**Proceedings of International Conference on Advances in Computer Technology and Management (ICACTM)**
***In Association with Novateur Publications IJRPET-ISSN No: 2454-7875***
**ISBN No. 978-81-921768-9- 5**
**February, 23rd and 24th, 2018**

# PHONET-A VOICE BASED WEB TECHNOLOGY

PROF. KIRAN S. PATIL
Assistant Professor, MCA Department, Alard Institute of Management Sciences, Pune, - 411057
Email: - patilkirans@gmail.com

**ABSTRACT**
Voice based web access is a rapidly developing technology. **PhoNET** is a solution for these and many other problems faced by the netizens. The basic idea is that using an ordinary phone to browse the web and the primary motivations are: to provide a widely available means for creating new interactive voice applications; addressing needs for mobility; and addressing issues inaccessibility. Basis of the idea are the age old IVR systems used to serve information for the dialers through a pre programmed process. Phonet is a very long journey from the IVRs; it involves the most complex technologies of the century Like Speech Recognition (SR), Text to speech (TTS) conversion and artificial intelligence (AI). This enables a user to be connected to internet as long as he has access to a phone. PhoNET uses the traditional HTML content so the web site need not be rewritten or redesigned. We present a detailed analysis in the most possible simplest way of how the technologies like SR, TTS and AI are integrated to develop a intelligent Platform (phoNET) to achieve voice based web access which involves Document processing and Document Rendering. In Document Processing we describe two approaches, telephone browsing and transcoding, focusing mostly on the former since that work is more mature. In Document Rendering we present the major problem i.e., the relevance of cognitive thought to text rendering along with its most suitable solution. In the end we examine the challenges and further developments involved in practical application of the proposed technology-The phoNET.

## 1. INTRODUCTION

Today's telecom business has seen recent growth, especially in bandwidth infrastructure for long distance (LD) and data. The industry is currently experiencing strong growth in the wireless segment as mobile devices prove to be very popular with both consumers and business. An evolving market segment is "Internet anywhere," and many companies are trying approaches to present viable products for this market. One approach is Internet access over wireless devices such as cell phones with a screen. However, this method has inherent limitations such as small screen size, lack of a keyboard, the need for a special device (web-enabled phone), the need to rewrite and maintain a special website, and severe bandwidth constraints using wireless data transfer protocols. Another approach that is becoming popular is voice-based limited Internet access, which overcomes all of the limitations of the wireless data devices but one; they still limit access to the few sites that are re-engineered for voice. They typically deliver content such as news, weather, horoscopes, and stock quotes, etc. over the phone. These companies are called "Voice Portals." Voice portals were the first web applications that tried to integrate websites with voice which gave birth to the enterprise based PBX systems.

Other solutions, such as Personal Digital Assistants, phones with display screens and other Internet appliances, are available, but have limitations. Users must have special hardware with intelligence built in, and often must view the Web through small, difficult-to-read screens. Such devices are often expensive, as well.

Our solution, which presents a third option, gives users all of the benefits of the voice portals, yet has complete access to the entire Internet without limitation. With our Voice Internet technology **phoNET**, anyone can surf, search, send and receive email, and conduct e-commerce transactions, etc. using their voice from anywhere using any phone, with the more freedom of movement than a standard Internet browser which requires a PC and an Internet connection.

**PhoNET** technology is faster and cheaper than existing alternatives. Today, only the largest of companies are making their Web sites telephone-accessible because existing technology requires a manual, costly and time-consuming re-write of each page. With the voice internet technology-**phoNET**, existing Web pages are used, allowing users to leverage their Web investment. The software dynamically converts existing pages into audio format, significantly lowering the up-front investment a business must make to allow users to hear and interact with their Web site by phone.

## 2. MOTIVATION

The primary method of access today continues to be the computer, which has certain advantages as well as some limitations. Computers offer a visual Internet experience that is usually rich in content. Some basic computer skills and knowledge are needed to access the Internet. But, computer-based access is proving insufficient for the professional on the move. When in the car or away from the office or computer, accessing the Web is difficult, if not impossible. And, an increasing number of people prefer an interface that allows them to hear and speak rather than see and click or type.

The computer-based Internet experience also does not meet the needs of another segment of the population – the visually impaired. Neither visual displays of information, nor do keyboard-based interactions naturally meet their needs, nor this segment. Often unable to benefit from all that the Information Age has to offer.

Some existing Internet users have also identified problems with the visual Internet experience. Pages are increasingly full of graphics, advertisement banners, etc., which move, flash, and blink as they vie for attention. Some find this "information overload" annoying, and lament the delays it creates by severely taxing the available bandwidth.

**The "Digital Divide"**

While computers and their use are on the rise, they're not ubiquitous yet. A large segment of the population still doesn't have access in the United and other parts of the world. Thus, Internet is limited to only a small fraction of the world

population; the majority is left out from the Internet. This gap between those who can effectively use new information from the Internet, and those who cannot is known as **The digital divide.** Bridging this digital divide is the key to ensure that most people in the world has the capability to access the Internet. Making computers ubiquitous is not a very attractive and feasible solution, at least in the near future, because of various barriers. One key barrier is cost, although the price of a computer has come down significantly in recent years. Insufficient visual Internet Infrastructure is another barrier in many countries and it will take a while to build such infrastructure. Other consumers have a basic distaste for complex technology, which prevents them from accessing Web-based information via a computer. A more natural, less cumbersome way to interface with the net would provide them an opportunity to experience the Internet as well, thus bridging the Digital Divide.

**The "Language Divide"**

Today more than eighty percent of website contents are written in English language. People in China, Japan and other countries in Asia, countries in Europe and Latin America speak a language other than English as their native language. These people are left out of a significant portion of the World Wide Web. For example, Japanese or a Chinese can not understand the content of CNN or New York Times. This gap of not having access to major part of the Internet because of language barrier is called **"The Language Divide"**.

**Bridging this Language divide is the key to ensure that most people in the world has the capability to access the major part of the Internet.** The demand for machine translation is growing phenomenally as more people each day embracing the Internet. A service that translates the accessed information into the desired language would clearly add value to these users.

As the need for alternative access to the Internet becomes more evident, several technology companies are pursuing solutions. Their products include "smart" cell phones with visual displays, intelligence built into the handset, and voice-activated Web sites. These products address different aspects of the problems outlined above.

While these alternative technologies are in the pipeline, few are ready for market. But the very existence of a race to market by many companies is evidence of a large potential market.

## 3. THE CHALLENGE

**To integrate existing technologies, or develop new technologies, to make simple, affordable, alternative Internet access possible.**

As the need for an alternative access method to the Internet has become evident, progress continues to be made by technologists to provide such solutions. One key area of focus has been voice-based technology, which would allow a very natural interface for most people, and address the limitations described earlier. A voice interface provides an alternative to the visually based interface. A device such as the telephone provides a readily accessible alternative to the computer.

Several technologies existing today are keys to the solution, but the problem lies in successfully integrating these technologies into useful applications of greater value than their individual components. These technologies include:

- Voice Extensible Markup Language (VXML) which is an extension of HTML, the normal language in which Web pages are created. This technology adds voice capability to a Web page. The page can then be displayed, as usual, over a computer, but it can also be presented in audio format with voice navigation.
- Speech Language Application Tags (SALT) specification for supporting multimodal communication from PCs, cell phones, PDAs and other handheld devices. For example, input can be voice (such as asking for directions) and output can be data (a map pops up). SALT is a lightweight set of extensions to existing markup languages, allowing developers to embed speech enhancements in existing HTML, xHTML and XML pages. As with VoiceXML, applications will be portable - thanks to the separation from the underlying hardware and platform.
- Speech recognition (SR), which allows computers, through the use of software, to recognize spoken language, eliminating the need for the computer keyboard as an interface. The vocabulary recognized in products using this technology tends to be limited.
- Text-to-speech (TTS), which allows text to be converted automatically to synthetic speech. It allows communications between computers and humans through a "natural" interface, such as speech.
- Telephone integration is the key to interface with computers from a remote location. A protocol is needed to communicate with the computer from a telephone using voice. This also includes multimedia integration (e.g. with .wav files).
- Intelligent software agents are needed to automate communication between a telephone and a computer, a computer and a Web site, to interpret the contents of a web page, to extract key information that makes sense in audio, to efficiently navigate through web pages, and to manage access to the Internet.
- Language processing allows translation to other languages, understanding and interpreting of structured sentences. Natural language processing allows us to understand and interpret human languages.

The first technology listed – VoiceXML – is a very elegant solution that leverages technology specifically developed for audio Internet access. However, it requires that Web sites be customized, or VXML-enabled. This means rewriting the web pages in VXML. According to analysts, today there are more than a billion web pages. Assuming that it takes one hour and costs about $100 to rewrite one page, the cost to voice-enable all sites would be about $100B. Clearly, it will take several years before the majority of popular pages are VXML-enabled. Today, only a very small portion of the total Web pages is voice-enabled using VXML.

The second technology listed – SALT – is another elegant technology that allows developers to embed speech enhancements in existing HTML, DHTML and XML pages. However, like VoiceXML, it requires that websites be

rewritten or enhanced with SALT and hence it will also take many years before majority of popular pages are SALT-enabled.

The paper presents a solution that successfully integrates the other technologies listed into a useful, audio-based approach for accessing the Internet today. It is independent of the timeline, interest and willingness of content providers to update their pages to be VXML or SALT-enabled.

Another approach is to provide Internet access over wireless devices such as palm pilot or a cell phone with a screen. However, this method has inherent limitations such as the small size of the screen and the need for a special phone. Also there is need to rewrite the website in WML. Today's wireless Internet industry is facing many challenges due to limitation of bandwidth and small screen. The cost of cell phone based Internet access is very high and users do not want to pay high service fee. Also our eyes and fingers are not changing but the devices are getting smaller and smaller. Thus, existing visual based access is going to be even more difficult in future.

## 4. THE PHONET SOLUTION

An audio Internet Technology that allows users to listen to email, buy on-line or surf and hear any Web site, using a simple and natural interface - an ordinary telephone. No computer is needed.

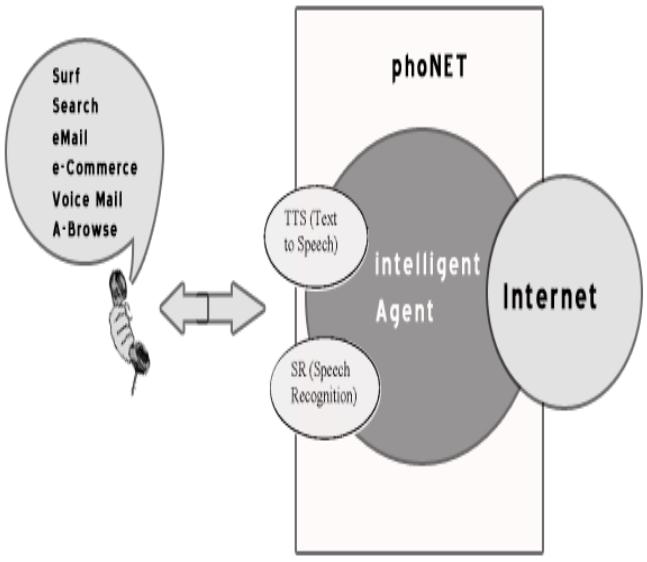


Fig. 1 Intelligent agent and Key features of phoNET

Subscribers dial a toll-free number, and start accessing the Internet using voice commands. Speech recognition technology in the company's system allows users to give simple commands, such as "go to Yahoo" or "read my email" to get to the Net-based information they want, when they want it, whether they're out on an appointment, stuck in traffic, sitting in an airport, or cooking dinner. They'll be able to quickly locate information, such as late-breaking news, traffic reports, directions, or anything else they're interested in on the World Wide Web. Our product **phoNET** has the capability to automatically down load web contents, filter out graphics, banners and images. It then renders extracted texts into concise, meaningful and very suitable in audio format texts before using TTS to convert into speech. **phoNET** also converts the rendered texts into other languages in real time. It can also be easily integrated with any back end application such as CRM/SCM, ERP etc. Thus, **phoNET** completely eliminates the need to rewrite any content in VXML, SALT or WML. So we strongly believe that our automation based approach will be very successful. Using text-to-speech technology, an "intelligent agent" will read the requested information out loud via a computerized voice, and process the user's voice commands.

## 5. TECHNOLOGY OVERVIEW.

The idea of listening to the Internet may at first sound a bit like watching the radio. How does a visual medium rich in icons, text, and images translate itself into an audible format that is meaningful and pleasing to the ear? The answer lies in an innovative integration of three distinct technologies that render visual content into short, precise, easily navigable, and meaningful text that can be converted to audio.The technologies and steps employed to accomplish this feat are:

**DOCUMENT PROCESSING**

1. Speech recognition
2. Text-to-speech translation, and

**Document Rendering**

3. Artificial Intelligence

The phoNET platform acts as an "Intelligent Agent" (IA) located between the user and the Internet (Figure 1). The IA automates the process of rendering information from the Internet to the user in a meaningful, precise, easily navigable and pleasant to listen to audio format. Rendering is achieved by using Page Highlights (a method to find and speak the key contents on a page), finding right as well as only relevant contents on a linked page, assembling right contents from a linked page, and providing easy navigation. These key steps are done using the information available in the visual web page itself and proper algorithms that use information such as text contents, color, font size, links, paragraph, and amount of text. Artificial Intelligence techniques are used in this automated rendering process. This is similar to how the human brain renders from a visual page; selecting the information of interest and then reading ii.

The IA includes a language translation engine that dynamically translates web contents from one language into another in real time. Thus, a Chinese speaking person can ask to surf an English website in Chinese - the Intelligent Agent would access the English website, extract the content of the website and translate it on the fly in Chinese and read it back to the user in Chinese.

The platform incorporates the highest quality speech recognition and text to speech engines from third party suppliers.

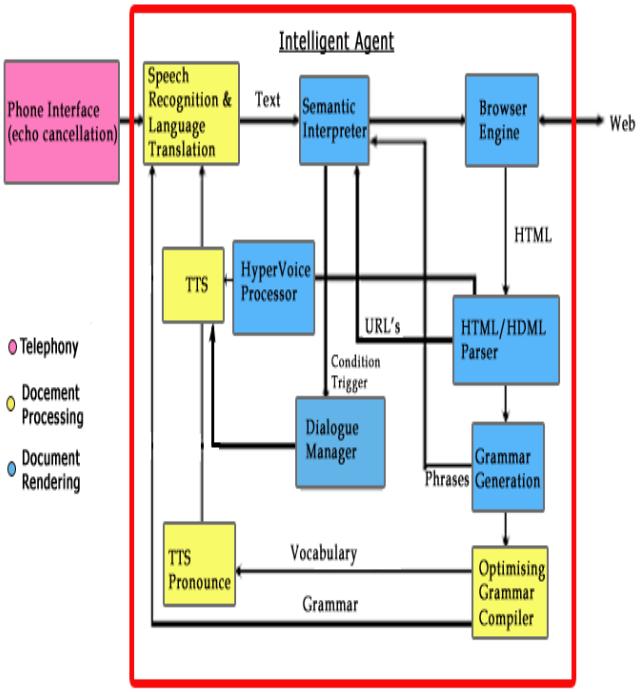


Fig. 2 PhoNET Architecture

The **PhoNET** architecture is shown in Figure 2. The process starts with a telephone call placed by a user. The user is prompted for a logon pass phrase. The users pass phrase establishes the first connection to the Web site associated with that phrase and loads the first Web page. The HTML is parsed, as described later, separating text from other media types, isolating URL from HTML anchors and isolating the associated anchor titles (including ALT fields) for grammar generation. Grammar generation computes combinations of the words in titles to produce a wide range of alternative ways to say subsets of the title phrase. In this process simple function words (i.e., "and," "or," "the," etc.) are not allowed to occur in isolation where they would be meaningless.

Browser control commands are mixed in to control typical browser operations like "go back "and "go home" (similar to the typical browser button commands). Further automatic expansion of the language models for link titles can include thesaurus substitution of words, etc. soon to be implemented .The current processing allows the user to say any keyword phrase from the title with deletions allowed to activate the associated link. Grammar processing continues with compilation and optimization into a finite-state network where redundancies have been eliminated. The word vocabulary associated with the grammar is further processed by a Text-To-Speech (TTS) pronunciation module that generates phonetic transcriptions for each word of the grammar. Since the TTS engine uses pronunciation rules it is not limited to dictionary words. The grammar and vocabulary are then loaded into the speech recognizer. This process typically takes about a second. At the same time, the Web document is described to the user, as discussed later.

The user may then speak a navigation or browser command phrase to control browsing. Each user navigation command takes the user to a new Web page. If the command is ambiguous the dialog manager collects the possible interpretations into a description list and asks the user to choose one. Recognizing that a collection of Web pages is inherently a finite-state network, it is easy to see that this mechanism provides the basis for a finite-state dialog and Web application control system.

## 6. APPLICATIONS

This technology can find applications for Service Providers, Businesses and governments in the following areas:

**For Service Providers** -

- Surf and browse the Web ·
- Email (send, receive, compose, copy, forward, reply, delete and more) ·
- Search the Web ·
- Voice Portal Features such as News, Weather, Stock Quotes, Horoscopes and more.

**For Businesses**

- Airline reservations and tracking
- Package tracker
- Reservations
- eCommerce
- Customer service
- Alert service
- Order Confirmation
- CRM Applications

**For Governments**

- All key benefits for Businesses Plus
- Easy Accessibility to all Government contents
- More efficient communication between Government and citizens, Government and businesses, and between Government departments.
- Meeting Government obligation to bridge the Digital and Language Divides (with automatic translation of Internet content from English to any other language and vice-versa.), and providing Internet to elderly, visually impaired and blind people in a very simple and cost effective way.

## 7. FUTURE WORK

Implementation and results of this new technology depends on the developments in the field of speech recognition and artificial intelligence. The major challenge of this technology is its complexity which comes with a high price tag. But it is not far more than the complexity and cost involved in voice enabling a web site using the present technologies. We believe that Voice/speech based interface options will become an important part of the overall solution to access the Internet content. And an automated approach to

Voice-Enable or create Voice Portals would be most practical and more common way than rewriting web contents in different languages and maintaining multiple version of the web sites.

## 8. CONCLUSION

we considered the possibility of accessing web through an ordinary phone . We presented a new technology which provides a true audio Internet experience. Using an ordinary telephone and simple voice commands, users will be able to surf and hear the entire Internet for the information they desire. A computer is not needed. Any web page will be accessible, but not limited to sites written with Wireless Application Protocol, and pages that are specially written in Voice Extensible Mark-up Language (VXML). We presented a detailed analysis of how the technologies like SR, TTS and AI are integrated to develop a intelligent Platform (phoNET) to achieve voice based web access. We presented the major problems involved in Document processing and document rendering along with solution.

## REFERENCES

1) http://www.w3.org/Voice/
2) http://www.voicexml.org/
3) Internet speech Inc.
4) Avaya Labs
5) http://www.lhs.com/
6) http://trqce.wisc.edu/world/web
7) http://www.dcp.ucla.edu/