IVR Junction: Building Scalable and Distributed Voice Forums in the Developing World

Aditya Vashistha
Microsoft Research India

William Thies
Microsoft Research India

Abstract

Interactive Voice Response (IVR) systems play an important role in collecting and disseminating information in developing regions. Recently, researchers have used IVR technology to build *voice forums*, in which callers leave messages that can be heard over the Internet and over the phone. However, despite their appeal, voice forums remain difficult to set up, and difficult to scale due to the overhead of moderating content and the cost of phone calls.

This paper discusses the challenges and opportunities in creating scalable voice forums. We also present a new open-source system, IVR Junction, that leverages existing free services and commercial tools to simplify the process of creating a voice forum. IVR Junction utilizes familiar cloud-based services to provide free content hosting and moderation, as well as a novel mechanism for automatically synchronizing content across geographically-dispersed offices, thereby enabling local access points with decreased calling costs.

1 Introduction

Most of the world's 3.6 billion mobile subscribers [3] live in the developing world, and they use their phones primarily for voice calls. As researchers seek to enable this population to access, report, and share information using the phone, the HCI and ICT4D communities have witnessed a surge of interest in Interactive Voice Response (IVR) systems. Recent voice-based services have spanned diverse domains, including citizen news journalism [13], agricultural discussion forums [15], community dialogue [1], user-generated maps [10], access to health information [19], outreach to sex workers [18], group messaging [14], feedback on school meals [5, 6], support for community radio stations [8, 9], and a viral entertainment platform [17]. Due to the success of these projects, there are numerous researchers, companies, and non-profit organizations (personally, we are in touch with over a dozen) that are actively seeking to establish their own phone-based voice applications in new domains.

However, despite the enthusiasm surrounding IVR systems, the unfortunate reality is that it remains quite complex to install and configure interactive voice forums. Systems that offer this functionality, such as CGNet Swara [13], Avaaj Otalo [15], and Phone Peti [9], utilize open-source platforms such as Asterisk or FreeSWITCH for the telephony interface, and require hosting a Web server to connect with moderators and an Internet audience. Though tractable for technology researchers, this process requires expertise that is usually beyond the reach of non-profit organizations and NGOs[1].

A second challenge that researchers and practitioners are facing in the IVR domain is how to scale their voice applications beyond the pilot stage. Though pilot deployments have shown that IVR applications are creating envisioned impact [13, 15, 18], their subsequent reach is often limited due to challenges in moderating content and managing operating costs at scale.

In this paper, we outline some new approaches for creating scalable voice forums. We also instantiate some of our ideas in a new open-source system, called IVR Junction, that integrates commercial tools (Voxeo Prophecy) and freely-available services (YouTube, SkyDrive) to enable rapid construction of interactive voice forums. Compared to prior solutions, IVR Junction offers two key novelties. First, it seamlessly connects Internet-based users with phone-based users. Both sets of users can contribute and retrieve audio messages from a repository that is hosted in the cloud. This enables rural users, whose access is limited to a phone, to gain an international audience for their recordings on the Internet. Likewise, Internet users can post audio recordings for automatic broadcast to mobile phones. All content can be reviewed online by a moderator prior to dissemination.

[1] In this paper, we use the term NGO (Non-Governmental Organization) to encompass non-profit organizations in developing regions.

The second novelty of IVR Junction is that it scales across geographically distributed access points, enabling affordable access via local phone calls. Prior tools for hosting one's own IVR server are centralized, assuming a single computer and telephony interface. In contrast, our system functions as a hub-and-spoke model in which audio recordings are stored in the cloud and telephony portals can be seamlessly added to connect to that repository from any location. Recordings that are contributed from one calling circle are automatically replicated to all other locations via the Internet. By using the Internet backbone to share audio content across different districts, states, or countries, we save users the cost of a long-distance phone call that is typically required today.

In addition to these novelties, we envision that the primary value of IVR Junction comes in its practical assembly of robust, off-the-shelf components into a coherent and usable whole. We are preparing IVR Junction for a public release that will be available at http://research.microsoft.com/ivrjunction/.

2 Related Work

While there exist several other toolkits for building IVR systems, none of the open-source platforms available to the research community offer distributed and scalable operations, catering to both local callers and a global audience on the Internet. ODK Voice [7] is a flexible platform that, like IVR Junction, runs on top of any Voice XML interpreter; however, as of yet it has not supported distributed access points. Freedom Fone [4] and Awaaz.De [16] are built on FreeSWITCH, and integrate an IVR system with an Internet site for viewing audio recordings. However, the audio hosting does not currently utilize free services in the cloud. The hosting could be done on an organization's own Web server, or using Awaaz.De's paid hosting service. Several projects use Asterisk to construct an IVR system and also make recordings available over the Web [9, 13], though they currently require setting up one's own Web server. Moreover, none of these projects support distributed access between synchronized servers.

The IBM Spoken Web project proposes a "World Wide Telecom Browser" that acts as a single access point as the user browses content hosted on separate servers [2]. As the hyperlinked voice services remain distributed, this solution could incur long-distance charges between the browser and the remote services. In contrast, IVR Junction pushes remote content to each local node.

In India, cloud telephony systems such as KooKoo and Exotel are accessible through centralized phone numbers. As far as we know, they do not yet synchronize content across distributed call centers.

3 Challenges and Opportunities in Creating Scalable Voice Forums

An increasing number of voice services for developing regions are focused not only on information dissemination, but also on information production. Analogous to the rise of Web 2.0 for Internet-enabled users, this new generation of voice services enables communities to create, share, and consume audio content using low-end mobile phones. We refer to such services as *voice forums*.

By way of example, CGNet Swara provides a voice forum for citizen news journalism [13]. Callers can either record a message, or listen to messages recorded by others. A trained moderator reviews submissions via the Internet and approves users' posts before they go live. Approved recordings are also available for listening via a website, enabling an international audience to access them. Since its launch in 2010, CGNet Swara has had a documented impact in addressing real issues in rural communities [13].

The success enjoyed by CGNet Swara in the domain of citizen news journalism is mirrored by similar voice services in several other domains, including Avaaj Otalo for agriculture [15], VoiKiosk for community dialogue [1], and Phone Peti for community radio [9]. However, to date none of these projects has scaled beyond the pilot stage. CGNet Swara has received about 1,500 posts and 100,000 phone calls in the course of a few years. At this volume, it remains feasible to enlist a dedicated employee to review and moderate the posts, and to provide a toll-free number for callers to access the system. But what does it take to scale this system across an entire country, where there could be millions of posts and hundreds of millions of phone calls? We believe there are two fundamental issues in scaling up voice forums: moderation of content and cost of calls.

3.1 Moderating Content at Scale

We foresee two basic solutions for monitoring content at scale. The first solution is to hire a large fleet of dedicated content moderators. While this might seem exorbitantly expensive in a Western context, in contexts such as India there are many services (such as Justdial) that run very large call centers at affordable costs. The drawback of this approach is that it is difficult to maintain consistent judgment, quality, and accountability across all of the moderators.

The second solution is to utilize community moderation. Web-based news forums such as reddit, Slashdot, and Digg have shown that users of a site can effectively rank and patrol the content, ensuring that it remains relevant and appropriate according to the interests and values of the community. This paradigm extends beyond news to Q&A systems such as StackOverflow, Quora, and Yahoo Answers. Community moderation on the Web can be extended to a voice service where users submit entries and also up-vote or down-vote other entries. Community moderators can flag the content which seems inappropriate to them. Once a threshold number of flags is reached, the flagged content can be automatically reported to a dedicated moderator who decides its fate. If the content is found to be inappropriate by the dedicated moderator, the caller's number can be blacklisted for some time: an operation that is perhaps easier to do on a phone-based service than on the Web, due to the difficulty of obtaining a new phone number.

Though community moderation is promising, it will be interesting to understand the motivation and incentives for the members of the community to moderate the content. In Avaaj Otalo, farmers started self-moderating the forum in a different way. Because they didn't have the ability to approve or reject posts (the partner NGO had this authority), farmers posted additional recordings that reprimanded those who deviated from the norm or posted irrelevant questions [15]. The success of community moderation at scale may depend on the application domain. We believe that community moderation is a good fit for the entertainment domain, e.g., where callers (say college students) are listening to and evaluating community-generated content like jokes, songs etc. But would this be true for socially relevant voice applications like CGNet Swara and Avaaj Otalo, in which the most important posts can also be the least entertaining?

It remains time-consuming to listen to posts on the phone, demanding more of the moderators than Internet users who can read (or skim) a post. To address this issue, it may be valuable to consider a hybrid voice/text moderation system, where posts are first transcribed into text by a crowd of distributed Internet users, and then moderated in textual form by stakeholders who are vested in the content. Also, a crowd of distributed Internet users can be used to translate the posts to other languages to boost global engagement.

One challenge in community moderation of voice content is preserving the privacy of contributors. For instance, it might be useful to protect the identity of whistleblowers that are using a voice forum for reporting wrongdoings. Similarly, a Q&A forum for sex workers to ask questions related to sexually transmitted diseases must preserve the privacy of callers. One opportunity could be to preserve the privacy of callers by masking their voices, i.e., by utilizing a voice anonymizer to distort the recording enough to prevent recognition of speakers.

3.2 Managing Call Costs at Scale

As a system scales to encompass millions of users, it becomes challenging or impossible to provide toll-free access numbers (or free return of missed calls) while maintaining a sustainable business model. While callers could potentially bear the cost of the calls themselves, this would likely make many services unaffordable for low-income users.

In order to reduce or eliminate the cost of calls, we are exploring three ideas. First, call charges can be reduced by leveraging a local call. While this may seem obvious, it is rarely implemented in practice because most voice services are operated out of a central location that is necessarily distant from some users. Local calls can be enabled via a distributed service that automatically distributes and caches content (via the Internet) at lightweight nodes in each calling circle. We have implemented this idea in IVR Junction.

The second option is to deliver audio via a cheaper channel than a voice call. In particular, with the growing use of data services, it could be much more affordable to broadcast audio via mobile Internet, once users have data-enabled handsets. In addition, an application running locally on the handset could cache content to prevent paying extra for a replay. In the case of CGNet Swara, there is anecdotal evidence that a significant fraction of calls from a given number are replaying the same content, as community activists replay the posts for different audiences in a village. A mobile application could also provide users with valuable metadata on each post, such as its language, subject, and place of origin, so that users could avoid downloading content that they are not interested in. Of course, this approach requires the target users to have access and familiarity with a feature phone or smart phone.

The final solution is to avoid all electronic transfers to a central location, and instead rely on offline, peer-to-peer connectivity to disseminate the audio content. There is already a rich ecosystem of peer-to-peer media sharing amongst low-income users in urban India [20]; they use Bluetooth and SD cards to transfer movies between handsets. However, this solution is valid only for those voice applications whose users are motivated to share the content with their peers, have Bluetooth enabled cell phones, and knowledge to send files via Bluetooth. We believe that this ecosystem could be strengthened and amplified by dedicated applications on the phone that make it easier to discover and transfer content amongst peers. Such application could also seed the offline content by recording audio from a live call, or downloading it via a data connection.

4 IVR Junction

IVR Junction is a new open-source system (still to be released) that combines other free and commercial tools into an integrated platform for building scalable voice forums. The system has two main goals. The first goal is to reduce the cost of calls by utilizing distributed access points that are automatically synchronized via the cloud. The second goal is to simplify the creation of voice forums, by leveraging a Windows-based installer for each client machine and familiar Internet services for hosting and moderation. While the architecture could be extended to support community moderation (as discussed in Section 3.1), this is not our focus at the current time.

4.1 System Architecture

Figure 1 depicts the overall architecture of our system. To set up IVR Junction, an organization needs to purchase three components: a laptop (or desktop) computer, a GSM (or landline) modem, and a VoiceXML interpreter with SIP support[2]. These components form a telephony access point which services phone calls from users. In addition to the telephony interface, the organization has to establish an account with a free cloud-based audio hosting service in order to provide storage and moderation via the Internet.

Some NGOs have geographically distributed branches. Even if these branches are in different regions, it may be desirable for them to share a single audio repository, as users (such as farmers) often speak the same local language, share similar local constraints and geographical features, and have comparable standards of living. However, sharing content via a single server can be expensive, because some users may need to dial a long-distance number if they are located in a different state or calling area. This expense could discourage people from using the service. IVR Junction overcomes this problem by utilizing distributed servers which can be set up in each calling area. If desired, local audio contributions from each region are automatically synchronized with an audio repository in the cloud, thereby facilitating increased knowledge sharing.

The overall functionality of the system can be best understood via a usage scenario. Consider a Q&A forum, in which callers can record audio questions as well respond to the questions that others have posted. A user interacts with the system by calling a local number connected to IVR Junction - say, in Jaipur, India (see Figure 1). The call is processed locally via the GSM modem and laptop, using a local audio repository that was previously synchronized with the cloud. When the call finishes, IVR Junction uploads the new recording (a new question) to the cloud. At this point, the question awaits approval by the moderator, who logs in via the Internet to listen to the question, categorize it, and summarize in textual form (for the benefit of Internet-based users). Following approval by the moderator, the question becomes live on a website, making it accessible to Internet users around the world. Also, at this point in time, the question automatically becomes visible to (i.e., appears in the local audio repository of) other branches of the NGO. For example, IVR Junction in Delhi would detect that a new question is available, download that question into its local cache, and play that question for callers in Delhi. If a caller in Delhi responds to the question, the response is later synchronized with the central server and upon moderator approval, becomes visible to Internet users and IVR users, and thus to the original questioner who is in Jaipur.

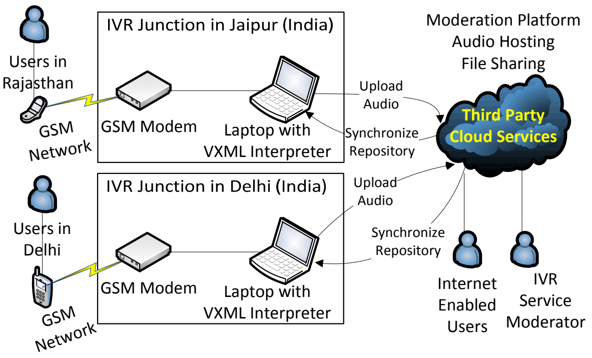

Figure 1: Architecture of IVR Junction. Our software runs on the laptops on the left. We rely on a commercial VoiceXML engine as well as free third-party services for hosting, moderating and sharing audio files in the cloud.

In India, many NGOs have offices and operations teams in small towns which have intermittent availability of electricity and Internet connectivity. IVR Junction is designed to run on a low-end laptop computer. The low cost, ease of deployment and mobility of the laptop enables non-expert users to set up the system by themselves. Also, while the GSM modem requires an electrical outlet, the remainder of IVR Junction is tolerant to intermittent power and Internet outages. Each local repository is synchronized with the cloud in an opportunistic manner, depending on the available connectivity.

The access points of IVR Junction can be scaled up to support many parallel callers. Modems can be added incrementally, with calls forwarded between numbers that are busy. This feature, known as “call hunting”, is commonly available for landlines in India. Mobile numbers can also support call hunting via a supplementary GSM service called “mobile access hunting” [11]. While we have been unable to obtain this service in India, it appears to be available elsewhere (e.g., in Nepal [12]).

[2] While there are free VoiceXML interpreters (e.g., VoiceGlue), they target Linux and would require work to integrate with IVR Junction.

4.2 Implementation

IVR Junction is fully implemented and working, and some deployments are currently in progress. The implementation of the system relies heavily on existing tools and Internet services. The key functionality provided by our software is to synchronize content across different platforms and to present a unified interface to the user. The specific functions we implement are as follows:

1. **Caller interface:** a template IVR call flow that presents users with audio prompts and control logic suitable for an audio discussion forum (700 lines[3] of ASP.NET, VoiceXML, and C#).

2. **Admin interface:** a set of configuration pages that enable non-expert users to set up a new audio forum and customize the sharing of content with audiences at other locations as well as the Internet (1500 lines of C# and ASP.NET).

3. **Synchronization with moderators:** a Windows service that opportunistically uploads new content to a cloud-based moderation system, polls that system for updates from moderators, and makes content available via the phone as soon as it is approved (3100 lines of C# code).

4. **Synchronization with other branches:** a Windows service that interfaces with a cloud-based file sharing system to automatically distribute approved posts to other branches (1300 lines of C# code).

One interesting design decision is regarding the interplay of synchronization activities: every recording needs to be reviewed by a moderator, but only approved recordings need to be broadcast to other branches. This implies that there is a tradeoff between bandwidth and latency in distributing posts between branches. To conserve bandwidth, each node can first wait for a reply from the moderator, and then initiate transfer of approved recordings between branches. To optimize for latency, recordings can be broadcast across all nodes in advance of moderator approval, and then made available more quickly following the moderator's decision. We currently opt for the former solution, which insulates the NGO's bandwidth expenditure from possible unwanted posts. The latter design would be preferable if most posts are guaranteed to be approved.

Third-party components

IVR Junction is agnostic with respect to the third-party tools and services that are used for telephony, moderation, and file sharing. Table 1 illustrates the specific hardware and software components that we utilize in the first version of the system.

IVR Junction Component	Software or Hardware	Cost
Laptop	Any	$500
GSM (or Fixed Line) Modem	Matrix ATA 211G	$100 (1 Port)
Internet	Any	$20/mo.
SIP and VXML Interpreter	Voxeo Prophecy 11	*requires quote*
Application Server	IIS Express	Free
Database Server	SQL Express 2008	Free
Moderation Platform	YouTube Channel	Free
Web Hosting of Content	YouTube Channel	Free
	Facebook Page	Free
Cloud Storage	Dropbox	Free
	SkyDrive	Free

Table 1: Services and tools utilized by IVR Junction.

The only component that could represent a significant expense for the organization is the VoiceXML interpreter. Currently we utilize Voxeo Prophecy, a robust enterprise-grade solution. Pricing is customized for each client, and Voxeo declined to provide us with a quote that we could include in this paper[4].

The other third-party software components are free. We use a YouTube channel as a free Web-based hosting and content moderation platform. Internet users can view the moderated content on YouTube, and also on a linked Facebook page. While YouTube is designed for video content, we found it to be preferable to current offerings (such as Audioboo and SoundCloud) that are tailored specifically for audio files. At the time of this writing, Audioboo has a work-in-progress API with very limited functionalities. Moreover, Audioboo cannot be used as a moderation channel because the audio files cannot be marked as public or private, unlike YouTube. Though the SoundCloud API has features suitable for moderation, and can also be used to upload and download files, only 120 upload minutes (which is equivalent to ~60 audio posts on CGNet Swara) are available for free, and additional space is relatively expensive.

While YouTube offers an API to upload content, it does not allow programmatic download of content. Thus, for replicating content across NGO branches we need to utilize a separate cloud-based service, such as SkyDrive, Dropbox, or Google Drive. While we anticipate that all of these services would work well, our current system uses Dropbox due to the maturity of its API.

[3] All line counts represent non-comment, non-blank lines of code.

[4] A two-port version is free for evaluation purposes, but a paid license is required for production environments.

4.3 Additional Applications

While we have focused our discussion on the scenario in which users are placing calls to the system, as in CGNet Swara and Avaaj Otalo, IVR Junction is equally suitable for making outbound calls to users. In this scenario, an NGO staff member could upload an audio message to the cloud storage unit, along with a list of phone numbers and scheduling constraints for broadcasting the message. The IVR Junction nodes at the branches of the NGO could automatically download the audio and broadcast it to the members registered in their calling area. This would represent a phone broadcasting system that is as simple and usable as a centralized system, but offers cost savings as distributed branches make calls at the local rate.

Another application of interest is to utilize IVR Junction to share audio content across wide geographical spaces, for example, across different countries. Because different branches are synchronized via the cloud, it becomes possible for communities to contribute to a shared audio forum even if it is very expensive for them to place direct phone calls. One usage scenario that leverages this capability could be to connect migrant workers with people in their home country, both for keeping in touch and for sharing information and advice with workers who are aspiring to relocate in the future.

5 Conclusions

While interactive voice response systems have rich potential to provide innovative information and communication services to mobile subscribers in the developing world, to date it has been difficult for researchers and practitioners to expand systems beyond the pilot stage due to challenges in moderating content and managing call costs at scale. This paper discusses various approaches for creating scalable and distributed voice forums. We also present the design and implementation of IVR Junction: a new and open-source system, built on the Windows platform, that aims to simplify installation and configuration. IVR Junction enables novel interactions in the design of IVR systems, including seamless interplay between Internet users and phone-based users, as well as a distributed architecture for affordable sharing of audio content across different locations.

6 Acknowledgments

We are grateful to Arjun Venkatraman, Shubhranshu Choudhary, and Ed Cutrell, as well as the anonymous reviewers, for helpful comments and feedback.

References

[1] Agarwal, S., Kumar, A., Nanavati, A. A., and Rajput, N. Content Creation and Dissemination by-and-for Users in Rural Areas. In *ICTD* (2009).

[2] Agarwal, S. K., Jain, A., Kumar, A., and Rajput, N. The world wide telecom web browser. In *DEV* (2010).

[3] Bouverot, A. Keynote Address, GSM Association Mobile World Congress, 2012.

[4] Clark, B., and Burrell, B. Freedom Fone: Dial-up Information Service. In *ICTD Demo Session* (2009).

[5] Using the mobile to track midday meal scheme. Economic Times, 2010.

[6] Grover, A., Calteaux, K., and Barnard, E. A voice service for user feedback on school meals. In *DEV* (2012).

[7] Hartung, C., Anokwa, Y., Brunette, W., Lerer, A., Tseng, C., and Borriello, G. Open data kit: Tools to build information services for developing regions. In *ICTD* (2010).

[8] Koradia, Z., Balachandran, C., Dadheech, K., Shivam, M., and Seth, A. Experiences of Deploying and Commercializing a Community Radio Automation System in India. In *DEV* (2012).

[9] Koradia, Z., and Seth, A. PhonePeti: Exploring the Role of an Answering Machine System in Community Radio. In *ICTD* (2012).

[10] Kumar, A., Chakraborty, D., Chauhan, H., Agarwal, S. K., and Rajput, N. FOLKSOMAPS - Towards Community Driven Intelligent Maps for Developing Regions. In *ICTD* (2009).

[11] 3rd Generation Partnership Project 2. Wireless Features Description: Mobile Access Hunting. 3GPP2 S.R0006-514-A, June 2007.

[12] United Telecom Limited. Additional Voice Facilities. http://www.utlnepal.com/serv_additional.php.

[13] Mudliar, P., Donner, J., and Thies, W. Emergent Practices Around CGNet Swara, A Voice Forum for Citizen Journalism in Rural India. In *ICTD* (2012).

[14] Odero, B., Omwengan, B., Masita-Mwangi, M., Githinji, P., and Ledlie, J. Tangaza: frugal group messaging through speech and text. In *DEV* (2010).

[15] Patel, N., Chittamuru, D., Jain, A., Dave, P., and Parikh, T. S. Avaaj Otalo - A Field Study of an Interactive Voice Forum for Small Farmers in Rural India. In *CHI* (2010).

[16] Patel, N., Klemmer, S., and Parikh, T. An Asymmetric Communications Platform for Knowledge Sharing with Low-end Mobile Phones. In *UIST* (2011).

[17] Raza, A., Rosenfeld, R., Sherwani, J., Milo, C., Alster, G., Saif, U., Pervaiz, M., and Razaq, A. Viral entertainment as vehicle for disseminating speech based services to low literate users. In *ICTD* (2012).

[18] Sambasivan, N., Weber, J., and Cutrell, E. Designing a Phone Broadcasting System for Urban Sex Workers in India. In *CHI* (2011).

[19] 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 *ICTD* (2007).

[20] Smyth, T. N., Kumar, S. K., Thies, W., Medhi, I., and Toyama, K. Where There's a Will, There's a Way: Mobile Media Sharing in Urban India. In *CHI* (2010).