Internet Access by Talking and Listening Using any Phone: Help Global Development by Effectively Bridging the Digital and Language Divides

Emdad Khan[1], Eisa Aleisa[2]
[1]Internet Speech Inc.
San Jose, California, USA
[1,2] College of Computer & Information Science
Imam University
Riyadh, Saudi Arabia
Emdad@internetspeech.com, aleisa@imamu.edu.sa

ABSTRACT: *The need to effectively bridge the Digital Divide is very clear. While Internet is accessed and widely used mostly by educated people via computers or high end phones, the Internet and its associated benefits are not accessible by many people on the other side of the Digital Divide, namely, illiterate people, people having no access to a computer or high end phone, elderly people and people with disabilities.*

In this paper, we propose Voice Internet based Internet access that uses a state of the art "rendering" technology to convert existing WWW (World Wide Web) content into Short, Precise, easily Navigable, Meaningful and pleasant to listen to content that can be heard over any phone in real time . Such a "rendering" technology also supports filling of various on-line forms by voice.

The rendering technology also bridges the Language Divides (over 70% Internet content today are in English; thus people in non-English speaking countries are left out from major part of the Internet which is called the Language Divide) by automatically translating the rendered content from one language into other languages.

Thus, our proposed Voice Internet technology bridges the Digital and Language Divides in a practical, cost effective, easy to use , easy to learn and in a natural way as users basically talk and listen. As phone is ubiquitous, over 5 billion people who have access to some phone can enjoy the benefits of the Internet using the proposed solution.

We also discuss how bridging the Digital & Language Divides, and focusing on Education, Innovation and Entrepreneurship can help the Global Development, especially the Economic, Social and Cultural Developments with increased world peace.

Keywords: Internet, Voice Internet, e-Services, Rendering, Intelligent Agent, Artificial Intelligence, User Interface, Education, Innovation, Entrepreneurship, Natural Language Understanding (NLU), Base of the Pyramid People (BOP)

Received: 16 November 2011, Revised 3 December 2011, Accepted 11 December 2011

1. Introduction

It is important that everyone can access the Internet and use all content and on-line applications easily and economically. The Internet and Internet based applications are important from various key aspects including education, employment, economic-social-cultural & other developments, and more. In fact, ***Internet access should be a "right" for all the citizens of planet earth. Unfortunately, only a small fraction of the market is served today mainly because of***

(a) Lack of computer or computer skills,

(b) Access through a personal device is difficult, visual, expensive and not usable by many people who do not know how to

read or write (such illiterate people usually belong to the Base of the Pyramid People [BOP]).

It is important to note that for the BOP, on-line applications, especially e-services are becoming even more important as such services are easier to learn and use by BOP. Besides, such services usually meet their needs more by providing specific desired information. For example, e-Farming can provide key farming related information to farmers.

Thus, it is very important that everyone **has access to the Internet.** Unfortunately, only a small fraction of population (about 14% globally [1]) can access the Internet today resulting a very **large Digital Divide**.

Besides, there is another large Divide, the Language Divide (over 70% Internet content today are in English; thus people in non-English speaking countries are left out from major part of the Internet which is called the Language Divide), similar in magnitude to the Digital Divide. Clearly, the Language Divide also needs to be bridged.

In [2] we argued that **existing approaches to bridge the Digital and Language Divides are good but not sufficient to completely bridge them**. We then showed a more practical solution using "**Voice Internet**" to provide the benefits of the Internet to over 5 billion people (out of about 7 billion people in the world) who have some access to a phone.

In this paper, we emphasize that for the people at the Base of the Pyramid **(BOP)**, we would need to use a different paradigm to provide Internet access and ensure its effective use. Our proposed paradigm is based on the following key ideas:

a. provide access using an ubiquitous device (like a simple phone),

b. use a natural user interface (like Voice),

c. no requirement to be literate (no need to know how to read or write; rather use talking and listening),

d. use existing content on the Internet (avoiding to re-write with another language like WML or VoiceXML), and

e. start with simple most desired content / services (e.g. portal-like content but directly from the Internet, and e-services).

It is important to note that usability, features, navigation and amount of content are very limited when Internet is accessed visually by a high end phone. Moreover, many BOP cannot afford a high-end phone and associated data package. **Hence, Voice Internet becomes even a more practical alternative considering these issues.**

Section 2 describes bridging the Digital Divide, Section 3 describes bridging the Language Divide, Section 4 describes our solution using Voice Internet, Section 5 describes the rendering algorithms, Section 6 emphasizes the need for NLU (Natural Language Understanding) and Section 7 describes how we can help global development by going beyond bridging the Digital and Language Divides and focusing on **Education, Innovation and Entrepreneurship**.

2. Bridging the Digital Divide

We have addressed bridging the Digital Divide issue in details in [1], [2]. Here we just briefly mention the highlights as a quick review as it is strongly related to provide the Internet to many people at the Base of the Pyramid.

The existing approaches of bridging the Digital Divide can be broadly classified into **three groups**:

(a) by providing computers or low cost simple computers or computer like devices to people who do not have one.

(b) by providing personal devices like PDAs and cell phones with good size display screen.

(c) by using TV with set-top box.

Today, computers connected to the Internet represent only about 14% of telephone population on the average worldwide [1]. This ratio is even worse in the developing countries. Improving this ratio to an acceptable desired figure, especially, in the developing world would take long time. Additionally, certain population (**like elderly, blind, visually impaired people and people unfamiliar with computers**) would have difficulty in learning how to use a computer and Internet.

Personal devices like a cell phone with good size screen or PDA are great devices to communicate via voice or text with small contents. But these are not good devices to do general computing or to access the Internet. The key reasons are:

• Difficult user interface because of small screen and small keypad.

• The content are limited as one would need to re-write the content in another language like WML (Wireless Markup Language) in case of cell phone viewing; or the content needs to be manually scrolled in case of a PDA.

• Visual access makes such devices difficult in an eyes busy-hands busy situation like while driving.

• Many people, especially, in the developing world do not know how to read or write. So, a visual display based access would not be very useful to such population.

Accessing the Internet via a TV and set-top box has not been very successful yet. However, it has a great promise. With TV becoming more interactive devices, people getting more and more familiar with the Internet and more attractive content becoming available through digital TV and IPTV, Internet access via TV shows great potential. The cost issue and fear of learning how to use a computer will be significantly minimized for many people.
However, most of the other key issues mentioned for computer or mobile phone based access will still apply – like digital TV will still be beyond reach for many people at the Base of the Pyramid, learning how to use complex features will still be there via a complex remote control.

2.1 Literacy Issue

To use a computer, PDA, cell phone or a TV for Internet access, one needs to know how to read and write. However, many people at the bottom of the pyramid do not know how to read or write. This is a huge problem as literacy rate for BOP, in general, is very low.

Thus, existing approaches will minimize the gap but are not sufficient to truly bridge the Digital Divide. To really bridge the Digital Divide, we would need to provide access using an ubiquitous device (like a simple phone), with a natural user interface (like Voice), without a requirement to be literate (no need to know how to read or write; rather use talking and listening), use existing content on the Internet (avoiding to re-write with another language like WML), and start with simple most desired content or services (like e-services). **These are well addressed with the solution presented in this paper.**

3. Bridging the Language Divide

The Digital Divide issue has been talked about in depth in many literatures. However, there is another Divide, called the Language Divide (as defined above), of similar magnitude which has not been really discussed or talked about. Due the Language Divide, people in countries like China, India, Japan, Russia are left out from the major part of the internet as they do not prefer to use English or many of them do not even know English. It has some overlap with multi-lingual support that has been talked about in some literatures.

Language Divide has other implications – in a few years Chinese content on the Internet will grow to a significant level. People not speaking Chinese will not be able to access the Chinese part of the WWW. This can be extended to other languages as well. In general, in this global world, people not only need to access websites in all languages (the main focus of the multi-lingual approach) but also need to understand the content of the websites in all other languages. Thus, bridging the Language Divide is equally important as bridging the Digital Divide. In other words, bridging the Digital Divide will not be complete if the "Language Divide" is not bridged as well.

To bridge the Language Divide, we would need automated translation to translate web content from one language into other languages, which is called **Machine Translation (MT)**. Automated translation, unfortunately, has not reached a point where the translation is highly accurate. This is more true when a book is translated from one language to another using a translation software. The error level, in most cases, is not acceptable. **This is mainly because of the fact that the machine translation (MT) is a very complex problem and the solution is not just there yet.**

However, in translating web pages, the problem is much less complex as a website usually has simple sentences and paragraphs. So, current MT can provide reasonably good results in many cases. But it is important to note that much more work is needed in machine translation to really bridge the Language Divide. **So, when we say bridging Language Divide, we mean bridging it using the state of the art technologies that exist today with limited translation capability that we can rely upon. But it is still better than not translating at all.**

The Voice Internet technology automatically translates the rendered content from one language into other languages.

4. An Attractive Solution Using Voice Internet

Voice Internet can truly bridge the Digital and Language Divides to anyone who has some type of telephone – wire-line phone, wireless phone, PDA and the like. Voice Internet does not need a computer. Users basically make a phone call and an **automated attendant** allows the caller to access the Internet and enjoy surfing, searching, email, e-Services and other features. Users basically talk and listen to the Internet. In other words, the telephone becomes the browser. Voice Internet can provide the benefits of the Internet to over 5.2 billion people who have access to some type of phone. **Voice Internet overcomes the difficulties mentioned above with existing approaches**:

- no need to buy a special device (thus allowing easy & affordable access to many more people)

- no need to deal with small screen or small key pad as users basically talk and listen

- much easier to learn as learning how to use a phone is much simpler than learning how to use a computer or personal device

- no need to re-write the content in another language. There are over 3 billions of websites on the Internet. Re-writing all of them would be very expensive and hence not practical.

- No requirement to know how to read or write

The **other key challenge that Voice Internet overcomes** is the ***"rendering"*** problem. The Internet was designed with visual access in a large display device in mind. Thus, all the information is laid out in a manner that attracts our eyes but not ears. **Rendering or converting such information into short, precise, easily navigable, meaningful and pleasant to listen to content** is a very complex problem that Voice Internet has overcome. These key features of rendering are very important as when listening, one does not have time to listen to everything on a page, would like to move around easily and quickly and make sure that content heard is the content that was desired.

5. How Rendering is Achieved

An **Automated Attendant (also called an Intelligent Agent, IA)** is used to perform the "*rendering*" function. IA (Figure 1) performs rendering by

(a) automatically generating important information of the page, called, "*Page Highlights*", presenting them in a small amount of information at a time that one can easily follow

(b) finding appropriate as well as only relevant content on a linked page selected by a Page Highlight, assembling the relevant content from a linked page, and presenting them

(c) and providing easy navigation.

Rendering allows users to easily navigate within and between pages using simple voice commands or key pad entries. The Intelligent Agent is capable of learning user preferences, to continually improve ease of access over time.

This Voice Internet **rendering technology** is very well suited to **render Internet content and applications** so that one can easily interact with such applications either to retrieve content or to fill forms. The following Section briefly describes the rendering algorithms.

5.1 Rendering Algorithms

Rendering is achieved by using algorithms similar to the algorithms used by sighted users. The key steps of rendering are done using the information available in the visual web page itself and employing appropriate algorithms to use all such information including text contents, color, font size, links, paragraph, amount of texts and meaning of the words. Some language processing algorithms are also used to further refine the rendering, navigation and filling of on-line forms (Form Filling). This is similar to how the brain of a normal sighted person renders information from a visual page by looking into the font size, boldness, color, content density, link, meaning of titles/labels, and then selecting a topic, going to the desired page and then reading only the relevant information on the desired page. Form filling is done by presenting forms as Form Page Highlights and also creating

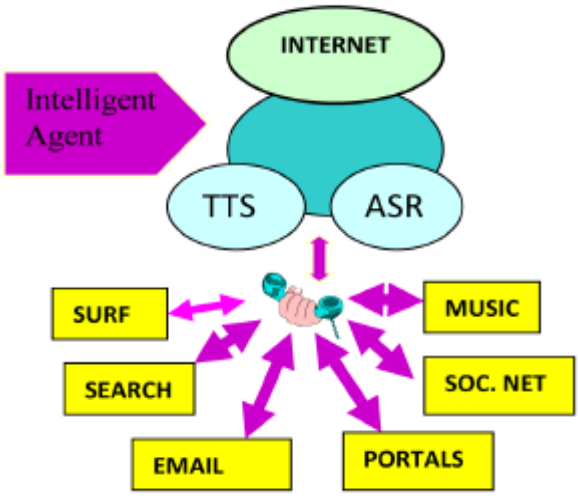

Figure 1. The Intelligent Agent (IA) and key features of Voice Internet

appropriate questions, taking the text/voice inputs from the user and then filling and submitting the form.

Thus, a user can seamlessly access any content on the Internet, interact with any forms and complete transactions like shopping, banking etc. using a simple phone and his/her own voice. Another key feature is that content can be translated in real time into another language, providing audio access to, for example, English-language web pages for those with limited English language skills, thus bridging the **Language Divide**. A good example of Voice Internet is net ECHO® from **InternetSpeech, Inc, a company based in California, USA (www.internetspeech.com).**

5.1.1 Rendering Algorithms for Content Retrieval & How Well Rendering Works

To answer how well Voice Internet can provide meaningful content from today's Internet, we need to answer the following questions:

(1) can the contents really be provided from any web site on the Internet ?

(2) can the existing Internet contents be rendered in a manner that the rendered content can be obtained in real time, is short, precise, easy to navigate, meaningful in audio and pleasant to listen?

The answers to both questions are "*yes*". Depending on the site, the "*yes*" can be a very strong "*yes*" or a strong "*yes*" a weak "*yes*". A content rich page with a small number of links makes rendering and navigation easy since there are only a few choices, and one can quickly select a particular topic or section. If the site is rich in content, links and images/graphics, the problem is more difficult but good solution still exists by carefully selecting a built-in feature called "*Page Highlights*". The most difficult case is when a page is very rich in images/graphics and links. In such cases, the main information is located several levels down from the home page and so navigation becomes more difficult as one has to go through multiple levels. Using multi-level Page Highlights and customized Highlights, the content can still be rendered well. But in this case, it is not as easy to navigate as the other two cases. Usually most of the Internet contents fall under the first and second categories.

Figure 2 shows the complexity of rendering content using all major cases. **Case (a)** is a simple case and the texts in the headline (i.e. Obama Visited Indonesia) for the calling page and the called page are same. So, simple string matching will easily find the desired content on the called page.

Case (b) is more difficult as the system needs to know that the name of the U.S. current President is Obama to find the desired content. **Case (c)** is even more difficult as the IA needs to know the history to derive that Indonesia is the country where Obama spent his Childhood. And, finally, **Case (d)** is also difficult as there are multiple stories about Obama. The IA needs to use the semantic meaning of words to determine the most relevant content on the called page.

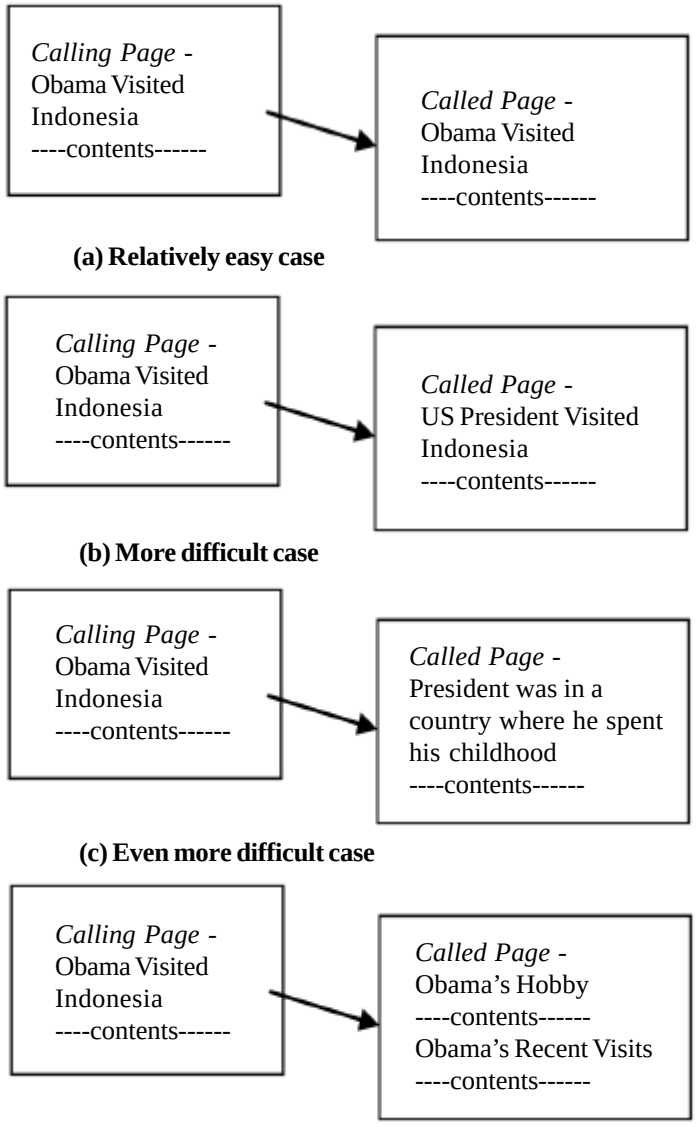

Figure 2. Complexity in Rendering Content

Case (b) is more difficult as the system needs to know that the name of the U.S. current President is Obama to find the desired content. **Case (c)** is even more difficult as the IA needs to know the history to derive that Indonesia is the country where Obama spent his Childhood. And, finally, **Case (d)** is also difficult as there are multiple stories about Obama. The IA needs to use the semantic meaning of words to determine the most relevant content on the called page.

There are some cases where all above approaches will fail (e.g. called page has title using image/video without ALT tag). For such cases, the IA uses some other algorithms including density of contents, font size, boldness etc. and the artificial intelligence (AI) rules to determine the desired content on the called page.

5.1.2 Rendering Algorithms for Form Filling

Generic form filling (basically allowing successfully filling and submitting any types of on-line forms) is needed for many applications including many **e-Services** (e-Learning, e-Health, e-Gov., e-Farming), **shopping, transactions, file downloads, gift registry and sending, online book reading, applying for jobs** and many more. We found that "*shopping*" applications are

usually very complex and includes most of the complex forms. Hence, we have used below the algorithm for a "*shopping*" application to describe the algorithms for general form filling.

Algorithms:

a) Determine form types (login, search, registration, shopping cart, form with multiple submission options like update form, submit form, continue form)

b) Determine element type (text, listbox, combo box, radio button)

c) Determine submission type (submission using image links, submission using form Action, handling multiple submission options like update cart, continue shopping or checkout)

d) Ensure linking and sequencing appropriate forms

e) automatically generate appropriate questions depending on the form type and element type.

For forms with file download and multi level security, there are additional steps to answer additional questions and also authentication. But basic form processing and sequencing still remains the same for such cases. For IM (Instant Messaging) forms, there are some additional buttons (like Transcript Pane, Hand up Button, Push to Talk, Ignore Button, Allow Button, and Input Area). In this case, form sequencing is simpler as the same form structure is used during communication, however form processing is different and complex as it needs to interact with multiple users and using multimedia inputs like voice or text.

Webpage forms have needed information to implement above mentioned algorithms. If a webpage has multiple forms or multiple applications, then forms are usually presented as "*Form Highlights*" (similar to content highlights described in 5.1.1) to make it simpler for the users to select corresponding form or application.

Note: Other approaches (e.g. **Ontology** [7]) would be very difficult to apply on today's WWW. **Natural Language Understanding** as described below will also significantly help the **rendering process.**

6. The Need for Natural Language Understanding (NLU)

It is important to note that the role of NLU is the key for the BOP to effectively use and interact with the Internet. A good example is "search". With today's string based search, we get hundreds (and sometimes thousands) of results**.** An educated user can select the desired results relatively easily by looking into a computer screen but have difficulty to do the same using a high end phone. Accessing the same via a phone and with audio is very difficult. An NLU based search will understand the meaning of the words (and sentences) and will be able to deliver very desired content with much smaller search results. NLU is also important as many people at the Base of the Pyramid will not use structured words or sentences. A good NLU system will significantly ease this process.

This has been verified by ***many users who have been successfully using the Voice Internet over 5 years - they are highly satisfied and recommended the service to many others. The key features used by users are Surfing, Searching, Email, Streaming Audio, and on-line book reading.***

Accordingly, we are doing some active research in this area using some bio-inspired and brain-like algorithms as humans are very good at NLU.

7. Going Beyond Bridging the Digital Divide

What's next after the Digital and Language Divides are Bridged from Connectivity Standpoint? Let's assume that together with conventional methods of using computers, PDAs, cell phones, and the proposed method of Voice Internet, the Digital and Language Divides are really bridged from connection standpoint. Now what? Well, the benefits of getting into the Internet need to be utilized properly to really help meet basic needs including food, shelter, education, communication, health, business and economy. To really bridge the Digital Divide, we would need to address the other key factors**: utilize the access to information to knowledge, use knowledge to drive innovation & entrepreneurship to finally drive the development and growth.**

Education is a very important key element in this process. To stimulate real economic growth, "education" needs to be highly

emphasized and targeted, especially with "**creativity", "productivity" and "resource" creation** in mind. Creativity will in turn **drive innovation, entrepreneurship, productivity** (and even resources – like discovering new oil reserve, alternate energy…) resulting successful business entities which in turn will create jobs and **drive economy reducing rich-poor gap**. In a nut shell, the key idea is to use the benefits of accessing the Internet and e-Services to create valuable resources.

By "education", we do not necessarily mean just formal education and degrees. Education can be as simple as training some people to do some specific job very well. For example, most workers in a garment industry do not have any formal education – they just know how to do their manufacturing job well (of course, some people with formal education are needed to manage, plan etc). So, the education needs to be creative and targeted to meet immediate needs. Creativity also depends on the need. If a country already has lot of skilled workers and doing well in improving economy, then the country's next goal is to keep on improving what it does better (to stay ahead of competition) and also innovate to diversify to move up the overall food chain of the economy.

Thus, education needs to be well planned as creativity is needed at various steps. And the creativity also needs to be well planned to meet specific countries needs with special emphasis on **immediate needs, entrepreneurship, innovation and valuable resource creation**.

8. Conclusion

Information should be for everyone and hence everyone should have the right to access the Internet. The Base of the Pyramid People (BOP) would need to start with simpler content (mainly portals and e-services) accessible via a ubiquitous device and without any need to become literate. Such population can gradually start using all other features of the Internet. Existing approaches to bridge the Digital Divide are good but not sufficient to completely bridge the Digital Divide. Our proposed solution using ***Voice Internet and its rendering capability*** *is a practical solution to not only completely bridge the Digital Divide but also bridge the Language Divide.*

The NLU is very much needed to provide the key benefits of the Internet to the BOP.

It is very important to properly exploit the benefits of bridging the Digital Divide through ***education with emphasis on entrepreneurship and innovation*** *to create valuable resources to drive* ***economic, social and other developments****. Our proposal to bridge the Digital and Language Divides, to do appropriate educational reform and make use of entrepreneurship and innovation can create an* ***enormous resource from the BOP who can*** *then actively participate in driving* ***global economy****. This would not only help the BOP to get out of the poverty but would also help top of the Pyramid people who can effectively use such an enormous low cost resource.*

References

[1] Khan, E. (2011). Internet for Everyone - Reshaping the Global Economy by Bridging the Digital Divide, Aug.

[2] Khan, E. (2010). Information for Everyone using any Phone – Global Development via Education, Entrepreneurship and Innovation, International Convention on Rehabilitation Engineering & Assistive Technology in Collaboration with ACM, July, Shanghai, China.

[3] Khan, E. (2009). Very Affordable and Easy to Use Internet for Everyone using any Phone: Ensuring Social Inclusion for People with Disabilities, ITU Asia-Pacific Forum, Aug 25- 27, Bangkok, Thailand.

[4] Khan, E. (2003). System and Method for Audio-Only Internet Browsing using a Standard Telephone, U.S. Patent Number 6, 606, 611, Aug 12.

[5] Williams, M. (1999). Telephone Speech Recognition and Web Access, *In:* Proceeding of AVIOS, May.

[6] [Stiglitz], Stiglitz, J. (2006). Making Globalization Work, W.W Norton & Company.

[7] Hammack, C. N. et al, (2002). Automated Ontology Learning for a Semantic Web, Dept of CS, Uni. Of Nebraska, Feb.