Real-time Automated Surveys among Low-literate Masses using Voice-based Telephone Services

Agha Ali Raza, Samia Razaq, Amna Raja, Rizwan Naru, Ali Gibran, Abdullah Sabri, Haroon Niaz, Muhammad Bilal Saleem, Umar Saif

Information Technology University, Lahore, Pakistan

ABSTRACT

This paper explores the use of *Interactive Voice Response* (IVR) systems for automatic surveys, data validation and prescreening. We report a deployment aimed at employing voice-based, telephone services to conduct automated, structured interviews of low-literate users and to advertise relevant development-related services to them. Survey calls were placed to 67,000 vocational training recipients to validate their phone numbers and to find out their current job status. Of these, 45,500 answered these calls and 11,500 (25%) responded to the survey questions. Manually conducted follow-up interviews found more than 70% of the survey results to be consistent and also revealed the impact of phone sharing (among family members), call timing, simplicity of interface and surveyor-participant interpretation mismatch regarding certain survey questions on participant involvement and the validity of survey results. The paper discusses the use of IVR to collect information, possible system design considerations and factors affecting the accuracy of gathered information.

CCS Concepts

• **Human-centered computing ~ Natural language interfaces** • **Human-centered computing ~ User interface design** • ***Human-centered computing ~ User studies*** • ***Human-centered computing ~ Accessibility systems and tools.***

General Terms

Human Factors, Languages.

Keywords

HCI4D; ICT4D; automated telephonic survey; information gathering; prescreening; data validation; IVR system; DTMF; speech interface; non-literate; low-literate; cell phone; mobile phone; telephone; job opportunities; employment; skill training.

1. INTRODUCTION

This paper reports an automated, voice-based telephonic survey involving 67,000 individuals in Pakistan. These individuals had received vocational skill training from various organizations and had provided their phone numbers for future contact. The objective of this survey was to find out if the phone numbers were still active and belonged to the same individuals and also to verify their current employment status. This data was required by two organizations (referred to as *Org-A* and *Org-B* for anonymity) that provide job placement services and connect unemployed, skilled workers with potential employers.

ACM DEV '16, November 17-22, 2016, Nairobi, Kenya
ACM 978-1-4503-4649-8/16/11.
http://dx.doi.org/10.1145/3001913.3006626

The reported research is a sub-component of a multi-module project called *Hello Rozgar* (Hello *Livelihood*), aimed to increase skill training and employment opportunities for low-literate, marginalized and remote communities in Pakistan. To this end, we have designed and developed IVR and Web portals for connecting job seekers with potential employers and skill training seekers with trainers, both in the formal and informal sector. Details regarding this portal would be reported in a separate manuscript. A subcomponent of this project was to facilitate other training and placement organizations in Pakistan in their outreach, matchmaking and placement efforts. We have worked with two such organizations to discover the bottlenecks where technology can help scale, speedup or improve the existing processes.

Org-A is a non-profit organization that provides vocational training services in 14 districts of the Punjab province in Pakistan. It also plays an active role in the job placement of its trainees. Org-A routinely follows up with the trainees regarding their employment status and shares the phone numbers of unemployed trainees with potential employers. Their representatives routinely call up trainees and gather the required information from them. In February 2015, Org-A had approximately 42,000 trained job seekers whom they were trying to place with employers. They face a major issue of scale as their placement service is run by a team of 2 to 3 individuals who cannot cope with such a large number of job seekers. Org-A does not have the capacity to hire more staff. It hurts their credibility if employers were to call a particular trainee and find the phone number to be inactive or no longer in use by the same individual or the individual is already employed. This happens often as switching phone numbers and service providers is a common practice in Pakistan where subscribers are constantly on the lookout for better and cheaper airtime/SMS deals and the actual process of switching is quick and straightforward. For Org-A, filtering out such cases is imperative to improving their credibility but their current capacity simply does not allow them to frequently follow-up with all trainees. Org-A had also tried an automated SMS based survey but the response was minimal. A smartphone based data gathering solution also had little chances of success as majority of the trainees are low-literate with low socio-economic backgrounds and mostly use simple or feature phones. A voice-based automated telephonic survey seemed like a good choice to verify if the phone number is active; in use by the same individual and to confirm the employment status.

Unlike Org-A, *Org-B* is a public sector organization that does not provide skill training but primarily works off the web using a dashboard where employers and job seekers can sign up, create profiles and be matched. Their team also routinely pulls information about skilled labor from government organizations as well as online sources and adds it to their pool of trained job seekers. The value addition opportunity for an IVR system was highlighted when we discovered that they do not take on any

responsibility for data validation. As a result, an outdated website could potentially list obsolete contact and employment status details and Org-B would simply gather and broadcast this information. Being a small public sector organization, Org-B does not possess the resources and manpower to perform data validation and were well aware of the problem and actively sought potential solutions. They welcomed the proposition of using an IVR system to validate their data.

2. LITERATURE REVIEW

Lack of scalability and the need for cost, time and human resource investment are some of the major drawbacks of traditional face-to-face surveys. Survey mechanisms not involving direct contact between participants and interviewers (e.g. Computer Assisted Telephone Interviewing (CATI), web-based forms, SMS, IVR systems or even pen-and-pencil) are preferred over face-to-face surveys not just because of the cost and scalability benefits but also because face-to-face interactions may result in more interviewer bias (where a participant is inclined to provide a more socially acceptable answer rather than a candid one) [8]. This is especially true in case of sensitive and personal questions. Social interface theory even suggests that humanizing computer interfaces may also lead to interviewer bias or social desirability bias; however, Tourangeau et al [8] did not find evidence for this hypothesis. They did find evidence of small effects of the gender of the interface on reported gender attitudes. Knapp et al. [3] compare the effectiveness of touch-tone (DTMF) and web-based surveys with traditional pen-and-pencil methods for acquiring personally sensitive information and find no significant differences in the veracity of data across these modalities.

Kreuter et al. [4] compare CATI, IVR and web surveys and find IVR to be mostly intermediate between the other two in terms of increased level of reporting of sensitive personal information and reporting accuracy. Galesic et al. [2] compare several modalities but their study reveals no clear winner for collecting sensitive data and highlights the need for methodologic research. Lerer et al. [5] conducted a study involving 150 teachers in rural Uganda using ODK-voice for data collection. Task success rates between 0-75 percent were achieved for various interface designs. Most of the users could not use the touchtone or touchtone-voice hybrid interface without prior training. Their study shows that users are not comfortable pressing buttons as they have to remove the phone from their ears to do so; and get confused when touch-tone and voice input modalities are mixed and also mistake the IVR system for an actual person. This was partially overcome using text messages prior to survey calls. They suggest recording voice input as opposed to touch-tone. Patnaik et al [6] compare electronic forms, SMS and operator-based telephonic surveys and show that operator based voice surveys outperform SMS and electronic forms by an order of magnitude in developing world context. Sullivan et al [1] employ an innovative technique called *flash voting* where poll participants use missed calls to express their opinion.

3. Methodology and Challenges

3.1 Survey Objectives

Following goals were finalized after consultation with each organization. It was decided that even partially collected information would be useful. Our partner organizations also emphasized that the surveys must be completed in minimum amount of time.

1. Is the phone number still valid / active?
2. Is the number still owned by the same individual?
3. Whether the user is employed or not?
4. Advertise *Hello Rozgar* to all survey participants and invite them to create profiles to look for potential openings.

Question 1 is simply decided if/when a user answers the outbound survey call. In case of users who do not answer, we implemented a back-off - retry mechanism that attempted to contact them for up to five times (detailed in section 3.6). There are multiple ways to answer questions 2. Ideally the IVR should play the recorded name of the individual and ask for confirmation. This would have required recording names of 32,302 individuals. However, as our partners needed the data on urgent basis and we were under-resourced in terms of voice artists, we found a workaround and simply asked: “Our records indicate that you or someone you know has recently completed skill training. If this is true, press one. Otherwise, press 2”. Question 3 was asked as a yes/no query. Goal 4 was achieved in a two-fold way:

- All outgoing calls displayed the phone number of the employment portal as caller ID, so that anyone calling back on this number would reach the portal.
- All users were provided brief details of the portal at the end of the survey (along with the phone number) and were encouraged to call and create profiles.

3.2 Major Contributions

The major contributions of this work include:

- Design and development of an IVR-based survey tool that also advertises relevant development related services;
- Analysis and verification of gathered data using manual follow up calls;
- Yield analysis in terms of users who later interacted with the advertised employment portal;
- Identification of challenges associated with automated voice surveys in the context of developing regions.

Following are some of our major challenges:

3.3 Low Yield of Cold Calls

Our survey mechanism was based on “cold calls” i.e. the recipients were not familiar with our phone number and were not expecting a call from us. Cold calling usually suffers from very low yield (less than 30% answered calls in case of [7]) as recipients often ignore these calls due to unknown caller IDs or suspect scams. This situation is especially true in regions with a lot of voice spam (telemarketing, publicity, scams etc.). Even when calls are answered, users are unlikely to provide real information as they do not trust the caller. Raza et al. [7] tried to overcome this hurdle by playing the recorded name of the recipient very early on in the interaction which makes the recipient comfortable and curious and gains some attention and trust. Unfortunately, in our case, there were 32,302 unique names (4,496 unique first names) in the database and we did not have the capacity to get these recorded quickly. Another strategy to attract user attention could have been to name the training organization early on in the interaction but we were not allowed to do so as it was not negotiated as part of the contract with our project funders.

As a result, we resorted to identifying ourselves with our employment portal. In addition, we chose a structured caller ID with repeated digits hoping that users may be less likely to ignore calls from “official-looking” numbers. This caller ID belonged to

the employment portal and any users calling back to verify the identity of the caller would end up reaching the portal. With these measures in place 68% of our cold-calls were answered.

3.4 Low-literate/Non Tech-savvy Users

The participants of these surveys were expected to be primarily low-literate and less familiar with technology. Presuming no familiarity with IVR systems and prior experience with browsing menus, we designed simple menus with a handful of input options and ample help in case of invalid key-presses or no input.

3.5 Data Cleaning

The phone numbers provided by both organizations required cleaning. Common issues included variations in phone number formats (non-uniform use of area code, country code, spaces, hyphens etc.), duplicates and invalid phone numbers. These instances were cleaned semi-automatically and we were left with 41,513 (out of 42,533) phone numbers for Org-A and 25,150 (out of 31,314) phone numbers for Org-B.

3.6 Temporarily Unresponsive Numbers

The first step of prescreening required us to verify whether a phone number is active or not. To separate a temporarily unresponsive number (busy, network congestion, turned off, out-of-signal range etc.) from an inactive (no longer in use) number, we devised a back off strategy where each number could be called for up to 5 times in case initial attempts are not answered. First retry is made 3 minutes after the original attempt; second 10 minutes, third 30 minutes and fourth 8 hours later. All calls are performed between 9:00am and 8:00pm only (around the week) and any retries falling beyond these hours are delayed until the next day.

3.7 Other Challenges

We did not have any way of automatically verifying the information gathered through the automated surveys and had to resort to manual follow up calls with a random sample of users to verify the provided information. And even that only captures the fraction of users who provide the same information to the human surveyor that they did to the automated service. True validation of the provided information is a bigger challenge and out of the scope of the current study.

Another open problem is identity verification: the phone number may now be owned by a new user or the number may be shared among multiple users and hence the person receiving the survey call may not be the intended recipient. Phone sharing is a common phenomenon in developing countries where sometimes the whole family uses the same phone number.

4. SETUP AND INTERFACE

4.1 IVR Setup

We used *Asterisk* as our IVR platform. Our servers were hosted in the datacenter of a local telco and our telephony bandwidth could support up to 90 concurrent calls.

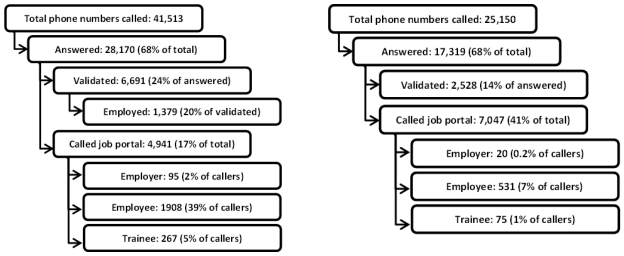
4.2 User Interface

All survey calls were outbound (i.e. from the IVR system to the survey call recipients) and hence free for the users. Even if users call back, their call is instantly disconnected without answering and they are called back within a few minutes by the employment portal (aka "missed call" mechanism). We chose Urdu as the interface language as it is the most widely understood language in Pakistan. A male voice artist recorded all prompts using a serious but friendly tone. The prompts were designed to be simple and informal yet polite and respectful. We refrained from using any difficult-to-understand and complex terms.

As soon as users answer the survey call, they are greeted and asked if they or someone they know recently completed vocational skill training. Users are directed to press 1 for yes or 2 for no. Following either response, users are asked whether they are currently employed. Interaction concludes by informing users about the employment and training portal and encouraging them to call and register. For phone numbers provided by Org-B, the employment-related question is not asked.

4.3 Automated Surveys

For the initial round of surveys Org-A provided us 42,533 records. As discussed, after cleaning we were left with 41,513 phone numbers. Annotation of a random sample of 1,000 records for gender (by looking at the names) showed 99% male and 1% female names. Between November 12 and 17, 2015 a total of 145,529 calls were placed to these numbers (including failed calls and retries) and the results are shown in Figure 1-a. Out of the 28,170 answered calls, only 6,691 pressed 1 to say that they have received skill training. Another 1,818 pressed 2 to indicate that they have not received skill training. The remaining 19,661 refrained from pressing any keys during the call. Out of the 6,691 validated trained users, 1,379 indicated that they are employed while 4,648 indicated that they are not, while 664 did not press any key. 17% of the contacted numbers called back the employment portal and several also created profiles as shown in Figure 1-a.



a. Org-A **b. Org-B**

Figure 1: Results of automated survey calls to trainees

The data provided by Org-B was noisier and after cleaning we were left with 25,150 (out of 31,314) phone numbers. The fraction of female records was found to be much higher (38%) with 62% male records. Between November 12 and 17, 2015, 107,341 calls were generated to these numbers (including failed attempts and retries) and the results are shown in Figure 1-b. Only 2,528 call recipients agreed to have received training, while 741 pressed 2 to indicate that they have not received any training. Remaining users did not press any key during the interaction. However, 41% of the contacted users called the employment portal but comparably fewer created profiles (Figure 1-b).

5. FOLLOW-UP SURVEYS

The objective of operator-based follow-up interviews was to validate the data gathered through automated surveys. A total of 452 individuals representing both organizations were randomly chosen from five categories based on their responses to the automated calls:

1. Had answered the call but provided no input
2. Had pressed 1 (yes), when asked regarding training

3. Had pressed 2 (no), when asked regarding training
4. Had pressed 1 (yes), when asked regarding employment
5. Had pressed 2 (no), when asked regarding employment

Five operators made the calls and tried to validate the findings of the automated surveys. The outgoing caller ID was same as was used for the survey calls. Participants were asked if they have received any training and whether they are employed. Individuals claiming to be trained or employed were further asked about the nature of their training/employment. When the responses did not match the survey results of the automated system, participants were asked further questions to find out why they responded differently. Table 1 summarizes the survey results.

Table 1: Follow up survey responses

Category	# calls placed	# of calls answered	Validated	
			Yes	No
Pressed 1 (yes), when asked about training	85	61	56	5
Pressed 2 (no), when asked about training	102	60	21	39
Pressed 1 (yes), when asked about employment	54	40	31	9
Pressed 2 (no), when asked about employment	69	44	36	8

Overall, 459 calls were placed out of which 265 were answered. For categories 2-5, 70% of the individuals provided exactly the same information to the operator as they had provided in the automated call. Of the remaining 30%, most (64%) belonged to the category of users who had chosen *No* to indicate that they have not received any training but responded to the operator that they have received training. Of these: 10 claimed that they did not receive any survey call; 9 said that they did not understand the instructions; 5 said that the imparted training was not good. The teachers did not pay attention and they had not received their diploma after a full year of training and also did not get any job, hence what use is such training; 5 said that may be someone else answered the automated calls.

For other inconsistencies, users indicated reasons like: Change of job status since the survey call (2); current job not based on the training that they had received (6); not being able to hear the instructions due to poor network quality (1) or that they do not remember receiving a call. Out of the interviewed users who had not pressed any keys during the interaction (60): 52 claimed to have received no such call or that someone else in their family might have answered the call; 2 remembered that the call caught them at a bad time; 2 did not understand the instructions; 2 were afraid that they will be asked to pay for the training while one thought that it was a prank call.

6. DISCUSSION AND FUTURE WORK

Our automated surveys achieved the goal of filtering out invalid/inactive phone numbers, and performed satisfactorily in finding true positives and bringing users to the employment portal. But the overall performance was adversely influenced by false negatives and a large fraction of users failing to provide any input. Following are some of the common issues that were identified through follow-up surveys (section 5).

The call recipients were often not the individuals listed as owners of the phone number, but someone from their family or friends. In this case, their response to automated calls was mostly to answer the call but not to provide any survey input. Our next step is to devise a mechanism to record most frequently occurring first names and then play those names during the interaction to confirm identity. User feedback also suggests that our first question regarding whether or not they received any training was confusing and users were not sure to what end were they being asked this question. Playing recorded names and identifying ourselves with the training organization may help establish trust. It is also interesting to note that some people responded to the spirit rather than form of the survey questions. For example, a person dissatisfied with the imparted training and its outcomes would deny receiving training at all. Individuals with employments not related to the imparted training would answer that they are unemployed. Such ambiguous connotations of the questions could potentially be resolved by stating the question more clearly and also by allowing users to record voice feedback. Using this, we expect to gather user opinions regarding interface, possible unresolved issues and hints regarding mismatches between interpretation of survey questions between surveyors and participants. The feedback recordings would need to be manually annotated and aggregated.

7. ACKNOWLEDGEMENTS

This work was funded under the FIT programme – a multi-lateral initiative co-funded by the European Union (EU), the Embassy of the Kingdom of the Netherlands and the Federal Republic of Germany – and managed by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). We are grateful to Dr. Susanna Adam, Constance Heinig- Augustini, Rubina Leslie, Waqas Haleem, Tahir Ashraf, Abdullah Arshad, Usman Ali, Tanzeel Ur Rahman, Muzammil Hussain, Suleman Khalid, Sohail Ahmad, Maryam Aslam, Adil Rana, Sher Afghan, Ramiz Javed, Mukhtar Ahmad, Faraz Ahmad, Zain Tariq and Shan Randhawa for their hard work and continued support.

8. REFERENCES

[1] Flashvoting: http://www.farmradio.org/ourblog/2013/06/10/-farm-radio-hangs-out-with-gates-foundation/. Acc: Jul 01, 2016.

[2] Galesic, M., Tourangeau, R., and Couper, M. P. Complementing random-digit-dial telephone surveys with other approaches to collecting sensitive data. *American journal of preventive medicine 31*, 5 (2006), 437–443.

[3] Knapp, H., and Kirk, S. A. Using pencil and paper, internet and touch-tone phones for self-administered surveys: does methodology matter? *Computers in Human Behavior 19*, 1 (2003), 117–134.

[4] Kreuter, F., Presser, S., and Tourangeau, R. Social desirability bias in cati, ivr, and web surveys the effects of mode and question sensitivity. *Public Opinion Quarterly 72*, 5 (2008), 847–865.

[5] Lerer, A., Ward, M., and Amarasinghe, S. Evaluation of IVR data collection UIs for untrained rural users. In *Proceedings of the First ACM Symposium on Computing for Development* (2010), ACM, p. 2.

[6] Patnaik, S., Brunskill, E., and Thies, W. Evaluating the accuracy of data collection on mobile phones: A study of forms, sms, and voice. In *ICTD* 2009, IEEE, pp. 74–84.

[7] Raza, A. A., Kulshreshtha, R., Gella, S., Blagsvedt, S., Chandrasekaran, M., Raj, B., and Rosenfeld, R. Viral spread via entertainment and voice-messaging among telephone users in india. In *ICTD* - 2016, ACM, p. 1.

[8] Tourangeau, R., Couper, M. P., and Steiger, D. M. Humanizing self-administered surveys: Experiments on social presence in web and ivr surveys. Computers in Human Behavior 19, 1 (2003), 1–24.