Viral Entertainment as a Vehicle for Disseminating Speech-Based Services to Low-Literate Users

Agha Ali Raza*, Christina Milo**, Guy Alster**,
Jahanzeb Sherwani*, Roni Rosenfeld*
Language Technologies Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA
*{araza, jsherwan, roni}@cs.cmu.edu
**{cmilo, galster}@andrew.cmu.edu

Mansoor Pervaiz, Samia Razaq,
Umar Saif
Department of Computer Science
School of Science and Engineering
Lahore University of Management Sciences
Lahore, Pakistan
{mpervaiz, 10030023, umar}
@lums.edu.pk

ABSTRACT
Entertainment has recently been shown to be a powerful motivator for mastering new technologies. We therefore set out to use viral entertainment to introduce telephone-based, speech-based services to low-literate people in developing countries. We describe Polly, a simple voice manipulation and forwarding system that went viral in Pakistan last year. Seeded once by 32 low-skilled office workers in a Pakistani university, in 3 weeks Polly amassed 2,032 users and 10,629 interactions. From analyzing the traffic and its content, it is evident that Polly has been used extensively for entertainment and social contact, but it has also been put to an unintended use as a voicemail and group messaging facility. This demonstrated the potential for speech based services, and the pent-up demand for entertainment, among our target population. Also of note, Polly's viral spread crossed gender and age boundaries and even established itself in a female population. However, it appears to have not crossed socioeconomic boundaries.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: *Human factors and Human information processing*

H.5.2 [**User Interfaces**]: *Natural language*

General Terms
Human Factors, **Languages**.

Keywords
Speech Interfaces, low literate, cell phone, telephone, viral, entertainment, communication services

ICTD '12, March 12 - 15 2012, Atlanta, GA, USA

1. Entertainment as Vehicle for Development
Most ICTD projects focus on core development areas – such as health, agriculture, and education – and design user interfaces suitable for their target users, who are often low literate and inexperienced with technology. Recently, however, Smyth et al. [1] described the remarkable ingenuity exhibited by such users when they are motivated by the desire be to entertained, and concluded that such powerful motivation "turns UI barriers into mere speed bumps" (ibid). Inspired by this powerful demonstration, we set out to systematically develop viral entertainment as a vehicle for dissemination of development related services.

Our ultimate goal is to disseminate speech-based communication services for low-literate telephone users throughout the developing world. Such services may include: buying and selling goods and services (i.e. a speech-based equivalent of Craig's List, craigslist.com); finding and communicating with others who share a common interest, and facilitating social and political activism (speech-based message boards); expressing opinions and making them broadly accessible to others (speech-based blogging and tweeting); sending and receiving group messages (speech-based mailing lists); broadcasting and receiving announcements in emergencies, and Citizen journalism. All these services are already available, in a textual form, to affluent people via the web, and some of them are also available to non-affluent but literate people via SMS. None are currently available to the low-literate.

To disseminate such services to low literate people throughout the world, we must first teach them how to use speech interfaces, and do so without the benefit of explicit user training, relying instead on peer training and viral spread. At the same time we must also advertise and promote our intended services. We can achieve all of this, and more, via viral entertainment. Specifically, our speech-based, telephone-based viral entertainment system has the following simultaneous, mutually reinforcing, goals:

Introducing and popularizing speech interfaces: Familiarizing a large number of people with automated dialog systems, and teaching them how to navigate menus, use DTMF (push button), provide speech input, etc.

A "hook" or 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.

Experimental testbed for testing speech interface choices: A stable viral entertainment system is immensely valuable for studying, via randomized controlled trials, the effect of interface design choices such as user input modality (voice vs. push-button); voice input style (keywords vs. unconstrained speech); optimal depth and breadth of menu trees; system prompt wording, and system voice characteristics.

Entertainment: A speech-based entertainment service has its own intrinsic development value by directly addressing the universal human need for entertainment.

In what follows, section 2 summarizes related work on the use of spoken dialog systems for development, and on viral services in the developing world. Section 3 describes Songline, our first abortive attempt at speech-based viral entertainment, and the lessons we learned from it. Section 4 describes Polly, a simple telephone-based voice modification and forwarding system, which as far as we know is the first successful viral system aimed at the low literate. Section 5 analyzes Polly's viral spread, and section 6 analyzes its usage patterns. We conclude in section 7 with lessons learned and future plans.

2. Related Work

Several attempts to design user interfaces for low literate users have been reported in the literature. Plauché et al [2] 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 [3]. In *Warana Unwired* [4], 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 [5] compares 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 *Health Line* ([6], [7]) 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 [8] is another successful example of a speech interface serving low literate 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. [9] 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. Wyche et al in their empirical study of professionals living in Nairobi, Kenya [10] highlight four factors to guide ICT work in infrastructure poor settings with an emphasis on collective consideration: limited band width; high costs of access; 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. To achieve virality, Baker [11] suggests (albeit in the context of literate users and web-based services) maximizing the product of (1) Install Rate (Percentage of invited users who install the applications); (2) Invite Sending Rate (percentage of users who invite at least one friend); and (3) Average Invites (average number of invites sent per user).

A successful example of cellphone based (though not speech-based) viral spread is *SMS-all* [12] (previously Chopal), a group text-messaging service in Pakistan. In addition to sending and receiving group messages, users can create new access-controlled groups and join already existing ones. As of last report [12], 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 ([13], [6] and [7]) 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, [6] found that low literate users preferred DTMF input over speech input, although they performed better on average with the latter. The results of user studies conducted in Botswana by Sharma et al [14] in the context of HIV health information systems for the semi and low literate population, also suggest user preference towards touchtone over speech while both systems perform comparably. In contrast, [8] and [15] (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 [15] 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 regions: 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 [16] can be used, as it provides high recognition accuracy in any language for up to several dozen words.

3. Songline: A Song Recording and Forwarding Application

Songline is a telephone-based, voice-based application which allows users to listen to songs recorded by others, as well as to record their own songs and to forward them to friends. We developed Songline as part of our exploratory study in Pakistan and adapted it in response to user feedback. We implemented it using the Tropo speech and telephony platform [17] due to the latter's reliability, ease of development and hosting service. Songline employs spoken prompts (in Urdu and English) for output and DTMF (push button) for input.

To promote Songline to low-literate, low-income people, we kept it free to the user via a "missed call" mechanism: a user calls Songline, hears a busy tone, and hangs up. The system then immediately calls the user back (incoming calls are free in Pakistan). The "missed call" mechanism is already familiar to telephone users in Pakistan and most other developing countries.

3.1 User Interface

When Songline returns a user's missed call, it offers them to either record a new song or listen to already recorded ones. Songs can be browsed by *most-popular* or *most-recent,* and the user can skip, listen, and vote for them. To record a song, the user is given up to 30 seconds to sing, but they can end earlier by pressing a button. The user may also enter phone number(s) of explicit recipients -- friends to whom their recorded song will be actively forwarded. In this case they may also record a brief introduction. Songline then calls the explicit recipient on behalf of the sender and plays them the sender's introduction followed by their recorded song. The recipient may choose to reply with a song, record their own song, or browse all the songs in the system.

All actions in Songline require explicit user confirmation.

3.2 User Feedback and Lessons Learned

We conducted focus groups in Lahore to gather feedback on Songline. The 10 participants, in two focus groups, were office workers at a local university, with eight years of formal education. Participants were first given a brief explanation of what Songline can be used for, then observed interacting with Songline. They were also encouraged to continue using Songline on their own after the focus group and their activity was logged. Following is a summary of users' feedback:

Feedback on Songline as an Application

- It was not clear to some of the users how such an application would be beneficial to them
- Songs and music of certain types are considered culturally immoral in Pakistan and hence the very theme seemed controversial to eight participants
- All ten participants were concerned about the fact that anyone can listen to their recording, considering it a breach of privacy. They also seemed shy to sing in front of our team or even their fellow workers

Feedback on the Interface

- Entering phone numbers of friends (rather than selecting from a built-in phone directory or call history) was reported to be a big hurdle by three
- Busy tone confused one user who thought that the number is actually "busy"
- Interface was reported as being confusing to two users due to detailed call tree and numerous options
- Prompts were unclear due to low audio volume as was reported by five participants
- At least two participants had difficulty entering international style phone numbers with country code and "+"

This was enough to convince us that drastic changes in theme and interface were required to make an application acceptable and compelling to our target population. This led to the development of *Polly*, as described below.

4. Polly: A Voice Manipulation and Forwarding Application

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 similar meaning to "Polly the Parrot". The theme of light entertainment using funny modifications of a voice recording is non-controversial and easy to understand (as we also confirmed later by analysis of user feedback). Recorded content in Polly is only available to the sender and intended recipients, in contrast to Songline where it was available to any caller.

Polly is also based on the Tropo platform and employs pre-recorded prompts for output. Standard system prompts play in Urdu or English depending on whether the call is made from Pakistan or the US. DTMF (push button) is used for input. The "missed call" mechanism is used to keep the service free to the user.

4.1 User Interface

Following feedback from Songline, we designed Polly's interface to be particularly simple, significantly reducing the depth and breadth of the call tree. When a user calls the Polly local number, instead of busy tone they hear a "caller tune" informing them in a parrot-like voice that Polly will call them back momentarily (caller tunes are mobile operator supported services for which the caller is not charged). Polly calls back the user and prompts them to record a short phrase. The recording terminates if the user remains silent for more than 4 seconds or when 15 seconds have elapsed. Immediately afterwards, a funny modified version of the user's voice is played back. This is done to engage the user early on, before they encounter any menus or need to press any key. At this point, the user is given an option either to try the next voice manipulation effect or to forward their modified voice to friends. We offered the following voice modifications effects, all achieved with a standard audio processing utility:

1) An *I-have-to-run-to-the-bathroom* effect, achieved by a gradual increase of the pitch,

2) A *drunk chipmunk* effect, achieved with pitch and pace modification,

3) Converting the voice to a *whisper*, achieved by replacing the excitation source of user's voice with noise

4) Adding *background music*.

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

Sample voice modifications and a detailed video demonstrating user interface are available online at [18].

If the user chooses to forward his recording to friends, he is prompted for the phone number and (optionally) name of his

friend and also (optionally) his own name for introduction. Only the phone number is confirmed for correction. The user is not required to press any keys to terminate recordings of names, which are terminated by silence detection and/or time-out after 4 seconds. The user is allowed to forward his voice to multiple recipients with the same or different modifications applied.

Polly calls the intended recipients to deliver the recorded voice and the sender's phone number is sent as the caller ID. The sender's name (in his own unmodified voice) is immediately played to the listener to prevent any confusion regarding the identity of the caller. A recipient can choose to send a reply, forward the recording to others or create their own recordings.

As an additional mechanism for viral spread, text messages containing Polly's contact information are sent to all of Polly's recipients on their first two interactions with the system.

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.

4.2 Distributed Setup

We use a distributed system setup for Polly. User initiated calls are received on a mobile phone in Pakistan which is attached to a PC, which in turn rejects the call and forwards the caller ID to the Polly application running on Tropo servers in US. The return call is made directly from US to Pakistan as an international Tropo call charged to the research project. Voice modification and file storage are done in the US, where a logging database is also kept.

4.3 Pilot Launch

In a pilot test of Polly. We gave Polly's phone number to two office workers at Lahore University of Management Sciences (LUMS) and asked them to call it, without explaining the details of the application. The test lasted two weeks during which the user base increased spontaneously to 32. We then stopped Polly and gathered feedback by interviewing the users. Most of the reported problems were minor software flaws which were fixed as we prepared for the main launch.

4.4 Main Launch and Shutdown

We launched Polly by seeding it with the 32 users who had participated in the pilot launch. We had our system call these users up and announce via a voice recording that Polly is back up. We made no further attempts to solicit users. Figure 1 shows the growth in system usage over the 22 days it was active. In what follows, we define "Polly Day" as a 24 hour period ending at 4am Pakistan time. "Polly Day 1" refers to the first day of Polly's launch, which was shorter than the other Polly Days since it started in midday.

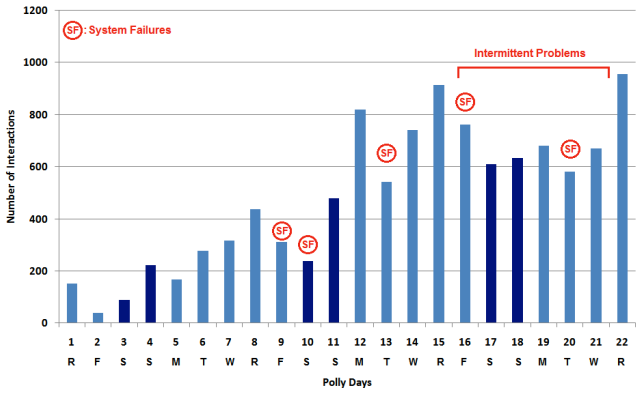


Figure 1: Polly Traffic Volume by day.

Unintended System Downtime: Polly experienced multiple down times due to power/ internet failures and other technical and administrative reasons. Major system failures occurred on Polly Days 9, 10, 13, 16 and 20, and repeated intermittent problems were experienced on Polly Days 16 − 21.

Shutdown: We shut down Polly at the end of Polly Day 22, for two reasons:

1. During the peak evening hours (see Figure 5 below) the system was saturating, with many callers receiving a busy signal due to contention over the single incoming phone line (this was later confirmed in the analysis of user feedback). We also started experiencing intermittent system problems, further exacerbating user frustrations. We did not want users, especially new ones, to be frustrated in their interactions with our system.

2. The international call charges were becoming a significant financial burden for us.

It thus became clear that, to support the ever increasing incoming call volume, we need to reconfigure Polly for a multiline setup operating entirely within-country.

We shut Polly down gradually and gracefully, for the next 10 days, we continued to use the Missed Call mechanism but the returned call played a message stating that Polly is offline temporarily to make improvements on the basis of user feedback. We also solicited further feedback by inviting the caller to record any additional suggestions and comments. This **Post shutdown feedback** will be discussed later on. After that period, we replaced Polly's caller tune with a brief message stating that Polly is temporarily offline and will return as soon as we are done improving it on the basis of user feedback.

Airtime Cost: We paid $0.12/minute to go from US to Pakistan over IP, with another $0.023/min for using a hosted solution. The total traffic for the 22 days amounted to 26,000 minutes (approx.) incurring a cost of around $4,000.

4.5 Annotations of Recordings and Feedback

Four student annotators (three male and one female) listened to the recordings and created detailed annotations based on their subjective assessments. All recordings were annotated in this manner, each by a single annotator. Each recording was first annotated as to the speaker's gender, estimated age; the language used, and net recording duration. Recordings were next categorized by their content (humor, information, romance, introduction, informational, profane, saying hello, meaningless, or other). Feedback recordings were similarly annotated, although with different content categories (request for continuation of service, appreciation for Polly, suggestion for improvement,

specific feedback, complaint, suggestions for new services, or other). Annotators also noted the location or geographic origin of the caller, if it were expressed in, or could be inferred from, the recording. Finally, the annotators were also encouraged to note in unstructured comments any interesting subjective observations, such as the functions for which Polly appears to be used.

For the purpose of annotation, recordings were sorted by the phone number used (which often, but not always, corresponded to a single user), and then by increasing date and time. Each annotator listened to all the recordings of a subset of the users, and all recordings of any user were listened to by the same annotator. This aided in making demographic assessments and enforcing annotation consistency. It also allowed the annotators to observe changes over time in a user's interaction style and usage patterns.

5. Analysis of Viral Growth

During the 22 days that Polly was active, it handled a total of 10,629 calls (interactions). We distinguish calls initiated by the user (albeit via the *missed call* mechanism) from those which were initiated by Polly (to deliver a recording from another user). We also distinguish whether any new delivery requests were made during the call (see Table 1). We consider an **Active Interaction** any call which was either initiated by the user or during which a new delivery request was made (all but the bottom right cell). Note though that, as our annotation shows, even interactions we deemed Not Active often involved significant user engagement: voice recording, listening to various modifications, and sharing with others nearby (203 such *non-active* interactions were made). Note further that many users who received deliveries chose to hang up and then place a separate user-initiated call. Taken together, this analysis shows that Polly simultaneously provided entertainment and promulgated viral growth.

	User Initiated	**System Initiated (delivery)**
User made new delivery requests during interaction	**4,340**	**699**
User made no delivery request during interaction	**2,444**	3,146
Total	**6,784**	**3,845**

Table 1: Breakdown of Polly's 10,629 interactions

The dual entertainment and viral-spread properties of Polly are also demonstrated in the growth of its user base. Table 2 shows the number of users who took part in each of the four types of interactions (the categories are not mutually exclusive). Note that 86% of Call Initiators and 17% of Call Receivers spread the service further to their contacts.

	Call Initiators	**Call Receivers**
New delivery requests made during call	**525**	**313**
No delivery request made during call	**476**	1,723
Total	**613**	**1,843**

Table 2: Breakdown of Polly's 2,145 Users by Interaction Type (not mutually exclusive)

Word-of-mouth spread: Of Polly's 613 Call Initiators, fully 291 (47.5%) placed their first call before receiving any calls from Polly. We conclude that a significant component of growth and adoption of a service like Polly can come from word of mouth or physical observation and emulation.

We define **Active User** as any user who participated in an Active Interaction

5.1 Rate of Adoption

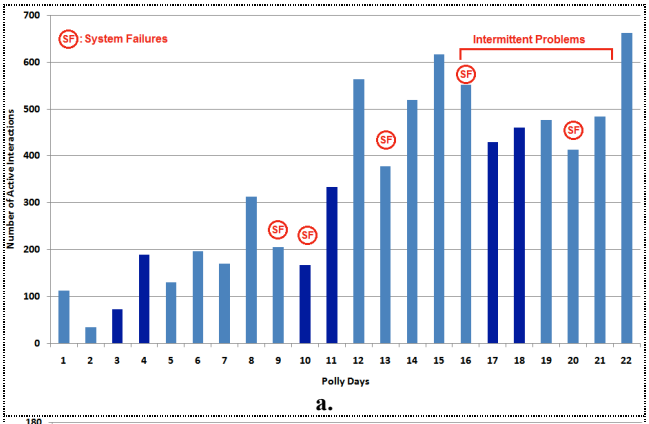
Another important measure of the viability of a service like Polly in the developing-world is its rate of adoption. Namely, how quickly did users adopt Polly once they learned about it? Of the **322** users who became Call Initiators when called by Polly

- **25** immediately placed delivery requests *during* the first call that they received from Polly (Instant Conversion)
- Out of the remaining **297**, 166 users started using Polly (called back Polly) just after receiving a single call. However, on average it took 1.83 calls to each of these **297** users before their first in-bound call

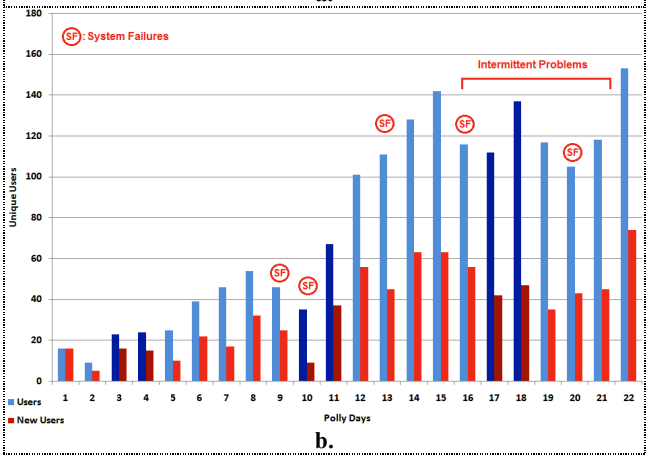
Therefore, while a very small percentage of users become active users in their first interaction, most became active users after fewer than two reminders. This relatively small number, 1.83, highlights the potential of, and the pent up demand for, value added voice-based services in the developing world.

5.2 Usage growth pattern

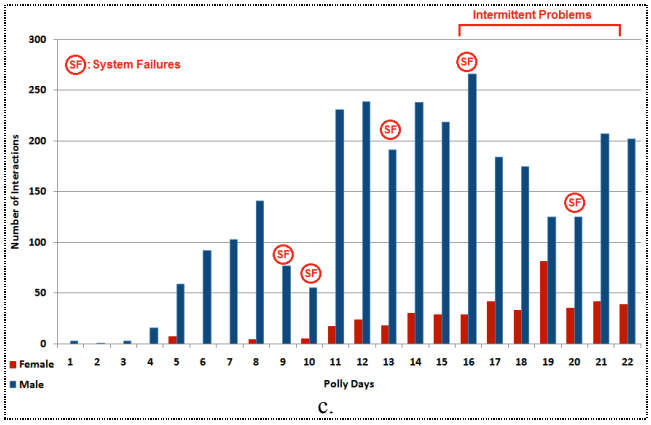
Figure 2 shows the daily growth in Polly's usage. Even though viral growth is evident, the dampening effect of system down times (Days 9, 10, 13, 16 and 20) is clearly visible, as is the impact of the intermittent disruptions during the last 6 days. Note also the impact on the number of new users following system down time, as many users are introduced to Polly via Polly deliveries.



a.



b.



c.

Figure 2 Polly's daily growth in terms of (a) interactions, (b) users, and (c) gender

5.3 Short-term vs. long-term users

Since the entertainment options provided by Polly did not vary over time, we expected that most users will tire of it within a short time. Figure 3 depicts user retention by means of sedimentation layers. Each layer represents the cohort of users who started using Polly on a given day. For example, on day 12 (yellow), there were 57 new users, out of a total of 99 users that day. The members of this cohort who remained active on subsequent days can be seen by following the shrinking color band. As expected, most users stayed with the system for only a few days. However, of note, a small but consistent fraction of users appeared to have settled into long-term usage. Figure 4 shows the percentage of active users who continue to use the system k days after their first interaction.

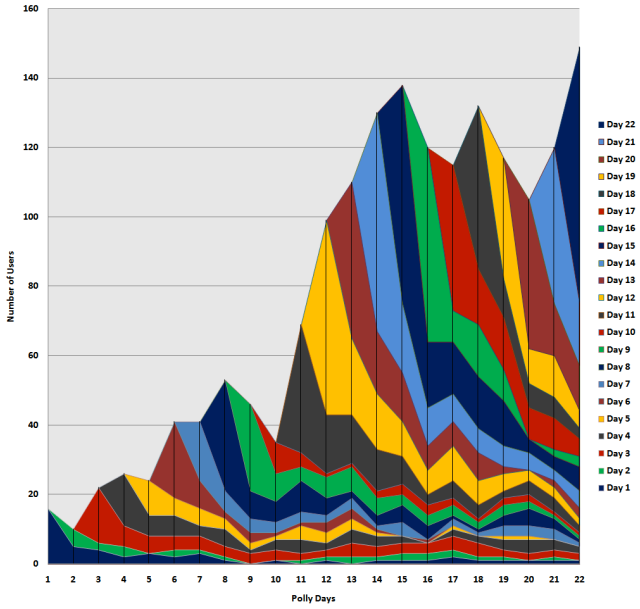


Figure 3: Number of Active Users each day. Each color corresponds to users who started using Polly on a given day. About 10% of users continue to use Polly long term.

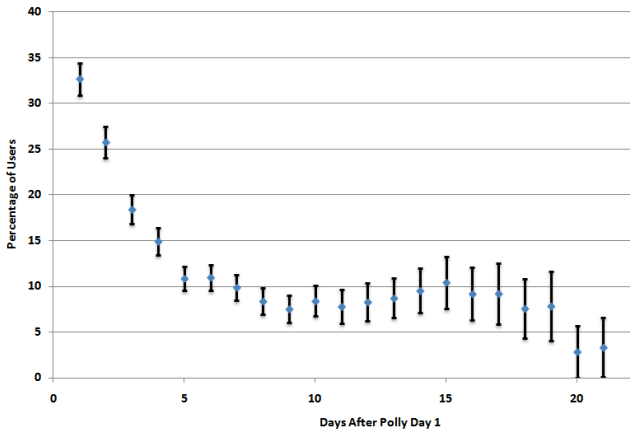


Figure 4: User retention in Polly

5.4 Activity by Time of Day

Figure 5 breaks down Polly's interactions by time of day. Although peak activity occurred in the evening hours, there was significant activity throughout the day, dropping down to a trickle only late at night.

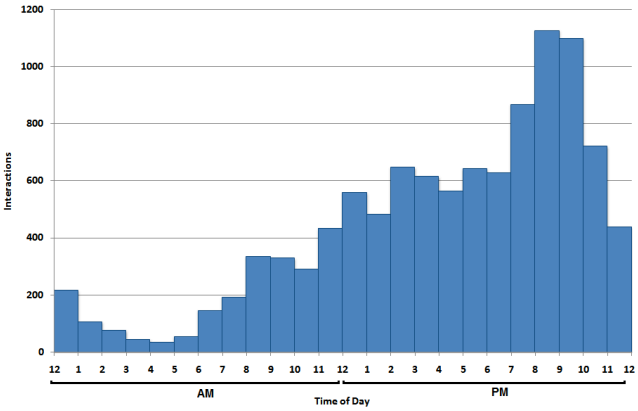


Figure 5: Polly activity by time of day

5.5 Post Shutdown Usage

Users were still calling Polly 40 days after its final shut down, when we finally stopped monitoring the calls. We received 1276 calls during this period made by 310 individuals. 117 out of these callers were new users who had never called Polly during its active period. A significant number of users kept calling repeatedly, as many as 46 times. This is perhaps the most compelling evidence for the potential of, and pent-up demand for, this type of service in developing countries.

5.6 Feedback

5.6.1 Feedback elicited during Polly calls

The feedback mechanism was implemented on Polly day 13. In the nine remaining days in which this option was available, 391 recordings of feedback were made by 264 unique users. Out of these, 189 recordings of user initiated feedback were made by 129 individuals. In addition, 202 system-prompted feedback recordings were provided by 169 users. 34 users recorded in both feedback categories. Out of the 391 recordings 118 were empty files.

Based on the remaining 273 files (191 distinct users) following is a brief summary of the main suggestions and complaints (the categories are not mutually exclusive):

Feedback Type	%
Interface/functionality related feedback and complaints: too long turn-around time of message delivery; poor call/sound quality; busy network; too short message recording time; increase/rearrange sound effects etc.	49
General appreciation including mentioned reasons such as: a way to connect to friends; a means of having fun; free service etc.	47
Confused users (pressing keys or saying "hello")	7
Irrelevant feedback including: songs; messages for friends; irrelevant messages for Polly etc.	5

In addition to these, several users recorded interesting and useful suggestions for the improvement of Polly as well as their ideas about other speech applications. A brief summary follows:

• Ideas for new voice modification effects in Polly including female/child voice modification, laughter and giggling, scary voice and background effects like sad music, rain drops, sound of a train, wind blowing etc.

• A application similar to Polly just for ladies

• Several suggestions to improve user interface including a rerecord option for messages, rapid access to effects of choice, options to go to the previous menu and to end the call etc.

• Suggestions to improve the wording of the prompts

Additional suggestions included:

• An accessibility software for the blind that could be used on less expensive mobile phones

• A software that could identify and filter out foul language in a message

5.6.2 Post Shutdown Feedback

As mentioned in 4.4, we continued to elicit feedback after shutdown. 565 post shut down recordings from the initial 11 post shutdown days were analyzed. 299 of these were empty files (noise, silence, button presses etc.). Out of the remaining 266 files 34% users asked to bring Polly back online as soon as possible. 8% were annoyed / angry because of Polly's shutdown and 8% explicitly expressed that this shutdown is creating problems for them and Polly was useful for them. 16% of the users simply stated that they want Polly to continue as it is a good service. Several files contained irrelevant feedback or recordings of users who were confused by the shutdown and were pressing keys or asking questions.

6. User Demographics and Usage Analysis

Gender: Among Polly's 773 Active Users, 74% were determined by the annotators to be male, 14% to be female, with the remaining 12% undetermined (young children, old people, too much background noise, etc.). As shown in Figure 2(c), sustained usage by female users coincided with an explosion in use by male users.

Age: Figure 6 gives the approximate distribution of the Active Users' age, as guesstimated by the annotators.

It is clear that most users are young men. This may partially reflect our seed population. Nonetheless, we were encouraged to observe participation by other demographic groups, especially women.

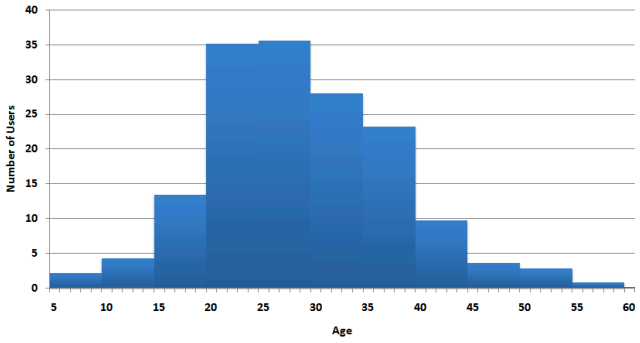


Figure 6: Age Distribution

Language of recordings: some 46% of the user recordings were in Urdu, which is Pakistan's national language and the most widely spoken and understood language in Pakistan. This is also the language we used for Polly's prompts, effectively priming for the use of this language, and excluding those who don't understand it. Of note, fully 38% of the recordings were in Punjabi, the regional language of Punjab province, where Polly was launched. Fewer than 3% of recordings were in mixed Urdu/Punjabi, in English, or in mixed Urdu/English. Interestingly, there was also a smattering of recordings in Potohari (17 recordings), Saraiki (6), Pashtu (2) and Arabic (2).

Socio-economic level and educational background: Although we have not rigorously verified this, the transcribers' best estimates of the socio-economic and educational level of the callers suggest that the vast majority come from a socio-economic class similar to that of the originally-seeded low-skilled office workers, and with an educational level that does not exceed theirs (approximately 5th grade and below). The fact that such users were able to successfully use and share the system with their friends is consistent with the finding by Smyth et al. [1] that low-literate users are able to learn potentially complicated user interfaces without any systematic training when the purpose of the system is entertainment.

6.1 Usage Analysis

Table 3 shows the distribution of voice recordings based on the type of content recorded. These message types are not mutually exclusive. In 2,138 interactions the users did not place any delivery requests and simply played with the system by recording their voice and listening to the modifications. 203 out of these were calls initiated by Polly to deliver messages. This brings forth two main uses of Polly: personal entertainment and free voice messages.

Content Type	% of Messages
Informational messages for friends	50.53
Hello/Hi and introductory recordings	13.01
Profane	6.26
Poetry/Songs/Whistling	2.60
Humorous recordings	2.07
Romantic	1.99
Political	0.02

Table 3: Distribution of recordings by content

Further analysis showed striking differences in usage by females and males. There were almost twice as many romantic recordings,

four times fewer profanity-laced recordings, and four times more song recordings by women than men.

6.2 Additional Findings

As discussed earlier, a significant fraction of Polly's users started using it without ever receiving a call from it, presumably having been introduced to it by word of mouth or by observing others. In addition, a large portion of user recordings (including feedback files) contained interactions where one user is training another to use the system. All this points to the important role of direct human interaction in the spread of Polly. This is not unexpected in a rural setting where people routinely gather face to face in the evenings.

Another interesting finding was that people kept recording detailed messages and complaints about problems with everyday necessities such as electricity outages. Some explicitly requested that their voice should reach government, relevant officials and people who can solve their problems. Some users even identified Polly as someone who can solve their issues or raise their voice.

7. Conclusion and Future Plans

Entertainment and free voice messaging both proved to be powerful motivators that attracted 2,032 users to Polly in a span of 22 days, compelling them to learn it and enabling them to find creative uses for it in their daily lives.

Although Polly was not designed to fan-out as much as SMS-all [12] (which was based on group messaging), it nonetheless became viral practically instantly. That this happened in spite of significant system failures and intermittent down times underscores the potential for speech-based services, as well as the pent up demand for entertainment, among our target audience.

Of particular note, Polly's viral spread crossed gender and age boundaries and even established itself in a female population. However, it appears to have not crossed socioeconomic boundaries. This could be due to the more insular nature of socioeconomic classes, or due to our service not being attractive to wealthier people, who may have other entertainment alternatives available via the Internet.

Stability and capacity: Much of the instability that Polly suffered from, and most of the airtime cost we incurred, can be attributed to the distributed and international setup we used. We are now working to set up Polly locally in Lahore, using robust multiline infrastructure. This will allow us to reliably support significantly larger call volume at the fraction of the cost.

Airtime charges: Even with our new setup, local airtime costs (currently about $0.02/minute) will still be incurred. Although we could shift most of this cost to the users by eliminating the missed-call mechanism, we are reluctant to do so at such an early stage because this may discourage the poorest users, who are the ones we are most interesting in reaching. One option we are considering is to offer two versions of Polly: a free barebone version, and a premium one which requires the user to pay their own airtime charges.

Help facility: Our annotators noted that some users found it difficult to master Polly on their own, and could be heard asking their friends for assistance. We also noticed that some features of Polly, like group messaging, were rarely used. In future versions, we will incorporate an intelligent help facility. Less frequently used features could be introduced after a user has had some success with the basic system (the callerID mechanism allows us to effectively retrieve at runtime the user's entire interaction history).

Speed dial: A frustrating bottleneck in requesting a forward is having to manually enter the friend's phone number. Since Polly is designed to work with any cellphone, it only uses the voice communication channel, and thus has no access to the phone's internal directory or dialing history. Our current system already elicits friends' names from our users during Polly interactions. In our next version, recipients could be specified by choosing from a short menu of recorded names.

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 views and conclusions contained in this document are those of the author 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 would like to thank Mr. Yasser Hashmi [19] for suggesting the use of voice modification applications.

9. REFERENCES

[1] 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, ACM, 2010.

[2] 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.

[3] 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*.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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, ACM, 2010.

[9] 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.

[10] 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, ACM, 2010.

[11] "Interview with edward baker about the viral factor | entrepreneurial minded." http://entrepreneurialminded.com/-business-interviews/ed-baker-viral-factor/. Accessed: Sunday, July 17, 2011.

[12] "Sms-all cheapest group sms service." http://-www.smsall.pk/. Accessed: Sunday, July 17, 2011.

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

[14] A. Sharma Grover, M. Plauché, E. Barnard, and C. Kuun, "Hiv health information access using spoken dialogue systems: touchtone vs speech," 2009.

[15] 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, ACM, 2009.

[16] 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, ACM, 2010.

[17] "Tropo - cloud api for voice, sms, and instant messaging services." http://tropo.com/. Accessed: Sunday, July 17, 2011.

[18] "Polly." http://www.cs.cmu.edu/polly/.

[19] "People - school of humanities, social sciences and law, lums." http://shssl.lums.edu.pk/people_detail.php?id=58. Accessed: Wednesday, July 20, 2011.