A Comparative study of SMS and IVR Interfaces for Crowdsourcing Water Availability Information

Thejo Kote
School of Information
University of California, Berkeley
Berkeley, CA 94720, USA
thejo@berkeley.edu

Shaon Barman
Department of Computer Science
University of California, Berkeley
Berkeley, CA 94720, USA
sbarman@eecs.berkeley.edu

Abstract
In this paper we present a study comparing SMS and dialed input voice user interfaces for crowdsourcing water availability information. The crowdsourcing process involves registration in the system, sending notifications of water availability and verification of accuracy. We have tested a working prototype of the system in Hubli, a city in southern India. We ran a controlled within subjects experiment with 19 people to test usability, task completion time and error rates for the SMS and IVR systems. In addition, we ran a between subjects experiment with 179 residents to test verification rates of the two systems.

Keywords
ICTD, SMS, IVR, DTMF, water availability

ACM Classification Keywords
H.5.2 User Interfaces: Voice I/O User Interfaces; H.5.2 User Interfaces: Evaluation; H.1.2 User/Machine Systems: Human Factors

General Terms
Design, Human Factors
Background and Introduction

Water availability is valuable information and not having it has a number of social implications [1]. In much of the developing world, piped water is available for only a few hours once every 4 to 8 days. Water can arrive at a different time each week, and the number of days between deliveries may also vary. Each city is broken up into sub-units called valve areas. Water is delivered in a rotational basis to each of these areas. An announcement listing the valve areas that will receive water is sometimes printed in the newspaper. But in many parts of the city, delivery is too erratic for these often out-of-date announcements to be of use. Households lose hours each week waiting for water, are stressed by water scarcity, and regularly need to substitute with unsafe sources of water. Hubli-Dharwad is one of over 400 cities in India, each with populations over 100,000, which face similar conditions of unreliable piped water supply.

NextDrop is an information service that generates and distributes real-time water timing updates and water delivery predictions to these households using SMS and interactive voice response (IVR) technology. NextDrop is presently in pilot mode, providing water timing information to around 200 households in Hubli-Dharwad across 7 valve areas that receive water at different times. Since the goal of the pilot is to study the efficacy of the model and to determine the best way to implement it, the families were recruited through a door to door sign up process.

NextDrop generates information through crowd-sourcing. Water board employees who operate the valves and residents in the valve areas are incentivized to immediately update our system when the water begins to flow. The validity of these updates is verified with inputs from other users in the same area, and a near real-time update is generated and shared with all households in the area.

We started our pilot by deploying a purely SMS based system to receive information from end users and to verify those updates. We noticed usability issues and tried a Wizard of Oz voice based system. This motivated us to build a fully automated IVR based application to perform the same tasks in the crowd sourcing system which we have not yet deployed in the field. In this paper, we will present the results of a study comparing the two systems.

We found that the tasks we defined were completed sooner using the IVR system, with more end users preferring it. We also found that IVR works better in verifying crowd sourced information in time sensitive applications. We also discuss our learning from conducting the test and provide hypotheses for our observations.

Related Work

Prior research in this area falls into two broad categories: research into the efficacy of IVR systems in various applications and crowdsourcing of information.

IVR on Mobile Systems

Research into IVR based systems has mainly focused on agriculture [5] or health related applications [7]. There has also been research into the usability aspects of IVR systems [3]. Specifically, researchers have compared speech based vs. touch tone based systems [4, 8], and the optimal design of menus in IVR systems.
**Crowdsourcing**
There has been a lot of research in the field of crowdsourcing information. One area of work that is relevant to our work is the concept of "participatory sensing". Researchers have tried to determine noise and air pollution levels using sensors on phones [2, 6]. But, many types of information cannot be generated through sensors either because they are too expensive, or because it needs human judgment. Our work attempts to create such a platform. Another challenge in any crowdsourcing application is the verification of the information. The work von Ahn has done [9] provides ways to address this.

We are not aware of prior work which has compared the performance of SMS and IVR systems in crowdsourcing of information.

**Prototype**
The NextDrop system works in three stages: notify, verify, announce. Each stage can be accomplished through either an IVR or SMS interface. During the notify stage, residents tell the system that water is flowing at their home. To verify the notification, other users in the same valve area are contacted and asked if their water is flowing. Currently, we randomly choose two other residents in the valve area. If we receive enough positive verifications to meet a threshold, then we progress to the announce stage where all the users in the valve area are contacted with the news. Otherwise, the resident who sent the original notification is informed of the unsuccessful verification.

The system knows the phone number and valve area for each family registered in the system, so the user does not need to provide this information during the notification step.

We encourage residents to use the system by providing monetary incentives for users who send in notifications.
and verifications. These incentives are used to offset the cost of outgoing calls/SMSs (incoming calls/SMSs are free). We are also in the process of testing other forms of incentives like recognition in the community.

We started with an SMS version but have now implemented an IVR version of NextDrop. Figure 3 shows the general information flow for both versions. Because many mobile phones do not have font support for the local language, Kannada, the SMS version uses English prompts and responses. For the IVR version, we recorded prompts in Kannada using a female voice.

NextDrop was built using the Django web framework and uses KooKoo, a service which enables integration of voice functionality through HTTP APIs.

**Deployment**
Currently, a non-automated Wizard of Oz system is deployed. Residents call an operator to notify water arrival. The operator then manually verifies the notification by calling two other residents. While this method has seen success, it is not scalable and cost-effective in the long-term. Our goal is to determine whether an SMS or IVR system works better. We have also partnered with the water board in Hubli to get accurate information about when water is available. The valvemen who control when water is turned on can call into the system to notify availability. Because the valvemen have exact information, we announce all notifications from them without going through the verify stage. We are not sure if this is the right design at this point and plan to study a valveman’s incentive to game the system in the future.

**Experiment**
Because the water for each valve area is only turned on once a week, we could not run an end-to-end test comparing the IVR and SMS implementations. Instead, we chose to focus on specific tasks of the system.

**Subscription/Notification**
We conducted a within-subjects controlled experiment to see how people use the SMS and IVR systems. We required each user to be a current subscriber in the NextDrop system. In addition, we further restricted our experiments to those users who regularly used SMS and had reading and writing comprehension of English, so the test wouldn't be biased.

Before the experiment, we explained what NextDrop is, and what the terms “subscribe”, “notification”, and “unsubscribe” are. Even though the user had already signed up to use NextDrop, it was likely that they would not be familiar with all the elements of the system and we found it useful to explain the system again so they could have a model of how the system worked.

We then explained we had two systems – one SMS and the other voice based – and were testing to see which one was easier to use. We stressed that there was no right answer and that they were not being graded. Since the participants were using their own phone, we gave them Rs. 20 (approx. $0.5). This would compensate the costs of sending SMS and making phone calls and for their time in helping us with the evaluation.

After explaining the system, we gave them 3 tasks to complete: subscribe to the NextDrop service, notify that water is flowing, unsubscribe from the service.
Each series of tasks was to be completed once on the IVR system and once of the SMS system. The order of systems was randomized to reduced affects from priming.

For each task, we measured the time to completion, the number of errors, and whether the user was able to successfully complete the task. If the user sent >3 SMSs or made >3 calls, we consider that task to be unsuccessful and ask the participant to move on the next task.

In addition to objectively measuring time and error rates, we asked each participant to fill out a brief survey. The survey asked each user which system they liked better and reasons for preferring that system and collected basic demographic information.

**Verification**
We conducted a between-subjects experiment to test the verification rate of SMS and IVR systems. We randomly divided the 179 families into 2 groups, one for SMS and IVR, and further broke each group into 2 subgroups, one for morning and afternoon. For each group, we used the system assigned to them to ask them to verify water availability during the time slot assigned to them. We collected data on how many people responded, what their response was and how many errors they made when using the system. We conducted the experiment when no water was available in the selected valve areas, so all responses should have been negative. In order to suppress other factors, we conducted this experiment on the same day for all groups.

<table>
<thead>
<tr>
<th></th>
<th>Subscribe</th>
<th>Notify</th>
<th>Unsubscribe</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SMS</strong></td>
<td>79.31 / 70</td>
<td>66.63 / 58</td>
<td>58.42 / 59</td>
</tr>
<tr>
<td><strong>IVR</strong></td>
<td>64.63 / 61</td>
<td>32.47 / 30</td>
<td>29.52 / 27</td>
</tr>
</tbody>
</table>

*Figure 1: Performance of SMS and IVR systems (Mean / Median in seconds)*

**Results**

**Subscription/Notification**
We conducted the within subjects test with 19 participants (14 male and 5 female). Figure 2 contains data about the task completion times of SMS and IVR systems in the experiment. The IVR system consistently outperformed SMS in all three tasks. With the alternate hypothesis that the mean time for completion of tasks is lower for the IVR system, we found that the results are highly significant (p < 0.01) for the notification and unsubscribe tasks, but not significant for the subscription task. This was expected since the complexity of subscribing to the service is about the same in voice and SMS. In the case of SMS, there is a cost of remembering the valve area ID and typing the message, and with IVR, we ask the user to verbally record their address and zip code, so that an operator can listen to it and add the user to the correct valve area.

We recorded 4 errors in a total of 57 SMS tasks. All of them were related to misspelling of the SMS keywords that had to be used. No errors were recorded in the use of the IVR system. Of the 19 participants, 13 preferred the IVR system, 5 preferred the SMS system and one said she was equally comfortable with both.
Verification
We ran the between subjects experiment on a Wednesday, at 10 AM for the morning group and at 3 PM for the afternoon group. We did not see any significant differences in results between the two groups. Figure 1 shows the result of the test. Users in the No response category either did not pick up the phone call or did not send an SMS back. Users in the Incomplete category for IVR picked up the call, but could not successfully complete the verification. For the SMS users, they returned the SMS but did not enter a message that was recognized.

In IVR calls, of a total of 90 calls, we did not receive a response in 48 cases, 42 people responded, of which 26 hung up immediately, 9 provided a valid response and 7 provided an invalid response.

In SMS verification requests, we sent a total 89 messages and received 2 responses, one of them not to the number we had mentioned in the SMS, but to the mobile phone of the operator we use for the wizard of Oz voice system. We provide a hypothesis for this very low response rate in the discussion section.

Discussion
Subscription/Notification
Most of our sample preferred the IVR system over SMS. The common reasons given were that it was easier to understand what they had to do in the voice based system and that it took time to type and send an SMS while remembering the keyword to be used, especially in English. Having the voice based system in the local language helped. Participants also noted that they liked the immediate feedback to an action in the voice based system, while they had to wait for a response in the SMS system. Some participants preferred the SMS system because they felt it is more convenient and faster, especially if there were in an environment where they did not want to disturb others. The interesting thing to note here is that the mean and median age of the group that preferred SMS was higher than that of the group that preferred voice.

We also found that 10 of the 19 participants had phones with support to send and receive local language SMS, 7 did not and 2 were not aware. While this is an improvement over the last few years, the support for local language SMS on phones and telecom networks is still spotty. The fact that it is not possible to determine support for local languages purely via SMS makes it harder to build robust systems that support them.

Verification
Overall, more users in the IVR system successfully responded to the call. We believe there are several possible explanations for this. One is the reliability of the services we are using affected the results. When conducting the experiment, we sent a verification message to ourselves using the SMS system. When we tried responding back, the message was not recorded. We hypothesize that the underlying SMS system may have resulted in lost verifications. Additionally, because of spam SMS, many people have become desensitized to messages they receive from numbers they are not familiar with. This problem has become especially acute over the last 18 months in India where around 15 million connections are being added every month, providing a huge incentive for telemarketers and spammers to send unsolicited messages.
While more users picked up the IVR call from NextDrop, many users immediately hung up without actually verifying water availability. We hypothesize that this is due to lack of familiarity with NextDrop. Even though all users voluntarily signed up with the system, we believe that many users forgot about how the system works and the incentive structure. One way of addressing this problem is to work towards a closer partnership with the city’s water board. This would give NextDrop both authority and publicity.

53.33% of users did not answer our call. Our hypothesis is that a major portion of these users were either unable to answer the call or unwilling to answer a call from an unknown number. But, the other reason is that since mobile number portability is still not available to subscribers in India, a non-trivial number of phone numbers get de-activated or reassigned regularly as people move between operators looking for the best deal in the world’s most competitive telecom market. Since the 179 numbers in our database were collected over 4 months before the experiment, at least some numbers would have been invalid.

In the case of the IVR system, we also found that more users indicated that water was flowing in their homes even though it wasn’t. We think this is because of the way the voice message was phrased. It translated to “is water flowing in your home?”, which for someone without the right context about NextDrop and our service may seem like an odd question if they have storage facilities at home. We did not catch this during our usability tests prior to conducting the experiment and plan to make it more explicit that we’re referring to supply from the water board in the future.

**Conclusion**

In this paper, we presented a comparative study of SMS and IVR implementations of a system to crowd source water availability information. We tested each part of the system to test user preferences and viability of the system as a whole. We found that IVR outperforms SMS in task completion times in the way we have defined tasks and that people prefer IVR systems to SMS. We also found that for verification in crowd sourcing systems, IVR is a better choice, especially in time sensitive applications. Regardless of the medium, recognition and familiarity with the system is very important to obtain good response rates. In the future, we plan to test the systems in a real world setting incorporating our learning from this exercise.

**Acknowledgements**

This work was supported by the UC Berkeley School of Information, the Centre for Information Technology Research in the Interest of Society at UC Berkeley and Clinton Global Initiatives. We thank Emily Kumpel and Madhusudhan B for their help in conducting the tests. We’d also like to thank Tapan Parikh and Bjoern Hartmann for their guidance.

**References**


