The Emerging Virtual Personal Assistant

*E-Business II*

By: Jianwei Yao

December 12, 2000
An Overview

With new vocabulary coming out everyday as a result of the recent phenomenal technological advancement, people will soon become outdated. “Vortal”, or more understandably, “voice portal” has caught people’s attention among this wave. For the past 20 months or so, despite the Nasdaq meltdown, up to a dozen vortals have been funded. Some industry analysts even predicted that this would become one of the four future trends, along with Bluetooth, Peer-to-Peer and XML, to rival the invention of web browser and broadband.¹

What is it, if it is really that significant? This paper intends to offer an overview of the technology of Voice Recognition/Virtual Personal Assistant, particularly with its potentially powerful application on the Net (i.e., vortal).

The paper will first explain Virtual Personal Assistant from a historical perspective and analyze its success factors and weaknesses that still haunt its further growth. Then the paper will review the potential market and various segments that could be attracted. Current players in this market are also examined. Finally, the paper will discuss potential threat to this newly-born technology.

Imagine This Scenario

This has been a horrendous Friday. You were away on a very demanding business trip. When you just sat down in the airport, you were notified that the flight was delayed for two hours. In contrast with the rest of the group who were complaining bitterly, you picked up your mobile and dialed into a toll-free number. Over the phone, you HEARD the dozens of e-mail messages accumulated in your inbox while you were away, and then you dictated some replies. After the boring work, you thought you might need to relax and then checked the scores for your favorite hockey team, top news lines on CNN, and even weather at home. When you eventually got home, you decided to take the family out for a nice dinner as a reward for your whole week’s hard work. In your car, you asked and obtained the direction to a newly opened Chinese restaurant in town...

Throughout the process, you asked questions over the phone, and there was a voice that confirmed or clarified the questions you asked.

The next day, you were shown a much better “device” than a mere mobile phone – something that has a “face” to show all the words, charts or graphs, an “ear” to listen to what you say, and even a “voice” to tell you all the information you want to know. And your inseparable personal assistant could easily fit into your pocket...

But…not yet. All these would be made possible by “Virtual Personal Assistant”, or “Voice Portal”, a technology being developed by several firms for the past 20 months.

¹ Four technologies that will shape the Net, Fortune Oct. 9, 2000. Eric Nee.
Although there have already been several players experimenting this technology, it is far from perfection.

**What is a Virtual Personal Assistant, anyway?**

As you can imagine, a virtual personal assistant is a combination of various MMI (Man-Machine Interface) software and hardware, including voice recognition technologies, and is supported by the powerful World Wide Web. The following graph could help explain the make up of a Virtual Personal Assistant (VPA) –

![Virtual Personal Assistant Diagram](diagram)

*Figure: how a Virtual Personal Assistant works.*

While the development of wireless internet access would definitely add to the positive trend of VPA, this paper intends to focus on the more recent development of voice recognition technologies as applied to this.

Voice recognition technology has been developed over a long period of time. Historically, however, the application has been used in call centers to replace much more expensive human operators. With the emergence of internet, voice recognition technology was quickly applied to PC as well. The quality of the technology has also been improving steadily. “The accuracy rates of dictation programs from industry leaders – Lernout & Hauspie, Dragon Systems, Microsoft and IBM – have been improving around 10% a year for a decade now. During the next year, voice software geared to their specialized vocabularies is expected to gain a foothold in the medical and legal professions.”

This technology is definitely no easy task, since we are measuring the most challenging speech recognition. Evidence has shown if the task is restricted in vocabulary and context, accuracy can be extremely high, much higher than general transcription. This could simplify the process and improve accuracy somehow. For instance, if the caller says, “I want to read sports news”, the assistant recognizes the key words – “read” and

---

“sports news” and can then perform its task. It does not necessarily need to understand what “to” here means.³

Another recent development is the shift from pure **dictation** to **dialog**. It is very critical to understand its significance to the vortal applications. Dictation is more word-for-word oriented, while dialog goes the other way – focusing on commands and natural language understanding. Also, Dialog allows the system to ask for clarification, which could improve the accuracy of voice input significantly. “Charles Schwab fielded more than 17 million telephone calls last year giving quotes for any of 13,000 stocks and mutual funds, with accuracy exceeding 97%, despite the problems of the telephone network’s distortion of the speech”.⁴

**“Good” and “Bad” about VPA**

**The thumb-up’s of VPA**

The distinct advantage of VPA is its **convenience**, and on certain occasions **safety**. Talking to a computer or over a phone is always easier than typing. Besides, situations like driving a car, waiting for a delayed flight normally make laptop access to internet very difficult, sometimes even dangerous. A recent study by Rochester Institute of Technology concluded that a person using a phone while operating a motor vehicle is 34% more likely to be involved in an accident than someone who is not using a phone. “Using voice commands instead of touch-tones is not only a convenience, it is being perceived by consumers as a safety necessity”.⁵

This benefit is best illustrated when you try to find local taxi services and restaurants. Normally nobody would have a clue where these places are when they are new to the locality.

The almost **unlimited information source** of the Internet provides VPA enormous power. With search engines much more developed and the ever expanding internet content, the value of VPA is increasing as well. Probably this is exactly what Tim Berners-Lee did when he invented the WWW – as long as you conform to the basic rules of the Net’s digital language, you can enhance it with all sort of wild new technologies.⁶

VPA could also become a **fashion** and then undeniable **trend** at some point. Metcalf’s’s Law equally applies here. In the next few years, some industry analysts expect to see the era of “pervasive computing”, when the focus of digital software will migrate from desktop PCs linked to the Internet by phone wire to a “plethora of newfangled”, Web-ready products ranging from TVs and cell phones to dashboards and Palm Pilots. Most

---

³ For more detailed explanation on this, please refer to “Wizzard Software and the PC/Internet speech recognition market”, by William Meisel.
⁴ Id.
⁵ Advanced Voice-Controlled Electronic Assistants Tutorial.
of these devices would be too small to have a keyboard. Even if handheld PCs and Palm
Pilot try to accommodate this shortfall somehow, voice recognition, if developed to a
more mature stage, could become a dominant force in the MMI systems. Dragon
Systems founder Janet Baker said, “The only way you’re ever going to get lots of data
into small devices is by talking to them.”

*The issues of VPA*

While VPA is a very promising technology, there are some weaknesses that we need to
bear in mind before embracing it fully.

The accuracy of voice recognition remains a central issue to be resolved, despite the
significant improvement as mentioned previously. In some cases, it could not recognize
some foreign language names, which is very common in restaurant name recognition.
Also, background noise could be another headache. In one extreme case, someone tested
the system in busy Penn Station. The systems he tested came up with wrong listings
because they could not recognize who was actually giving instructions. Even Mike
McCue, the CEO of Tellme, one of the pioneering vorters, admitted that one problem is
that “the voice recognizer thinks the band is talking” when it occurred that the system
mistakenly messages from the Bar late at night.

Another problem has to do with insufficient range of offering at the current stage at any
of the leading vortals. Jupiter Research analyst Seamus McAteer commented, “It’s more
like a collection of services than a complete offering.” It seems this is NOT going to be
changed in the immediate future.

As a matter of fact, vortals may never offer exactly what the WWW has to offer. From
consumer behavior perspective, it takes time to speak to the system or to have computers
read out information. This could counter some of the convenience factors discussed
earlier. As a result, users are more likely to use voice portals mostly for things they can
do quickly, such as checking weather, e-mail titles, news headlines, or essential
information like restaurant address. In addition, currently most phones do not have
graphic displays. Therefore, it is even more limited than the WWW in showing off
products. Any potential e-commerce would be confined to more commodity-type goods.
Of course this would improve when more user-friendly hardware is available in the
marketplace.

Another potentially important issue is people are normally more used to writing out a
message rather than reading a message. This is particularly problematic in a dictation
process. Speech recognition would be hard, in part because many people are not
experienced with composing documents loud and do not speak fluently with trying to do
so. Therefore, a VPA would be best used for getting one’s ideas down in rough form or
for informal documents like email, where a few awkward constructions or errors are often

---

8 Four technologies that will shape the Net, Fortune Oct. 9, 2000. Eric Nee.
ignored. Dialog, as described earlier, more resembles the normal conversation and thus could mitigate this issue to certain extent.

**VPA’s Target Markets**

The target markets of VPA could be generally categorized into the following potential segments –

- **Remote and Mobile workers**

There are probably over 40 million remote and mobile workers in the U.S. today.\(^{10}\) Essentially all of these individuals heavily rely on telephones, pagers and laptop computers to become informed about ongoing changes that affect their work. All mobile workers share the need to receive and respond to telephone and email messages, easily conduct multiparty conversations whenever necessary and manage their appropriate calendars and To Do lists as the need arises. To them, of great importance is to have a device that could help them achieve all these tasks while they are on the move.

VPA could fit this need nicely. It could combine voice recognition and text-to-speech technologies, a dialog management engine, and advanced human interaction techniques to keep the mobile worker in constant touch.

- **Business Travelers**

Similar to the previous category, business travelers also do not have easy access to Internet when on the road. This is a huge segment which potentially includes almost all the workforce. However, obviously travelers with different income levels and different travel frequency would have different desire for VPA. While they are less reliant on mobile devices to obtain information, they care more about business information and are also generally less sensitive to price.

- **Vision-restricted Personnel**

Until 1998, there were 10.5 – 17 million vision-restricted people in North America, with the vast majority in the U.S.\(^ {11}\) This is a particular group that has not benefited as much from the Internet revolution. While we could have some reservation about their financial ability, this group still represents real opportunities for VPA. This is particularly true as they are, for the first time, offered the opportunity to “surf” the Internet.

The above are just some typical examples of the target segments. Actually other segments, even students could become users of VPA as a cheaper alternative to PCs or laptops. In the future, when VPA develops into a “brain” as described at the beginning of

this article, the sheer convenience (without typing in words, clicking search and then looking for the item you are looking for) and responsiveness would make it a great hit in the marketplace.

Current Players

The relatively recent development of both voice recognition on the PC and Voice Portals means much fewer players than other normal dot.com’s in the marketplace currently. I summarize below the current players in these two markets.

- **Voice Recognition on PCs**

Voice recognition on the PC comprises mostly dictation products from IBM, Dragon Systems, and Lernout & Hauspie (L&H), all of whom use their own underlying voice recognition technology. L&H’s recently purchased Dragon, after which there are two primary vendors of retail dictation products. Philips Speech Processing, which also has a dictation product, has been active in the marketplace. These vendors provide a large-vocabulary speech recognition engine that can be licensed independently. In addition, Carnegie Mellon University provides its Sphinx voice recognition engine and soon its Festbox TTS (text-to-speech) system free as source code under the Open Source Speech Initiative. IBM and Dragon have both provided their voice recognition engines at reasonable license fees.

- **Voice Portals**

Telephone voice recognition is growing rapidly and being considered an extension of the Internet to any phone. Up to now, several national consumer voice portals have emerged, including Tellme Networks, BeVocal, and Quack.com, which are all well-financed. Existing web portals, such as AOL, are responding quickly with an acquisition, intending to create a general voice portals to its services. On the other hand, telephone service providers also recognize this opportunity to distinguish their services. BellSouth has launched a voice portal service in Atlanta.

How about Business Model?

The early days of surging dot.com’s with no clear path to profitability have long gone. All the start-ups need to have a viable business model before they even try to secure VC funding. So what are the business models for VPA firms and vortals? Currently there are three broad models as follows –

---

12 For more detailed explanation on this, please refer to “Wizzard Software and the PC/Internet speech recognition market”, by William Meisel.
13 Id.
14 Id.
• **Core software + add-on modules**

This business model is to make initial purchase of core voice recognition software an affordable impulse purchase. The firms can then add add-on modules and sell upgrades to this customer base. Obviously these add-on modules could be expensive upgrades to create a “deluxe” version. This sale process could be conducted on the WWW as downloads with very low overhead.

• **Co-operation with various software, e-mail application providers**

This approach calls for close relationship with relevant software, e-mail application providers. Sometimes, the VPA firm could be bought out completely, or a joint-venture is formed with these providers. Then the business model of these VPA firms is combined with its strategic buyers. For instance, Microsoft, a big sponsor of voice recognition on PCs, bought Entropic, a voice recognition company, outright, and have a substantial investment in L&H. They will put voice recognition in Microsoft Office, and in a version of Windows, probably in 2001.\(^{15}\)

• **Next-generation phone services?**

Unlike pure VPA software firms, vortals are marketing themselves as next-generation phone services, which people will pay for, rather than Web services, which people tend to want for free. For instance, AOL charges $4.95 a month on top of its regular $21.95 fee. Several of the companies charge/plan to charge for premium services. Yahoo offers free basic service but bills for extras such as disk space or long-distance calls.\(^{16}\)

**Future Trends and Potential Threat**

As you would conclude from the above analysis, VPA applied to PC and the Web are still in its early stage. Although we can expect a bright future for this technology, it is not without risks. However, I believe with the advent of broadband, improved Graphic User Interface (GUI), and also availability of high-quality voice recognition, VPA would advance toward a more sophisticated level in the near future.

Opportunity is presented by mutual development of both voice recognition software and wireless hardware. If we can integrate voice interface with wireless internet access technologies, plus the option of graphic and text interface, we will really have a portable personal assistant that is powered by the endless World Wide Web.

Another trend is the shift from dictation to dialog in the voice recognition technologies. This would not only enhance the accuracy of voice input, but also make the device a real interactive personal assistant, through whom you can have access to the entire world!

\(^{15}\) Id.

A potential threat of voice recognition technologies, is also the fast development of wireless internet access. When people can read their e-mails, or other Web information through their mobile phone, they may not need to hear it any more. However, they may still need to input using voice recognition software, and after all, some people would prefer to hear rather than read under certain circumstances.

I would boldly predict that in the future, this virtual personal assistant is going to be an essential part of our life. And it does have ear, voice, face, (you can choose whatever communication media) and most importantly, its brain is empowered by the ubiquitous and boundless Internet.