 2, 2012.
6. M. Plauché and U. Nallasamy, "Speech interfaces for equitable access to information technology," *Information Technologies and International Development*, vol. 4, no. 1, pp. 69–86, 2007.
7. M. Plauché, U. Nallasamy, J. Pal, C. Wooters, and D. Ramachandran, "Speech recognition for illiterate access to information and technology," in *Proc. of 2006 International Conference on Information and Communication Technologies and Development*.
8. R. Veeraraghavan, N. Yasodhar, and K. Toyama, "Warana unwired: Replacing pcs with mobile phones in a rural sugarcane cooperative," *Proceedings of ICTD*, 2007.
9. I. Medhi, A. Sagar, and K. Toyama, "Text-free user interfaces for illiterate and semiliterate users," *Information Technologies and International Development*, vol. 4, no. 1, pp. 37–50, 2007.
10. N. Patel, "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*, CSCW, 2011.
11. N. Patel, K. Savani, P. Dave, K. Shah, S. Klemmer, and T. Parikh, "Power to the peers: Authority of source effects for a voice-based agricultural information service in rural india," in *International Conference on Information and Communication Technologies and Development (ICTD)*, vol. 2, 2012.
12. P. Mudliar, J. Donner, and W. Thies, "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)*, vol. 2, 2012.
13. Z. Koradia, C. Balachandran, K. Dadheech, M. Shivam, and A. Seth, "Experiences of deploying and commercializing a community radio automation system in india," in *Proceedings of the 2nd ACM Symposium on Computing for Development*, p. 8, 2012.
14. S. Wyche, T. Smyth, M. Chetty, P. Aoki, and R. Grinter, "Deliberate interactions: characterizing technology use in nairobi, kenya," in *Proceedings of the 28th international conference on Human factors in computing systems*, pp. 2593–2602, 2010.
15. "Interview with edward baker about the viral factor | entrepreneurial minded." http://-entrepreneurialminded.com/business-interviews/ed-baker-viral-factor/.
16. "Sms-all cheapest group sms service." http://-www.smsall.pk/.
17. J. Sherwani, "Speech interfaces for information access by low-literate users in the developing world.," *PhD Thesis*, May 2009.
18. A. Sharma Grover, M. Plauché, E. Barnard, and C. Kuun, "Hiv health information access using spoken dialogue systems: touchtone vs speech," 2009.
19. N. Patel, S. Agarwal, N. Rajput, A. Nanavati, P. Dave, and T. Parikh, "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*, pp. 51–54, 2009.
20. F. Qiao, J. Sherwani, and R. Rosenfeld, "Small-vocabulary speech recognition for resource-scarce languages," in *Proceedings of the First ACM Symposium on Computing for Development*, p. 3, 2010.
21. "Polly." http://www.cs.cmu.edu/~Polly/.
22. "Voxeo prism." http://www.voxeo.com/prism/.

23. “Ivr platforms / ivr hosting / ivr development.” http://-www.voxeo.com/tropo/.

24. “Wateen - best broadband in pakistan.” http://-www.wateen.com/.

25. Z. Javed, B. Khilji, and M. Mujahid, “Impact of education on socio-economic status of villagers life: A case study of shrien wala village of faisalabad district,” in *Pakistan Economic and Social Review*, vol. 46, 2008.

26. A. Raza, F. Haq, Z. Tariq, U. Saif, and R. Rosenfeld, “Spread and sustainability: The geography and economics of speech-based services,” in *DEV*, 2013.