Behavior Analysis of Low-literate Users of a Viral Speech-based Telephone Service

Haohan Wang, Agha Ali Raza, Yibin Lin, Roni Rosenfeld
Language Technologies Institute
School of Computer Science
Carnegie Mellon University
{haohanw, araza, yibinl, roni}@cs.cmu.edu

ABSTRACT

We studied the behavior of users of a simple telephone-based voice modification and forwarding system, which has operated in Pakistan for about a year, attracting 165,000 users who interacted with the system by telephone over 636,000 times and generated very rich interaction data. Trying to cluster the users based on their activity profile, we found that they form a continuum rather than truly distinct clusters. We did discover that, with experience, users respond faster to menus (using more barge-in) and make fewer mistakes and abortive attempts. Finally we studied how users' choice of activity evolved over time, and found that with experience users show an increasing interest in message sending, become more explorative of the system's capabilities, and better adapt themselves to its constraints. Many new users seem to arrive with some preexisting knowledge of Polly's functionality, presumably through some back-channel information from their friends. Long-term users engage in lengthier calls from the start, and take a more active interest in voice modification and forwarding features.

Categories and Subject Descriptors

H.1.2 [**Models and Principle**]: User/Machine Systems *Human factors and Human information;* H 5.2 **[Information Interfaces and Presentation (e.g., HCI)]** User-Interfaces – *natural languages*

General Terms

Human Factors, Languages

Keywords

Speech interfaces, dialog systems, illiterate, low literate, developing countries, human factors, user behavior, human computer interaction.

1. INTRODUCTION

In this paper we analyze the data collected by Raza et.al through *Polly* [5, 14, 24], a telephone-based, voice-based entertainment service, which serves as a viral conduit for spreading development related services among low-literate users. The ultimate goal of the Polly project is to promote speech-based, development-related information services among low-literate telephone users throughout the developing world. Polly allows its users to record a short audio message, listen to various manipulated versions of their recording (such as faster/slower pace, high/low pitch etc.), and optionally forward a version of their choice to their friends. Users can also audio-browse a menu of job ads, which can also be forwarded to friends. User input to Polly is in the form of DTMF (push buttons) for menu navigation, and short recorded messages which are not interpreted.

The first, pilot deployment of Polly was launched in 2011. Over a period of 3 weeks it attracted 2,000 users and resulted in 10,000 interactions. The second deployment of Polly remained on-line for a year starting in May 2012. It attracted some 165,000 users and resulted in over 636,000 interactions, including 200,199 forwarded voice messages and 22,104 forwarded job ads.

In both deployments, system activity was limited only by the available telephone bandwidth (one phone line in the pilot, 30 phone lines in the main deployment). While Polly has indeed spread to users of intermediate (more than 12 years) and higher level (university) education, the vast majority of users were low literate [14].

The large Polly user base resulted in various creative uses of the system. Polly was used mostly for entertainment and job browsing, but some users also found Polly useful as a voicemail and group messaging service and even in telemarketing-like activities. All user interactions were recorded in the form of log files, database and audio files.

Polly's rich interaction data can be mined for many purposes. In this paper, we try to answer a few simple questions: (1) Do users fall into distinct categories based on their pattern of interactions with Polly? (2) Do users' interaction skills improve with experience? (3) Do users' activity patterns within Polly evolve over time?

2. RELATED WORK

Speech-based interfaces are commonly used in designing communication services for low-literate users ([1, 2, 3, 6, 7, 8, 13]). Medhi et al [25] found that abstracted non-textual and voice based systems are favored by low-literate users over text-based ones. Voice-based media has helped to promote social inclusion among underserved communities [9, 10, 11]. In terms of general user interface for developing regions, Wyche et al [15] reported that people in a Nairobi slum use Facebook for employment and entrepreneurial opportunities, and suggested that an interface allowing access to low-literate users be built on top of the

ACM DEV-4, December 06–07 2013, Cape Town, South Africa

http://dx.doi.org/10.1145/2537052.2537062.

existing, familiar ones, such as a Facebook website. [23] discussed various user interface modes, such as spoken dialog, text-based, Interactive Voice Response (IVR) and live operator in mobile applications.

User behavior in IVR systems in the developing world was studied in Project Gurgaon Idol [16], a telephone-based singing competition. In this project, over 80 participants were trained to use the IVR system to record singing by four methods: training over radio, repeated calls, over the phone and in-person handholding. No significant differences were found between repeated calls and a single call in terms of task completion rate, while the in-person handholding, which costs more, significantly improved task completion rate. [19] studied the time-of-day periods when 51 users call a self-reporting system of tobacco and alcohol consumption. In project Avaaj Otalo [3], 51 small-scale farmers joined an interactive voice application. There was no evidence of error prevalence decreasing with user experience.

More broadly, behavior analysis has wide application in real life. User segmentation, for example, has been an area of research with applications such as behavioral targeting [21, 22, 23, 27], and detecting social spammers [26]. Jeon et, al [20] performed cellphone user segmentation by analyzing smart phone logs. Three types of users were proposed: communicative-use type, entertainment-use type and restricted-use type. Ozer [21] proposed fuzzy clustering method to classify users with different goals in an online music service.

3. USER INTERFACE DESCRIPTION

Polly has two major features. It is a voice messaging service as well as a job market.

The most common way to contact Polly is via a "missed-call" mechanism which shifts the airtime cost to the system. A user calls Polly's phone number (a "Call-Me-Back" request), and Polly immediately hangs up the call. Shortly afterwards, Polly calls the user to start the interaction.

Once a user is connected with Polly, they are prompted to record a short voice message, limited to 15 seconds, and can opt to terminate the recordings before that by pressing the # button. Polly then applies its first voice modification to the recorded message (raising its pitch, thereby creating a male-to-female effect) and plays it back to the user. At this point the user arrives at the *Main Menu*, which gives him the following choices:

- "To re-record, press 0" (re-record a voice message, and return to main menu). This option is removed from the main-menu after the user chooses to forward the recorded voice to a friend.

- "To repeat, press 1" (replay the manipulated voice message, and return to main menu)

- "To forward (the manipulated recording) to friends, press 2" (navigate to the forwarding menu)

- "To try another effect, press 3" (play another voice modification effect, and return to main menu)

- "To listen to job ads for free, press 5" (navigate to the job browsing menu, from which the main menu is no longer reachable)

- "To provide feedback by recording comments and suggestions about the system, press 8" (only available from the user's 5th call onwards)

Note in particular that the user is prompted to record a short audio message, and then presented with a manipulated version of their recording -- all before he is presented with any action choices, including access to the job ads. This is because we view the entertainment aspect of Polly as crucial to achieving viral spread.

If the user chose to forward a message to friends, they are prompted to enter the phone number of their first friend, and then to confirm it. They are then prompted to record their name and their friend's name. This is repeated for any number of friends desired, after which they are returned to the Main Menu.

Polly only allows one voice message to be forwarded in one call, although it can be forwarded to multiple friends. The "re-record" option goes away when the user navigates back to main menu after the forwarding menu.

In the Job Ads menu, users can browse, listen to and forward job information to their friends. However, they cannot return to the Main Menu.

In any of the Polly menus, if a user presses an *invalid button* (defined as a button that is not allowed in that menu), Polly will play the menu options again.

Barge-in (pressing a button before the menu finished being played out) is allowed in all Polly menus.

4. USER CLUSTERING

In this section, we try to determine whether users fall into naturally distinct groups by attempting to cluster them based on their activity profile. We focus on the 63,023 users who had at least one active interaction with the system.

4.1 Representing Users by Features

We first extracted features based on the interaction log to capture users' behaviors (Table 1). Features 1-4 capture the different call types. Note that the majority of users interacted with Polly only once or twice, but some users used the system many hundreds of times. About three quarters of the total calls are user-initiated Call-Me-Backs calls. The rest are mostly message or job ad deliveries (initiated by a friend of the current user), with a small number of "unsubsidized" (a.k.a. caller-paid) user-initiated non-CMB calls to a second phone number where Polly picks up the phone and engages immediately.

Features 5-7 attempt to capture users' perseverance. More than half the users interacted with Polly for only one day, but some stayed active for many months. Interestingly, some users may stay away from the system for a very long time (up to 314 days), then come back.

Feature 8 shows that users make calls at any time of day. When calculating average time-of-day of calls, we "linearize" this value by consider each day to start at 6:00am and end at 5:59am.

Features 9-13 capture users' choices during their interactions. Interestingly, job ads (utility) have been played significantly more often than voice effects (entertainment) (feature 11 vs. feature 12).

Feature 14-17 categorize users' first experience with Polly. About two thirds of users initiated a call, presumably after being told about Polly by a friend through a "back channel": e.g. in person, by SMS, or perhaps via a regular (non-Polly) phone call. The other third of users first interacted with Polly when they received a delivery from it, although it is possible that they too were previously told about Polly (and alerted to the coming call) by a similar back channel.

Features 18-21 capture aspects of the users' social connectivity pattern. More than half of the users received messages but fewer than half sent messages out. Interestingly, the average branch-out (mean of feature 20) is greater than the average branch-in (mean of feature 19).

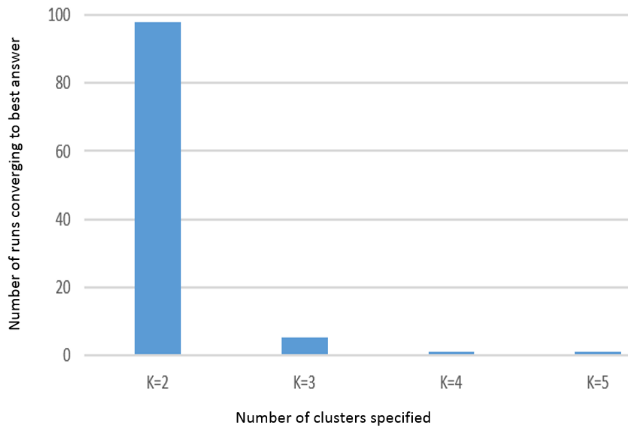


Figure 1: Numbers of clustering runs (out of 100) that converged to the best answer achieved, for different number of pre-specified clusters K

4.2 Clustering By Users' Behavior

We used the K-median [17] algorithm for clustering based on the above features. We chose medians instead of means to reduce the influence of outliers. We used city-block distance [18] rather than Euclidean distance because we felt that Euclidean distance weighed some features too much even after normalizing the scale.

The K-median algorithm is heuristic. Even when pre-specifying the desired number of clusters K, the outcome depends on initial choices. Typically, the algorithm is run multiple times using different random initializations, and the best solution (by some measure of separation) is reported. One way to assess the stability of the best solution is by how often it is found. For each value of K=2,3,4,5, we ran the algorithm 100 times. Figure 1 shows how many of these runs converged on the same best solution. It seems that for K>2, the solution space has so many local maxima that we hardly scratched its surface. We therefore settled on K=2 clusters, which resulted in two clusters roughly comparable in size. Figure 2 shows the result of our clustering, which was reduced to two dimensions with PCA. There is no natural separation between the clusters; rather, the users form a continuum.

To better understand these two clusters, Table 2 shows the difference between them in terms of the features' first two moments. Also shown is a simple measure of separation: the difference between the means, divided by the sum of the standard deviations. By this measure, by far the most informative feature is the average time of day of the calls (f#8). Also of possible significance are the average call duration (f#9), the number of calls (f#1) and the number of CMB calls (f#2). Figure 3 show the distribution of these features across the two clusters in more detail. We also inspected the scattergrams of every pair of features, but did not observe any interesting interactions (results not shown).

Feature number	Feature Description	Min	Max	Median	Mean
1	Overall number of calls by this user	1	794	2	5.6
2	Number of Call-me-back calls	0	624	2	4.3
3	Number of Delivery calls	0	684	0	0.8
4	Number of Job Delivery calls	0	127	0	0.3
5	Number of days user was active	1	134	1	2.9
6	User "life span" (from first to last interaction, in days)	1	338	1	20.8
7	Largest gap in usage (in days)	0	314	0	13.8
8	Average time of day of calls (in 6:00am--5:59am range)	6:00am	5:59am	4:24pm	4:42pm
9	Average call duration (in seconds)	10	762.1	187	208.6
10	Average time per call spent interacting with the job menu (in seconds)	0	589	0	21.2
11	Average number of job ads played per call	0	102	0.3	1.4
12	Average number of effects played per call	0	19.2	0.7	0.8
13	Average number of forwards sent per call	0	24	0	0.6
14	Was the user's first interaction a user-initiated Call-Me-Back? (1 yes, 0 no)	0	1	1	0.67
15	Was the user's first interaction a Message Delivery from the system?	0	1	0	0.2
16	Was the user's first interaction a Job Delivery from the system?	0	1	0	0.1
17	Was the user's first interaction a user-initiated unsubsidized call?	0	1	0	0.01
18	Number of delivered messages re-forwarded by this user	0	280	0	0.7
19	Number of distinct people (phone numbers) from which user received messages	0	20	1	1.0
20	Number of distinct people (phone numbers) to which user sent messages	0	60	0	1.5
21	Number of messages the user sent out	0	465	0	2.7

Table 1: User features used for clustering

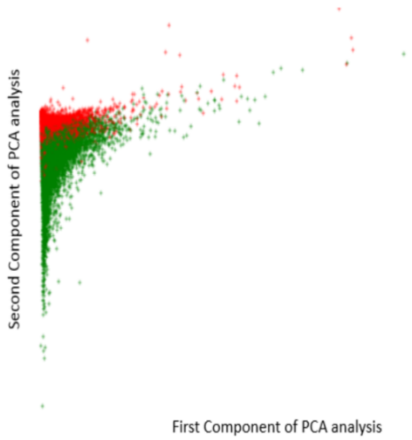


Figure 2: Clustering of the 63,023 active users into K=2 clusters. The lack of clear separation suggests that user space is a continuum.

Based on the above analysis, it seems that users can be categorized, albeit weakly, as "mid-day users" (green cluster) and "evening users" (red cluster). By comparing the centroids of the two clusters, we found that mid-day users tend to make more phone calls and generally interact with more of the system. They also show higher social connectivity (in both sending and receiving messages). Evening users might be thought of as more casual users.

	Red Cluster		Green Cluster		
Feature #	**mean**	**stdev**	**mean**	**stdev**	**Separability**
1	4.04	9.56	6.76	14.07	0.12
2	3.23	7.56	5.25	9.71	0.12
3	0.53	3.65	1.03	6.93	0.05
4	0.22	0.95	0.40	1.84	0.06
5	2.18	2.90	3.53	5.10	0.17
6	14.41	39.86	25.65	53.65	0.12
7	10.09	31.69	16.54	38.99	0.09
8	8:51pm	161min	1:39pm	157min	1.36
9	195.30	111.43	218.48	108.65	0.11
10	18.31	53.77	23.49	56.03	0.05
11	1.23	2.89	1.46	2.83	0.04
12	0.73	1.02	0.88	0.99	0.08
13	0.58	1.51	0.71	1.51	0.04
14	0.67	0.47	0.66	0.47	0.01
15	0.20	0.40	0.21	0.41	0.01
16	0.12	0.32	0.12	0.32	0.00
17	0.02	0.13	0.01	0.11	0.02
18	0.39	3.66	0.88	5.03	0.06
19	0.87	1.18	1.11	1.51	0.09
20	1.20	2.55	1.74	3.12	0.10

Table 2: feature value distribution in the two user clusters. Separability is calculated as the difference in means divided by the sum of the standard deviations

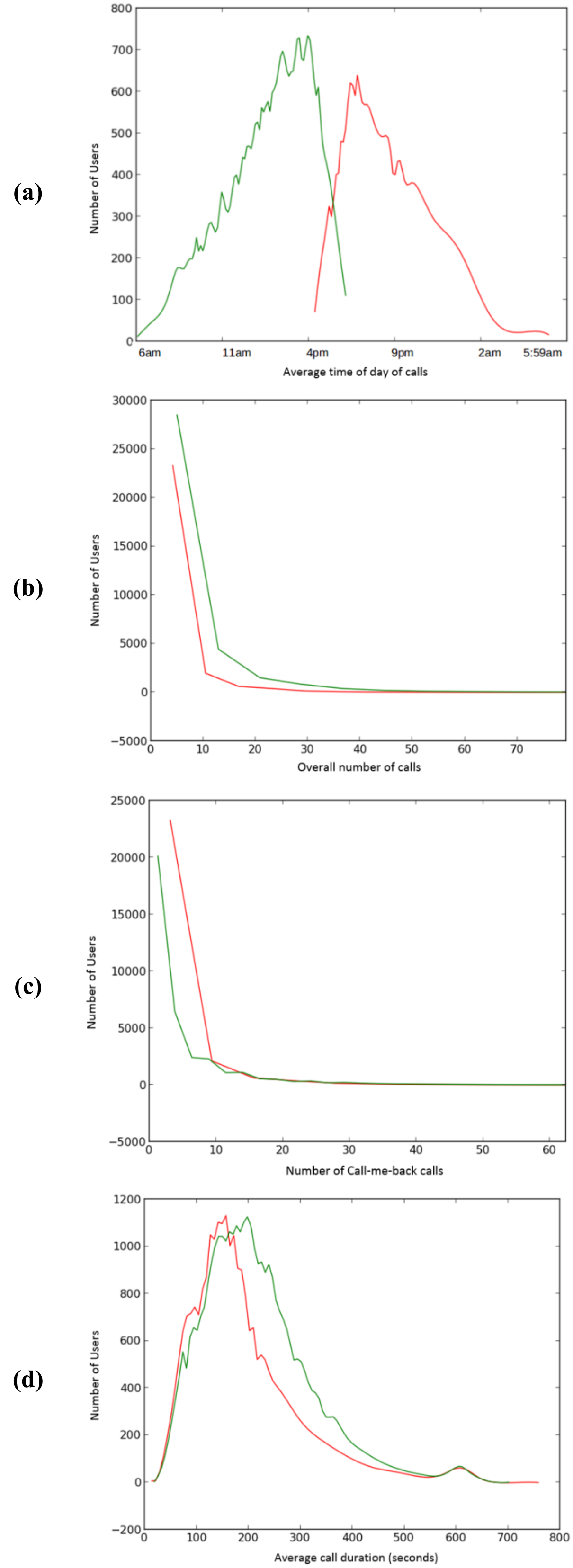
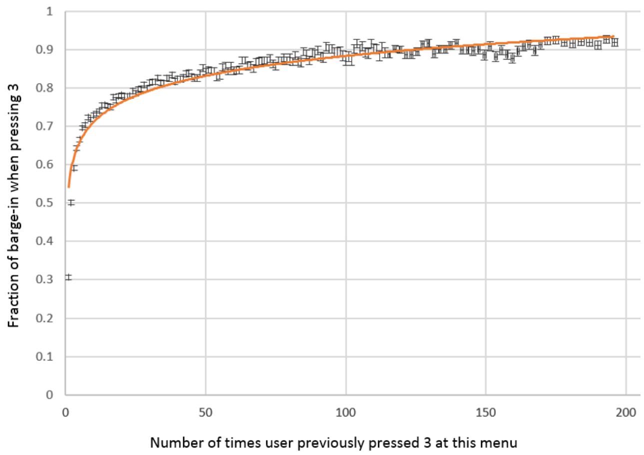
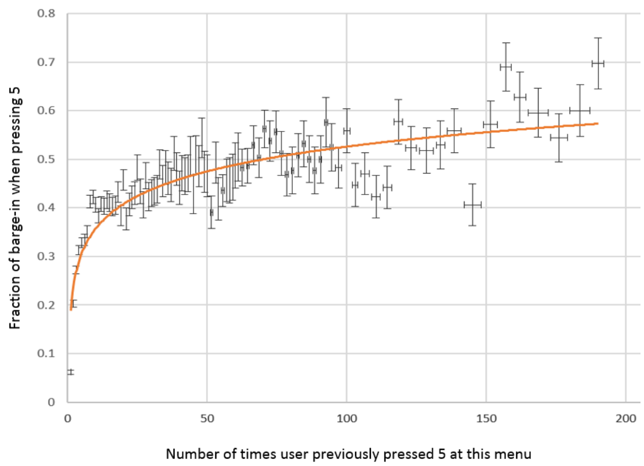


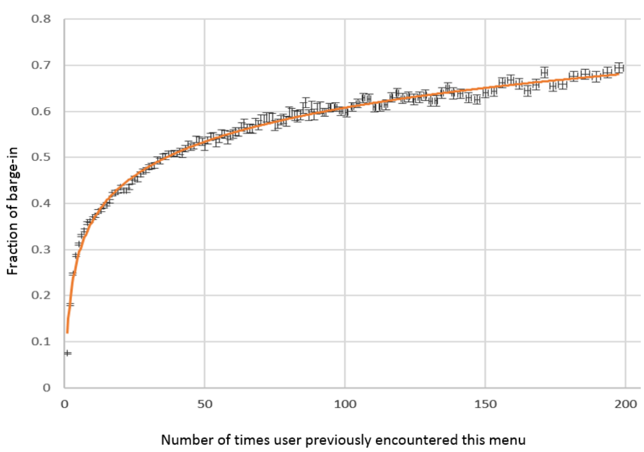
Figure 3: Cluster-specific distribution of the most informative features.



(a)



(b)



(c)

Figure 4: Prevalence of barge-in by user experience level. (a) for pressing 3 (next voice effect); (b) for pressing 5 (listen to job ads); (c) for any button. Horizontal bars correspond to binned values.

5. IMPROVEMENT IN USERS' INTERACTION SKILL

Since most of Polly's users were low literate, many of them may not have been familiar with speech interfaces prior to their interaction with Polly. In this section, we investigate if users' skill at interacting with our system improves as they gain experience with the dialog interface.

We focused on users' interactions with the first (main) menu of Polly. A user's menu-interaction is the process from hearing the menu options to pressing any button, hanging up, or a timeout (which only happens after the menu is played a second and then a third time). There may be none to several menu-interactions within one call. We analyzed 934,742 main-menu interactions from 292,951 calls during which there was at least one main-menu interaction. These calls were made by a total of 50,414 users, assuming that each phone number corresponds to a single, distinct user. We investigated users' interaction skill by tracking the prevalence of three phenomena: barge-in, invalid button presses, and unsuccessful forwarding attempts.

5.1 Barge-in Behavior

Barge-in, in speech interface terminology, occurs when a user interrupts a menu by pressing a button before the instructions end. It often happens when a user is familiar with the speech interface and needs fewer or no reminders.

We analyzed the changes in barge-in behavior as one indication of learning by the users. We tracked the prevalence of barge-in as users' experience level increases. Figure 4 shows the fraction of barge-in as a function of user experience level, separately for button 3 ("next voice effect"), button 5 ("go to job ads menu") and for any button. Here, user experience level is defined as the number of times that user encountered the same menu and made the same choice, before the current interaction. Whenever there were fewer than 100 such interactions at a particular experience level, we combined (binned) the data with the next higher experience level (horizontal bars). This tended to happen only at the higher experience levels.

We observe that barge-in prevalence increases with experience in all three categories, with a roughly logarithmic growth rate in the range studied. Fitting a logarithmic regression line results in $y = 0.075 \log x + 0.58$ for (a); $y = 0.074 \log x + 0.19$ for (b); and $y = 0.10 \log x + 0.11$ or (c). Thus the "learning curve" appears to be steeper for pressing button 3 (next voice effect, which is often repeated multiple times in order to reach the desired voice effect) than for button 5 (going to the job ads menu, which never returns to the main menu and hence can only be done once per call).

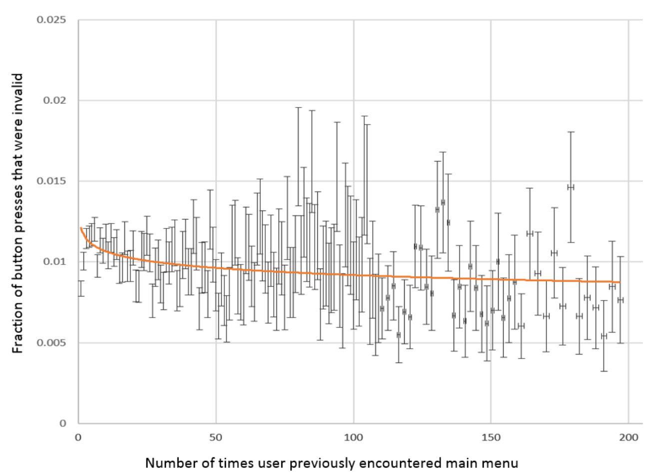


Figure 5: Prevalence of invalid buttons presses as function of user experience. Horizontal bars correspond to binned values.

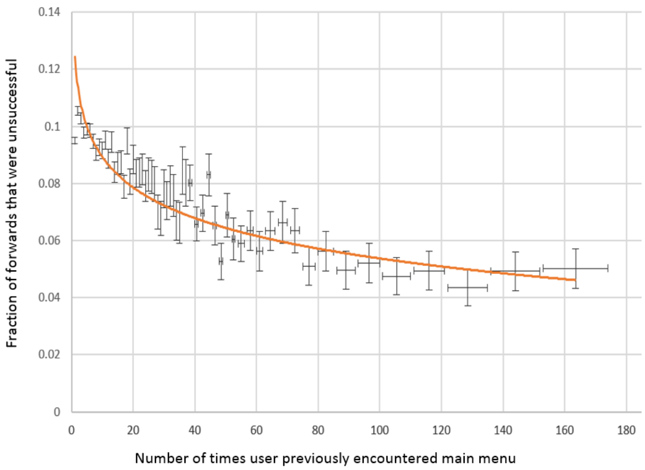


Figure 6: Prevalence of unsuccessful forwarding attempts as function of user experience. Horizontal bars correspond to binned values.

5.2 Invalid Selections

People can respond very quickly to voice instructions, but about 0.8% of all button presses in our data were invalid (not one of the offered choices). We studied how the frequency of these invalid selections changed as users' experience with the system increased. We continued to focus on the main menu of Polly. Figure 5 tracks the fraction of invalid selections as function of user experience. Whenever there were fewer than 1000 total button presses at a given experience level, we combined (binned) the data with the next larger experience level. The (negative) logarithmic regression line suggests a very mild improvement rate, contrary to our expectations.

5.3 Complex Functions

Perhaps the most complex function available from the main menu is "forward your recording to friends" (button 3). Forwarding requires a user to key in a receiver's phone number, to listen to the number being read back and press a button to confirm it (or else re-key the phone number), to record the sender name (for first recipient only), to record the receiver's name, and then to repeat the process for a an additional recipient or else indicate that there are no more recipients. This requires more interactions with the system than any other choices. As a result, there were numerous attempts to forward a message that failed.

Here, we focused on the interactions in which forwarding was requested. We binned the data as before to achieve a minimum of 1000 forward attempts per bin. Figure 6 shows how the fraction of unsuccessful forwarding attempts changes as user experience increases. The (negative) logarithmic regression line ($Y = -0.014\log x+0.12$) shows good agreement with the data. There is significant improvement even after more than 20 successful interactions.

5.4 Caveat

The analyses in this section were based on a very large number of interactions, which was made possible by combining the interaction data of many users, most of whom were short-term users, but many of whom were long-term users. Because of that, the analysis confounded true learning by any one user with differences between the different user types (short-term vs. long-term). It is reasonable to suspect that long-term users may tend to be more adept at using IVR system to start with. To isolate this effect, in the next section we study the experience-dependent behavior of specific, controlled user groups.

6. USAGE CHANGE WITH EXPERIENCE

In this section, we analyze specific sets of users and explore changes in their usage patterns as a function of their experience with Polly. To simplify things, we continue to concentrate on the main menu interactions only. We define an *interaction* with the main menu as the response of a user when faced with this menu: pressing any of the valid or invalid buttons, failing to press any button, or hanging up. We define *user experience* as the number of Polly calls this user experienced prior to the current call.

6.1 First Main Menu Interaction in a Call

For the analysis in this subsection, we focused on the first 30 calls of the 1,523 users who experienced at least that many calls. Our analysis is thus based on 45,690 calls. During these calls, these users interacted with the main menu on average 3.4 times per call, but here we look only at the first main menu interaction of each call.

Figure 7 depicts user responses when encountered with the main menu for the first time in a call, as a function of their experience (number of prior calls). The most striking finding is that, some 35%-50% of the time, users' very first chosen action in a call is to forward their recorded message (by pressing 2). Even on their very first interaction with Polly (their very first call), 35% of users jump straight to forwarding, without exploring any other options. We believe this may be indicative of preexisting knowledge of Polly's functionality among new users through a "back channel". This occurs when a new user of Polly gets informed about Polly and its use by a friend via means other than Polly itself (see discussion in section 4.1). In fact, we have anecdotal evidence in the recorded messages that some of Polly's users inform their friends about Polly and even give them basic usage tips before sending them a Polly message or giving them Polly's phone number. An early jump to delivery even by new users also suggests that Polly has a reputation as a "voice messaging system".

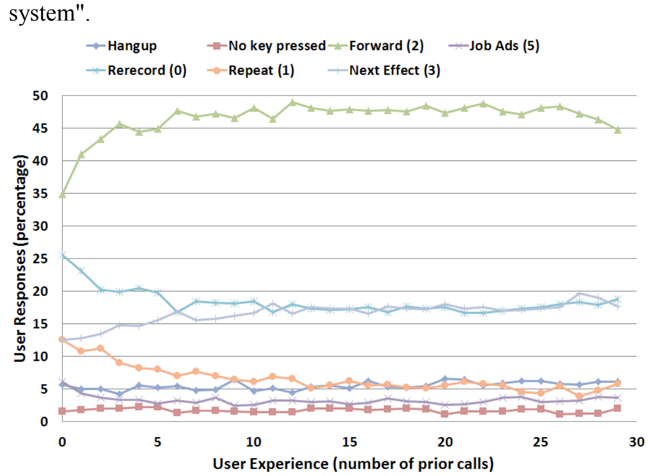


Figure 7: User response when encountered with the main menu for the first time in a call, among 1,523 long-term users

An increase in the use of forwarding indicates users' growing interest in this feature. This increase is accompanied by a corresponding decrease in the use of keys 0 (rerecord) and 1 (repeat). The repeat and rerecord functions are used a lot in the initial few calls. Our annotators were able to find two main reasons for this behavior: a) many users find it hard to compress their complete message in the 15 seconds allowed for recording, so they have to retry a few times; b) users play with the voice modifications by repeatedly recording themselves and listening to their modified voice, sometimes even playing out loud the modifications to family or friends around them. An increase in the use of key 3 (next voice modification) supports the latter hypothesis and also indicates users' increasing interest in exploring Polly's features. However, there can be a second reason behind this: some users prefer to send their voice using a particular voice modification, or even completely unmodified. As the unmodified voice is offered in the fourth place among other modifications, key 3 must be repeated pressed in order to reach it. The use of key 5 also gradually drops over the initial few calls. The frequent use of keys 2 and 5 may indicate that Polly is perceived by long-term users primarily as a "message sending" or "job ad browsing" system.

6.2 All Main Menu Interactions

In this section we compare the usage pattern of different types of users. The user types are defined in terms of the overall number of calls made by each user (feature 1 in Section 4). In the following subsections, we further define and analyze these user sets.

6.2.1 Short-term users

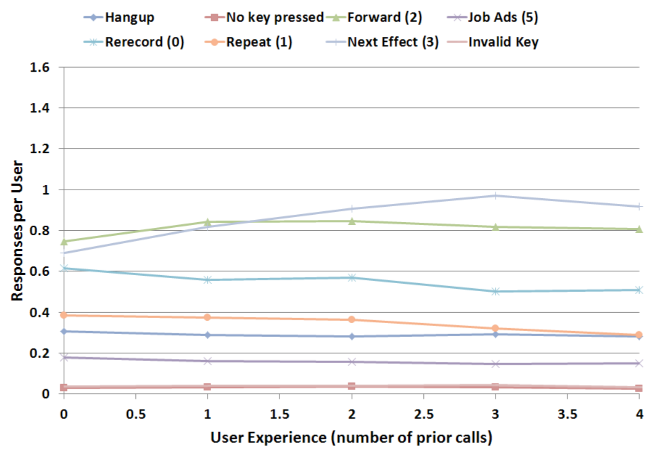
We define our *short-term* set as the 2,701 users who stopped using Polly after exactly 5 calls (we chose 5 because we were hoping this is a large enough number to observe trends). In this set, users encountered the main menu on average 3.09 times per call. Figure 8-a shows, at each experience level, the average per-user number of responses of each response type.

The most common user responses are keys 2 (forward) and 3 (next effect), both of which initially increase. The number of forwards starts decreasing slightly from the second call while the number of "next effect" choices decreases after the fourth call, before all these users stopped using the system altogether after 5 calls. Use of "rerecord" and "repeat" also show a decreasing trend. The prevalence of "switch to job ads" option is low and gently decreasing . Note that once a user chose to switch to the job ads menu, they cannot return to the main menu, so there can be at most one such response per call, and the average number of such responses corresponds to the fraction of calls that end up in the jobs menu.

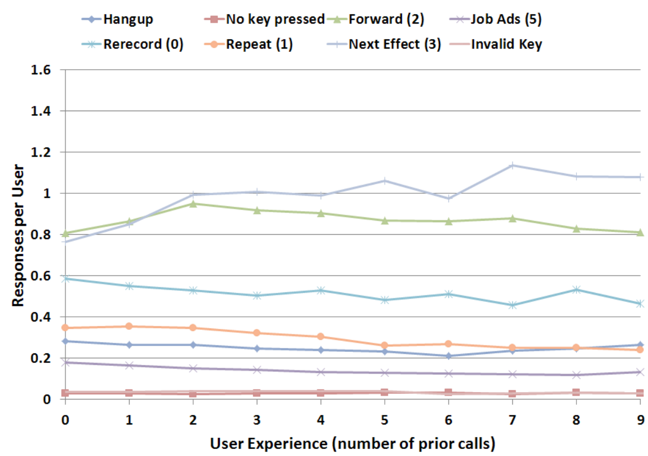
6.2.2 Intermediate-term users

We define the intermediate-term set as the 1,862 users who stops using Polly after exactly 10 or exactly 11 calls. We lumped the two experience levels together to arrive at a large enough sample, comparable to the short-term set (while we do not show the error bars, the large samples guarantees that changes of 0.1 are statistically significant). These users had encountered the main menu on average 3.15 times per call. Figure 8-b shows, at each experience level, the average per-user number of responses of each response type.

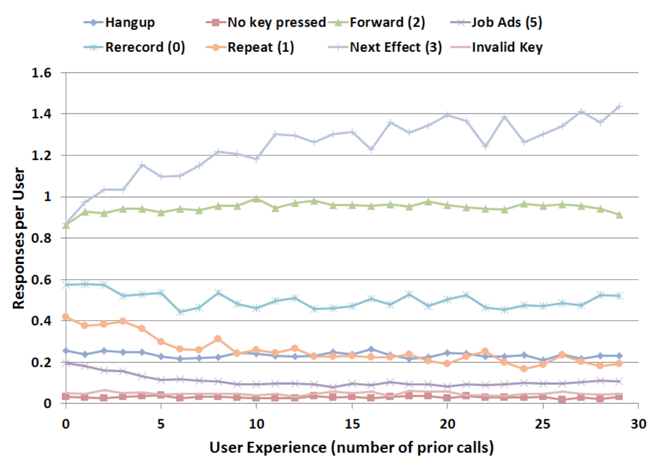
We find that the most prominent of user responses are again keys 2 (forward) and 3 (next effect). However, as compared to the short-term users *over the same range of 5 initial calls*, activity not only begins at a higher level but also climbs higher. While the average number of 'next effect' choices continues to climb until the experience level of 7 calls, the number of forwards starts showing a decline much sooner. Finally, use of both these options declines during the last two calls. Use of rerecord and repeat is very similar to that of the short-term users during the initial five calls. The option to switch to job ads declines gradually as in the short-term case, but here we can see the trend persisting through the users' 10th (and last) call.



(a) Short-term users (2,701 users who stopped interacting with Polly after exactly 5 calls)



(b) Intermediate-term users (1,862 users who stopped interacting with Polly after exactly 10 or 11 calls)



(c) Long-term users (1,523 users who continued interacting with Polly for 30 calls or more)

Figure 8: Average per-user number of responses of each response type in main menu, as function of user experience.

6.2.3 Long-term users

Finally, we define the *long-term* set as those 1,523 users who continued using Polly for 30 calls or more (we also studied separately the 508 “very long-term users” with 50 calls or more (not shown), but did not find any significant changes in their behavior past the experience level of 30). Users in the long-term set, during their first 30 calls, encountered the main menu on average 3.4 times per call. This is the same set of users that was studied in section 6.1. Figure 8-c shows, at each experience level, the average per-user number of responses of each response type.

Once again the most prominent of user responses are keys 2 (forward) and 3 (next effect). The prevalence of these two choices is higher than in the intermediate-term users *over the same range of 10 initial calls*. The prevalence of “next effect” climbs up more steeply during the first 10 calls, and then continues to climb till the end. It should be noted that not all users of this group stopped using Polly after 30 calls; therefore we don’t see the usual tapering off of the number of “next effect” choices towards the end.

The increase in the use of “next effect” with experience may due to users who (1) are exploring different voice modifications; and/or (2) are looking for a particular voice modification. We have anecdotal evidence about this latter scenario, where we found people telling their friends in delivered messages to keep pressing 3 until the voice "clears up", and people complaining that by the time they cycle through all effects to reach the unmodified voice (fourth) the call sometimes disconnects, so they had to be content with the first modification (male-to-female). Also, a common feedback request was to bring the unmodified voice to the first position.

The prevalence of forwarding is more or less stable after the initial 10 calls. Use of “rerecord” and “repeat” shows no major differences from intermediate-term users during the first 10 calls, and continues its gradual decline afterwards. The prevalence of switching to job ads shows the same initial decline as in the intermediate set, but then stabilizes or even begins to rise.

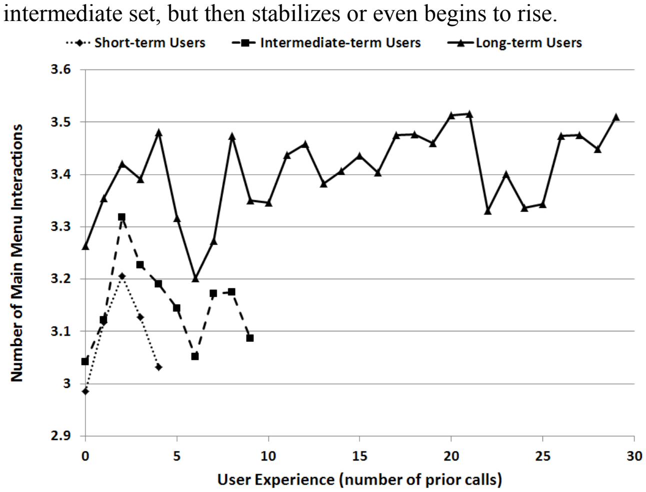


Figure 9: **Average call complexity as function of user experience for short-term (N=2,701), intermediate-term (N=1,862) and long-term users (N=1,523).**

Overall, we find that the most frequently used keys are 2 (forward) and 3 (next effect), and that long term users not only start out using these options more times, but also continue to increase their use of “next effect” over time. In all cases “rerecord” and “repeat” are used more in the beginning (presumably to calibrate to the 15 second limitation, and to share the fun with nearby friends, respectively) and their use declines as user experience increases.

6.3 Early Differences in Call Complexity

By aggregating the different response types in the graphs of Figure 8, we can compare the overall call complexity (number of menu interactions) of the different user sets at similar levels of experience (Figure 9).

We observe that short-term users tend to make the shortest, simplest calls (with fewest main menu interactions) throughout their entire “life span” (first five calls). Similarly, intermediate-term users make significantly simpler calls than long-term users throughout their own “life span” (first 10 calls). Put another way, users who make simpler calls are more likely to stop using the system sooner. Thus call complexity can be useful in predicting user dropout, which can then be acted upon in a variety of ways.

Another striking observation is the persistent pattern of steep rise in call complexity from the first to the third call, followed by an equally steep decline to the sixth call. Since the three user sets are large and mutually exclusive, we believe this represents a real and persistent phenomenon, although we are not sure how to explain it.

7. ACKNOWLEDGEMENTS

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.

8. REFERENCES

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

[2] Sherwani, J., Tongia, R., Rosenfield, R., Ali, N., Memon, Y., Karim, M., & Pappas, G. “HealthLine: Towards speech-based access to health information by semi-literate users Proc.” *Speech in Mobile and Pervasive Environments, Singapore,* 2007.

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

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

[5] Raza A., Milo C., Alster G., Sherwani J., Pervaiz M., Razaq, S. Saif, U. and Rosenfeld R. “Viral entertainment as a vehicle for disseminating speech-based services to low-literate users,” in *International Conference on Information*

and Communication Technologies and Development (ICTD), vol. 2, 2012.

[6] Plauché M. and Nallasamy U. "Speech interfaces for equitable access to information technology," *Information Technologies and International Development, vol. 4, no. 1*, pp. 69–86, 2007.

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

[8] Patel, N., Savani, K., Dave, P., Shah, K., Klemmer, S. and Parikh, T. "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.

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

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

[11] Wyche, S., Smyth, T., Chetty, M., Aoki, P., and Grinter, R. "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.

[12] "Interview with edward baker about the viral factor | entrepreneurial minded." http://entrepreneurialminded.com/business-interviews/ed-baker-viral-factor/.

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

[14] Raza, A., Haq, F., Tariq, Z., Pervaiz, M., Razaq, S., Saif, U. and Rosenfeld, R. "Job Opportunities through Entertainment: Virally Spread Speech-Based Services for Low-Literate Users", in *Proceedings of the 2013 ACM SIGCHI Conference on Human Factors in Computing Systems.* April 27–May 2.

[15] Wyche, S. P., Forte, A., & Yardi Schoenebeck, S. "Hustling online: understanding consolidated facebook use in an informal settlement in Nairobi". In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (pp. 2823–2832), 2013.

[16] Koradia, Z., Aggarwal, P., Seth, A., & Luthra, G. "Gurgaon Idol: A singing competition over Community Radio and IVRS." In *Proceedings of the 3rd ACM Symposium on Computing for Development* (p. 6), 2013.

[17] Bradley, P. S., Mangasarian, O. L., and Street, W. N. "Clustering via Concave Minimization," in *Advances in Neural Information Processing Systems*, vol. 9, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, pp. 368–374, 1997.

[18] Krause E. Taxicab Geometry. *Dover*. ISBN 0-486-25202-7.

[19] Mundt, J. C., Perrine, M. W., Searles, J. S., & Walter, D. An application of interactive voice response (IVR) technology to longitudinal studies of daily behavior. *Behavior Research Methods, Instruments, & Computers*, 27(3), 351–357, 1995.

[20] Jeon, M. H., Na, D. Y., Ahn, J. H., & Hong, J. Y. User segmentation & UI optimization through mobile phone log analysis. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (pp. 495–496), 2008. ACM.

[21] Ozer, M. User segmentation of online music services using fuzzy clustering. *Omega*, 29(2), 193–206, 2001.

[22] Yahoo! Smart Ads, http://advertising.yahoo.com/article/smart-ads.html

[23] Medhi, I., Patnaik, S., Brunskill, E., Gautama, S. N., Thies, W. and Toyama, K. "Designing mobile interfaces for novice and low-literacy users." *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, no. 1, 2011.

[24] Raza A., Pervaiz, M., Milo, C., Razaq, S., Alster, G., Sherwani, J., Saif, U., and Rosenfeld, R. Viral entertainment as a vehicle for disseminating speech-based services to low-literate users. In *Proceedings of the Fifth International Conference on Information and Communication,* 2012 *Technologies and Developmen*t,. ACM, New York, NY, USA, 350-359. DOI=10.1145/2160673.2160715 .

[25] Medhi, I., Sagar, A., and Toyama, K. "Text-free user interfaces for illiterate and semiliterate users," *Information Technologies and International Development*, vol. 4, no. 1, pp. 37–50, 2007.

[26] Lee,, K. Caverlee, J. and Webb S. "Uncovering social spammers: social honeypots+ machine learning." *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval,* 2010.

[27] Ahmed A., Low Y., Aly M., Josifovski, & Smola, A. J. Scalable distributed inference of dynamic user interests for behavioral targeting. In *KDD* (pp. 114–122), 2011.