Natural language based human computer interaction: a necessity for mobile devices

Emdad Khan

***Abstract*–**-Human Computer Interaction (HCI) with any computing device is becoming more important with the rapid growth of the use of such devices worldwide. Computing devices range from a computer (large, medium or small) to small devices, like a mobile phone. Since mobile devices are ubiquitous and many computing functions are moving to mobile devices, the need for a very good HCI is becoming even more important for such devices. While existing mechanisms are good for high end mobile devices (e.g. smart phones and PDAs), they mainly serve the technically literate people. Nontechnical, semi-literate and illiterate people have great difficulty in using existing interfaces, mainly because of the complexity in learning and the literacy needs. Besides, many people cannot afford such high end mobile devices. The User Interface (UI) for low end and medium end mobile devices are even more difficult as such devices do not have some of the nice features that high end devices have (e.g. touch screen, scrolling, larger size screen/keypad). However, complexity of learning is relatively less as such devices mainly have simple features and do not have complex features like accessing the Internet and interacting with Internet applications.

Clearly, to provide the key benefits of this Information Age to many more people including largely dominated Base of the Pyramid People (BOP), low and medium end mobile devices should have all the key features that high end mobile devices, tablets and computers have. And such key features should have a very good UI. Two technologies are essential to achieve these goals – **a Natural Language Interface and an Intelligent Agent.**

In this paper, we propose a Natural Language Understanding (NLU) based Intelligent Agent (IA) that overcomes the existing problems by automating the key tasks while allowing a natural user interface using user's voice (or typing). Our solution not only makes it much easier to effectively use mobile devices by non-technical, semiliterate and illiterate people, but also by technical people as the key tasks in a high end mobile device now handled by manual scrolling and rendering by user's eyes & brain are automated via the use of IA. We argue that such automation is, in fact, becoming a necessity when one would like to use various complex on-line applications including e-Services using low & medium end mobile devices. We also argue that our proposed approach makes the UI much simpler, easier and more effective for all other devices including high end mobile device, tablets and computers.

Emdad Khan is with the College of Computer & Information Science, Imam University **,** P.O. BOX 5702, Riyadh , Saudi Arabia. He is also with InternetSpeech, Inc, San Jose, CA, USA (Phone: 408-532-9630, fax: 408-274-8151, email: emdad@internetspeech.com)

***Keywords* ----**Base of the Pyramid People (BOP), Economic & Social Development, Human Computer Interaction (HCI), Intelligent Agent, Natural Language Understanding (NLU), User Interface (UI), Rendering.

I. INTRODUCTION

There are over 4 billion mobile phones out of about 7 billion people in this world. There are also over 1 billion landline phones [1]. Thus, phone is a good ubiquitous device and hence use of phone is the key way to effectively bridge the Digital and Language Divides, and thus provide all the key benefits of the Information Age to most of the people in the world. Information access for all people in the world, dominated by the Base of the Pyramid People (BOP), is the key for Economic, Social, Cultural and other developments including increased world peace [1].

Thus, it is very important that users can use a mobile device for not just talking to another person, but also to access the Internet and various on-line applications in easy, effective and efficient ways. This means that we need a very good Human Computer Interaction (HCI) method for such devices. In order to have a very good HCI, we need to address the following two broad issues:

(a) Easily accessing and interacting with the Internet and Internet applications

(b) Easily using the on-device content and applications

While high end mobile devices (like smart phones and PDAs) are reasonably good in addressing these issues, they mainly serve the technically literate people, leaving out majority of the phone users. Nontechnical, semi-literate and illiterate people have great difficulty in using existing interfaces, mainly because of the complexity in learning and the literacy needs. Besides, many people cannot afford such high end mobile devices. The User Interface (UI) for low end and medium end mobile devices are even more difficult as such devices do not have some of the nice features that high end devices have (e.g. touch screen, scrolling, larger size screen/keypad).

As stated above, since low and medium end mobile devices would need to provide easy access to the Internet and also should have all the key features that high end mobile devices, tablets and computers have, we need a good solution for the two broad issues (a) and (b). Two technologies are essential to achieve these goals – a **Natural Language Interface and an Intelligent Agent.**

Existing solutions for (a) and (b) use three major approaches –

(a) **manual scrolling** which is used in PDAs and smart phones
(b) **touch screen** to easily select and navigate (mainly for high end phones) and
(c) **re-writing the content** in small amount that can fit on the screen of a mobile device, mostly used in low or medium end mobile phones.

While manual scrolling provides a nice way to go over the text content, it is more difficult to understand the content of the whole page when a page is rich with images, pictures or multimedia content. This process of grasping what's on the whole page by looking at a small amount of content at a time and then making an understanding of what's really on the page is a very difficult task and very subjective as it depends on brain's ability to correctly assemble various views in a coherent overall page content. Many users cannot do this function well. Besides, visually searching content by scrolling is also difficult.

There are also some major problems in rewriting the content into small chunks to fit on a low or medium end mobile phone screen. First of all, there are over **three billion** websites. Re-writing them with another language (for example, WML or CHTML) so that the content can be viewed on a cell phone screen would cost a lot, and hence not practical. Secondly, if a website is re-written into several websites of small chunks of content, the navigation would be much more difficult. For example, if a regular website has 100 pages and these pages are broken into, say 1000 small pages, then to navigate to 999^{th} page would take a very long tree of selection process; thus making it almost impossible to use. Hence, websites on the Wireless Web (that mainly use WML or CHTML pages) usually take top 1% or 2% content of a regular website on WWW and re-write that in WML or CHTML. Today, there are only a few million websites on the Wireless web as opposed to over 3 billion websites on WWW. For problem (b), there is no good solution for low and medium end phones.

Thus, it is very important to solve these problems associated with all mobile devices in an effective, efficient and easily affordable way. In **this paper**, we propose a **solution to these problems using an Intelligent Agent that uses the features on the page itself, Artificial Intelligence and Natural Language Processing / Understanding.** Our approach uses similar algorithms as our eyes and brain use when we access and navigate through the Internet. Our proposed NLU and Intelligent Agent (IA) based approach also works well with any basic phone, like a land-line phone. It also works with tablets and computers.

Apart from rendering and displaying desired content, another key role that an IA plays is easy interaction with the user (problem (b) mentioned above). The IA ensures that all key UI related issues including having a simple, easy to use, easy to learn, and easy to navigate (with a short navigation tree) UI are addressed properly. This in turn ensures that the IA provides a very good UI for non-Internet based applications like various local applications in a mobile device itself.

Section II describes the "rendering", the core concept to our approach. Section III & IV describe an elegant solution to the rendering problem. Section V describes how an Intelligent Agent (IA) uses the rendering technique in a MicroBrowser – an Intelligent Agent based browser for a mobile device and solves the problems associated with rendering and navigation. It also describes how IA significantly improves the ease of use by using a simpler UI in a mobile device. Section V also elaborates how Intelligent Agent based User Interface significantly eases both content retrieval and interaction via form filling; thus allowing easy and effective use of all on-line applications.

Section VI describes how rendering can be applied to a land-line phone. Section VII describes how rendering can be used in computers and tablets. Section VIII describes how the Intelligent Agent addresses key UI related issues in a very effective way, Section IX describes the benefits of our approach for various populations and Section X provides Conclusion.

II. THE RENDERING PROBLEM

As we know, the Internet was designed with visual access in a relatively large display screen (like a 8.5 inch x 11 inch page) in mind. Thus all the content are laid out on any website and webpage in a manner that attracts our eyes in a large screen. Retrieving the desired content (which is much smaller in size than the total content on a webpage or website) from a typical webpage / website and displaying that into a much smaller screen (like in a cell phone or PDA) is a challenging task. This process of retrieving and converting most desired content from a large source of content into a much smaller display is called **"rendering".** It is important to note that such rendered content are also good to play in audio as the content is short and most desired. The rendering process significantly eases the navigation, especially within a page as rendering includes finding the most desired content in an automated way. Because the most desired content can also be searched from multiple web pages, rendering also to some extent helps navigation between pages. Clearly, finding most desired content is very difficult as just string matching does not produce most desired content in a reliable way. Accordingly, more than string matching, including Natural Language Understanding(NLU), Natural Language Processing

(NLP),and Artificial Intelligence algorithms are needed; thus making the process very complex.
There are other key issues with rendering – e.g. how to find most desired content when we have multiple similar content found on a page and how to re-assemble all related desired content in a logical manner.

Apart from rendering most desired content, we also need to render most desired forms associated with many Internet applications. Form rendering includes accessing, filling, navigating and submitting sequence of appropriate forms associated with an Internet Application. **Since the need for on-line transactions is increasing steadily, the need for a solution for automated generic form filling has become very crucial.** Automated generic form filling means automatically getting a form from a website for a particular application and filling all the required fields by voice or keypad entry, submitting the form, retrieving the next form and continuing the process until all the required forms in an application are filled, submitted and final result or a notification is obtained.

In addition to rendering appropriate content & form, we also need to render streaming audio and video content as such content is increasing rapidly on the Internet.

Apart from rendering to a small screen, we have rendering problems for rendering to a Tablet PC, regular PC, TV or other devices (see [1] for more). In such cases the issue is not to find the most desired content that would easily fit on a small screen, rather, finding the most desired content that eases the viewing and minimize information overload. For many users (especially those not familiar with the Internet), the amount of content on a webpage is too much (even if they can see all content easily) and retrieving the most desired content and displaying that nicely, significantly eases the usability.

III. AN ATTRACTIVE SOLUTION USING VOICE INTERNET

Voice Internet can truly bridge the Digital and Language Divides ([1], [2], [5]) 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**:

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

b. **no need to deal with small screen or small keypad as users basically talk and listen.**

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

d. **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.**

e. 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 & quickly and make sure that content heard are the content that were desired.

IV. HOW RENDERING IS ACHIEVED

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

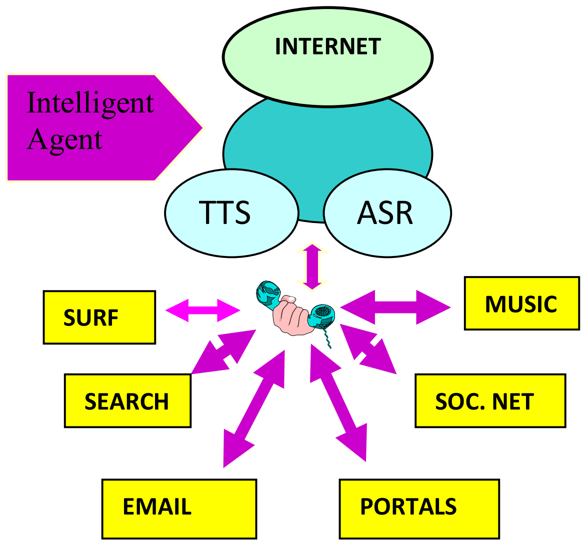


Fig. 1: The Intelligent Agent (IA) and key Features of Voice Internet

(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 keypad entries. The Intelligent Agent is capable of learning user preferences, to continually improve ease of access over time.

Thus, 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.

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

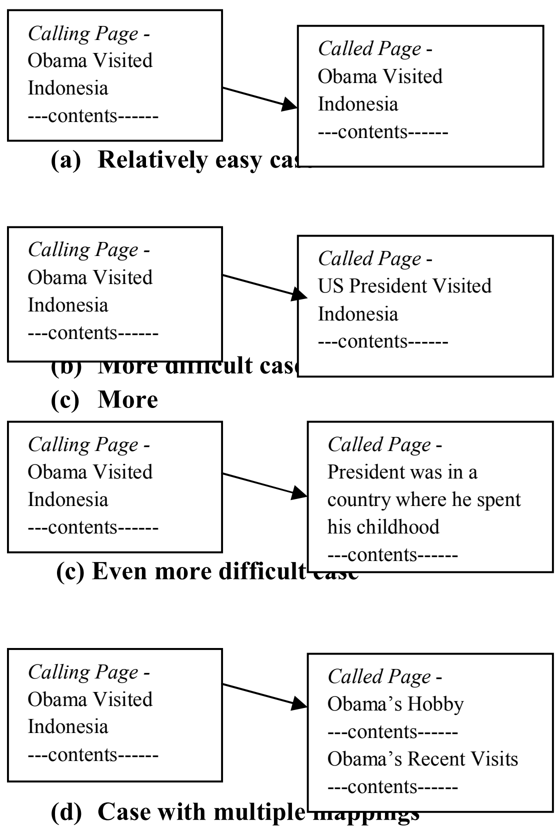
Thus, a user can seamlessly access any content on the Internet, interact with any form(s) and **complete transactions** like shopping, banking etc. using a simple phone and his/her own voice. Another key feature is that content can be translated in real time into another language, providing audio access in a local language translated from, e.g. English-language web pages, to 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).**

4.1.1 *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 content really be provided from any web site on the Internet?
(2) can the existing Internet content be rendered in a manner that the rendered content can be obtained in real time, are 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 content fall under the first and second categories. **Fig. 2 shows** the complexity of



(a) Relatively easy ca

(b) More difficult cas

(c) More

(c) Even more difficult

(d) Case with multiple

Fig. 2 Complexity in Rendering Content

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.

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.

Below is a short description of general algorithms to determine Page Highlights, perform Rendering and Navigation for Cases (a) to (d) described above.

A- Algorithms for Rendering, Navigation and Selection of Appropriate content

Various algorithms are used [10] by the Intelligent Agent IA to do rendering, navigation and selection of appropriate content. All these algorithms use the information already available on the visual web page including columns, tables, frames, color, boldness, various fonts size, positioning, pop up windows and **semantic meaning** of words.

A.1 Page Highlights

"Page Highlights" provides important information or highlights of the page referred by the URL. A small number of highlights (e.g. 3) are read at a time and users are given with options to select any one highlight/topic at a time. Once a highlight is selected, the content that describes the highlight is read. The basic assumption here is that the related content exists and is somewhere in the current page or in linked pages a few levels down. There are a few possibilities as described below:

1) Related content is on the same page as the highlight:
The IA will start reading the selected content from the current page. Selected content is found by using a few algorithms called **"Algorithms for Related Content Selection"** as described below.

2) Selected highlight is a link:
In such a case, IA will go to the linked page and will find the relevant content and read it using "Algorithms for Related Content Selection" as mentioned above.

3) A few related content on the linked page:
If the linked page has several content related to the selected highlights, IA will provide an option to further fine tune the selection. And then select and read the desired content.

4) Several content on the linked page but none of them can be easily identified and selected:
IA will provide highlights of the linked page so that user can make further selection and the process continues until the related content is found.

The page highlights are determined using algorithms that we use to find a highlight on a page by looking at the page content. Thus, it is based on font size, link, color, size of the content, semantics & language understanding. IA looks into the HTML/XML tags and determines boldness, font size, link, length of the contents and calculates the highlights. Other algorithms are also used to determine which highlights are more important and hence should be read first. Fig. 3 shows basic algorithms to determine highlights. Clearly, the highlight algorithms described in Fig.3 provide key information in a logical manner that we normally use when we look into a page visually. It also provides good navigation to the next page and provides only the relevant content from the linked page (using the algorithms described in *Algorithms for Related Content Selection below*). Thus, the algorithms "renders" information from visual web page into audio that is concise, desirable and pleasant to listen to.

Figure 3: Algorithms to determine page highlights (part 1)

If the content is with largest font size (largest font on the current page but not part of a banner)

Then

{

this is highlight #1.

If this content is a link, then related content on the linked page will be read when this highlight is selected

Else

Associated content on the current page will be read when this highlight is selected. **(Continued to next page)**

In this case association is determined by next paragraph or table or frame etc. that is directly related to this highlight.

If there are more than one content with largest size and none of them are links, then

priority is assigned to the highlight with largest content associated with it. If they

are all links, the one with highest # of words has the highest priority. If they are

mixture of links and non-links, then priority is assigned to the links.

}

else

If the content is flash but not part of a banner

Then

{

this is highlight #2.

If there are more than one flash content, the priority is decided based on the same algorithm outlined above for the largest font size.

}

else

{

use 2nd largest font, followed by 3rd largest font etc to determine the priority. When font sizes become same, then priority is determined using same algorithm as for the largest font size except that a content with bold feature has the higher priority.

}

Note: Minor variations of these algorithms are possible. E.g. Flash content might be treated with highest priority instead of 2nd highest priority. The goal is to use an algorithm that well represents how a human selects highlight when look at a visual website.

Figure 3: Algorithms to determine page highlights (2nd part)

A.2 Key words, query based rendering

Apart from using highlights, rendering can be done by providing key words of a page and then using queries. In general, queries should include one or more of the key words. Queries can be a simple sentence or just a word or a few words. Using "Algorithms for Related Content Selection" and the key words, the related content is selected.

User might already know a few key words for some website and might just try with those words without listening to key words of the page. Alternatively, simple key word may not be found in a page but a user still can say any word(s) as query and if there is a match the relevant content would be read out. If there are confusion (e.g. multiple matches), the IA will ask more questions to the user to minimize ambiguity and improve relevant content selection. If there is no match for the word(s) asked (i.e. for more difficult cases), semantic analysis and language understanding can be used to find and select relevant content. For example, user might say "U.S President" but the page may not contain this sentence at all. Instead the page might have the name of one U.S President (say Obama) and so the language understanding unit will match this with the query and will read the content associated with this. If the page contains "Obama" in multiple paragraphs which are not associated i.e. not next to each other i.e. are under separate topics, the IA will read the 1st sentence of each topic and ask the user which content he/she would like to hear.

A.3 Algorithms for Related Content Selection

Related contents are selected based on various algorithms depending on the level of complexity.

a) Parsing and word matching

These are used to match words in the label of the highlight with the label of the highlights on the linked page. After a match is found, the content associated with the match is selected. Association is based on frames, tables, semantics etc. If multiple associations are found, then most important association is selected first. Importance of association can be determined with semantic meaning, language understanding and processing, and simpler algorithms like paragraph size. To save computations, matching for all words in a sentence is not usually needed. The relevant content can be found after matching a few words as the page may not have more than one instance of the selected words in the desired sequence. If similar sequence of words are found more than once, content are read based on the priority determined by the size of the paragraph associated with the matches (to speed up using simple computation) or by use of semantic meaning.

b) If no matching found

If no matching is found, the page is tested for whether it is a content rich page or link rich page. If it is a content rich page, the key body of the content is identified and selected. Key body is identified by length of the content, font size, color, table, frame, semantics (see following sub-sections for more). If it is a link rich page, then the highlight of the page is calculated and presented so that user can go down to the next level to find the desired content. Content richness or link richness is determined by calculating the content density and link density.

Content Density:

Content density is determined by counting total number of words (or letters as appropriate e.g. for Chinese or Japanese languages) without considering links divided by total number of words (or letters) considering both links and non-links.

Link Density:

Link density is determined by either counting total number of words (or letters as appropriate e.g. for Chinese or Japanese languages) in all links or counting total number of links divided by total number of words (or letters) considering both links and non-links.

c) Semantic analysis, language processing and understanding

The computational effort needed in (b) is not that high and hence preferred. However, if good content cannot be found in (b), a more compute intensive algorithm can be used to find more relevant content. These algorithms are basically based on **semantic analysis, language processing and understanding** using context information. **Machine Learning** algorithms can be used to improve semantic analysis and language understanding. With much improved language understanding, it will also be possible to make a summary of long paragraph or content. In such a case the key concept or statements in the 1st (and sometimes 2nd) and last paragraph are noted. Content with similar meaning (direct or implied) are gathered and duplications are removed resulting in a summary of the Section. This is just an example. Any other language understanding based "summary" computation can be used.

4.1.2 Rendering Algorithms for Form Filling

As mentioned, Page Highlights finds key information on the page using algorithms similar to the algorithms used by sighted users. Form filling is done in a similar way by presenting forms as a key Page Form Highlights and also creating appropriate questions, taking the text/voice inputs from the user and then filling and submitting the form. These key steps are done using the information available on the visual web page itself and employing appropriate algorithms to use all such information including <form> and related tags, text content, color, font size, links, paragraph, and amount of texts as described for rendering regular text content. Some language processing algorithms are also used to further refine the form filling process. This is similar to how the brain of a normal sighted person renders information from a visual page.

Generic form filling (basically allowing successfully filling and submitting any type 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 before) to make it simpler for the users to select corresponding form or application.

It is possible that all above mentioned algorithms will fail to select related content (or sequencing appropriate forms). This is because while semantic analysis and language understanding has come a long way, these have still long way to go to make it comparable to human understanding of language (more in 4.1.3). In such case the whole page will be read so that user can do the rendering directly.

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.**

4.1.3 *The Need for Natural Language Understanding (NLU) in UI and Rendering*

Use of NLU is already mentioned in Sections 4.1.1 & 4.1.2 in help determining Page Highlights and appropriate content (or form) when a page highlight is selected. In such cases, mainly semantics of words are used. However, NLU plays a major role in the UI to improve the HCI as mentioned in the Abstract and Introduction. As also mentioned, the role of NLU is the key for the BOP (Base of the Pyramid People) 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 information 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 BOP will not use structured words or sentences. A good NLU system will significantly ease this process. Clearly, NLU is needed for a very good UI as well as better rendering so that the retrieved content is more appropriate. In fact, a very good NLU is key for a Question and Answer System which is very much needed for the BOP. **Thus, NLU is a necessity for the BOP to really easily and effectively use low and medium end mobile devices; and NLU based HCI is a great useful feature for high end phones, tablets and computers.**

This has been verified by ***many users who have been successfully using the Voice Internet (from InternetSpeech, Inc) over 5 years - they are highly satisfied and recommended the service to many others.***

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.

V. INTELLIGENT AGENT (IA) BASED USER INTERFACE (MICROBROWSER)

As mentioned, an Intelligent Agent based User Interface is the key to enable any user (including illiterate people) to easily and effectively use mobile devices. The IA overcomes the existing problems by automating the key tasks while allowing a natural user interface using user's voice (or typing) and Natural Language Understanding. The IA includes the rendering feature described above. It also includes the general UI related issues including having a simple, easy to use, easy to learn, and easy to navigate (with a short navigation tree) UI and other related issues. This in turn ensures that the IA provides a very good UI for non-Internet based applications like various local applications in a mobile device itself (more details in Section 8).

IA based solution not only makes it much easier to effectively use mobile devices by non-technical people but also by technical people as the key tasks now handled by manual scrolling and rendering by user's eyes and brain are automated.

Using the Page Highlight feature of Voice Internet Rendering solution, users see only a small number of highlights (usually 3) which easily fit on a mobile device screen. Also, the rendered content when a Highlight is selected, are usually a column of text information which easily fit on a cell phone screen. Hence, the rendered content are viewable at ease. It can also be heard at ease.

The displayed content can be automatically scrolled using desired speed if audio option is selected (i.e. if audio is turned on). If audio is turned off, then auto scrolling will be off and user will scroll vertically at user's choice. This is what results a MicroBrowser or "true" wireless Internet accessibility, viewing & navigation that does not need any re-writing of the website, and present content at ease in a meaningful way on a cell phone or PDA screen. Because, **rendering does data mining** i.e. finds desired data from the page as opposed to re-writing the page into several small tiny pages, the navigation becomes much easier and practical. Besides, all content are accessible to the user as there is no re-writing of top 1% or 2% of the whole content as usually done in Wireless Web. Thus, a MicroBrowser enables anyone to view and navigate any website at ease. It also enables a PDA user to get to any desired content without manually scrolling and trying to figuring out what's on the whole page.

Apart from accessing desired content (including multi-media content), MicroBrowser allows one to fill various types of forms and hence support all on-line applications including online transactions. This way **existing wireless web automatically get extended to the World Wide Web (WWW) i.e. whole WWW becomes available on any cell phone screen at ease with easy navigation.**

Thus, a MicroBrowser allows any user to easily enjoy the content from the Internet as well as interact, easing the use of all on-line content and applications.

Apart from providing an easy and effective interface for the Internet, an IA based UI also significantly helps users to enjoy non-Internet based features. For example, the IA provides simpler interface to fill various forms, easy access and interactions with all on-device applications.

VI. HOW RENDERING IS APPLIED FOR LANDLINE PHONE

The Voice Internet rendering process described in Sections II – V can be used by any phone including a basic landline phone like POTS (plain old telecom service). This is because the rendering engine is running on a server and clients just take user's input (using voice or keypad) and provides output using audio (usually .wav file). Thus, it does not matter whether a user uses a landline phone or mobile phone – what is needed is just a phone that can make a phone call and establish a phone connection. The rendered content is heard over the phone. While on a page, a user can move around easily by pressing appropriate keys or using voice commands although in most cases voice command feature is turned off while on a page to avoid unwanted effect from noise.

VII. HOW RENDERING WORKS FOR COMPUTERS AND TABLETS

Rendering to computers and tablet PCs is a bit different than rendering to small screen (as in a mobile phone) or to audio. In such cases the issue is not to find the most desired content that would easily fit on a small screen, rather, finding the most desired content that eases the viewing and minimize information overload. We still need to calculate Page Highlights, find most appropriate content when a Page Highlight is selected. But we don't need to worry about UI issues related with small screen and small keypad. However, we still need to address UI issues from a different angle to ensure that the HCI is very good. For many users (especially those not familiar with the Internet), the amount of content on a webpage is too much (even if they can see all content easily) and retrieving the most desired content and displaying that nicely, significantly eases the usability. For example, many people get confused with navigation links, tables, pictures, frames, toolbars, pop-ups, links displayed with an ad versus regular links, how many levels one would need to go down to get to the desired content, streaming files, various series of confusing links when one would like to download something and the like.

Thus, instead of displaying all content on a page, we need to display only the Highlights or relevant content or a small set of content PLUS simple text or buttons to point (may be briefly describe) the content and show how to easily navigate. Besides, for many people, use of a regular keyboard is complex and difficult. A good size simpler touch screen can be used to make various selection and navigation easier.

And the value of NLU (especially with speech recognition) is also very important here, not from small size keypad standpoint but from complexity of the keyboard and typing standpoint. Other advantages of NLU in rendering still apply.

One key difference worth mentioning is the fact that the IA and rendering engine can reside in the computer or tablet (as opposed to in the server, especially for low and medium end mobile devices). Since such devices have more computing power, performance of IA and associated UI is usually better in a computer or tablet.

VIII. HOW IA ADDRESSES KEY UI RELATED ISSUES

Apart from providing an easy and effective interface for the Internet, an IA based UI also significantly eases Human Computer Interaction (HCI) i.e. it helps with general UI features. Such features include basic interfaces and interactions (e.g. manipulating files, finding on-device information, on-device forms & applications). These also include inputting data, viewing on-device desired content (part of the rendering), and navigation within the device. It is important to note that just providing a simple interface to begin with can solve some major problems in using such devices. E.g. an elderly person or a person new to use such devices, are not usually comfortable when he/she sees too many options, too many icons, too many forms and the like. They also do not like to see features in text written on buttons. For some people even visual symbol is not clear. E.g. for an illiterate person (say a farmer), a symbol may mean something else than what it means to a more educated technical person. Thus, a symbol for an apple will really mean something about an apple (not Apple's logo), and a symbol for an orange will mean something about an orange. Additionally, if there is a flashing light and voice saying something like "press me or touch if you like to know more about me", it will be much more clear and will be more appealing to such user group.

The key idea is to address all basic HCI needs using good models (like cognitive complexity model) and framework (like Interaction Framework). Moreover, Intelligent Agent based UI addresses other missing issues - like minimizing total number of steps in completing a transaction, e.g. Interactive Framework addresses language issues between User, Input, State and Output; but it does not addresses how to minimize the total number of cycles of this loop needed to complete a transaction. As today's applications are getting more capable and complex, it is very important to minimize the total number of interactions needed. An IA based UI, especially using NLU, minimizes the total number of interactions in an optimal way.

Let's use an example using shopping on Amazon. With conventional method, a user would need to go to www.amazon.com, search for the desired book, then fill forms to buy the book etc. While the process is relatively easier in a computer, it is more difficult to do the same on a high end mobile device or smart phone. And, of course, it is not possible using low or medium end phones (may be done in a medium end phone using a special app). With IA based UI using NLU, user can use any phone (low, medium or high end) and just say

"Go to Amazon, search the desired book and if the price is within the budget, then buy the book using the credit card on file".

The IA will parse this sentence into smaller parts, fill all necessary forms and then will complete the transaction. Such a UI will be attractive to almost anyone (including illiterate, elderly, Digital Divide and other people); and thus will make mobile devices more appealing and successful.

In fact, as already mentioned, and as more clear from this example, **NLU and IA based UI & HCI is a necessity to do this type of applications using a low or medium end mobile phone. And, of course, such NLU & IA based interface will also significantly improve the interface for high end phones, tablets and computers.**

IX. BENEFITS TO VARIOUS POPULATION

Today, there are over 5 billion phones (over 4 billion mobile phones and over one billion landline phones) out of about 7 billion people in the world. Most people (especially at the BOP) use just the basic features of a phone i.e. making / receiving a call and talking. Not many people even use the text messaging (not to mention other features like setting alarms, making schedules etc) as many of them are illiterate. These "other features" are increasing rapidly as many more computing functions are moving to mobile devices. As already mentioned, existing mechanisms are good for high end mobile devices (e.g. smart phones and PDAs), and serve mainly the technically literate people. Nontechnical, semi-literate and illiterate people have great difficulty in using existing interfaces. With more features moving to mobile devices, it is key that all such features are accessible and usable by a much larger population including BOP.

Our solution presented in this paper allows any user (especially from the BOP) to use above mentioned key computing features moving to mobile devices. Thus, illiterate, semi-literate, Digital Divide, Elderly, Blind, Visually Impaired, Highly mobile, children and all other kinds of users would be able to use all key features and hence enjoy the benefits of the computing as well as Internet applications easily in a natural way.

Our solution not only makes it much easier to effectively use mobile devices by non-technical people but also by technical people as the key tasks now handled by manual scrolling and rendering by user's eyes and brain are automated. As we argued, such an automation is, in fact, becoming a necessity when one would like to use various complex on-line applications including e-Services. This is true for all people – literate, illiterate, technical and non-technical.

X. CONCLUSION

Today, mobile devices are highly ubiquitous and used by majority of the world population. Many computing functions/features are moving from desktop to mobile devices very rapidly. However, existing user interface (UI) for such mobile devices is difficult, cumbersome and require some literacy. Thus, such devices are mainly effectively usable by some educated technical people. Since information should be for everyone in this Information Age, it is important that we provide all the benefits of the Information Age (i.e. Internet / non-Internet Applications) to everyone in a simple, effective, and easy to use way that is also easily affordable.

We have presented an elegant Human Computer Interaction (HCI) and associated User Interface (UI) solution using Natural Language Understanding (NLU) and Intelligent Agent (IA) so that anyone including illiterate, semi-literate, Digital Divide, Elderly, Blind, Visually Impaired, Highly mobile, children and Base of the Pyramid people (BOP) can easily & effectively use all computing features on a mobile device.

Through the use of NLU & IA, our solution also addresses all basic HCI issues while minimizing the total number of interaction cycles needed in completing information retrieval, transaction or interacting with on-device applications. As NLU advances, the interaction and usability will be further enhanced.

We have argued and shown how NLU and IA based HCI & UI is becoming a necessity to effectively and easily use various complex on-line applications including e-Services using low & medium end mobile devices. Our proposed approach also makes the UI much simpler, easier and more effective for all other devices including high end mobile device, tablets and computers.

Because of the importance of NLU and its advancement, we are doing some active research in this area using some bio-inspired and brain-like algorithms as humans are very good at NLU, and hence anything that we can learn & borrow from our brain should result a better NLU system.

Our proposed solution using **Voice Internet and its rendering capability** is a practical solution to provide Internet to many people, especially BOP and effectively bridge the **Digital & Language Divides**, thus helping many people around the world in their **Economic, Social, Cultural and other developments including increased world peace.**

REFERENCES

[1] E. Khan, "Internet for Everyone – Reshaping the Global Economy by Bridging the Digital Divide", *Book- ISBN*

978-1-4620-4251-7(Soft Cover), 978-1-4620-4250-0 (Hard Cover), Aug 2011.

[2] E. Khan & E. Aleisa, "e-Services using any Phone & User's Voice: Bridging Digital Divide & help Global Development", *IEEE International Conference on Information Technology and e-Services*, March 24-26, 2012, Tunisia.

[3] E. Khan, "Information for Everyone using any Phone – – Global Development via Education, Entrepreneurship and Innovation", *International Convention on Rehab. Engg. & Assistive Technology in Collaboration with ACM*, July 2010, Shanghai, China.

[4] E. Khan, "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, 2009, Bangkok, Thailand.

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

[6] J. Stiglitz, "Making Globalization Work", *W.W Norton & Company*, ISBN 13:978-0-393-06122-2, 2006.

[7] C. N. Hammack et al, "Automated Ontology Learning for a Semantic Web", *Dept of CS, Uni. Of Nebraska*, Feb2000.

[8] D. Ciobanu et al, "Predicting the Next Page that Will be Visited by a Web Surfer Using Page Rank Algorithm", *International Journal of Computers and Communication*, Issue 1, Volume 6, 2012.

[9] E. Khan et al, "Internet Access by Talking and Listening Using any Phone: Help Global Development by Effectively Bridging the Digital and Language Divides", *International Journal of Information Studies*, Vol. 4, No 1, 2012.

[10] E. Khan," Method for Facilitating Internet Access with Vocal and Aural Navigation, Selection and Rendering of Internet Content", *US Pat* # 07712020, May 4, 2010.

[11] S. Pastore, "The Technology Evolution in Web Design, Development and Deployment Supporting Web Scientific Applications", *International Journal of Computers and Communication*, Issue 1, Volume 5, 2011.

[12] Nikola Vlahovicet al, "Managing Online Learning Resources Using Web 2.0 Tools – a Croatian Experience", *International Journal of Education & Information Technology*, Issue 1, Vol 5, 2011.

[13] Ivaylo I. Atanasov, "Reasoning on Service Interaction in Mobile Networks", *International Journal of computers and communication*, Issue 2, Vol 5, 2011.