Economical Global Access to a VoiceXML Gateway Using Open Source Technologies

Kulwinder Singh, Dong-Won Park
Dept of Information & Communications Engineering
PaiChai University, Daejeon, South Korea
{singh, dwpark}@pcu.ac.kr

Abstract

Voice over IP and the open source technologies are becoming popular choices for organizations. However, while accessing the VoiceXML gateways these systems fail to attract the global users economically. The objective of this paper is to demonstrate how an existing web application can be modified using VoiceXML to enable non-visual access from any phone. Moreover, we unleash a way for linking an existing PSTN-based phone line to a VoiceXML gateway even though the voice service provider (VSP) does not provide a local geographical number to global customers to access the application. In addition, we introduce an economical way for small sized businesses to overcome the high cost of setting up and using a commercial VoiceXML gateway. The method is based on Asterisk server. In order to elucidate the entire process, we present a sample Package Tracking System application, which is based on an existing website and provides the same functionality as the website does. We also present an online demonstration, which provides global access to commercial voice platforms (i.e. Voxeo, Tellme Studio, Bevocal and DemandVoice). This paper also discusses various scenarios in which spoken interaction can play a significant role.

1 Introduction

The end of the 20th century witnesses an explosive growth in Internet usage. We have seen an explosion in the number of browser-based visual applications, from the broad examples we use every day, such as e-commerce, movie or flight schedules, and financial information. The most common means for accessing information residing on many websites across the globe is still the dominating interface of point and click with a mouse using the graphical user interface (GUI). Additionally, telephone is also widely used to access information. Still, in densely populated countries it seems to be difficult to handle large amounts of calls simultaneously, which leads to long call queues and frustrated customers.

However, the challenge that is presented to the present Internet world is to make the enormous web content accessible to users who don't have the computers or maybe don't have the money to buy as well as visually impaired users. Since speech is the most natural means of communication, especially for these users, voice will be a dominating mode in newly designed multi-modal (Oviatt, S.L., 1999) user interfaces for future devices. This calls for a revolutionary design of a voice user interface (VUI) to supplement the conventional GUIs. Internet and telephony used to be two separate technologies to build applications accessible over the phone. VoiceXML bridges the gap; it leverages the existing web infrastructure and enables web developers to build voice-enabled web applications accessible from any telephone, by anyone, anywhere, anytime. A major advantage of VoiceXML is that it provides web content over a simple telephone device, making it possible to access an application even without a computer and an Internet connection. VoiceXML finds ready acceptance in the business world due to the following reasons.

Providing a voice-based interface with the web interface is an advantage to the visually challenged who are unable to use a visual interface. It is also possible to use the application for accessing a web-based interface even while on the move through a mobile phone, which is much easier to carry around than a personal computer. Phone applications are more spontaneous. Most people these days always have their phone on their hip. In many cases, the phone transaction can be completed before the PC even boots or you can log in. Lastly, there is no possibility of a virus from a phone call and it is typically much more secure.

The number of telephone users is far greater than the number of people who use personal computers or the Internet.

Thus, by using VoiceXML applications, we can reach out to more customers than is possible by using the Internet. Voice portals put all kinds of information at a consumer's fingertips anytime, anywhere. Customers just dial into the voice portal's 800 number and use simple voice commands to access whatever information they need. It's quick, easy, and effective, even from a car or the airport. However, it still fails to attract the huge global customers as they have to pay the long distance calling charge to access the information. Hence, this paper is an attempt to peep behind the curtain and analyze the market trends and thereby proposes a solution to resolve the current issues and satisfy the global customers by providing them a solution to access the VoiceXML gateway economically. The structure of this paper is as follows. In the

Coling 2008: Proceedings of the workshop on Speech Processing for Safety Critical Translation and Pervasive Applications, pages 17–23
Manchester, August 2008

next section we present the voice hosting infrastructure. We then discuss our experimental results and finally conclude by presenting the scenario for using Voice User Interfacing followed by the summary of the outcome.

2 Voice Hosting Infrastructure

A voice hosting infrastructure requires many interlocking components such as telephony hardware, software: TTS (text to speech, ASR (automatic speech recognition), networking technology, monitoring and administrative services. We discuss all the essential elements bellow.

2.1 Linking

Most of the VoiceXML gateways (Ruiz, Q. Sanchez, M. 2003) can operate VoiceXML speech applications on any standard web server and can support both static and dynamic content, and provide a high degree of scalability and platform-independence. Also, voice applications can be seamlessly integrated into existing enterprise web and IT infrastructure. There are two ways to accomplish the task:

-Link your existing web server with VSP's voice gateways.

-Port your web applications to VSP's web server.

Linking an existing web application with VoiceXML gateways is fairly straightforward. As you see in figure 1, when a VoiceXML gateway receives a phone call, it looks at the number dialed to lookup the URL of the web server, then sends the HTTP request. You need to provide the URL of your web server to VSP. One VSP provides Web-based GUI for linking an application as shown in Figure 1.

There may be some changes required to your Web server before you connect with your VSP. Changes vary from VSP to VSP, or depending on your service provider and type of Web server. As an example, our application residing on an Apache HTTP Server, according to Bevocal, must modify the httpd.conf file to add the new MIME type in the following way.

```
# AddType allows you to add to or override the MIME configuration.
# file mime.types for specific file types.
# MIME types for VoiceXML-related content.
AddType application/voicexml+xml                .vxml
AddType application/srgs                        .gram .srgs
AddType application/srgs+xml                    .grxml
AddType application/x-nuance-gsl                .gsl .grammar
AddType application/x-nuance-dynagram-binary    .ngo.
```

2.2 Mapping

Speech-enabled Internet portals, or voice portals, are quickly becoming the hottest trend in e-commerce-broadening access to Internet content to everyone with the most universal communications device of all, the telephone. Currently, voice hosting providers set up local toll free numbers or DID (direct inward dialing) numbers in order to access voice applications through their VoiceXML gateways. If the VSP is unable to provide the local DID numbers in the desired country, the users from that country have to pay international calling charges, which is sometimes quite expensive. We propose our idea to resolve this issue as follows.

SIP Mapping: It totally depends upon the telephony infrastructure of the VoiceXML gateway. If it is asterisk-based (Meggelen, J. V. Madsen, L. Smith J. 2007) then the job is fairly easy, otherwise it could be a tedious task to configure a VoiceXML gateway with a remote telephony network. Our proposed idea is independent of any kind of telephony infrastructure, provided it supports SIP signaling.

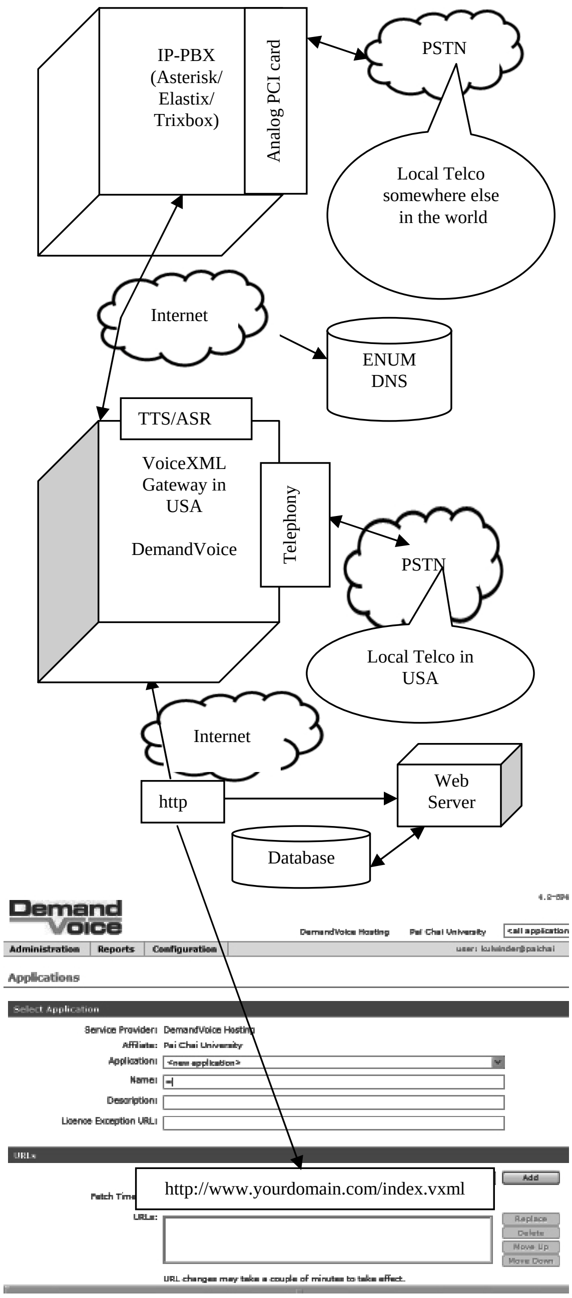


Figure 1. Linking a web server with a VoiceXML gateway

The most promising way to connect a VoiceXML gateway with a third party's Asterisk server (any IP-PBX) is to use the ENUM service. In order to use ENUM DNS efficiently, there are few steps needed to be followed. First of all, at e164.org (Ruiz, Q. Sanchez, M. 2003), in the ENUM database, we need to register the IP address and DID number, which is landing on your SIP extension of VoiceXML Gateway, as depicted in the figure 2.

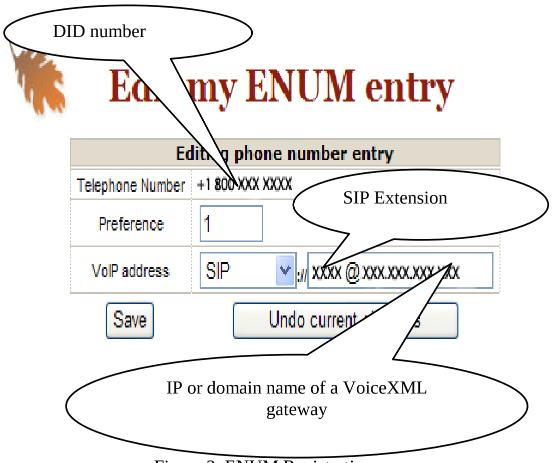


Figure 2. ENUM Registration

After editing the ENUM (tElephone NUmber Mapping) entry, we set up the ENUM trunk and outbound route on the remote IP-PBX machine.

We are running Elastix IP-PBX (elastix.org) on the remote side because it is easy to manage the configuration through GUI on Elastix. Moreover, it is an open source technology, and comes with a self-installing package that installs a complete operating system, Asterisk PBX, FreePBX etc.

9|.XXXXXXXXXXXXX (X matches any digit from 0 to 9)

According to our dial plan shown above, let's assume that we need to dial the American DID number 641-543-6745, and dialing pattern would be like: 916415436745.

Our DID number 641-543-6745 is registered at e164.org. This means that when someone calls the DID, the call will land on the SIP number instead of DID number, as the *e164.org* DNS zone will provide the IP addressing and protocol information needed to connect to your VoiceXML gateway . In other words, the call will not go over the DID provider's network (see figure 3).

There would be a native or Packet2Packet SIP bridging between the VoiceXML gateway and remote IP-PBX. Ultimately, VSP and remote client will not pay any toll to PSTN operator or ITSP (Internet Telephony Service Provider) because the call bypasses their network. Moreover, the VSP does not need to open all the credentials of telephony setup of the VoiceXML gateway. So, most of the information will be isolated from the remote client. This is attractive to the VSP that does not want to register the sip number and IP address of the gateway in the ENUM (tElephone NUmber Mapping) database, (because some people are afraid to disclose their IP addresses to others). Moreover, they do not want to accept anonymous SIP calls, and want to run their own IP-PBX instead of using client's IP-PBX. In that case, we propose a very easy solution to set up the SIP extension on the VoiceXML

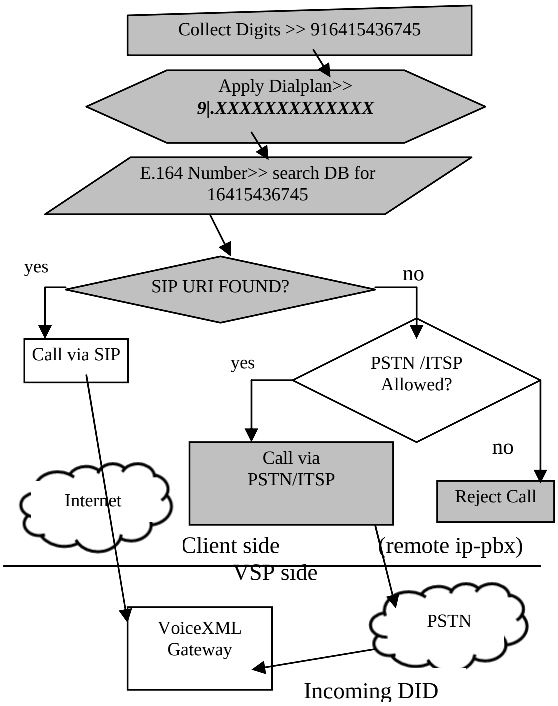


Figure 3. Flow chart of the call logic

gateway and configure it on the remote IP-PBX in the custom extension as shown in figure 4.

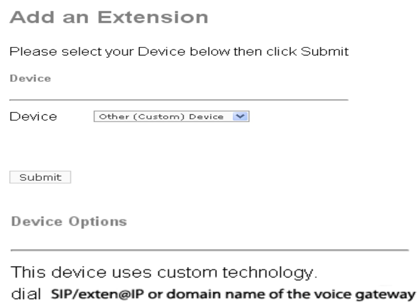


Figure 4. Custom extension settings

Our IP-PBX is connected with Bevocal, Tellme Studio, Voxeo and DemandVoice. So, our Device Options look like as follows

SIP/8773386225@voip.cafe.bevocal.com
SIP/8005558965@sip.studio.tellme.com

Both of the above mentioned methods are really good if VSP does not want to use a remote IP-PBX for outbound calls.

On the contrary, when VSP wants to setup outbound calls on the remote machine, we propose another idea to accomplish the task. Fortunately, this is very easy to configure the

machines on both sides, if a telephony infrastructure uses an asterisk-based PBX on both ends.
In this scenario, we can register the machines with each other using username and secret or we can use IP-based authentication without registering with each other. Actually, it is very easy on Elastix because it uses a Freepbx for configuring most of the tasks of Asterisk server.
In other words, it's becoming less and less common to have static IP addresses. So, if you have a dynamic IP address it is good to go with username and secret. Typically, we have to deal with sip.conf and extensions.conf on Asterisk, provided you use sip protocol. For a sample configuration code (Meggelen, J. V. Madsen, L. Smith J. 2007) see subsection *DID Mapping*.

DID Mapping: We have two scenarios to deal with: a) When a VoiceXML gateway does not support SIP signaling. b) When VSP wants to land the calls only on a DID number assigned for your application execution.

First, if it is a toll free DID number then the remote client can dial through ENUM in order to connect with a toll free gateway, and call will land on the toll free network, which is connected with a VoiceXML gateway (see figure 5). It means a toll free subscriber has to pay for it, and the call between a remote IP-PBX and the toll free gateway would be free, because it will go over the internet.

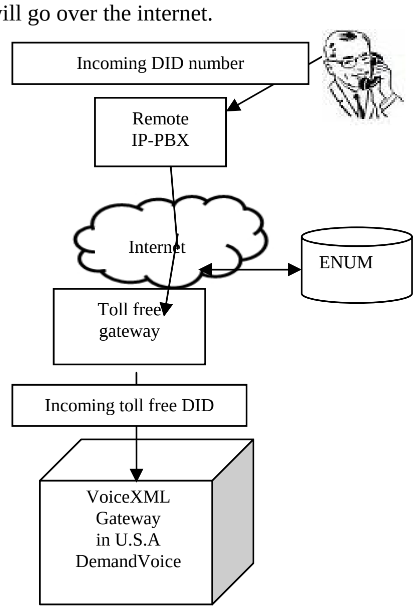


Figure 5. Remote toll free connectivity

For example, we connect DemandVoice's voice gateway using a toll free DID number remotely as follows:
Set up the custom extension as we discussed in subsection *SIP Mapping* , and it will directly connect with a toll free gateway (see figure 6).

SIP/8008042865@sip.tollfreegateway.com

Or you can dial through ENUM as we discussed in subsection *SIP Mapping*.
If it is a DID number and has no registration in the ENUM database then you need to originate the call using your ITSP, and the call will directly land on your DID assigned for your application by VSP. With the advent of VOIP technology,

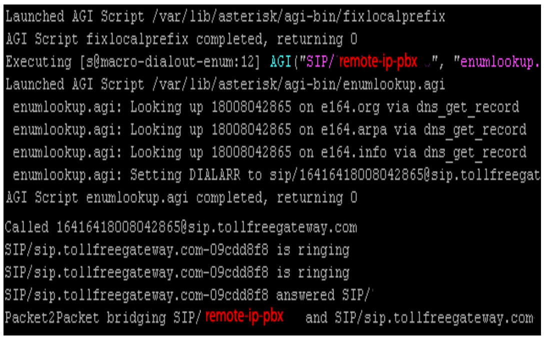


Figure6. Asterisk CLI

there has been a flood of ITSP (Internet Telephony Service Provider) all over the world. It is really hard to choose one. We have tested the following configuration using our Static IP address on Elastix with VTWhite (Internet Telephony Service Provider) for VOIP termination and origination.

Peer Details:
allow=ulaw
canreinvite=no
context=from-pstn
disallow=all
dtmf=rfc2833
dtmfmode=rfc2833
host=sip.vtwhite.com
insecure=very
nat=yes
qualify=yes
sendrpid=yes
type=peer

Since our IP address is registered with VTWhite.com, there is no need for more typical authentication or registration parameters.

Inbound Routes:

DID number: 1XXXXXXXXXX (11 digits)

Set destination for incoming calls landing on your DID. If you are dialing out through VTWhite you must set your outbound CID as follows:

"1XXX-XXX-XXXX"<1XXXXXXXXXX>

We have tested the following configuration with voiptalk.org (Internet Telephony Service Provider) using username and secret.

Peers Details:
host=voiptalk.org
insecure=very
secret=XXXX
type=peer
username=XXXX
username:secret@voiptalk.org/username

2.3 Porting

Many organizations have their existing toll free phone numbers, and they want to connect their existing numbers with a voice portal, and don't like to get a new phone number. Luckily, it is very easy in the United States to port the number from one carrier to another carrier. There are many independent "RESPORG" (RESPonsible ORGanization) companies, which help for porting the numbers.

If there are issues for porting the existing number, we propose a very simple idea to install an asterisk-based IP-PBX at your premises and route the calls landing on your existing number to VoiceXML gateway using a sip or ITSP as we have discussed in section 2.2 *Mapping*.

2.4 Editing

Adding VoiceXML interface (Tsai, M.-J. 2005) (Kenneth, R. A. 2001) (Yankelovich, N., 2000) to web contents presents unique challenges to designers. Complexity depends upon the web application's architecture. In this section, we demonstrate how to modify an existing package tracking web site powered by a relational database. We use PHP, MySQL, Apache web server, and these tools are widely used in web applications development, because these are cross-platform and open source technologies. There are a couple of ways to add voice user interfacing (VUI). It is possible to add VoiceXML tags either on the fly when the VoiceXML interpreter extracts the contents from the web server or in other case tags can be embedded into an existing web page. However, we concentrate only on the latter case. First of all let's have a look on a web application (see figure 7) (Tracking number: 6754356786). This application is available on the following URL for demonstrating the task. http://biometrics.pcu.ac.kr/demo/search1.php

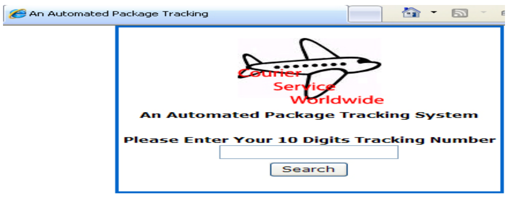


Figure 7. WEB-GUI for tracking the package

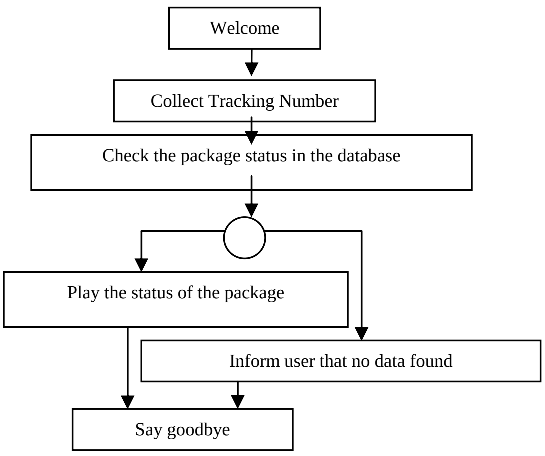


Figure 8. Call flow diagram for a VUI design

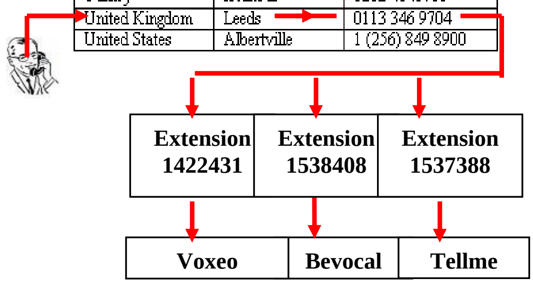
We design the call flow diagram (see figure 8) of the package tracking application as follows: According to our flow chart we need to make two VoiceXML documents. Before adding the VoiceXML tags into your webpage you must check with your VSP how to specify the document type definition (DTD) in your VoiceXML documents. Since our application is linked with Bevocal platform (BeVocal Café, 2007), we do the following way.

Collecting the Tracking number-Voicexml Document-1 (index.vxml) (see Appendix A).

Tracking Report-Voicexml Document-2 (track.php) (see Appendix B).

Table 1. Geographical Distribution of Phone Numbers for Accessing VoiceXML Gateways

Argentina	Buenos Aires	011 50314720
Australia	Sydney	02 90372744
Bahrain	National	0 16199009
Belgium	Brussels	02 7470254
Brazil	Campinas	019 31192787
Bulgaria	Sofia	02 4917134
Canada	Toronto	1 (647) 723 3640
Chile	Santiago	02 5821844
Cyprus	Nicosia	22 022630
Czech Republic	Prague	02 46019148
Denmark	National	77345753
Estonia	National	6681372
Finland	Helsinki	09 42597847
France	Paris	01 72898101
Germany	National	0180 35350033089
Guatemala	Guatemala City	02 3560986
Hungary	Budapest	01 9994901
Ireland	Dublin	01 6575613
Israel	Jerusalem	02 5695205
Italy	Rome	06 99268160
Japan	Tokyo	03 45903116
Latvia	Riga	7 661061
Lithuania	Vilnius	05 2111750
Luxembourg	National	2 0202381
Mexico	Mexico City	055 11689854
Netherlands	Amsterdam	020 8908243
New Zealand	Auckland	09 4427385
Norway	Oslo	02 1543295
Pakistan	Islamabad	0 51 8080931
Peru	Lima	01 7061950
Poland	Warsaw	022 3988047
Portugal	Porto	022 1451091
Romania	Bucharest	021 5398124
Spain	Barcelona	93 3905484
Sweden	Stockholm	08 52500225
Switzerland	Geneva	022 5330324
Thailand	Bangkok	02 1013109
Turkey	Istanbul	0212 4141710
United Kingdom	Leeds	0113 346 9704
United States	Albertville	1 (256) 849 8900



Now, it is time to call the application using a phone. We provide PSTN numbers from 40 courtiers to access the VoiceXML gateway of Bevocal, Tellme Studio, DemandVoice and Voxeo. In order to test the sample package tracker you need to dial extension (1538408) for Bevocal after dialing the local number as depicted in Table 1. You need PIN: 1234 and Developer ID: 5369574 to access our application. User can also call our application from the following numbers without dialing any pin or extension numbers.

Direct Numbers:

Italy	Rome	06 916507970
United Kingdom	Leeds	0113 350 8176
United Kingdom	Manchester	0161 660 4556
United States	Bellevue	1 425 998 0503

We will try to keep alive these Geographical Distributed numbers for public use. Developers and researchers can test their applications by paying just local charges applied by the terminating PSTN operator.

3. Scenarios for Using Voice User Interfacing

Despite the availability of various media of communication utilized in human computer interaction, people tend to prefer the more social medium of communication such as voice. With the advent of the Internet, the PC has become the most preferred device which people turn to when they need to enquire for information. On the interaction side, the telephone seems to remain the best example for usability preferred by the various classes of users. So, to power of voice communication, with the richness of the Internet on one side, and the usability of the phone device on the other side, we present various situations in which VUI can be of great utility.

Situations:

- Driving
- No internet service
- Visually Impaired persons
- Replacement of human operators

VUI is the most time efficient modality for input, because voice input is nimbler than typing. VUI can be used to check and answer web emails while driving a vehicle. Another class of situations is when there is no Internet or PC available and the user needs to access web applications such internet banking, parcel tracking, directory assistance, online reservation, order status enquiry, instant messaging, electronic voting, dating/chat services, and information services. Moreover, visually impaired people can take advantage of the above mentioned services just over the regular phone. Furthermore, in many situations cost efficiency can be increased by replacing human operators in call centers and offices with a VoiceXML-based interactive voice response system.

4. Experimental Results

To verify the performance of our proposed idea, we implemented an IP-PBX, an automated package tracker and the business listing search using VoiceXML, PHP, and MySQL. Then, we linked remotely with various VoiceXML gateways, and tried to call the application using different codecs (ulaw, g729, gsm). We found that ulaw codec is much better for interacting with the ASR engine, and also it provides the best voice quality since it uses no compression. This means that it has the lowest latency and the highest MOS (Mean Opinion Score) because there are no additional processing delays. However, it requires high bandwidth, and this can be easily managed via proper network provisioning. The compression has very adverse affect on speech recognition when it comes to deal with the ASR engine. The more compression is used, the more characters will be lost. Fortunately, ulaw is immune to this effect. Table 2 shows the call volume according to bandwidth and codec. Table 3 shows the hardware and software specifications.

Table 2. VOIP codec and their utilization

Codec	Bandwidth used per Call	Calls per megabit
ulaw	79.7kbps	14
g729	29.0 kbps	100
gsm	34.2 kbps	67

Table 3. Hardware and software specifications

Component	Description
CPU	Intel(R) Xeon(TM) 2.80GHz
RAM	1 GB
Telephony Board	Sangoma A200/REMORA 4 port FXO/FXS
OS	Linux centos kernel 2.6.18
Web Server	Apache
VoiceXML platform	DemandVoice, Tellme, Bevocal and Voxeo

5. Conclusion

In this paper we have targeted the large number of international users who are deprived of taking the advantage of using the toll free number remotely, and have introduced an economical way to access VoiceXML gateways globally. Moreover, our globally distributed PSTN numbers are available to access VoiceXML platform for only research, test and educational purpose. We conclude that the call quality may differ depending upon the different feature sets (e.g., codecs) and network bandwidth available. In order to get a nice connectivity with a VoiceXML gateway, the call should pass through minimum VOIP gateways. Currently, we are developing a virtual user agent based on ATOM/RSS protocol, which can be accessed by phone globally for accessing information.

Appendix

A VoiceXML Document-1

```
<?xml version="1.0"?>
<!DOCTYPE vxml PUBLIC "-//BeVocal Inc//VoiceXML 2.0//EN"
 "http://cafe.bevocal.com/libraries/dtd/vxml2-0-bevocal.dtd">
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml">
   <form id="login">
      <field name="t_number" type="digits">
```

```
        <prompt>
            Welcome to <emphasis>Department of Information and
Communication Engineering, PaiChai
University, South Korea</emphasis>.This demo version of Parcel
tracking system is developed by
<emphasis>Mr. Singh </emphasis>.This research work is partially
sponsored by<emphasis>Demand voice dot com
</emphasis><break
size="medium"/>
Please enjoy the music while I connect you with a package tracking
system.
<audio src="http://biometrics.pcu.ac.kr/demo/m3.wav"></audio>
Welcome to an automated parcel tracking system. Please tell me the
10 digits tracking number of your
package.
        </prompt>
        <filled>
           <prompt>
           The tracking number you entered is
           <say-as type="number:digits"> <value
expr="t_number"/></say-as>
           Please wait while I'm checking this package's status.
          <audio
src="http://biometrics.pcu.ac.kr/demo/wait.wav"></audio>
         </prompt>
            <submit next="http://biometrics.pcu.ac.kr/demo/track.php"
method="post"
namelist="t_number"/>
        </filled>
        <noinput>
           I'm sorry, I am not familiar with your accent.  Now you can
just type the 10 digits tracking number from the key pad of your
phone.
           <reprompt/>
        </noinput>
        </field>
    </form>
</vxml>
```

B Voicexml Document-2

```
<?xml version="1.0"?>
<!DOCTYPE vxml PUBLIC "-//BeVocal Inc//VoiceXML 2.0//EN"
"http://cafe.bevocal.com/libraries/dtd/vxml2-0-bevocal.dtd">
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml">
<form><block> <prompt> <voice gender="male">
<?php
header("Content-type: application/voicexml+xml");
$number = trim($_POST['t_number']);
    $host = "hostname";
    $user = "db_user";
    $pass = "user_pass";
    $db = "db_name";
    $link = @mysql_connect($host, $user, $pass, $db) or die
("Unable to connect.");
    mysql_select_db($db) or die ("Unable to select database!");
$sql = "SELECT * from track WHERE t_number = '$number' ";
$result = mysql_query($sql);
if (!$result) {
   echo "Could not successfully run query ($sql) from DB: " .
mysql_error();
}
elseif(mysql_num_rows($result) == 0)
{
   echo "I could not find any information for that package. Thank
you for using the telephone package tracker.Good bye";
}
else
{
while ($Row = mysql_fetch_assoc($result))
       {
echo "The following events were reported for package number.";
?>
<say-as type="number:digits">
<?php
echo " $Row[t_number]";
?>
</say-as>
<break size="medium"/>
<?php
echo "$Row[t_status]";
?>
<break size="medium"/>
<?php
echo "$Row[t_address]";
?>
<break size="medium"/>
<say-as type="date:ymd">
<?php
echo "$Row[t_date]";
?>
</say-as>
<?php
echo "Thank you for using the telephone package tracker. Good
bye";
       }
}
mysql_free_result($result);
mysql_close($link);
?>
    </voice> </prompt></block>
    </form>
</vxml>
```


Acknowledgment

We would like to express our gratitude to Ashraf Alattar, PaiChai University, South Korea, and Mark Rayburn, Demandvoice.com, USA for their help in designing the network and for participating in many useful discussions.

References

Tsai, M.-J. 2005. *The VoiceXML Dialog System for the E-Commerce Ordering Service,* IEEE Proceedings of the Ninth International Conference.

Ruiz, Q. Sanchez, M. 2003. *Design of a VoiceXML Gateway*, Fourth Mexican International Conference on Computer Science p. 49.

Meggelen, J. V. Madsen, L. Smith J. 2007. *Asterisk: The Future of Telephony*, Second Edition. O'Reilly.

BeVocal Café, 2007. *VoiceXML development environment*

Kenneth, R. A. 2001.*Voice Enabling Web Applications: VoiceXML and Beyond*".Apress; 1 edition.

Yankelovich, N., 2000. *Designing Effective Speech Interfaces*, John Wiley & Sons, Inc.

Oviatt, S.L., 1999. *Ten myths of multimodal interaction* Communications of the ACM, 42 (11), November